

---

# Application of Sensitivity and Uncertainty Analyses to Linear Time-Invariant Compartmental Models

---

Suzan Gazioglu

*A dissertation submitted to the  
University of Glasgow  
for the degree of  
Doctor of Philosophy*



**UNIVERSITY**  
*of*  
**GLASGOW**

Department of Statistics  
April 2002

ProQuest Number: 13834184

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



ProQuest 13834184

Published by ProQuest LLC (2019). Copyright of the Dissertation is held by the Author.

All rights reserved.

This work is protected against unauthorized copying under Title 17, United States Code  
Microform Edition © ProQuest LLC.

ProQuest LLC.  
789 East Eisenhower Parkway  
P.O. Box 1346  
Ann Arbor, MI 48106 – 1346

GLASGOW  
UNIVERSITY  
LIBRARY:

THESIS 12736 - COPY 2

*in memory of my Mother*

1920-1921  
1922-1923  
1924-1925  
1926-1927

# Preface

Chapter 1 reviews the fundamental aspects of modelling and introduces sensitivity and uncertainty issues.

Chapter 2 first introduces and reviews the theory of linear, time-invariant compartmental models, then describes a number of methods used to solve model state equations analytically and numerically in order to make predictions. This chapter also describes the methodology of numerous sensitivity analysis methods.

In Chapter 3, application of various sensitivity analysis techniques to two 8-compartment global carbon cycle models is presented. For ease of comparison, a measure of similarity between the sensitivity conclusions from different methods is defined based on the top 10 ranked input factors according to each method for each output variable (i.e. for each compartment at chosen time points).

Chapter 4 presents the results of the application of various sensitivity analysis methods including non-parametric methods to a more complex 25-compartment global carbon cycle model.

An overall informal comparison indicates that the 8-compartment global carbon cycle models used in Chapters 3 and 4 are optimal with respect to efficiency (i.e. both are simple and model codes are not very time-consuming to run), but in return do not have a high degree of stability and reliability since they do not

adopt biological and chemical processes. As for the 25-compartment model, it is more complex and more costly to run. These chapters review the applicability of the sensitivity analysis methods to these models which has steady-state constrain.

Chapter 5 explores various sources of uncertainty and presents results of uncertainty analysis applied to the three global carbon cycle models that are used in Chapters 3 and 4. Here, we partition the overall prediction uncertainty of an output variable into different components of uncertainty.

Finally, Chapter 6 presents conclusions and main findings of the thesis.

# Acknowledgements

It is my great pleasure to record thanks to the many people who helped in many ways to contribute to this thesis.

First and foremost, I would like to express my deepest thanks to my supervisor, Professor E. Marian Scott, for suggesting the problems we studied and for all her advice, support and patience during the course of this research. Without her constant inspiration, guidance and endless encouragements, this thesis would not have existed.

For financial support, I am grateful to the Turkish Higher Education Council and Gaziosmanpaşa University for giving me the graduate scholarship.

I would like to thank all the staff members and postgraduate students in Statistics Department of Glasgow University. The coffee breaks and enjoyable yet relaxing trips around Scotland with the postgraduate students will always be part of my treasured memories.

My gratitude also goes to my friends Filiz Doğru and Christina Yap for their constant encouragement and support during the trying times.

My thanks are also due to all my friends in Turkey and the ones I have had the privilege to meet and develop close friendships with during my studies overseas. Their support and encouragement have made a great difference. I would never forget the friends whom I have shared my life with within and outwith the walls of the university.

Last, but by no means least, special thanks to my family, especially my dad, for their understanding, patience, support and faith in me. Their confidence in me and expressions of love have made this thesis possible.

# Contents

<b>Preface</b>	<b>i</b>
<b>1 Introduction: Modelling, Compartmental Modelling, Sensitivity and Uncertainty</b>	<b>1</b>
1.1 Introduction . . . . .	1
1.1.1 Conceptual, Mathematical and Computational Models . .	3
1.1.2 Sensitivities in Modelling World . . . . .	5
1.1.3 Uncertainties in Modelling World . . . . .	5
1.2 Modelling Compartmental Systems . . . . .	6
1.3 Aims and Objectives of this thesis . . . . .	9
<b>2 Methodology: Linear, Time-Invariant Compartmental Systems; Sensitivity and Uncertainty Analyses</b>	<b>10</b>
2.1 Introduction . . . . .	10
2.2 Linear, Time-Invariant Compartmental System Theory . . . . .	12
2.3 Analytical Solution of State Equations . . . . .	15
2.3.1 Classical Approach . . . . .	15
2.3.2 Laplace Transformation Approach . . . . .	16
2.4 Numerical Solution of State Equations . . . . .	17
2.4.1 Runge-Kutta Method . . . . .	18
2.4.2 Adams Methods . . . . .	19
2.5 Sensitivity Analysis . . . . .	23
2.5.1 Graphical Sensitivity and Uncertainty Methods . . . . .	23
2.5.1.1 Scatter Plots . . . . .	24
2.5.1.2 Star Plots . . . . .	24
2.5.1.3 Pie Charts . . . . .	25

2.5.1.4	Other Graphical Methods . . . . .	26
2.5.2	Local SA Methods . . . . .	26
2.5.2.1	The Sensitivity Function / Matrix . . . . .	27
2.5.2.2	Numerical Methods for Calculating Local Sensitivities . . . . .	29
2.5.3	Sampling based Global SA Methods . . . . .	33
2.5.4	Steps in Performing Sensitivity and Uncertainty Analyses .	34
2.5.4.1	Designing the Computer Experiment . . . . .	34
2.5.4.2	Assigning <i>pdf</i> 's to each input factor . . . . .	35
2.5.4.3	Generating the design matrix . . . . .	38
2.5.4.3.1	Simple Random Sampling . . . . .	38
2.5.4.3.2	Importance Sampling . . . . .	39
2.5.4.3.3	Latin Hypercube Sampling . . . . .	40
2.5.5	Screening Methods . . . . .	42
2.5.5.1	One-at-A-Time Designs . . . . .	43
2.5.5.2	Morris Design . . . . .	44
2.5.6	Methods of Analysis . . . . .	48
2.5.6.1	Correlation Measures . . . . .	48
2.5.6.2	Regression Analysis . . . . .	49
2.5.6.3	Stepwise Regression . . . . .	49
2.5.6.4	Rank Transformation . . . . .	50
2.5.6.5	SRC and SRRC . . . . .	50
2.5.6.6	PCC and PRCC . . . . .	51
2.6	Non-parametric SA Methods . . . . .	52
2.6.1	Smirnov Test . . . . .	52
2.6.2	Cramér-von Mises Test . . . . .	53
2.6.3	Mann-Whitney Test . . . . .	54
2.7	Implications on SA Methods . . . . .	54

<b>3</b>	<b>Application of Sensitivity Analysis Techniques to Compartmental Models with Specific Application to Global Carbon Cycle Models</b>	<b>56</b>
3.1	Introduction . . . . .	56
3.2	The Global Carbon Cycle . . . . .	57
3.2.1	Observed Atmospheric CO <sub>2</sub> since 1744 . . . . .	61

3.2.2	IPCC and Emission Scenarios . . . . .	63
3.2.3	CO <sub>2</sub> Inputs . . . . .	65
3.3	Modelling the Global Carbon Cycle . . . . .	65
3.3.1	The Two 8-Compartment GCC Models . . . . .	66
3.3.1.1	Description of Model I . . . . .	66
3.3.1.2	Description of Model II . . . . .	70
3.4	Extension and Application of SA to GCC Model I . . . . .	73
3.5	Results and Discussion on 8-Compartment GCC Model I . . . . .	80
3.5.1	Screening Methods . . . . .	80
3.5.1.1	Standard OAT Design on Initial Conditions . . . . .	80
3.5.1.2	Morris Design on Initial Conditions . . . . .	91
3.5.1.3	Standard OAT Design on Transfer Coefficients . . . . .	94
3.5.1.4	Morris Design on Transfer Coefficients . . . . .	104
3.5.2	Local SA . . . . .	110
3.5.2.1	Local SA on Initial Conditions . . . . .	111
3.5.2.2	Local SA on Transfer Coefficients . . . . .	113
3.5.3	Global SA Methods Applied to Initial Conditions . . . . .	115
3.5.3.1	Windowing Analysis . . . . .	115
3.5.3.2	Examination of Scatterplots . . . . .	125
3.5.3.3	Regression Methods . . . . .	127
3.5.3.4	Stepwise Regression . . . . .	128
3.5.4	Global SA Methods Applied to Transfer Coefficients . . . . .	130
3.5.4.1	Examination of Scatterplots . . . . .	132
3.5.4.2	Regression Methods . . . . .	137
3.5.4.3	SRC and PCC . . . . .	139
3.5.4.4	Stepwise Regression . . . . .	144
3.5.5	Discussion on the Results . . . . .	148
<b>4</b>	<b>Sensitivity Analysis Techniques Applied to a 25-Compartment GCC Model</b> . . . . .	<b>151</b>
4.1	Introduction . . . . .	151
4.2	The Model . . . . .	152
4.3	Results from Screening Methods . . . . .	159
4.3.1	Standard OAT Design . . . . .	159
4.3.1.1	Sensitivity Index . . . . .	163

4.3.1.2	Standardised Range . . . . .	172
4.3.2	Morris Design . . . . .	180
4.4	Global SA Methods . . . . .	187
4.4.1	Examination of Scatterplots . . . . .	189
4.4.2	Regression Methods . . . . .	201
4.4.3	SRC and PCC . . . . .	203
4.4.4	Stepwise Regression . . . . .	212
4.4.5	Rank Transformation . . . . .	216
4.4.6	Two-sample Nonparametric Tests . . . . .	220
4.4.6.1	Smirnov Test . . . . .	221
4.4.6.2	Cramér-von Mises Test . . . . .	223
4.5	Discussion on the Results . . . . .	227
<b>5</b>	<b>Uncertainty Analysis</b>	<b>231</b>
5.1	Introduction . . . . .	231
5.2	Model Intercomparison and Model Validation . . . . .	234
5.3	Uncertainties in GCC Models . . . . .	236
5.4	Main Sources of Uncertainty . . . . .	237
5.5	Input Factor Uncertainty . . . . .	238
5.5.1	Effect of Sample Size and Sampling Technique on Model Predictions . . . . .	240
5.5.2	Effect of Input Factor Distribution on Model Predictions .	248
5.6	Scenario Uncertainty . . . . .	253
5.7	Model Uncertainty . . . . .	264
5.8	Partitioning Uncertainty . . . . .	270
5.9	Modeller Uncertainty . . . . .	277
5.9.1	General Background . . . . .	277
5.9.2	Description of the Study and Data . . . . .	278
5.9.3	Testing Modeller Predictions . . . . .	280
5.9.3.1	Exploratory Data Analysis . . . . .	281
5.9.3.2	Modelling the Modeller Response . . . . .	282
5.9.3.3	Clustering . . . . .	284
5.10	Discussion . . . . .	285
<b>6</b>	<b>Conclusions</b>	<b>288</b>

<b>References</b>	<b>296</b>
<b>A Simulation Results of Model I Summarised in Chapter 3</b>	<b>1</b>
A.1 Model I Initial Conditions . . . . .	2
A.2 Results of Morris Design on Model I Initial Conditions . . . . .	9
A.3 Model I Transfer Coefficients . . . . .	10
A.4 Results of Morris Design on Model I Transfer Coefficients . . . . .	17
<b>B Summary of Model II Simulation Results</b>	<b>20</b>
B.1 Screening Methods . . . . .	21
B.1.1 Initial Conditions . . . . .	21
B.1.2 Transfer Coefficients . . . . .	34
B.2 Local SA . . . . .	47
B.2.1 Initial Conditions . . . . .	47
B.2.2 Transfer Coefficients . . . . .	47
B.3 Global SA . . . . .	49
B.3.1 Initial Conditions . . . . .	49
B.3.2 Transfer Coefficients . . . . .	57

# List of Tables

3.1	IPCC 1992 CO <sub>2</sub> Emission Scenario Results . . . . .	64
3.2	Model I reference case initial compartment contents. . . . .	69
3.3	Model I reference case transfer coefficients for carbon transfer among compartments. . . . .	69
3.4	Model II reference case initial compartment contents. . . . .	72
3.5	Model II reference case transfer coefficients ( <i>in units of yr<sup>-1</sup></i> ) for carbon transfer among compartments. . . . .	72
3.6	Rankings of Model I compartmental output sensitivities to the range of the initial conditions based on SI. . . . .	86
3.7	Results of Morris experiment on Model I. Initial conditions are ranked in order of importance according to the SA measures of Morris mean $\mu$ . . . . .	94
3.8	Three most effective transfer coefficients from SI - Transfer Coefficients of Model I are varied OAT. . . . .	100
3.9	Results of Morris experiment on the transfer coefficients of Model I.	109
3.10	Rankings of the initial conditions to Model I based on standardised sensitivity coefficients evaluated in years 1900, 2000 and 2100. . .	112
3.11	Rankings of the transfer coefficients to Model I for each compartmental output based on sensitivity coefficients evaluated in years 1900, 2000 and 2100. . . . .	114
3.12	Classification performance on (a) Fisher's discriminant method and (b) Quadratic discriminant method assessed on the original and cross-validated sample. . . . .	119
3.13	Classification performance on (a) Fisher's linear discriminant method and (b) Quadratic discriminant method used for stepwise discriminant function analysis. . . . .	121

3.14	Stepwise regression analyses for output variable $y_1$ of Model I in years 1900, 2000 and 2100. . . . .	129
3.15	Rankings of absolute Pearson correlation coefficients (CC) for the outputs of Model I. . . . .	136
3.16	Summary of regression analysis with Model I output variable $y_1(t = 2100)$ and input factors $k_{12}, k_{23}, k_{75}, k_{76}, k_{86}, k_{17}, k_{87}$ and $k_{18}$ . . . .	138
3.17	Rankings of absolute Standardized Regression Coefficients (SRC) and Partial Correlation Coefficients (PCC) for outputs of Model I. . . . .	143
3.18	Summary of stepwise regression analyses for Atmospheric $\text{CO}_2$ content of Model I in years 1900, 2000 and 2100. . . . .	145
3.19	Predicted error sum of squares (PRESS) values for the regression models summarized in Table 3.18. . . . .	146
3.20	Three most important transfer coefficients identified by stepwise regression procedure on Model I. . . . .	147
4.1	Model input factors selected for sensitivity analysis . . . . .	155
4.2	Dynamic equations of 25-compartment global carbon cycle model . . . . .	157
4.3	Rankings of Sensitivity Indices of the compartmental outputs in years 1900, 2000 and 2100 resulting from evaluating the model input factors at their max and min values OAT . . . . .	171
4.4	Rankings of Standardized Ranges of the compartmental outputs in years 1900, 2000 and 2100 resulting from varying the model input factors OAT . . . . .	179
4.5	Results of Morris experiment on the 25-compartment model . . . . .	186
4.6	Rankings of absolute CCs for the model outputs . . . . .	200
4.7	Summary of regression analysis for $y_{Atm}(t = 2100)$ and all input factors . . . . .	202
4.8	Rankings of absolute SRCs for the model outputs . . . . .	211
4.9	Summary of stepwise regression analyses for $y_{Atm}(t), y_{SO}(t), y_{DO_5}(t)$ and $y_{DO_{13}}(t)$ . . . . .	214
4.10	Summary of stepwise regression analyses for $y_{NWPT}(t), y_{WPT}(t), y_{GV}(t), y_{DD}(t)$ and $y_{ASC}(t)$ . . . . .	215
4.11	CCs, RCCs, SRCs, SRRCs, PCCs and PRCCs with raw and rank-transformed data for $y_{Atm}, y_{SO}, y_{DO_5}$ and $y_{DO_{13}}$ in year 2100. . . .	219

4.12	Comparison of Stepwise Regression Analyses with Raw and Rank-Transformed Data for $y_{Atm}$ , $y_{SO}$ , $y_{DO_5}$ and $y_{DO_{13}}$ in year 2100. . . .	220
4.13	Rankings of Smirnov test statistics for the model outputs . . . . .	222
5.1	Uncertainty importance rankings obtained by metric distance measure. . . . .	253
5.2	Scenario baseline values in 2100; uncertainty ranges of compartmental predictions from 2100 as a result of scenario uncertainty; and uncertainty ranges of compartmental predictions from 2100 as a result of both scenario and input factor uncertainties. . . . .	262
5.3	Comparison of Stepwise Regression Analyses with three IPCC emission scenarios for various output variables calculated from the three GCC models. . . . .	264
5.4	Comparison of models in terms of estimated mean and coefficient of variation (CV) values based on year 2100 CO <sub>2</sub> contents of Atmosphere, Ocean and Terrestrial Ecosystem components calculated using the three GCC models with the three IPCC-1992 scenarios. . . . .	270
5.5	Estimated scenario-specific means and standard deviations of Atmospheric CO <sub>2</sub> content in 2100 from all three models, together with two sets of scenario probabilities. . . . .	273
5.6	Results from partitioning the total uncertainty in predicted Atmospheric CO <sub>2</sub> content in 2100 into 'between scenarios' and 'due to input factors within scenarios' components as a function of scenario probabilities. The results are given for all three models. . . . .	274
5.7	Estimated scenario-specific means and standard deviations of Atmospheric CO <sub>2</sub> content in 2100, together with two sets of scenario probabilities. . . . .	276
5.8	Results from partitioning the total uncertainty in predicted Atmospheric CO <sub>2</sub> content in 2100 into 'between scenarios ( <i>BS</i> )', 'between models within scenarios ( <i>BMWS</i> )' and 'between predictions within models and scenarios ( <i>BPWMS</i> )' components as a function of scenario probabilities. . . . .	276
A.1	Morris estimated means and standard deviations associated with the initial conditions of Model I. . . . .	9

A.2	Morris estimated means and standard deviations associated with the selected transfer coefficients of Model I. . . . .	17
A.3	Pearson correlation coefficients (CC) for the outputs of Model I. . . . .	18
A.4	Order of importance between the independent transfer coefficients resulted from stepwise regression on Model I. . . . .	19
B.1	Sensitivity Rankings of Model II initial conditions from SI . . . . .	29
B.2	Sensitivity Rankings of Model II initial conditions from SR. . . . .	30
B.3	Results of Morris experiment on initial conditions and output variables of Model II. . . . .	33
B.4	Sensitivity Rankings of Model II transfer coefficients from SI. . . . .	42
B.5	Sensitivity Rankings of Model II transfer coefficients from SR. . . . .	43
B.6	Results of Morris experiment on initial conditions and output variables of Model II. . . . .	46
B.7	Rankings of the transfer coefficients of Model II based on standardised sensitivity coefficients evaluated in 1900, 2000 and 2100. . . . .	47
B.8	Rankings of the transfer coefficients of Model II for each compartmental output based on sensitivity coefficients evaluated in 1900, 2000 and 2100. . . . .	48
B.9	Rankings of absolute Pearson correlation coefficients (CC) for the outputs of Model II. . . . .	54
B.10	Rankings of absolute Standardized Regression Coefficients (SRC) for the outputs of Model II. . . . .	55
B.11	Stepwise regression analyses for Circulating carbon-(NH) compartment of Model II in 1900, 2000 and 2100. . . . .	56
B.12	Rankings of absolute Pearson correlation coefficients (CC) for the outputs of Model II. . . . .	61
B.13	Rankings of absolute Standardized Regression Coefficients (SRC) for the outputs of Model II. . . . .	62
B.14	Stepwise regression analyses for all compartments of Model II in 1900, 2000 and 2100. . . . .	63

# List of Figures

1.1	A diagram of Modelling World . . . . .	2
1.2	Two-compartment model diagram . . . . .	8
2.1	An example of a star plot. . . . .	25
2.2	A hypothetical model with three input factors ( $X_1, X_2, X_3$ ) and two output variables ( $Y_1, Y_2$ ). . . . .	33
2.3	An illustration of an LHS of size $N = 5$ from a two-dimensional input space. . . . .	41
3.1	A schematic diagram of the Global Carbon Cycle . . . . .	59
3.2	CO <sub>2</sub> emissions from fossil fuels. . . . .	60
3.3	The annual atmospheric CO <sub>2</sub> concentrations from Mauna Loa Observatory, Hawaii: 1958-2000; South Pole: 1973-1993; Siple Station: 1744-1953. . . . .	62
3.4	Historic and future CO <sub>2</sub> emissions. . . . .	64
3.5	Compartment diagram of Global Carbon Cycle Model I. . . . .	67
3.6	Compartment diagram of Global Carbon Cycle Model II. . . . .	70
3.7	Atmospheric CO <sub>2</sub> predictions resulting from varying initial compartmental content $x_i^{\circ}$ of compartment $i$ ( $i = 1, 2, \dots, 8$ ) OAT. . . . .	82
3.8	Sensitivity indices of compartmental outputs to the range of initial conditions. . . . .	85
3.9	Dotcharts showing how each compartmental output of Model I in years 1900, 2000 and 2100 is effected by the variation in $x_1^{\circ}, x_2^{\circ}, x_3^{\circ}$ and $x_4^{\circ}$ in terms of Standardised Ranges. . . . .	89
3.10	Dotcharts showing how each compartmental output of Model I in years 1900, 2000 and 2100 is effected by the variation in $x_5^{\circ}, x_6^{\circ}, x_7^{\circ}$ and $x_8^{\circ}$ in terms of Standardised Ranges. . . . .	90

3.11	Morris screening results on Atmosphere compartment of Model I in years 1900, 2000 and 2100. . . . .	93
3.12	Atmospheric CO <sub>2</sub> predictions resulting from varying transfer coefficients $k_{ij}$ OAT. . . . .	96
3.13	Sensitivity indices of compartmental outputs to each transfer coefficient. . . . .	98
3.14	Dotcharts showing how output of Atmosphere, Surface ocean, Deep ocean and Nonwoody parts of trees compartments in years 1900, 2000 and 2100 are effected by the variation in the transfer coefficients $k_{ij}$ in terms of Standardised Ranges. . . . .	102
3.15	Dotcharts showing how output of Woody parts of trees, Ground vegetation, Detritus/decomposers and Active soil carbon compartments in years 1900, 2000 and 2100 are effected by the variation in the transfer coefficients $k_{ij}$ in terms of Standardised Ranges. . .	103
3.16	Morris screening results on Atmosphere and Nonwoody parts of trees compartments of Model I in 1900, 2000 and 2100. . . . .	105
3.17	Morris screening results on Woody parts of trees and Ground vegetation compartments of Model I in 1900, 2000 and 2100. . . . .	106
3.18	Morris screening results on Detritus/decomposers and Active soil carbon compartments of Model I in 1900, 2000 and 2100. . . . .	107
3.19	Illustration of windowing N=100 atmospheric CO <sub>2</sub> predictions resulted from varying Model I initial conditions ( $x_1^o, \dots, x_8^o$ ) simultaneously. . . . .	117
3.20	Classification tree showing the performance of the windowing data.	122
3.21	A pruned version of the classification tree given in Figure 3.20. . .	123
3.22	Dependent variables predicted by Model I following windowing analysis: CO <sub>2</sub> content of (a) Atmosphere, (b) Surface ocean, (c) Deep ocean, (d) Nonwoody parts of trees, (e) Woody parts of trees, (f) Ground vegetation, (g) Detritus/decomposers, and (h) Active soil carbon compartments as a result of varying all input factors ( $x_i^o$ s) simultaneously. . . . .	124
3.23	Scatterplots of predicted Atmospheric CO <sub>2</sub> content in years 1900, 2000 and 2100 versus each compartment's initial condition ( $x_i^o$ ). .	126

3.24	Dependent variables predicted using Model I: CO <sub>2</sub> content of (a) Atmosphere, (b) Surface ocean, (c) Deep ocean, (d) Nonwoody parts of trees, (e) Woody parts of trees, (f) Ground vegetation, (g) Detritus/decomposers, and (h) Active soil carbon compartments as a result of varying selected 8 transfer coefficients ( $k_{ij}$ s) simultaneously. . . . .	131
3.25	Scatterplots of predicted Atmospheric CO <sub>2</sub> content in years 1900, 2000 and 2100 versus each transfer coefficient $k_{ij}$ . . . . .	133
3.26	Scatterplot matrix of predicted Atmospheric CO <sub>2</sub> content in 2100 and each transfer coefficient $k_{ij}$ . . . . .	134
3.27	Standardized regression coefficients (SRCs) for the eight independent transfer coefficients, with the predicted CO <sub>2</sub> concentrations in each compartment (i.e, with the dependent variables $y_1(t), \dots, y_8(t)$ ), N=5,000. . . . .	140
3.28	Partial correlation coefficients (PCCs) for the eight independent transfer coefficients, with the predicted CO <sub>2</sub> concentrations in each compartment (i.e, with the dependent variables $y_1(t), \dots, y_8(t)$ ), N=5,000. . . . .	141
4.1	Diagram of 25-Compartment Global Carbon Cycle Model (adapted from Emanuel <i>et al.</i> (1984) . . . . .	154
4.2	Atmospheric CO <sub>2</sub> predictions resulting from varying the initial conditions OAT . . . . .	160
4.3	Atmospheric CO <sub>2</sub> predictions resulting from varying the forest clearing and the reforestation input factors OAT . . . . .	161
4.4	Atmospheric CO <sub>2</sub> predictions resulting from varying the chemical and the physical ocean input factors OAT . . . . .	162
4.5	Sensitivity Indices of compartmental outputs due to the range of the initial conditions (CA0, CF0, CW0, CG0, CD0, CSL0). . . . .	164
4.6	Sensitivity Indices of compartmental outputs due to the range of the forest clearing input factors (PHIA, PHID, PSIS, SXIT) and the reforestation input factors (SIG, SS, EPS). . . . .	165
4.7	Sensitivity Indices of compartmental outputs due to the range of the chemical ocean input factors (SIGB, TEMP0, CL, RELHUM) and the physical ocean input factors (HM, AREA, DELTP). . . . .	167

4.8	Sensitivity Indices of compartmental outputs due to the range of the terrestrial turnover times input factors (TF, TW, TG, TD, TSL), the soil-forming fractions input factors (THW, THG, THD) and the intrinsic recovery times input factors (TT2, TV2). . . . .	169
4.9	Bar charts showing how each compartmental output of the model in years 1900, 2000 and 2100 is effected by the variation in the initial conditions (CA0, CF0, CW0, CG0, CD0, CSL0) in terms of standardized ranges . . . . .	174
4.10	Bar charts showing how each compartmental output in years 1900, 2000 and 2100 is effected by the variation in the forest clearing input factors (PHIA, PHID, PSIS, SXIT) and the reforestation input factors (SIG, SS, EPS) in terms of standardized ranges . . .	175
4.11	Bar charts showing how each compartmental output in years 1900, 2000 and 2100 is effected by the variation in the chemical and physical ocean input factors in terms of standardised ranges . . .	176
4.12	Bar charts showing how each compartmental output of the model in years 1900, 2000 and 2100 is effected by the variation in the terrestrial turnover times input factors in terms of standardized range . . . . .	177
4.13	Bar charts showing how each compartmental output in years 1900, 2000 and 2100 is effected by the variation in the soil-forming fractions and the intrinsic recovery times input factors in terms of standardized ranges . . . . .	178
4.14	Morris screening results on Atmosphere, Surface ocean and Deep ocean (layer5) compartments in years 1900, 2000 and 2100 . . . . .	183
4.15	Morris screening results on Deep ocean (layer13), Nonwoody parts of trees and Woody parts of trees compartments in years 1900, 2000 and 2100 . . . . .	184
4.16	Morris screening results on Ground vegetation, Detritus/decomposers and Active soil carbon compartments in years 1900, 2000 and 2100	185
4.17	Time dependent behaviour of the nine dependent variables predicted using 25-compartment model as a result of varying all 30 input factors simultaneously . . . . .	188
4.18	Scatterplots of $y_{Atm}(t = 2100)$ versus each input factor . . . . .	191
4.19	Scatterplots of predicted $y_{DO13}(t = 2100)$ versus each input factor	192

4.20	Scatterplots of predicted $y_{NWPT}(t = 2100)$ versus each input factor	193
4.21	Scatterplots of predicted $y_{WPT}(t = 2100)$ versus each input factor	194
4.22	Scatterplots of predicted $y_{GV}(t = 2100)$ versus each input factor.	195
4.23	Scatterplots of predicted $y_{DD}(t = 2100)$ versus each input factor	197
4.24	Scatterplots of predicted $y_{ASC}(t = 2100)$ versus each input factor	198
4.25	Time-dependent behaviour of the SRCs for the atmosphere and three ocean compartments	204
4.26	Time-dependent behaviour of the SRCs for the terrestrial compartments	206
4.27	Time-dependent behaviour of the PCCs for the atmosphere and the three ocean compartments	208
4.28	Time-dependent behaviour of the PCCs for the terrestrial compartments	209
4.29	Scatterplot for $y_{Atm}(t = 2100)$ versus AREA with raw and rank transformed data	217
4.30	Star plots of Cramér-von Mises test statistics for the output of the atmosphere, surface ocean and the two deep ocean compartments in year 2100 - with the 50th & 90th quantile partitioning	224
4.31	Star plots of Cramér-von Mises test statistics for the output of the nonwoody parts of trees, woody parts of trees, ground vegetation, detritus/decomposers and active soil carbon compartments in year 2100 - with the 50th & 90th quantile partitioning	226
5.1	A diagram of sources of uncertainty.	232
5.2	Comparison of historical atmospheric CO <sub>2</sub> predictions of the three GCC models with the historical CO <sub>2</sub> record.	235
5.3	A simple example of prediction uncertainty bands.	239
5.4	Example CDFs for $y_{Atm}(t)$ from Model I estimated with random samples of size 10 and 100 under the assumption that the initial conditions of the compartments are uniformly distributed on their assigned ranges.	241

5.5	Summary of distribution of CDFs for Atmospheric CO <sub>2</sub> predictions in 2100 from one of the 8-compartment GCC model (Model I) estimated with 3 replications of 100 simple random samples and 3 replications of 100 Latin hypercube samples of size 10 and 100 under the assumption that the initial conditions are uniformly distributed on their assigned uncertainty ranges. . . . .	242
5.6	Summary of distribution of CDFs for $y_{Atm}(t = 2100)$ from the 25-compartment GCC model estimated with 3 replications of 100 simple random samples and 3 replications of 100 Latin hypercube samples of size 10 and 100 under the assumption that the model input factors are uniformly distributed on their assigned uncertainty ranges. . . . .	243
5.7	Comparison of estimated CDFs for output variable $y_{Atm}(t = 2100)$ from Model I based on SRS and LHS with different sample sizes under the assumption that the model input factors: (a) Initial conditions, and (b) Transfer coefficients are uniformly distributed on their assigned uncertainty ranges. . . . .	245
5.8	Comparison of estimated CDFs for output variable $y_{Atm}(t = 2100)$ from the 25-compartment model based on SRS and LHS with different sample sizes under the assumption that the model input factors are uniformly distributed on their assigned uncertainty ranges.	246
5.9	Comparison of PCC values based on different sampling technique and sample sizes. The output variable considered here is $y_{Atm}(t = 2100)$ calculated from Model I with the transfer coefficients varied simultaneously under the condition that all factors follow uniform distribution on their assigned ranges. . . . .	247
5.10	CDFs showing the effect of input factor - (a) Initial conditions, (b) Transfer coefficients - distributions on the estimated distribution of the output variable $y_{Atm}(t = 2100)$ from Model I. . . . .	249
5.11	CDFs showing the effect of input factor distribution on the estimated distribution of the output variable $y_{Atm}(t = 2100)$ from 25-compartment model. . . . .	250
5.12	An example showing the metric distance measure between two CDFs.	252
5.13	The range of CO <sub>2</sub> baseline predictions of each compartment of Model I, based on IPCC-IS92a,c,e emission scenarios. . . . .	255

5.14	The range of CO <sub>2</sub> baseline predictions of each compartment of Model II, based on IPCC-IS92a,c,e emission scenarios. . . . .	256
5.15	The range of CO <sub>2</sub> baseline predictions of the nine compartments of the 25-compartment model, based on IPCC-IS92a,c,e emission scenarios. . . . .	257
5.16	Boxplots showing the distribution of CO <sub>2</sub> predictions in year 2100 from Model I considering three IPCC emission scenarios. . . . .	259
5.17	Boxplots showing the distribution of CO <sub>2</sub> predictions in year 2100 from Model II considering three IPCC emission scenarios. . . . .	260
5.18	Boxplots showing the distribution of CO <sub>2</sub> predictions in year 2100 from the 25-compartment model considering the three IPCC emission scenarios. . . . .	261
5.19	Comparison of predicted atmospheric CO <sub>2</sub> concentrations from the three GCC models under the base-case scenario and the measured atmospheric CO <sub>2</sub> concentrations from Mauna Loa Observatory: 1959-2000. . . . .	266
5.20	Baseline predictions from the three GCC models with the IS92a,c,e emission scenarios for the time period 2000-2100. . . . .	268
5.21	Maximum uncertainty range of the Atmospheric CO <sub>2</sub> predictions based on the three GCC models with the IS92a,c,e emission scenarios for the period 1995-2100. . . . .	271
5.22	Measured concentrations, and model predictions calculated with <i>CHERPAC</i> on daily concentrations of <sup>137</sup> Cs on Pasture from May 5 - June 27, 1986. . . . .	281
5.23	Log-transformed measured concentrations, and modeller predictions calculated with <i>CHERPAC</i> on daily concentrations of <sup>137</sup> Cs on Pasture. . . . .	282
5.24	Scatterplots of estimated model parameters (Intercepts vs. Slopes) of log-transformed data. . . . .	283
5.25	Dendrogram of log-transformed measured concentrations and modeller predictions of daily <sup>137</sup> Cs concentrations on Pasture. . . . .	284
A.1	Surface ocean CO <sub>2</sub> predictions resulting from varying initial compartmental content $x_i^o$ of compartment $i$ ( $i = 1, 2, \dots, 8$ ) OAT. . . . .	2

A.2	Deep ocean CO <sub>2</sub> predictions resulting from varying initial compartmental content $x_i^o$ of compartment $i$ ( $i = 1, 2, \dots, 8$ ) OAT. . . .	3
A.3	Nonwoody parts of trees CO <sub>2</sub> predictions resulting from varying initial compartmental content $x_i^o$ of compartment $i$ ( $i = 1, 2, \dots, 8$ ) OAT. . . . .	4
A.4	Woody parts of trees CO <sub>2</sub> predictions resulting from varying initial compartmental content $x_i^o$ of compartment $i$ ( $i = 1, 2, \dots, 8$ ) OAT. . . . .	5
A.5	Ground vegetation CO <sub>2</sub> predictions resulting from varying initial compartmental content $x_i^o$ of compartment $i$ ( $i = 1, 2, \dots, 8$ ) OAT. . . . .	6
A.6	Detritus/decomposers CO <sub>2</sub> predictions resulting from varying initial compartmental content $x_i^o$ of compartment $i$ ( $i = 1, 2, \dots, 8$ ) OAT. . . . .	7
A.7	Active soil carbon CO <sub>2</sub> predictions resulting from varying initial compartmental content $x_i^o$ of compartment $i$ ( $i = 1, 2, \dots, 8$ ) OAT. . . . .	8
A.8	Surface ocean CO <sub>2</sub> predictions resulting from varying transfer coefficients $k_{ij}$ OAT. . . . .	10
A.9	Deep ocean CO <sub>2</sub> predictions resulting from varying transfer coefficients $k_{ij}$ OAT. . . . .	11
A.10	Non-woody parts of trees CO <sub>2</sub> predictions resulting from varying transfer coefficients $k_{ij}$ OAT. . . . .	12
A.11	Woody parts of trees CO <sub>2</sub> predictions resulting from varying transfer coefficients $k_{ij}$ OAT. . . . .	13
A.12	Ground vegetation CO <sub>2</sub> predictions resulting from varying transfer coefficients $k_{ij}$ OAT. . . . .	14
A.13	Detritus/decomposers CO <sub>2</sub> predictions resulting from varying transfer coefficients $k_{ij}$ OAT. . . . .	15
A.14	Active soil carbon CO <sub>2</sub> predictions resulting from varying transfer coefficients $k_{ij}$ OAT. . . . .	16
B.1	Circulating carbon-(NH) CO <sub>2</sub> predictions resulting from varying initial condition $x_i^o$ of compartment $i$ ( $i = 1, 2, \dots, 8$ ) OAT. . . . .	21
B.2	Surface ocean-(NH) CO <sub>2</sub> predictions resulting from varying initial condition $x_i^o$ of compartment $i$ ( $i = 1, 2, \dots, 8$ ) OAT. . . . .	22
B.3	Deep ocean-(NH) CO <sub>2</sub> predictions resulting from varying initial condition $x_i^o$ of compartment $i$ ( $i = 1, 2, \dots, 8$ ) OAT. . . . .	23

B.4	Humus-(NH) CO <sub>2</sub> predictions resulting from varying initial condition $x_i^o$ of compartment $i$ ( $i = 1, 2, \dots, 8$ ) OAT. . . . .	24
B.5	Circulating carbon-(SH) CO <sub>2</sub> predictions resulting from varying initial condition $x_i^o$ of compartment $i$ ( $i = 1, 2, \dots, 8$ ) OAT. . . . .	25
B.6	Surface ocean-(SH) CO <sub>2</sub> predictions resulting from varying initial condition $x_i^o$ of compartment $i$ ( $i = 1, 2, \dots, 8$ ) OAT. . . . .	26
B.7	Deep ocean-(SH) CO <sub>2</sub> predictions resulting from varying initial condition $x_i^o$ of compartment $i$ ( $i = 1, 2, \dots, 8$ ) OAT. . . . .	27
B.8	Humus-(SH) CO <sub>2</sub> predictions resulting from varying initial condition $x_i^o$ of compartment $i$ ( $i = 1, 2, \dots, 8$ ) OAT. . . . .	28
B.9	Morris screening results on Circulating carbon-(NH), Surface ocean-(NH), Deep ocean-(NH) and Humus-(NH) compartments of Model II in 1900, 2000 and 2100. . . . .	31
B.10	Morris screening results on Circulating carbon-(SH), Surface ocean-(SH), Deep ocean-(SH) and Humus-(SH) compartments of Model II in 1900, 2000 and 2100. . . . .	32
B.11	Circulating carbon-(NH) CO <sub>2</sub> predictions resulting from varying transfer coefficients $k_{ij}$ OAT. . . . .	34
B.12	Surface Ocean-(NH) CO <sub>2</sub> predictions resulting from varying transfer coefficients $k_{ij}$ OAT. . . . .	35
B.13	Deep Ocean-(NH) CO <sub>2</sub> predictions resulting from varying transfer coefficients $k_{ij}$ OAT. . . . .	36
B.14	Humus-(NH) CO <sub>2</sub> predictions resulting from varying transfer coefficients $k_{ij}$ OAT. . . . .	37
B.15	Circulating Carbon-(SH) CO <sub>2</sub> predictions resulting from varying transfer coefficients $k_{ij}$ OAT. . . . .	38
B.16	Surface Ocean-(SH) CO <sub>2</sub> predictions resulting from varying transfer coefficients $k_{ij}$ OAT. . . . .	39
B.17	Deep Ocean-(SH) CO <sub>2</sub> predictions resulting from varying transfer coefficients $k_{ij}$ OAT. . . . .	40
B.18	Humus-(SH) CO <sub>2</sub> predictions resulting from varying transfer coefficients $k_{ij}$ OAT. . . . .	41
B.19	Morris screening results on Circulating carbon-(NH), Surface ocean-(NH), Deep ocean-(NH) and Humus-(NH) compartments of Model II in 1900, 2000 and 2100. . . . .	44

B.20	Morris screening results on Circulating carbon-(SH), Surface ocean-(SH), Deep ocean-(SH) and Humus-(SH) compartments of Model II in 1900, 2000 and 2100. . . . .	45
B.21	Classification tree showing the performance of the windowing data.	49
B.22	A pruned version of the classification tree given in Figure B.21. . . . .	50
B.23	Time dependent behaviour of all output variables predicted by Model II following windowing analysis as a result of varying all initial conditions ( $x_i^o$ , $i = 1, \dots, 8$ ) simultaneously. . . . .	51
B.24	Scatterplots of predicted CO <sub>2</sub> content of Circulating carbon-(NH), Surface ocean-(NH), Deep ocean-(NH) and Humus-(NH) compartments in 2100 versus each compartment's initial condition. . . . .	52
B.25	Scatterplots of predicted CO <sub>2</sub> content of Circulating carbon-(SH), Surface ocean-(SH), Deep ocean-(SH) and Humus-(SH) compartments in 2100 versus each compartment's initial condition ( $x_i^o$ ). . . . .	53
B.26	Time dependent behaviour of all output variables predicted by Model II as a result of varying selected 11 free transfer coefficients simultaneously. . . . .	57
B.27	Scatterplots of predicted CO <sub>2</sub> content of Circulating carbon-(NH), Surface ocean-(NH), Deep ocean-(NH) and Humus-(NH) compartments in 2100 versus each free transfer coefficient. . . . .	58
B.28	Scatterplots of predicted CO <sub>2</sub> content of Circulating carbon-(SH), Surface ocean-(SH), Deep ocean-(SH) and Humus-(SH) compartments in 2100 versus each free transfer coefficient. . . . .	59
B.29	Scatterplot matrix of predicted CO <sub>2</sub> content of Circulating carbon-(NH) in 2100 and each transfer coefficient. . . . .	60

# Chapter 1

## Introduction: Modelling, Compartmental Modelling, Sensitivity and Uncertainty

### 1.1 Introduction

Why model? We need models to mathematically and simply represent reality, to understand, test and explain observed behaviours, and to make predictions of future behaviours. Many studies of systems and processes begin with the construction of a model. Models serve as tools for formalizing how we think a system or a process is working. They function as investigative tools and play an important role in many disciplines of modern science. In science, many processes dealt with are so complex that physical experimentation is too costly, too time consuming or even impossible. In such cases, investigators often turn to mathematical or computational models [97]. Such models are used to describe the relationships between the system variables in terms of mathematical expressions which can then be used for the system's future predictions. Mathematical models may be characterised as linear or nonlinear, deterministic or stochastic, time dependent or time independent, continuous or discrete, *etc.*, and they are extensively used in

many fields of science, such as engineering, biology, economy, chemistry, ecology, and physics. From a statistical point of view, we use models to make predictions, test hypotheses, manage and make decisions about a system under investigation.

In the modelling process, some details about the system being modelled are lost, but the model is still expected to preserve the essential features, be simple, realistic, efficient, reliable, accurate and precise, in short, be a “good” model. Modellers have been using mathematical and statistical tools to check those expectations about their models, and Sensitivity and Uncertainty Analyses are two of the most important tools used. Consider the diagram presented in Figure 1.1. The problems which scientists aim to solve exist in the physical world. In reality, we do not know all the properties of this world. First, this complicated physical world is simplified and a model world is created. A model is developed for a specific problem. Then, by using relevant mathematical and statistical techniques and tools, the model is analysed. Next, the model is tested to see if it is a good model, i.e., provides an answer to the problem and behaves as expected. If this is the case, then it is used in the physical world, otherwise the modelling process

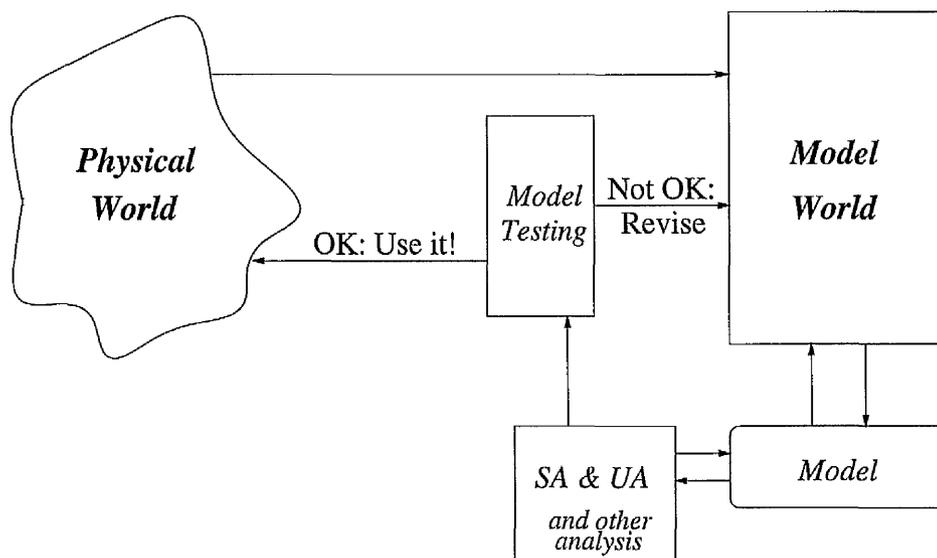


Figure 1.1. A diagram of Modelling World

is revised.

### 1.1.1 Conceptual, Mathematical and Computational Models

#### *Conceptual Models*

Conceptual models can simply be defined as the extraction of the essence. Construction of a conceptual model is a way of determining how best to sort, group, and annotate the information contained in a system.

A conceptual model refers to concepts that offer a range of suggestions which concentrate on the examination of structures and their function. Using conceptual models, a system can be presented visually to the users. Also the ease of use and understanding of the system can be optimized with a conceptual model.

#### *Mathematical Models*

Often, a conceptual model which is a simplified model of the variables and interactions present in the system is formulated to develop a mathematical model that can be used to examine the system behaviour. In a mathematical model the content, relation, structure and decision factors/variables all appear in equations.

#### *Computational Models*

Computer models (often referred to as *computer codes*) that implement the mathematical models are widely used in science as well as in industry. Common characteristics of these computer codes are:

- many input variables (often numbered in hundreds),
- the values and distribution functions of these variables are frequently not well known,
- the relationships among the variables are usually complex, modeled only by systems of differential equations which are not mathematically tractable.

Computational models are computer implementations of mathematical models. They are important tools in science as well as in practical life. The terminology ‘*computational model*’ is used to refer to representations of physical or other systems of interest that are first expressed in mathematical terms often as large and/or complicated systems of differential equations, and then implemented in the form of computer programs. Consider a system which is modeled by means of a compartmental model. Provided that the computer implementation of the model is described, the transport of material in the system can be simulated, and the outputs of interest which result from given inputs can be evaluated.

Suppose  $\mathbf{y}$  is an output of a computational model, and it is a function of  $k$  inputs denoted by  $x_1, x_2, \dots, x_k$ . For a given  $k$ -element input vector  $\mathbf{x}$ , the model can be used to evaluate  $\mathbf{y}$  without error; that is, replicate evaluations of  $\mathbf{y}$  obtained running the model’s computer code with the same  $\mathbf{x}$  will be identical. In other words, the model is deterministic [64]. Here to simplify notation we assume that the model produces only one output, but later in the thesis we deal with several dependent output variables. A computational experiment will consist of  $n$  model runs; the  $i$ th row of  $n \times k$  dimensional sample matrix  $\mathbf{X}$  corresponds to the set of input values  $\mathbf{x}_i$  for the  $i$ th model run. After  $n$  model runs, the experiment will result in a model output vector  $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n]^T$ . Then, to investigate the model behaviour, sensitivity and uncertainty analyses are carried out on the results. Since in this analysis we are taking into consideration not only model parameters but also initial conditions, input variables, physical conditions, *etc.*, as do Campolongo & Saltelli [12], we use the statistical design terminology ‘*factor*’ to refer to all conditions and variables unless stated specifically. Thus, a factor can be defined as a quantity that can be changed in the specification of the model prior to its run.

The model result while all parameters are held constant at their reference values, which are gathered from the literature and often called ‘nominal values’, is referred to as the ‘*base-line case*’.

### 1.1.2 Sensitivities in Modelling World

The basic idea of sensitivity theory is based on the evaluation of model performance resulting from the changes in model parameters. Eslami defines sensitivity as behavioral study of system performance under parametric variations, and certain unwanted exogenous input acting on the system [34].

Sensitivity Analysis (SA) was originally created to deal with uncertainties in the input variables and model parameters. SA of models has been advanced and is commonly employed in the modelling world to increase the confidence in the model and its predictions by providing an understanding of how the model response variables respond to changes in the inputs, data used to calibrate it, model structures, or factors, i.e. model independent variables [97].

### 1.1.3 Uncertainties in Modelling World

Uncertainty exists because models are imperfect mimics of reality [55]. Often, it is not possible to know many parameters that are required for the solution of problems with great accuracy, that is, there is always uncertainty involved in the system being studied. As Saltelli [97] puts it “Models and uncertainty go hand in hand”.

SA is closely linked to Uncertainty Analysis (UA) which aims to quantify the overall uncertainty associated with the response as a result of uncertainties in the model input parameter (often addressed as parametric uncertainty) and model structure itself (model structural uncertainty) [97]. Parametric uncertainty is quantified in a distribution of parameter values. Model uncertainty becomes an issue when more than one model can adequately reproduce the observed data. Another source of uncertainty is scenario uncertainty which reflects uncertainty on future conditions particularly when dealing with future predictions.

The use of systematic sensitivity and uncertainty analysis on large, complex computational models plays an important role in science. As complex numerical models are being increasingly applied for problem solving in many application

areas, such as atmospheric science, combustion physics and engineering, biological systems, *etc.*, the need for sensitivity and uncertainty analysis is becoming more and more apparent. Saltelli [97] considers SA as a prerequisite for model building in any discipline where models are used, and since sensitivity questions arise when uncertainty is present, UA can also be considered as a prerequisite in the modelling process.

Below we give basic definitions for sensitivity and uncertainty analysis to summarize what has been said above:

**Sensitivity Analysis** is a procedure of determining and quantifying the change in model behaviour as model factors change.

**Uncertainty Analysis** is an assessment/quantification of the uncertainties associated with the factors, the data and the model structure and their effect on the model output.

## 1.2 Modelling Compartmental Systems

In order to evaluate an experimental investigation concerning transport of material in a system of the real world, an appropriate model of the observed system is considered in the model world. The most common type of model used for that purpose is a compartmental model.

Compartmental models are used to approximate the systems of real world phenomena. They have been widely used to model systems in biomedicine, biology, pharmacokinetics, ecology, chemistry and engineering ([21], [38], [66], [82]). The books by Jacquez (1985), and by Godfrey (1983) (see [66], [38]) describe both the theory and the application of compartmental models. Articles by Brown (1980) and Zierler (1981) are excellent reviews of compartmental analysis. Brown lists around 140, and Zierler more than 50 further references on this topic (see [6] and [110]). According to Zierler, the origin of compartmental analysis may go as far back as 1822, when Fourier conjectured that heat flow may be proportional to a

temperature gradient. Godfrey, in his book, notes that because many first-order differential equations are compartmental without necessarily being described as such, this makes tracing the exact origins of compartmental models difficult. He also gives a date as early as 1923, when radioactive tracers were first applied to biological systems. Mulholland & Keener also trace the origins of the application of compartmental analysis to biological systems back to 1923 with the work of Hevesey, and to ecological systems back to the work of Kostitzin (1935) [88].

A compartmental system is a system divided into a finite number of subsystems called 'compartments', and it is assumed that each of these compartments is homogeneous and well-mixed. Furthermore, it is assumed that the 'material' under study is neither destroyed nor synthesized in any compartment. The compartments of a system interact with each other and with the environment by exchanging material.

The rate at which the quantity of material changes in the  $i$ th compartment can be written as the difference between the sum of all inputs into and the sum all outputs from that compartment as follows:

$$\frac{dx_i}{dt} \equiv \dot{x}_i = \left( q_{i0} + \sum_{\substack{j=1 \\ j \neq i}}^n q_{ij} \right) - \left( \sum_{\substack{j=1 \\ j \neq i}}^n q_{ji} + q_{0i} \right), \quad (1.1)$$

$$0 \leq t < \infty, \quad x_i(0) = x_i^0, \quad i = 1, 2, \dots, n$$

where  $x_i$  is the state variable associated with compartment  $i$ ;  $\dot{x}_i$  is the derivative of  $x_i$  with respect to time  $t$ ;  $x_i^0$  is the initial value of  $x_i$ ;  $q$  is the rate with which material is transferred and with the subscript  $ij$  read as 'into compartment  $i$  from compartment  $j$ ', and subscript 0 refers to outside the system. If there is no material leaving the system (i.e.,  $q_{0i} = 0, i = 1, 2, \dots, n$ ), then the system is said to be 'closed or blind', otherwise it is an 'open' system.

A general two-compartment model is shown in Figure 1.2.

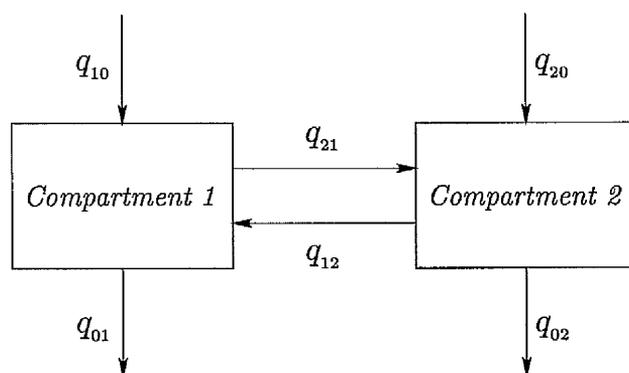


Figure 1.2. Two-compartment model diagram

Depending on the nature of the system being modelled, compartmental models can have different characteristics. However, the majority of the compartmental analysis literature is on linear, time-invariant compartmental systems [38]. In spite of the basic simplicity, there is no doubt that this type of model appears to be the most important and commonly considered model in practice as well as in the literature [77]. Hallstadius emphasizes that the most usual type of compartmental modelling in studies on the transport of material in physical or biological systems is linear [44]. In a paper by O'Neill, which is a review of linear compartmental analysis in ecosystem modelling, it is noted that linear time-invariant compartmental modelling has played an important role in the development of systems ecology since the inception of the area in the late 1950s, and linear compartmental models continue to find a significant number of applications in ecology [91]. In this review O'Neill also provides a long list of references available in the literature on this class of compartmental models. Because our attention in this thesis is confined to linear, time-invariant ecosystem models, applied to global carbon cycle models in particular, the dynamics of linear time-invariant compartmental systems is introduced in more detail in the next chapter.

Like any modelling techniques, in compartmental modelling the output from

the model is the item of interest. The knowledge we have about the model parameters, coefficients, and so on, is not perfect. As a result, there always exists some degree of uncertainty about the model inputs which is reflected in the model response(s). This, as it is expressed by Helton & Davis, leads us to two closely related questions: (1) ‘What is the uncertainty in the model response given the uncertainty in the model inputs?’, and (2) ‘How sensitive is the model response to the uncertainty in each model input?’ [48]. These two questions highlight the important role of ‘sensitivity’ and ‘uncertainty’ concepts in the modelling process very well.

### 1.3 Aims and Objectives of this thesis

The aims and objectives of this thesis are:

- (i) review and testing of different sensitivity analysis approaches,
- (ii) application/modification of sensitivity analysis techniques to compartmental models with steady-state constraint,
- (iii) presenting various graphical methods used as visualization tools to support sensitivity and uncertainty analyses,
- (iv) uncertainty analysis - model comparability and benchmarking,
- (v) validation in environmental radioactivity.

The previous sections have emphasised the relevance of sensitivity and uncertainty analyses as modelling tools. In particular, we illustrate and develop such methods as are required in a common applied class of models, namely radioactivity modelling. Such models have been used in the field of global carbon cycle modelling and climate change predictions.

## Chapter 2

# Methodology: Linear, Time-Invariant Compartmental Systems; Sensitivity and Uncertainty Analyses

### 2.1 Introduction

One of the commonest types of model used in the study of the transport of material in a system is a compartmental model. Even though compartmental models are often good approximations to many systems, no compartmental model is ever exact [66]. There are always discrepancies between the actual system and its mathematical model. The reasons for these discrepancies, noted by Frank [36], are as follows:

- restricted accuracy of the measuring devices or methods,
- the behaviour of any real system changes with time due to some natural, economic or man-made effects,
- mathematical models are often simplified or idealised intentionally so that

the mathematical problem can be made simpler or even soluble.

Because of these reasons the results of mathematical syntheses need not necessarily be practicable. The appropriateness of any mathematical model is judged by how closely it reproduces and predicts the observed behaviour of the actual system. The results produced by the mathematical model may not be satisfactory; they may even be very poor. For instance, there may be a considerable amount of parameter deviation between the actual system and the mathematical model; the solution may be very sensitive to the parameters; or the output may not lie in an acceptable bound. Therefore, sensitivity consideration of models and assessment of uncertainties become important issues. Since it is often unclear which one of the many model parameters is going to change and when this change will occur, it is important to develop techniques that enable us anticipate changes that might occur in the system due to changes in any model parameter. It is also essential to develop techniques that allow us to obtain certain specific bounds on model parameters in order to hold the model output within certain bounds. Hence, in order to respond to such important demands, sensitivity and uncertainty analysis techniques have been developed and are commonly employed.

One of the main objectives of sensitivity and uncertainty analyses is to cope with discrepancies between the actual system and its mathematical model, and ensure that the model is reliable.

Sensitivity analysis(SA) studies the effects of the variations in model parameters on the behaviour of the model. SA is typically applied to initial conditions, time-invariant or time-variant coefficients, sampling interval, sampling instant, characteristic frequencies, input frequency, temperature effect, delay, *etc.* [105]. As we have noted in Chapter 1, we use the terminology 'input factor' to refer to all these characteristic elements (parameters, variables, coefficients, *etc.*).

Uncertainty Analysis(UA) studies the effect of uncertainties inherent in model factors, model structure, scenarios, *etc.* on the model output and aims to quantify its uncertainty with a view to evaluating confidence and prediction ranges.

Both of these analyses play an important role in the overall study of modelling systems.

In this chapter, we next introduce and review the theory of linear, time-invariant compartmental models, then give a description of methods used to solve the model equations analytically and numerically. Sensitivity theory and calculation of the sensitivity matrix as applied to compartmental modelling is explained. We then introduce and identify the steps in performing a sampling based SA. This is followed by a description of the methodology of numerous SA methods. Finally we conclude the chapter with a critical review of the advantages and disadvantages of the different approaches.

## 2.2 Linear, Time-Invariant Compartmental System Theory

For compartmental systems the basic equations can be developed either in terms of concentrations or in terms of total amounts in each compartment<sup>a</sup>. Such quantities are called the *state variables* of the system and typically denoted by  $x_1, x_2, \dots, x_n$ . In Chapter 1 we have introduced the most general form of compartmental equations for a system consisting of  $n$  compartments as

$$\frac{dx_i}{dt} = q_{i0} + \sum_{\substack{j=1 \\ j \neq i}}^n q_{ij} - \sum_{\substack{j=1 \\ j \neq i}}^n q_{ji} - q_{0i}, \quad i = 1, 2, \dots, n. \quad (2.1)$$

Now, suppose we have a linear compartmental system modelled by an  $n$ -compartment model. Changes in such systems are represented by differential equations which describe how the rate of change of one state variable depends on the current values of each of the state variables. Hence, the equation for  $x_i$  might be written

---

<sup>a</sup>Sometimes quantity in each compartment considered in terms of concentration rather than absolute amount of material. In such cases, the concentration is defined as the amount of material in compartment  $i$  at time  $t$  divided by the volume of compartment  $i$ , i.e.  $c_i = x_i(t)/v_i$ .

as

$$\frac{dx_i}{dt} = f_i(x_1, x_2, \dots, x_n).$$

For a system being modelled using a linear compartmental model it is assumed that the amount of material transferred between compartments follows linear kinetics, so that at any time point the rate at which the material leaves a *source* (or *donor*) compartment is a linear function of the amount present in the compartment. Hence, the possible flow of material from compartment  $j$  to compartment  $i$  which is indicated by  $q_{ij}$  in Equation (2.1) is defined as

$$q_{ij} = k_{ij} x_j(t), \quad i = 0, 1, \dots, n, \quad j = 1, 2, \dots, n, \quad j \neq i$$

where  $k_{ij}$  is the proportionality parameter which characterizes the rate of transfer from compartment  $j$  to compartment  $i$ , and we use terminology '*transfer coefficient*' to refer to the  $k_{ij}$ 's. As the name implies, the  $k_{ij}$ 's are constants, they are not time dependent, and when the transfer of material can be described by a fixed set of transfer coefficients, the system is said to be *time-invariant*. Hence, the set of  $n$  differential equations, which represents an  $n$ -compartment model given as Equation (2.1) takes the form

$$\frac{dx_i}{dt} = \dot{x}_i(t) = \sum_{\substack{j=1 \\ j \neq i}}^n k_{ij} x_j(t) - \sum_{\substack{j=1 \\ j \neq i}}^n k_{ji} x_i(t) - k_{0i} x_i(t) + u_i(t), \quad (2.2)$$

$$0 \leq t < \infty, \quad x_i(t=0) = x_i^0, \quad i = 1, 2, \dots, n$$

where  $u_i(t)$  is used instead of  $q_{i0}$  to conform with the change of notation.

The state variable representation of a linear, time-invariant multicompartment model with  $n$  states,  $m$  inputs and  $p$  ( $p \leq n$ ) outputs in matrix notation is

described by

$$\begin{aligned}\dot{\mathbf{x}}(t) &= \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t), \\ \mathbf{x}(t_0) &= \mathbf{x}^\circ\end{aligned}\tag{2.3}$$

and the model output vector by

$$\mathbf{y}(t) = \mathbf{C}\mathbf{x}(t)\tag{2.4}$$

where  $\mathbf{x}$  is a  $n$ -dimensional vector and represents the state of the system with the initial state  $\mathbf{x}(t_0) = \mathbf{x}^\circ$ ;  $\mathbf{u}$  is a  $m$ -dimensional input vector;  $\mathbf{A}$ ,  $\mathbf{B}$  and  $\mathbf{C}$  are system matrices. These time-invariant matrices show distinct and particular characteristics.

The order  $n$  of matrix  $\mathbf{A}$  is basically the number of compartments, and its entries are defined as follows:

$$\begin{aligned}a_{ij} &= k_{ij}, \quad i \neq j \\ a_{ii} &= -\left(\sum_{\substack{j=1 \\ j \neq i}}^n k_{ji} + k_{0i}\right).\end{aligned}$$

Important characteristics of  $\mathbf{A}$  can be summarised as: (i) its elements are all real; (ii) the elements on the main diagonal are negative; (iii) the off-diagonal elements are positive, and (iv) the column sums are negative.

Matrix  $\mathbf{B}$  is of dimension  $n \times m$  where ( $m \leq n$ ). If the inputs enter compartments  $i_1, i_2, \dots, i_b$  say, the  $[b_{ij}]$  can take value 1 or 0 subject to the following condition:

$$\sum_{i=1}^n b_{ij} = 1 \quad \forall j; \quad \sum_{j=1}^m b_{ij} = \begin{cases} 1 & , i = i_1, \dots, i_b \\ 0 & , i \neq i_1, \dots, i_b. \end{cases}$$

Matrix  $\mathbf{C}$  is of dimension  $p \times n$  with ( $p \leq n$ ). If the outputs leave from compartments  $j_1, j_2, \dots, j_c$  say, the  $[c_{ij}]$  can take value 1 or 0 subject to the following

condition:

$$\sum_{j=1}^n c_{ij} = 1 \quad \forall i; \quad \sum_{j=1}^p c_{ij} = \begin{cases} 1 & , j = j_1, \dots, j_c \\ 0 & , j \neq j_1, \dots, j_c. \end{cases}$$

Equation 2.3 will be referred to as the *state equation* and Equation 2.4 as the *observation equation* of the system. Note that Equation 2.4 is simply an algebraic equation not a differential equation. Having defined a linear, time-invariant compartmental system, we know that  $\mathbf{A}$ ,  $\mathbf{B}$ ,  $\mathbf{C}$  matrices are time independent, and the state variables  $\dot{\mathbf{x}}$  and the input  $\mathbf{u}$  are time dependent. For the sake of simplicity in notation, we drop the time element ( $t$ ) from the state and observation equations.

Having defined the state equations of an  $n$ -compartment model, we now can introduce methods for solving these model equations analytically and numerically.

## 2.3 Analytical Solution of State Equations

There are various ways to solve the set of  $n$  state equations analytically. The book by Jacquez [66] provides a useful reference to solutions in many books and articles (for example, the book by Godfrey [38], the articles by Matis & Wherly [82], O'Neill [91], Zierler [110]). We will discuss two of these methods which are found to be more appropriate when linear compartmental models are concerned.

### 2.3.1 Classical Approach

The well-known general solution of equation (2.3) is given by

$$\mathbf{x} = e^{\mathbf{A}t} \mathbf{x}^o + \int_{t_o}^t e^{\mathbf{A}(t-\lambda)} \mathbf{B} \mathbf{u}(\lambda) d\lambda$$

where  $\lambda$  is a time variable ranging over the time interval over which the model input has been applied. This solution requires the calculation of an exponential

of the parameter matrix which can be a very time consuming task, and also the accuracy of this calculation can be questionable because the matrix  $e^{\mathbf{A}t}$  which is called the *transition matrix* is the sum of the infinite series

$$e^{\mathbf{A}t} = \mathbf{I} + \mathbf{A}t + \frac{(\mathbf{A}t)^2}{2!} + \frac{(\mathbf{A}t)^3}{3!} + \dots$$

This requires calculation of the powers of  $\mathbf{A}$  and there is uncertainty about how many terms need to be computed for acceptable accuracy [38].

### 2.3.2 Laplace Transformation Approach

Much of the literature on compartmental analysis and especially on linear systems uses a particular type of linear transformation called the Laplace transformation to solve the model equations [66]. In comparison to the general solution approach, the Laplace transformation method provides simpler and more accurate results [38].

Taking Laplace transforms on both sides of Equations (2.3) and (2.4), and assuming that the system is not empty at time  $t = 0$ , gives

$$s \cdot \mathbf{X}(s) - \mathbf{x}^\circ = \mathbf{A}\mathbf{X}(s) + \mathbf{B}\mathbf{U}(s) \quad (2.5)$$

$$\mathbf{Y}(s) = \mathbf{C}\mathbf{X}(s) \quad (2.6)$$

where  $\mathbf{X}$ ,  $\mathbf{Y}$  and  $\mathbf{U}$  are the Laplace transforms of  $\mathbf{x}$ ,  $\mathbf{y}$  and  $\mathbf{u}$ , respectively;  $s$  is the transformation domain variable;  $\mathbf{x}^\circ$  represents the state of the system immediately before any input enters the system.

Rearranging Equation (2.5) gives

$$\begin{aligned} (s\mathbf{I} - \mathbf{A})\mathbf{X}(s) &= \mathbf{x}^\circ + \mathbf{B}\mathbf{U}(s) \\ \mathbf{X}(s) &= (s\mathbf{I} - \mathbf{A})^{-1}\mathbf{x}^\circ + (s\mathbf{I} - \mathbf{A})^{-1}\mathbf{B}\mathbf{U}(s). \end{aligned} \quad (2.7)$$

Taking inverse Laplace transformation results in

$$\mathbf{x} = \mathcal{L}^{-1}[(s\mathbf{I} - \mathbf{A})^{-1}\mathbf{x}^{\circ}] + \mathcal{L}^{-1}[(s\mathbf{I} - \mathbf{A})^{-1}\mathbf{B}\mathbf{U}(s)] \quad (2.8)$$

which is then substituted into Equation (2.4) to obtain the model output vector  $\mathbf{y}$ . Alternatively, substituting Equation (2.7) into Equation (2.6) yields

$$\mathbf{Y}(s) = \mathbf{C}[(s\mathbf{I} - \mathbf{A})^{-1}\mathbf{x}^{\circ} + (s\mathbf{I} - \mathbf{A})^{-1}\mathbf{B}\mathbf{U}(s)], \quad (2.9)$$

and then the inverse transformation gives the following model output vector:

$$\mathbf{y} = \mathbf{C}\mathbf{x} = \mathbf{C}\{\mathcal{L}^{-1}[(s\mathbf{I} - \mathbf{A})^{-1}\mathbf{x}^{\circ}] + \mathcal{L}^{-1}[(s\mathbf{I} - \mathbf{A})^{-1}\mathbf{B}\mathbf{U}(s)]\}. \quad (2.10)$$

For more complicated models, using an analytical method to solve the state equations may be impractical and computationally very expensive, it may not even be possible. In such situations, numerical methods are preferred over analytical methods.

## 2.4 Numerical Solution of State Equations

Numerical methods are alternative approaches to solving differential equations and with the use of these methods we aim to obtain an accurate approximation to the solution of the state equations. The main idea of numerical methods is that starting with the initial condition, time is incremented in small steps and the changes in model output are calculated for each step. In general, the smaller the time increment, the more accurate the approximate solution becomes.

There is a large literature on the numerical solution of differential equations. The methods used to solve systems of first-order ordinary differential equations from initial conditions (also referred to as ‘first-order initial value problem’ in numerical analysis and differential equations terminology) can be found in every extensive numerical analysis and differential equations book, and routines which

can be used to carry out these standard numerical analysis procedures are available in the open literature. As Zwillinger comments in his book, numerical codes are available for solving nearly any type of ordinary differential equations, and one should use prepared software packages whenever possible [111].

The particular methods used in this thesis are the fourth-order Runge-Kutta method and Adams methods. Runge-Kutta methods are called single-step (or one-step) methods since they use only the information from the previous step. They have the ability to perform the next step with a different step size and are ideal for beginning the solution when only the initial conditions are available. Adams Methods, on the other hand, are multi-step methods because they make use of some of the information obtained at a few points beyond the initial point by utilizing the past values of the function and its derivative to construct an interpolating polynomial that approximates the derivative function, and extrapolate this into the next interval. Runge-Kutta method of order four which is the most popular Runge-Kutta method has been used to solve the state equations of the two 8-compartment global carbon cycle models, and Adams PECE method has been used in the computer implementation of the 25-compartment global carbon cycle model (see Chapter 3) used in this thesis.

In the following two sections we describe how these two methods are derived.

### 2.4.1 Runge-Kutta Method

Suppose we have a single compartment model with the state variable denoted by  $x$ , and its initial condition by  $x^\circ$  (for a multi-compartment model the solution can readily be obtained for each state equation). The mathematical model of the system is then given by the first-order differential equation

$$\dot{x}(t) = f(t, x(t)) \quad \text{with the initial condition } x(t_0) = x^\circ. \quad (2.11)$$

As with all numerical methods, the Runge-Kutta method also involves finding approximate solutions at  $t_0, t_1, \dots$ , where the difference between any two successive

$t$ -values is a constant,  $h$ ; that is,  $t_{n+1} - t_n = h$ , ( $n = 0, 1, 2, \dots$ ). The approximate solution at  $t_n$  will be designated by  $x(t_n)$ . Note that once  $x(t_n)$  is known, Equation (2.11) can be used to obtain  $\dot{x}(t_n)$  as

$$\dot{x}(t_n) = f(t_n, x(t_n)).$$

The classic fourth-order Runge-Kutta method (often referred to simply as *the* Runge-Kutta method) is based on a weighted average of values of  $f(t, x)$  at different points in the interval  $t_n \leq t \leq t_{n+1}$ . It is given by

$$x(t_{n+1}) = x(t_n) + \frac{h}{6} (F_1 + 2 F_2 + 2 F_3 + F_4),$$

where

$$\begin{aligned} F_1 &= f(t_n, x(t_n)) \\ F_2 &= f(t_n + \frac{1}{2}h, x(t_n) + \frac{1}{2} h F_1) \\ F_3 &= f(t_n + \frac{1}{2}h, x(t_n) + \frac{1}{2} h F_2) \\ F_4 &= f(t_n + h, x(t_n) + h F_3). \end{aligned}$$

Different orders of Runge-Kutta methods can be found in many differential equations and numerical analysis books including [18], [86] and [111]. FORTRAN code for the fourth-order Runge-Kutta algorithm is given in [94].

### 2.4.2 Adams Methods

Any solution of the initial value differential equation given in Equation (2.11) can be written as

$$x(t_{n+1}) - x(t_n) = \int_{t_n}^{t_{n+1}} \dot{x}(t) dt = \int_{t_n}^{t_{n+1}} f(t, x(t)) dt.$$

The main idea of the Adams methods is to approximate this solution by replacing  $f(t, x(t))$  with a polynomial interpolating to computed derivative values,  $f_i$ , and

then integrating the polynomial. Note that we denote  $f(t_i, x(t_i))$  by  $f_i$  for an integer  $i$ .

The Adams-Bashforth formula of order  $c$  at  $t_n$  uses a polynomial  $P_{c,n}(t)$  interpolating the computed derivatives at the  $c$  preceding points, i.e.

$$P_{c,n}(t_{n+1-j}) = f_{n+1-j}, \quad j = 1, 2, \dots, c.$$

These derivatives and  $x(t_n)$  have to be stored from the preceding step. Then, an approximation to the solution at  $t_{n+1}$  is obtained as

$$x(t_{n+1}) = x(t_n) + \int_{t_n}^{t_{n+1}} P_{c,n}(t) dt. \quad (2.12)$$

In more general terms, an approximation can be obtained for all  $t$  near  $t_n$  using

$$x(t) \approx x(t_n) + \int_{t_n}^t P_{c,n}(t) dt.$$

Several ways of representing the interpolating polynomial,  $P_{c,n}(t)$ , given in the above equations exist. Here we only give the Lagrangian form, and the other forms of  $P_{c,n}(t)$  (for example, divided difference form) can be found in Ref. [101]. Lagrange's form of the interpolating polynomial is as follows

$$P_{c,n}(t) = \sum_{i=1}^c l_i(t) f_{n+1-i}$$

where

$$l_i(t) = \prod_{\substack{j=1 \\ j \neq i}}^c \frac{t - t_{n+1-j}}{t_{n+1-i} - t_{n+1-j}}, \quad i = 1, 2, \dots, c.$$

Substituting into Equation (2.12) yields the following Adams-Bashforth formula

$$x(t_{n+1}) = x(t_n) + \sum_{i=1}^c f_{n+1-i} \int_{t_n}^{t_{n+1}} l_i(t) dt.$$

which is usually written as

$$x(t_{n+1}) = x(t_n) + h \sum_{i=1}^c \alpha_{c,i} f_{n+1-i}, \quad (2.13)$$

where  $h$  is a constant step size and

$$\alpha_{c,i} = \frac{1}{h} \int_{t_n}^{t_{n+1}} l_i(t) dt$$

A variation on the derivation of the Adams-Bashforth formulae gives another set of formulae called the Adams-Moulton formulae. These formulae of order  $c$  at  $t_n$  use a polynomial  $P'_{c,n}(t)$  that also interpolates to  $c$  derivative values given below:

$$\begin{aligned} P'_{c,n}(t_{n+1-j}) &= f_{n+1-j}, \quad j = 1, 2, \dots, c-1, \\ P'_{c,n}(t_{n+1}) &= f(t_{n+1}, x(t_{n+1})). \end{aligned}$$

Hence, the approximate solution is given by

$$x(t_{n+1}) = x(t_n) + \int_{t_n}^{t_{n+1}} P'_{c,n}(t) dt. \quad (2.14)$$

By following the same procedure as in the Adams-Bashforth method, one can find the Lagrangian form of the Adams-Moulton equation as

$$x(t_{n+1}) = x(t_n) + h \sum_{i=0}^{c-1} \alpha'_{c,i} f_{n+1-i} \quad (2.15)$$

where

$$\alpha'_{c,i} = \frac{1}{h} \int_{t_n}^{t_{n+1}} l'_i(t) dt$$

with

$$l'_i(t) = \prod_{\substack{j=0 \\ j \neq i}}^{c-1} \frac{t - t_{n+1-j}}{t_{n+1-i} - t_{n+1-j}}.$$

As noted by Boyce & DiPrima [5] the Adams-Bashforth formulae are explicit and faster, and the Adams-Moulton formulae on the other hand are more accurate but implicit and slower. To achieve both simplicity and accuracy, numerical analysts have combined the two types of formulae in the so-called *Predictor-Corrector Method*. Once  $x(t_n), x(t_{n-1}), \dots$  are known we can calculate  $f_n, f_{n-1}, \dots$ , and then use the Adams-Bashforth (predictor) formula (2.13) to obtain a value for  $x(t_{n+1})$ . Then  $f_{n+1}$  is computed and using the Adams-Moulton (corrector) formula (2.15) an improved value of  $x(t_{n+1})$  is obtained. This predictor-corrector procedure is called Adams PECE method. The acronym *PECE* is derived from the description of how the computation is done. This is described above but in summary: we *Predict*  $x(t_{n+1})$ , *Evaluate*  $f_{n+1}$ , *Correct* to get  $x(t_{n+1})$ , and *Evaluate*  $f_{n+1}$  to complete the step.

The book by Shampine & Gordon [101] completely explain Adams methods (see Ch. 3), and they also provide computer codes (written in FORTRAN) for these methods.

Before proceeding further, in the context of this thesis, the matrix  $\mathbf{C}$  of the observation equations (see Equation (2.4)) is considered to be the  $p \times n$  dimension identity matrix  $I$ . In other words, the model output from the  $i$ th compartment is taken to be the amount of material present in that compartment ( $y_i = x_i$ ). Taking that into account, in the following sections the notation  $\mathbf{x}$  ( $n$  vector of  $x_i$ 's) is dropped and the notation  $\mathbf{y}$  ( $n$  vector of  $y_i$ 's) used to denote both compartmental

contents and compartmental outputs, and referred to simply as the vector of model outputs or (in statistical terminology) the vector of response variables.

## 2.5 Sensitivity Analysis

The basic idea of sensitivity is to calculate the change in the system behavior due to the factor variations.

As suggested by Saltelli [97] and Campolongo *et al.* [14] sensitivity analysis techniques may be grouped into the following three main categories:

**Local SA methods** which concentrate on the local impact of the model input factors on the model response;

**Global SA methods** which concentrate mainly on apportioning the model output uncertainty to the uncertainty in the input factors; and

**Screening methods** which are used to identify the most influential factors on the model output.

In the following sections we describe the methodology of various SA techniques, but the above classification is further extended for the purpose of giving an ordered presentation of the techniques. First, graphical methods for sensitivity and uncertainty analysis are presented. Second, a differential analysis based method and a number of numerical methods for the calculation of local sensitivities are described. Then, before including the methodology of global SA methods some fundamental elements of these methods, design of the computer experiment for instance, is introduced. Finally, the factor screening procedure is explained.

### 2.5.1 Graphical Sensitivity and Uncertainty Methods

A literature search had recently been done by Cooke & van Noortwijk [23] which shows very little theoretical development for graphical methods in sensitivity and uncertainty analysis. There exist reference books in the literature, for example

[20] and [28], which study visualizing univariate, and multivariate data. As Cooke & van Noortwijk mention, when graphical methods are used in sensitivity and uncertainty analysis the focus is not visualizing data as it is in the general sense, but rather visualization to support SA and UA.

A number of graphical tools which may have an application in sensitivity and uncertainty analysis techniques have been presented below.

#### **2.5.1.1 Scatter Plots**

A useful non-quantitative screening technique is a sequence of scatterplots in which each response variable (model prediction) appear on the vertical axis and each explanatory variable (model input factor) appear in turn on the horizontal axis. In SA and UA, scatterplots of the input-output relationships are used as a guide to better understanding of the model behaviour. If the relationship is strong this indicates that the considered model input has significant effect on the model output.

A scatterplot matrix (often called matrix plot) can also be produced. This type of plot displays the main features of the 2D relationships between each pair of variables without reference to the other variables. However as the number of variables increases, it becomes harder to interpret the set of plots and obtain an overall sense of the data configuration.

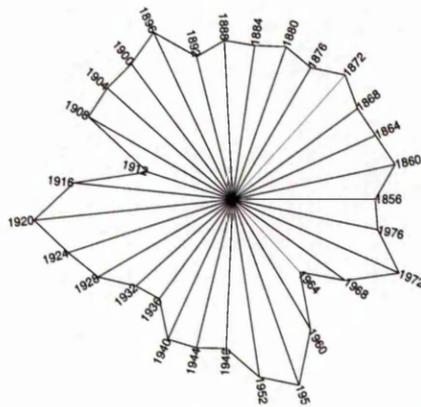
There is no doubt that generating scatterplots is the simplest SA method. One disadvantage of this technique is that it requires drawing and inspecting a large number of plots, at least one plot for each model input factor. Considering that we may need to analyse several model outputs which may also be time dependent, then the number of plots we need to generate becomes quite large.

#### **2.5.1.2 Star Plots**

Star plots are used for representing multivariate data in two dimensions. Star representations of the data can be obtained by utilising the 'stars' function of S-PLUS programme package. Stars represent the several measurements of a case on

equally spaced radii extending from the centre of a circle. Then the measurements are linked to form a star. The values of the measurements occupy a circle, and the fact that the starting points correspond to the end points facilitates comparison between cases. We use star plots to visualize different sensitivity measurements of the model response(s) to the uncertainties about the model inputs. Figure 2.1 shows an example of a single star plot.

However if the number of cases presented in a star plot is very large, star representations may not provide a clear illustration.



**Figure 2.1.** An example of a star plot showing Republican votes for New York between 1856 and 1976. (This plot is produced using the dataset `votes.repub` available in S-PLUS).

### 2.5.1.3 Pie Charts

In a pie chart, each of the values in the range is represented by a slice of the pie. A pie chart is used to compare parts to the whole. The wedges of the pie are labeled and often coloured. For single diagrams the use of pie charts is desirable, but when the comparison involves different times or scenarios, the change in size of the pie charts (radius proportional to the square root of total amount) may not be easily comprehended. In such cases, bar charts may be more preferable since the change in the length of the bar charts, which is directly proportional to total amount, may be more easily visualized.

#### 2.5.1.4 Other Graphical Methods

There are other graphical techniques, such as tornado graphs, radar plots and cobweb plots used for uncertainty and sensitivity analysis, which are more complex and less familiar than other techniques discussed above. They are described in detail by Cooke & van Noortwijk in Part II-11 of Ref. [98]. Here we give a brief description of these methods and refer the interested reader to Ref. [98].

Tornado graphs are basically bar graphs of any global sensitivity measures (for example, rank correlation coefficients) arranged in decreasing order of absolute value.

Radar plots are similar to star plots. Each variable is represented by a ray in a radar plot, it shows x-axis values as imaginary lines radiating from a common centre, like 'spokes of a wheel'. y-axis values are plotted on each of the spokes. The variable with the highest sensitivity measure is plotted furthest from the centre and the variable with the lowest sensitivity measure is plotted closest to the centre. Radar graphs can highlight trends, depending on the shapes drawn by the plot lines.

Cobweb plots are used in identifying local probabilistically important factors. These plots give a picture of the joint distribution of the percentiles of up to 20 variables. Each parallel vertical line in a cobweb plot represent one variable, and the possible values of these variables are given on these lines. Then each set of values are marked on the vertical lines and connected by a jagged line.

Other than the graphical methods described above, histograms, bar charts, dot plots are also used for sensitivity and uncertainty analysis.

### 2.5.2 Local SA Methods

Local SA methods focus on input factors which are varied within a small interval around a specified value. Local sensitivities provide the slope of the model output in the input space for a given set of factors [107]. Local sensitivities are exactly defined, and they depend neither on chosen distributions for input factors nor

any choice of a calculation method.

The evaluation of the sensitivity function and the sensitivity matrix (this method is often called ‘differential analysis’), and the use of numerical methods for the calculation of local sensitivities are discussed in the following subsections. A recent review written by Turányi & Rabitz [107] discusses both analytical and numerical local SA methods and their applications. They also provide further references to local methods. In this review, they point out several advantages and disadvantages of local sensitivities which include:

- when dealing with large models, local sensitivities can provide useful information on the performance of the model near the nominal values of model factors, but calculation of global sensitivities, which are based on studying the model in a wide range of factors, is computationally prohibitive;
- calculation of local sensitivities are much faster than that of global sensitivities;
- when the uncertainty about the model factors is very high, local sensitivities are totally incapable of providing information on the effect of significant factor changes, i.e. changing factors within a wide uncertainty range can give a qualitatively different model and this may result in a completely different sensitivity pattern; and
- local sensitivities are really local, and the information they provide is related to a single point in the input space.

### 2.5.2.1 The Sensitivity Function / Matrix

The sensitivity function is one of the main concepts in sensitivity theory. If the sensitivity function is known, then it is simple to calculate the change in model behaviour with respect to a given model factor. The sensitivity function is obtained by solving differential equation (2.3), evaluating the algebraic observation

function (2.4) and taking the partial derivative of the solution with respect to the factor which is then evaluated at a given time and condition.

Consider a single compartment, multi-factor model whose output function is  $y = y(t; \boldsymbol{\theta})$ . Here  $\boldsymbol{\theta}$  is the  $k$ -vector of model factors, and associated with a reference value (often called *nominal value*)  $\boldsymbol{\theta}^\circ$ , and the corresponding output function has a nominal value  $y^\circ = y(t; \boldsymbol{\theta}^\circ)$  which is often referred to as 'base-case' or 'base-case scenario'. Now, suppose that  $\boldsymbol{\theta}$  changes from  $\boldsymbol{\theta}^\circ$  to  $\boldsymbol{\theta}^\circ + \Delta\boldsymbol{\theta}$ . Using sensitivity analysis we aim to find out what happens to  $y$  when  $\boldsymbol{\theta}^\circ \rightarrow \boldsymbol{\theta}^\circ + \Delta\boldsymbol{\theta}$ . The first derivative of  $y$  with respect to the factor  $\theta_i$ ,  $\partial y / \partial \theta_i$ , provides a measure of sensitivity, that is, it is a local sensitivity index measuring the effect on  $y$  of perturbing  $\theta_i$  around nominal value  $\boldsymbol{\theta}^\circ$ . We can expand  $y(t; \boldsymbol{\theta}^\circ + \Delta\boldsymbol{\theta})$  around  $\boldsymbol{\theta}^\circ$  by a Taylor series expansion as follows:

$$y \approx y^\circ + \sum_i^k \left. \frac{\partial y}{\partial \theta_i} \right|_{\boldsymbol{\theta}^\circ} \Delta\theta_i + \frac{1}{2} \sum_j^k \sum_i^k \left. \frac{\partial^2 y}{\partial \theta_j \partial \theta_i} \right|_{\boldsymbol{\theta}^\circ} \Delta\theta_j \Delta\theta_i + \dots \quad (2.16)$$

The expression in Equation (2.16) is typically truncated after the first or second order derivatives. In general, the first and second order derivatives are of more interest to the system or process investigators than any other higher order derivatives [62]. The partial derivatives  $\partial y / \partial \theta_i$  are called first-order local sensitivities, and  $\partial^2 y / \partial \theta_j \partial \theta_i$  second-order local sensitivities.

Now, suppose we have a multi-output, multi-input model which is represented by the following system of time-dependent ordinary differential equations

$$\frac{d\mathbf{y}}{dt} = \mathbf{f}(\mathbf{y}, t, \boldsymbol{\theta}) \quad (2.17)$$

where  $\mathbf{y}$  may be thought of as an  $n$ -vector ( $y_i$ ) and  $\boldsymbol{\theta}$  as a vector of  $k$  model input factors. The solutions of Equation (2.17) may be thought of as functions of two variables,  $t$  and  $\boldsymbol{\theta}$ ; that is,  $\mathbf{y}(t, \boldsymbol{\theta})$ . The initial conditions are also treated henceforth as factors,  $\boldsymbol{\theta}$ , in  $\mathbf{y}(t, \boldsymbol{\theta})$ .

Having defined the model equations, we now want to calculate the first-order local sensitivity of the  $i$ th component of the model response  $\mathbf{y}$ ,  $y_i$ , in terms of the

$r$ th component of the model factor  $\theta$ ,  $\theta_r$ .  $\partial y_i / \partial \theta_r$  gives the sensitivity function of  $y_i$  in terms of  $\theta_r$  which is then evaluated at a given nominal condition  $\theta^\circ$  and at a given time  $t$ .

In the case where there are  $n$  model responses  $\mathbf{y} = [y_1, y_2, \dots, y_n]^T$  and  $k$  factors  $\theta = [\theta_1, \theta_2, \dots, \theta_k]^T$ , the first-order local sensitivities form the  $n \times k$  dimensional sensitivity matrix  $\mathbf{S}$  as given below:

$$\mathbf{S} = \begin{bmatrix} \frac{\partial y_1}{\partial \theta_1} & \frac{\partial y_1}{\partial \theta_2} & \dots & \frac{\partial y_1}{\partial \theta_k} \\ \vdots & \vdots & & \vdots \\ \frac{\partial y_n}{\partial \theta_1} & \frac{\partial y_n}{\partial \theta_2} & \dots & \frac{\partial y_n}{\partial \theta_k} \end{bmatrix} \quad (2.18)$$

Each column of matrix  $\mathbf{S}$  is called the vector-sensitivity function of  $\mathbf{y}$  with respect to  $\theta_j$ , ( $j = 1, \dots, k$ ).

Differential analysis techniques have been widely used in uncertainty and sensitivity analyses and introductory theory can be found in Refs. [34], [36], [107], [105]. Examples of the use of differential sensitivity analysis are given by Iman & Helton [63], and Hamby [46]. As noted by Iman & Helton [63] the Taylor series approximation given in Equation (2.16) is the starting point for uncertainty and sensitivity analysis techniques based on differentiation. The first step in such an analysis is generating the partial derivatives required in the series. When the model function given in Equation (2.16) is a relatively simple function it may be possible to generate the required derivatives analytically or by simple differencing schemes, but more complicated models often require complex numerical procedures.

### 2.5.2.2 Numerical Methods for Calculating Local Sensitivities

When the analytical solution of the model equations given in Equation (2.17) is known, then sensitivity functions may easily be found by direct differentiation. In some cases, however, it is easier to solve the sensitivity equations directly by using numerical methods, such as indirect method, direct method, the green function

method, *etc.*. Here we give a brief discription of these three most commonly employed methods.

**Indirect Method** also called *brute force method* or *finite difference approximation* is the simplest numerical method to calculate local sensitivities. It is based on changing one input factor at a time and rerunning the model code with the new set of input factors. Then, the elements of the sensitivity matrix are approximated by the following equation:

$$\frac{\partial \mathbf{y}}{\partial \theta_j} \approx \frac{\mathbf{y}(\theta_j + \Delta\theta_j) - \mathbf{y}(\theta_j)}{\Delta\theta_j} \quad j = 1, \dots, k$$

where  $\Delta\theta_j$  denotes the change in factor  $\theta_j$ . For the calculation of local sensitivities using this method  $k + 1$  model runs are required, but if central differences are used then the number of simulations required is  $2k$ . The main advantage of this method is that it is easy to implement and no extensive modification to the original model code is needed. On the other hand, compared to more advanced methods, this method is slower and less accurate [107].

**Direct Method** In the direct method, in addition to the system of  $n$  differential equations, for each factor  $n$  additional differential equations which describe the sensitivity of the original system with respect to the chosen factor are defined. This second set of equations called sensitivity differential equations are obtained by differentiating both sides of Equation (2.17) with respect to a chosen factor, say  $\theta_r$ :

$$\frac{d}{dt} \frac{\partial \mathbf{y}}{\partial \theta_r} = \frac{\partial \mathbf{f}}{\partial \theta_r} + \mathbf{J} \frac{\partial \mathbf{y}}{\partial \theta_r} \quad (2.19)$$

or in matrix notation

$$\dot{\mathbf{S}} = \mathbf{F} + \mathbf{J} \mathbf{S}$$

where  $\mathbf{F}$  is a vector of length  $n$  whose components are  $\partial f_i / \partial \theta_r$ , and  $\mathbf{J}$  is

an  $n \times n$  matrix recognized as Jacobian and its elements are  $\partial f_i / \partial y_l$  with  $i = 1, 2, \dots, n; l = 1, 2, \dots, n$ .

Direct methods are based on numerical solution of Equation (2.19) which requires knowledge of the values of matrices  $\mathbf{F}$  and  $\mathbf{J}$  at each step of the solution. In order to evaluate these values, we have to know the actual values of system variables. This can be achieved by a simultaneous solution of Equation (2.17). Turányi & Rabitz [107] note that in the early realizations of the direct method, the system of original model equations and the system of sensitivity equations, i.e. Equations (2.17) and (2.19), were solved independently but simultaneously and the solution of Equation (2.17) was used to set up Equation (2.19). They also note that direct methods based on this algorithm were relatively slow.

Another algorithm called *decoupled direct method (DDM)*, which is based on a special relation between Equations (2.17) and (2.19) that allows a numerical shortcut, was first introduced by Dunker [29]. Dunker shows that Equation (2.17) and Equation (2.19) have the same Jacobian matrix and therefore the spectrum of time steps on which the elements of  $\dot{\mathbf{y}}$  change will be the same as the spectrum of time steps on which the elements of  $\mathbf{y}$  change. In Dunker's method the matrix  $\mathbf{J}$  is evaluated only once, and then at each time-step Equation (2.17) is solved and then Equation (2.19) is solved with all factors one after the other. Because the evaluation of the Jacobian matrix, which is the most time consuming part of solution, is reduced considerably with this method, computational cost for calculating sensitivities using the *DDM* method is relatively low. According to Dunker, the *DDM* method is a very efficient form of direct method especially for models based on complicated equations [29].

**Green's Function Method** In the Green's function method, we first differentiate Equation (2.17) with respect to the initial values  $\mathbf{y}^\circ$  which gives

$$\frac{d}{dt} \mathbf{K}(t, t_1) = \mathbf{J}(t) \mathbf{K}(t, t_1), \quad (2.20)$$

where  $t$  and  $t_1$  are the observation and perturbation times, respectively, and  $\mathbf{K}$  is the initial value sensitivity matrix with  $\mathbf{K}(t, t_1) = \partial c_i(t) / \partial c_j^0(t_1)$  and  $\mathbf{K}(t_1, t_1) = \mathbf{I}$ ,  $t \geq t_1$ . The linear non-homogeneous system of differential equations given in Equation (2.19) can be solved by first determining the homogeneous part given in Equation (2.20) and then calculating the following particular solution:

$$\mathbf{S}(t_1, t_2) = \int_{t_1}^{t_2} \mathbf{K}(t_2, s) \mathbf{F}(s) ds. \quad (2.21)$$

Here  $\mathbf{K}$  is called the *Green's function* (or *kernel*), and the method based on the solution of Equation (2.21) is known as *Green's function method* [107]. Dougherty *et al.* have applied the Green's function method to atmospheric chemical reaction models and compared this method with the direct method for the same type of models. They concluded that Green's function method should be more efficient than the direct method if the number of factors of interest is greater than the number of variables [25].

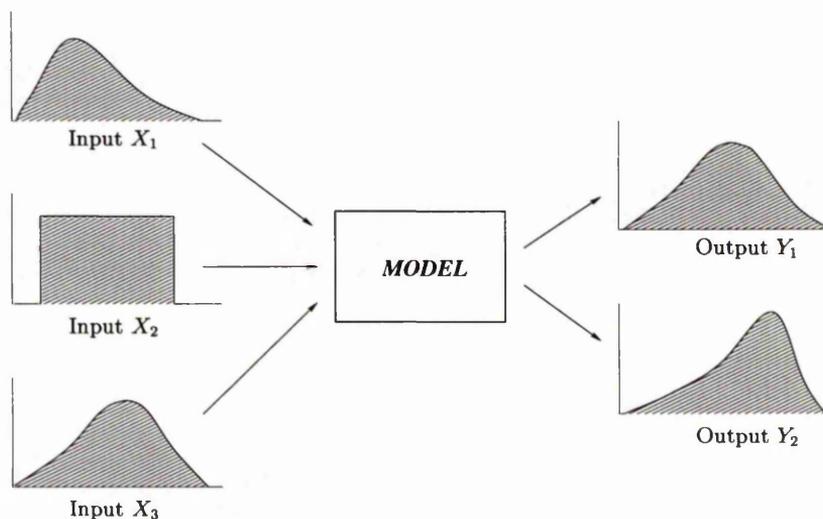
When the model under investigation is a linear and relatively simple model, local sensitivities can be obtained easily, and can provide information about the effect of changes in the input factors on the model output(s). On the other hand, if the model is known to be non-linear, and the input factors are affected by large uncertainties, the appropriateness of local methods is questionable, and in such cases, as Campolongo *et al.* [14] suggest, a global method should be used instead. Global SA methods are discussed in Section 2.5.3. As complex large models are increasingly applied for problem solving in many areas which include atmospheric science, engineering, biological systems, the need for global sensitivity analysis methods are becoming increasingly apparent.

### 2.5.3 Sampling based Global SA Methods

The procedures followed for global SA of large computational models which are described in the following sections may be summarized as a two-step procedures; the generation or sampling of input factor values, and the statistical analysis of the resulting output(s).

The SA and UA methods based on sampling are widely used in the analysis of large complex computational models, and they involve generation and exploration of mappings from model inputs to model predictions. In these sampling-based SA and UA methods after the values for input factors are obtained using a chosen sampling scheme, the model is run for a specified number of times  $N$ , and one or more output variables for each run are recorded. This input-output process is illustrated in Figure 2.2 for a hypothetical 3-inputs, 2-outputs model.

Following this process, appropriate statistical analyses are performed on the output variables as dependent variables and the input factors as independent variables to assess the input-output relationship and the effect of the uncertainties in inputs on the output variables.



**Figure 2.2.** A hypothetical model with three input factors ( $X_1, X_2, X_3$ ) and two output variables ( $Y_1, Y_2$ ).

These processes summarized above are broken into more steps in the following section to give a clearer and more detailed explanation of these essential steps.

## 2.5.4 Steps in Performing Sensitivity and Uncertainty Analyses

The basic steps in conducting a SA process can be identified as follows:

1. defining the model and determining which of the inputs and outputs of the model need to be included in the sensitivity analysis, in other words designing the experiment,
2. assigning probability density functions and variability ranges to each of the identified factors,
3. generating an input matrix using an appropriate sampling technique,
4. evaluating the model and obtaining the model response(s), and
5. analysing the relationship between the induced output distribution and the input sets to assess the effect and relative importance of each input factor on the output(s) of the model.

In Ref. [98] Saltelli provides a diagram to illustrate these steps (see Figure 1.2 in given reference).

### 2.5.4.1 Designing the Computer Experiment

As the first step, the model investigator defines the model, its response (model output(s)) and explanatory (input factors) variables. This step also involves deciding which factors are to be used in the analysis and which output variable(s) are to be considered. The selection of inputs at which to carry out a physical experiment or to run a computer implemented model code is a design of experiment problem. According to Campolongo & Saltelli [13], design of experiment can be considered as one of the forefathers of sensitivity analysis.

In a computer experiment, input(s) are obtained by a sampling procedure which can be done either systematically (factorial design, fractional factorial design) or randomly (Monte Carlo simulation). In this section we give general descriptions of factorial and fractional factorial designs, and three types of random sampling are described in Section 2.5.4.3.

Factorial design can be used to obtain a given number of samples for each input factor and running the model code for all combinations of the samples. The model output obtained in this fashion is then used to estimate the effects of each factor on the model response. When input factors are considered at several values, the number of factors play an important role. If we want to investigate  $k$  factors, and factor  $i$  ( $i = 1, 2, \dots, k$ ) has  $\ell_i$  levels, then the number of possible combinations is  $\ell_1 \times \ell_2 \times \dots \times \ell_k$ . Even for a small number of factors this number of combinations can be quite high. For instance, if we have only  $k = 10$  factors and each factor has a minimum number of levels, i.e.  $\ell_i = 2$  (high and low) then the number of combinations we have is  $2^{10} = 1024$ . As the number of input factors increases, the number of factor combinations rapidly increases. In such cases, fractional factorial design proves to be a useful alternative.

Fractional factorial design is based on the idea of assuming some of the higher order interactions are unimportant. Again, considering the simplest case, two levels for each of the  $k$  factors, a fractional factorial design involves running the model only a fraction ( $1/2$  to some power  $s$ ) of the total possible runs  $2^{k-s}$ .

#### 2.5.4.2 Assigning *pdf*'s to each input factor

In order to find the effect of a factor on the model output, this factor has to be varied from a specified probability distribution, or considered at several levels within a known range.

As Haimes & Lambert [42] point out the question 'How can we specify, generate, and use more appropriate probability distributions for model input?' still remains important for the scientific community especially from the risk assessment point of view where examinations of health and environmental effects, economic

impacts, energy impacts, technical feasibility *etc.* are concerned.

In an article, Hoffman & Kaplan [55] suggest how probability assessments can best be representations of the full state of knowledge of uncertain inputs. They provide three approaches for obtaining distributions for uncertain inputs, and the classification of these methods is based on the availability of the resources. These three approaches and their descriptions given by the authors are as follows:

- *Classical statistical method.* This is appropriate if there is a large data set obtained using an appropriate sampling scheme and the data are sufficiently relevant for addressing the assessment endpoint. Then, using statistical methods, empirical distributions are obtained. With this method, two analysts using the same data sets and the same statistical techniques should produce the same results.
- *Analyst judgement using all sources of information.* This approach is taken when there is no data available or data exist but it is only partially relevant to the assessment endpoint. This is the method most commonly employed. In this case, summarizing the state of knowledge and specifying the subjective probability distribution depend on the analyst and his fellow reviewers. Because individual interpretation of the evidence is involved in this approach, it is very likely that two different analysts given the same assessment will describe the present knowledge differently and produce two different probability distributions.
- *Formal expert elicitation.* In this approach the analyst is required to identify the individuals known as “experts” in the area of concern, and bring these experts together. After the analyst describes the assessment problem to the experts, and provides them with all relevant information and data, the experts are asked to formalize and document their rationales. The experts are then interviewed and asked to defend their rationales before committing any specific probability distribution. The experts specify their own subjective probability distributions by estimating quantiles. This

method is considered to be a strong approach, but time and finance-wise it can be very expensive. With this approach it is also likely to get different answers from two or more independent analysts or two or more different expert groups involved in the assessment.

Because of the lack of information about the distributions followed by the model input factors, in many works related to the sensitivity analysis experiments input factors are assumed to be uniformly/loguniformly distributed. For example, Campolongo & Saltelli consider a uniform distribution for all 32 model input factors of a well known model for the production of a key sulphur bearing compound from algal biota [12]. Another example, in the application of several SA techniques to the MACCS model of the early health effect associated with a severe accident at a nuclear power station, Helton *et al.* [52] assign uniform and loguniform distributions to the 34 imprecisely known input factors. White [108] suggests that in the case of minimum knowledge a uniform distribution over the maximum conceivable range can be considered for the input factors. Campolongo *et al.* [15] also note that when the knowledge of the input factors is quite poor, assuming a uniform distribution for each input factor is acceptable.

As for the ranges of variability, if no such information exists in the literature, then different criteria are adopted to obtain variability ranges for the model factors. For instance, Campolongo & Saltelli [12] use  $\mp 20\%$  of the nominal value to derive a range of variability for some of the model input factors they consider in their work. In the same article they also mention another criterion for calculating variability ranges, which is  $(1/2K_0; 2K_0)$ , where  $K_0$  is described as the nominal value of the input factor. In this study, the former criterion is used in the analysis where no reference values for the variability ranges are available. In another article by Reed *et al.* the nominal values of the 15 model parameters are taken as the mean values, and by assuming 10% coefficient of variation about those mean values they calculated the standard deviations for the parameters [95].

Different distributional assumptions on the model input factors may have a (significant) effect on the distribution of the model predictions, and therefore on

the outcomes of the uncertainty and sensitivity analyses. The effect of changing the distribution of the input factors on the model predictions will be discussed in Chapter 5.

After assigning a range and an appropriate probability distribution to each input factor we then need to generate a sample. In the following section three ways of producing input samples are described.

### 2.5.4.3 Generating the design matrix

In the sensitivity and uncertainty analyses framework, simple random sampling, Latin hypercube sampling and importance sampling appear to be the most commonly used sampling techniques used to generate the input matrix. The purpose of all sampling techniques is the same: to obtain a better coverage of the sample space of the input factors. Comparison of these sampling methods, and their effect on the results have been discussed by Helton & Davis in [49].

#### 2.5.4.3.1 Simple Random Sampling

Simple random sampling (SRS), sometimes also called random sampling, is the simplest and most widely used random sampling method. In SRS each member of the population has an equal probability of being included in the sample, and each selection is independent of the previous drawings. By applying this procedure to each of the  $k$  input factors the following random sample input matrix is obtained:

$$\mathbf{X} = \begin{bmatrix} X_{11} & X_{12} & \cdots & X_{1k} \\ X_{21} & X_{22} & \cdots & X_{2k} \\ \vdots & \vdots & & \vdots \\ X_{N1} & X_{N2} & \cdots & X_{Nk} \end{bmatrix}$$

where  $N$  is the sample size and  $k$  is the number of input factors.

This sampling technique gives unbiased estimates of the means, variances, and

distribution functions of the output variables. Therefore, whenever the computer resources are not limited, random sampling is the best technique to use [56].

For notational convenience, we shall use

$$\mathbf{X} = [X_1, X_2, \dots, X_k]$$

to represent the model input factors under consideration, and

$$\mathbf{X}_i = [X_{i1}, X_{i2}, \dots, X_{ik}], \quad i = 1, 2, \dots, N$$

to represent the observations.

#### 2.5.4.3.2 Importance Sampling

Unlike simple random sampling, importance sampling can assure that sampled values do not fall very close together in the sample space. In importance sampling the sample space, say  $\Omega$ , is divided into a number of subregions (often called *strata*)  $\Omega_j$ ,  $j = 1, 2, \dots, l$  which typically have unequal probabilities, and assure inclusion of specific regions of sample space in the analysis. After dividing the sample space into strata, we then sample  $l_j$  values for  $\mathbf{X}$  from strata  $\Omega_j$  using random sampling. Hence, the following vectors form a sample obtained using importance sampling

$$\mathbf{X}_i = [X_{i1}, X_{i2}, \dots, X_{ik}], \quad i = 1, 2, \dots, \sum_{j=1}^l l_j.$$

If one value is sampled from each strata, then the sample has the form

$$\mathbf{X}_i = [X_{i1}, X_{i2}, \dots, X_{ik}], \quad i = 1, 2, \dots, N.$$

In importance sampling, the partition of the sample space is based on how important the  $\mathbf{X}$ 's contained in each set are to the final result of the analysis. This sampling technique is often used to assure the inclusion in an analysis of

$\mathbf{X}$ 's that have high consequences but low probabilities, that is the probabilities  $p(\Omega_j)$  are small for the  $\Omega_j$  which contain such  $\mathbf{X}$ 's. Helton & Davis give several examples of importance sampling in [49]. These examples consider both equal and unequal strata probabilities.

### 2.5.4.3.3 Latin Hypercube Sampling

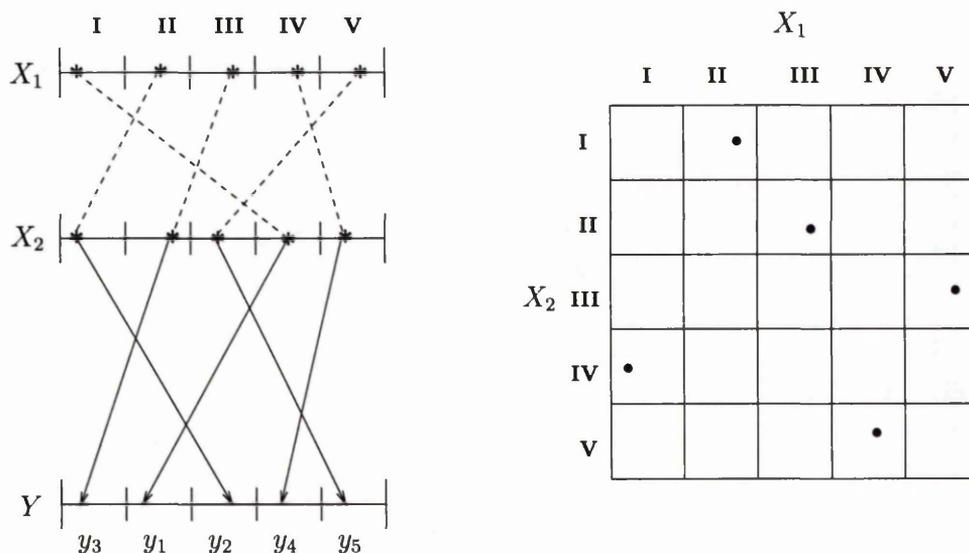
The Latin Hypercube Sampling (LHS) technique, which was originally introduced in 1979 by McKay *et al.* [84], has the objective of space covering sampling, and is sought in order to reduce the number of model runs  $N$ .

Iman & Shortencarier [65] and Stein [102] define LHS as a particular case of stratified sampling. In stratified sampling, the aim is to obtain a better coverage of the sample space by dividing the sample space into various subintervals such that units within each subinterval are as homogeneous as possible. Thus the sample space of the input vector  $\mathbf{X} = [X_1, X_2, \dots, X_k]$  is divided into  $l$  non-overlapping strata  $\Omega_1, \Omega_2, \dots, \Omega_l$  of sizes  $N_1, N_2, \dots, N_l$ ; such that  $N = \sum_{i=1}^l N_i$ . If  $l = 1$ , the result is a simple random sample over the entire sample space.

In LHS, to generate a sample of size  $N$  from  $k$  input factors  $X_1, X_2, \dots, X_k$ , first the range of each uncertain input factor is divided into  $N$  non-overlapping intervals on the basis of equal marginal probability  $1/N$ . Then one value from each interval is selected at random with respect to the probability density in the interval. The  $N$  values obtained for the first variable  $X_1$  are paired with the  $N$  values of  $X_2$  randomly. Then, these  $N$  pairs are randomly combined with the  $N$  values of  $X_3$  to form  $N$  triplets, and so on, until finally  $N$   $k$ -tuplets are obtained. Thus, for a given sample size( $N$ ) and number of input factors( $k$ ), there is  $(N!)^{k-1}$  possible interval combinations for a Latin hypercube sample. With this method, we gain the advantage of making sure that all portions of each factor's distribution is represented in the sample. An example of a LHS with  $N = 5$  sample size, and  $k = 2$  inputs  $\mathbf{X} = [X_1, X_2]$  with a known probability density function is given in Figure 2.3. The left-hand side of the figure shows a random pairing of each element of  $X_1$  and  $X_2$ , and the corresponding output variable,

whereas the right-hand side of the figure shows the cartesian coordinates of the randomly matched pairs. Note that in the cartesian coordinates presentation which consists of  $N^k = 5^2 = 25$  cells there is only one observation in each row and column.

As in SRS, introducing unwanted pairwise correlations between some or all of the input factors is also possible in LHS because of the random pairing of intervals in the process. This is more likely to occur if the sample size is small. A method introduced by Iman & Conover [61] can be used to avoid unwanted correlations between factors, it can also be used to induce known correlations between input variables. This method often referred to as 'restricted pairing technique' is based on the rank correlation structure of the input variables, and its desirable properties are: it is distribution free (i.e. it may be used on all types of input distribution functions); is simple to use, does not require unusual mathematical techniques; it can be applied to any sampling technique for which



**Figure 2.3.** An illustration of an LHS of size  $N = 5$  from a two-dimensional input space.

correlation of input variables is a meaningful concept, while preserving intention of the sampling technique; and it preserves the exact marginal distributions. This restricted pairing technique is discussed at length in [61], and examples are also given.

Iman & Conover's restricted pairing technique is implemented in the computer program written by Iman & Shortencarier to generate Latin hypercube and random samples (see [65]). Using this program one can generate random samples, random samples with restricted pairing, simple LHS, and LHS with restricted pairing.

If the sample size is small, SRS may produce clusters of observations anywhere in the range of the input factors but LHS produces observations which are spread over the entire range of each input factor. More detailed description of these two techniques with their properties may also be found in [48], [60], [65] and [84].

LHS has been proven to work well to considerably reduce the computer cost required to obtain a complete analysis of the model. LHS design appears as the most promising design as far as sensitivity analysis is concerned [63], [100]. A discussion on the advantages of LHS, along with a comparison of other sensitivity and uncertainty analysis methods can be found in [63]. A comparison of SRS and LHS techniques used for selecting values of input factors in the analysis of model output variables is presented in Chapter 5.

In cases where a large number of factors are involved, based on the computational cost we may decide to concentrate on the most important factors that control most of the output variability. To determine which factors among the many (potentially) important factors really are important '*screening methods*' presented in the following section are used.

### 2.5.5 Screening Methods

The term '*screening design*' is used to indicate any preliminary activity which aims to find out which of the model input factors are important [11]. Screening

designs are organised to deal with models containing hundreds of input factors. Therefore, they should be economical. They provide information about the sensitivity of the model to its inputs, while keeping computational cost of the experiment low. Sensitivities obtained from screening methods are qualitative, that is, screening methods rank the input factors in order of importance but do not quantify how much a given factor is more important than another one.

Screening methods are used for physical experiments as well as computer experiments. In this thesis, we focus on screening methods developed in the context of computer experiments which include standard One-at-A-Time (OAT) design and Morris's OAT design. Cotter's design, Andres's iterated fractional factorial design and Bettonvil's sequential bifurcation design are also among the screening techniques used in SA. Detailed description of these methods are presented by Campolongo & Kleijnen in Ref. [11]. Here, we discuss standard OAT which is computationally not expensive but provides only local information, and Morris OAT design which is more expensive but the information it provides is considered global.

### 2.5.5.1 One-at-A-Time Designs

In a standard OAT (sometimes also called elementary OAT) design, each input factor is changed one at a time and SA is performed to quantify the change in the model output. The literature value of each factor (often called '*nominal value*'), and a specified range in which the nominal value is normally between the two extremes are used in a standard OAT approach. This standard strategy is not the only one followed when an OAT design is conducted. Campolongo & Kleijnen list the five categories of OAT designs given by Daniel (1973), and they are as follows:

- Standard OAT designs where each factor is varied from a standard condition,
- Strict OAT designs where each factor is varied from the condition of the

last preceding experimental run,

- Paired OAT designs where two observations are produced and one simple comparison is made at a time,
- Free OAT designs where each new run is made under new conditions,
- Curved OAT designs in which a subset of results is produced by varying only one easy-to-vary factor (see [11]).

Since this type of analysis only addresses sensitivity relative to the point estimates chosen it is considered as a *'local'* sensitivity experiment [13], [45].

According to Campolongo & Saltelli [13], neither an elementary OAT nor a derivative based SA (discussed in Section 2.5.2) should be used to determine which input factors are more influential on the model output(s) than others, unless the model is known to be linear or the range of variation is small. Another OAT design, introduced by Max D. Morris in 1991, has advantages over standard OAT design and is described below. This design is recommended for use instead of standard OAT design by Campolongo & Saltelli (see [13]).

#### 2.5.5.2 Morris Design

This method is basically an individually randomized one-factor-at-a-time design in the input factors. Using this method, the effect of changing the value of each factor is evaluated in turn. The guiding philosophy of this method proposed by Morris (see [87]), is that a major role of a preliminary experiment is to determine, within reasonable uncertainty, which model input factors may be considered to have effects which are (a) negligible, (b) linear and additive, (c) nonlinear or involved in interactions with other inputs. Since Morris's experiment covers the entire input space  $\Omega$ , over which the factors may vary (unlike a local experiment in which the factors vary only around their nominal values), Campolongo & Kleijnen [11] consider this design as a *'global'* sensitivity experiment, and considering Daniel's terminology, they categorize the Morris design as a strict OAT design.

The two main advantages that this design provides are: (i) it is 'economic' in terms of number of model evaluations needed; (ii) no simplifying assumptions regarding the form of the model are needed [87].

Morris estimated the main effect of a factor by first evaluating a number, say  $r$  ( $r \in \{4, \dots, 10\}$ ), of local measures, at different points  $\mathbf{x}_1, \dots, \mathbf{x}_r$  in the experimental region, and then taking their average. By taking the average he reduces the dependence on the specific point that a local experiment has. For a given number of levels, say  $p$ , for each of the  $k$  factors, the input space  $\Omega$  is a  $k$ -dimensional  $p$ -level grid, where each component of the  $(1 \times k)$ -dimensional input vector  $\mathbf{x}$ ,  $x_i$ , may take values in the set  $\{0, 1/(p-1), 2/(p-1), \dots, 1\}$ . In practical applications, the values sampled from  $\Omega$  are first scaled to a suitable input vector for the model as follows

$$x_i = A_i + x_i(C_i - A_i)$$

where  $A_i$  and  $C_i$  are the extreme values of the variability range of  $x_i$ . For a given value of  $\mathbf{x}$ , Morris defines the elementary effect of the  $i$ th input as

$$d_i(\mathbf{x}) = [y(x_1, x_2, \dots, x_{i-1}, x_i + \Delta, x_{i+1}, \dots, x_k) - y(\mathbf{x})]/\Delta \quad (2.22)$$

where  $\mathbf{x} \in \Omega$  such that the perturbed point  $(\mathbf{x} + \Delta)$  is still in  $\Omega$  and  $\Delta$  is a predetermined multiple of  $1/(p-1)$ . A finite distribution of  $p^{(k-1)}[p - \Delta(p-1)]$  elementary effects for the  $i$ th input factor, which is denoted by  $F_i$ , is estimated by sampling  $\mathbf{x}$  from  $\Omega$ . Analysis of the distribution  $F_i$  through its mean  $\mu$  and standard deviation  $\sigma$  gives us useful information regarding the relative importance of the  $i$ th input factor. A large mean value indicates a factor with a high overall influence on the output; a high standard deviation indicates an input factor which is interacting with other factors or whose effect is non-linear [12]. In the simplest case, the total computational cost for obtaining a random sample of  $r$  values from distribution  $F_i$  is  $n = 2rk$  computer runs, that is, each elementary effect requires two model runs: one at the selected values  $x_1, \dots, x_k$ ; and one after  $x_i$

is increased by the quantity  $\Delta$ ,  $x_1, \dots, x_{i-1}, x_i + \Delta, x_{i+1}, \dots, x_k$ .

Morris defines the *economy* of a design as the number of elementary effects it produces divided by the number of experimental runs. So the economy of the Morris design under the assumption that all  $rk$  observed elementary effects are independently drawn is  $rk/2rk = 1/2$ . As Campolongo & Kleijnen point out, the larger the economy for a particular design, the better it is in terms of providing information for sensitivity and uncertainty analysis [11].

The main idea of Morris design is based on construction of a sampling matrix which is called the '*Orientation Matrix*' and denoted by  $\mathbf{B}^*$ . This  $(k+1) \times k$  dimensional  $\mathbf{B}^*$  matrix has the property that for every column  $i = 1, 2, \dots, k$ , there are two rows of  $\mathbf{B}^*$  that differ only in their  $i$ th entries. With this particular property,  $(k+1)$  rows of  $\mathbf{B}^*$  produce  $(k+1)$  output values for the model, allowing the calculation of  $k$  elementary effects, one for each input factor, from  $(k+1)$  runs. If  $r$  is the size selected for the sample of the elementary effects, the experiment requires construction of  $r$  orientation matrices. Therefore, the total computational cost for the experiment is  $n = r(k+1)$  model runs.

The first step in constructing an orientation matrix  $\mathbf{B}^*$  is to select a  $(k+1) \times k$  matrix  $\mathbf{B}$  with elements of 0's and 1's, such that for every column  $i = 1, 2, \dots, k$  there are two rows of  $\mathbf{B}$  that differ only in their  $i$ th entries; for example,  $\mathbf{B}$  may be chosen to be a lower triangular matrix of 1's as given below

$$\mathbf{B} = \begin{bmatrix} 0 & 0 & 0 & \dots & 0 \\ 1 & 0 & 0 & \dots & 0 \\ 1 & 1 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 1 & 1 & 1 & \dots & 1 \end{bmatrix}.$$

Then, the matrix  $\mathbf{B}'$  is given by

$$\mathbf{B}' = \mathbf{J}_{k+1,1} \mathbf{x}^* + \Delta \mathbf{B}$$

where  $\mathbf{J}_{k+1,1}$  is a  $(k+1) \times 1$  matrix of 1's;  $\mathbf{x}^*$  is a randomly chosen starting vector of  $\mathbf{x}$ ; and  $\Delta$  is the selected increment for the components of  $\mathbf{x}$ . Morris also introduces the possibility of using  $\Delta\mathbf{B}$  as a design matrix for which the corresponding experiment would provide  $k$  elementary effects, one for each input, based on only  $(k+1)$  model runs, but as he points out these would not be random selections from the  $F_i$ 's. In order to obtain random selections, a randomized version of the sampling matrix  $\mathbf{B}^*$ , which is called 'a random orientation of  $\mathbf{B}$ ', is employed for the design, and it is given by the following equation:

$$\mathbf{B}^* = (\mathbf{J}_{k+1,1}\mathbf{x}^* + (\Delta/2)[(2\mathbf{B} - \mathbf{J}_{k+1,k})\mathbf{D}^* + \mathbf{J}_{k+1,k}])\mathbf{P}^*$$

where  $\mathbf{D}^*$  is a  $k$ -dimensional diagonal matrix in which each diagonal element is either -1 or +1 with equal probability;  $\mathbf{J}_{k+1,k}$  is a  $(k+1) \times k$  matrix of 1's; and  $\mathbf{P}^*$  is a  $k \times k$  matrix obtained by randomly permuting columns of a  $k \times k$  identity matrix, where each such matrix has an equal probability of selection. If a sample of  $r$  effects is required from each distribution  $F_i$ , then  $r$  different orientation matrices  $\mathbf{B}^*$  have to be selected for the design matrix of the entire experiment

$$\mathbf{X} = \begin{bmatrix} \mathbf{B}_1^* \\ \mathbf{B}_2^* \\ \vdots \\ \mathbf{B}_r^* \end{bmatrix}.$$

Over the entire design  $\mathbf{X}$ ,  $r$  elementary effects can be produced for each input factor, and total computational cost would then be  $n = r(k+1)$  model runs.

Morris also compares his OAT design and the Latin hypercube design in the input screening context, and shows that the Morris design may prove to have more advantage over the Latin hypercube design [87].

Since both 8-compartment global carbon cycle models considered later in this thesis are linear models with reasonably small number of factors (less than 20 in

each model), we have used both the standard OAT design and Morris design as well as other local and global sensitivity analyses techniques to present similarities and dissimilarities of results obtained following different approaches.

## 2.5.6 Methods of Analysis

These design methods are basically Monte Carlo based methods. In the global sense, all model input factors are varied simultaneously and the sensitivity is measured over the entire range of each factor. These methods are based on performing multiple model evaluations with randomly selected input factors by using an appropriate sampling technique.

It should be noted that when we describe the methods in the following sections, by the term ‘explanatory variable’ we mean model input factors, i.e., initial conditions, transfer coefficients etc., and by the term ‘response variable’ we mean model predictions, i.e. model output variables. Now, let us concentrate on a model with a single output variable, say  $Y$ , and  $k$  input factors  $X_j$  ( $j = 1, 2, \dots, k$ ). If there are several outputs considered, the techniques described in the following sections can easily be applied per output.

### 2.5.6.1 Correlation Measures

Input factors can be ranked in order of importance based on the value of the correlation coefficient between the output variable and each input factor. This ranking assumes that the greater the correlation coefficient, the more controlling influence that input factor has on the model behaviour [95].

Here, it is important to emphasize that the Pearson product-moment correlation coefficient only picks up linear association, and it may miss more complicated relationships where the input and output are related in a non-linear fashion.

When dealing with non-linear models, utilizing the Pearson correlation coefficients for sensitivity ranking is not appropriate. In such cases, the Spearman rank correlation coefficient is used.

The Pearson moment-correlation coefficient and the Spearman rank correlation coefficient are often utilized for sensitivity studies (for example, see [74], [75], [96]).

### 2.5.6.2 Regression Analysis

Regression analysis is a more formal investigation of the relationship between the input factors  $X_{ij}$  and the output  $Y_i$  ( $j = 1, \dots, k, i = 1, \dots, N$ ).

The use of regression technique allows the sensitivity ranking to be determined based on the relative magnitude of the regression coefficients. This is discussed further in Section 2.5.6.5.

### 2.5.6.3 Stepwise Regression

When a model involving a large number of input factors is under consideration, constructing a regression model with all input factors may not be appropriate due to reasons such as the possibility of overfitting of the model, or only small number of input factors having a significant effect. In such cases, stepwise regression which creates a model by selecting the 'best' input factors from all original independent variables one at each step, may be used.

Helton & Davis note three aspects of stepwise regression analysis that give us insights on the importance of the individual input factors [49]. One of these aspects is the order in which the factors are selected in the stepwise procedure. The most important input factor is selected first, the next most important factor is selected second, and so on. Another aspect is the  $R^2$  values at successive steps of the analysis that provide a measure of input importance by showing how much of the uncertainty in the output variable is accounted by all input factors selected at each step of the analysis. A third aspect of this analysis mentioned by Helton & Davis is the absolute values of the standardized regression coefficients (SRCs) in the individual regression models which can provide an indication about the input factor importance. Examples of the stepwise regression analysis in sampling-based sensitivity analysis can be found in various articles by Helton *et al.*, for

example [51], [52], [53] and [49].

#### 2.5.6.4 Rank Transformation

The rank transformation is a simple procedure where the raw data is simply replaced with their ranks, i.e., if there are  $N$  observations of the model output,  $Y$ , these observations are replaced by their corresponding ranks 1 to  $N$ , where  $R(Y_i)$  is the rank assigned to the  $i$ th value of  $Y$ . Similarly, each of the model input factors is replaced with its corresponding ranks 1 to  $N$ . For ties the average ranks are assigned.

If the model output is a monotonic function of the input factors, then the rank transformation is used to linearize relationships and also to reduce the effects of extreme values. In Section 2.5.6.1, we have already mentioned the use of ranking technique for calculating a correlation between ranked variables, and in the following sections we provide other statistics based on the rank transformation.

#### 2.5.6.5 SRC and SRRC

The estimated coefficients of a linear regression model show the effect of one unit change in each  $X_j$  on the model predictions assuming that the other input factors are held constant. If the input factors are not in equivalent units, which is often the case, then the estimated regression coefficients, which depend on the units in which  $X_j$ 's and  $Y$  are expressed, do not provide a useful indication of model input factor importance, unless the effect of scale is removed. For linear regression, this involves standardizing the  $X_j$ 's and  $Y$  to mean 0 and standard deviation 1. The size of the standardized regression coefficients (SRCs) then provide a more meaningful indication of input factor importance.

The standardized rank regression coefficients (SRRCs) are simply the SRCs calculated on ranks.

### 2.5.6.6 PCC and PRCC

In some cases, the correlation between an input factor and an output variable, say  $X_1$  and  $Y$ , may be partially due to the correlations of other input factors, say  $X_2, \dots, X_k$ , with both  $X_1$  and  $Y$ . In such a case, we may want to find out what the correlation between  $X_1$  and  $Y$  would be if the effect of  $X_2, \dots, X_k$  on each of  $X_1$  and  $Y$  were eliminated. This measure of correlation is called the partial correlation, and the correlation coefficient between  $X_1$  and  $Y$  after the linear effect of  $X_2, \dots, X_k$  on both  $X_1$  and  $Y$  has been eliminated is called the partial correlation coefficient (PCC) [41].

As described by Kleijnen & Helton [74] PCCs can be calculated using a sequence of regression models. To calculate PCC between  $X_j$  and  $Y$  first the following two regression equations are constructed to correct for the linear effects of other variables

$$\hat{X}_j = c_0 + \sum_{\substack{p=1 \\ p \neq j}}^k c_p X_p, \quad \hat{Y} = b_0 + \sum_{\substack{p=1 \\ p \neq j}}^k b_p X_p,$$

then the sample correlation on the residuals  $(X_j - \hat{X}_j)$  and  $(Y - \hat{Y})$  is calculated.

The PCCs performed on the ranks instead of the raw data provide partial rank correlation coefficients (PRCCs). The PCCs provide a ranking of the input factors by indicating the strength of the linear relationship between  $X_j$  and  $Y$ , and with the PRCC's the linear relationship between the ranks of  $X_j$  and  $Y$  is measured.

According to Saltelli & Homa (see [99]) non-parametric statistics based on ranks, such as the SRRCs and PRCCs appear to be among the most robust and reliable SA methods.

## 2.6 Non-parametric SA Methods

Because of our limited knowledge of the input factors and their associated distributions it is often desirable to use some nonparametric statistical tests.

The application of these three tests to SA comes from the idea of partitioning the sample of an input factor under consideration into two sub-samples according to the quantiles of the output distribution. If the distributions of the input factor in the two sub-samples can be proven to be different then the input is identified as an influential input. Using the test statistics as a sensitivity measure, the relative importance of the input factors for each output variable can be obtained (i.e., the higher the test statistic calculated between an input factor and an output variable, the more influential the input is on that output variable).

### 2.6.1 Smirnov Test

The Smirnov test is used to test if two different samples belong to the same population. In order to be able to carry out a Smirnov test, the two samples must satisfy the assumptions of: 1) the samples are random samples; 2) the two samples are mutually independent; 3) the measurement scale is at least ordinal; and 4) the random variables are continuous.

In the context of sensitivity analysis, the sample of a model input factor, say  $X$ , is partitioned into two sub-samples, say  $X_i$  and  $X_j$ , according to the quantiles of the distribution of the model output, say  $Y$ . Let  $X_i$  be size  $N_1$  and  $X_j$  size  $N_2$ , and  $S_1(x)$  and  $S_2(x)$  be the empirical distribution functions of these two sub-samples, respectively. The greatest vertical distance between the two empirical distribution functions gives us the Smirnov test statistic denoted by  $T_S$ , that is,

$$T_S = \sup_x |S_1(x) - S_2(x)|.$$

This test statistic can be used to rank the model input factors, the higher the  $T_S$  value the more influential the input factor on the model output.

For hypothesis testing, if this test statistic exceeds its  $1 - \alpha$  quantile obtained from the available tables (see [22]), the hypothesis of ‘the two sub-samples belong to the same population’ is rejected at significance level  $\alpha$ . When the test results show that the distributions of the two sub-samples are different, it can be said that the investigated model input factor is an influential factor on the model output.

### 2.6.2 Cramér-von Mises Test

Like the Smirnov test the Cramér-von Mises test is also used to determine whether two empirical distributions are statistically identical. The assumptions for this test are: 1) the samples are independent random samples; 2) the measurement scale is at least ordinal; and 3) the random variables are continuous. The Cramér-von Mises test statistic  $T_{CM}$  is defined as

$$T_{CM} = \frac{N_1 N_2}{(N_1 + N_2)^2} \left( \sum_{i=1}^{N_1} [S_1(X_i) - S_2(X_i)]^2 + \sum_{j=N_1+1}^{N_1+N_2} [S_1(X_j) - S_2(X_j)]^2 \right)$$

where  $N_1$  is the size of the first sub-sample,  $N_2$  is the size of the second sub-sample, and the squared differences in the summation is computed at each  $X_i$  and at each  $X_j$ .

The Cramér-von Mises and Smirnov tests are very similar; however the calculation of the test statistic for the former is slightly more difficult since it makes more effective use of the data. According to Conover [22] there is little difference in power between the two tests. Saltelli & Marivoet [100] note that because the Cramér-von Mises statistic depends upon the total area between the two empirical distributions, it may be more appropriate for SA when the model output function is a non-monotonic function.

### 2.6.3 Mann-Whitney Test

This non-parametric test is also based on two samples. The assumptions must be satisfied by the two samples under consideration are: 1) both samples are random samples; 2) the two samples are mutually independent; 3) two populations from which the two samples are taken have the same shape, hence the same variance; 4) the measurement scale is at least ordinal.

The Mann-Whitney test essentially looks at the difference between the means of ranks of the  $X$  values in the two sub-samples. The values of one of the two sub-samples, say  $X_i$ , are first ordered and ranks are assigned based on the ordering. Then, the ranks  $R(X_i)$ 's are used to compute the test statistic  $T_{MW}$  as

$$T_{MW} = \sum_{i=1}^{N_1} R(X_i)$$

where the summation is extended to the elements of one sub-sample only.

## 2.7 Implications on SA Methods

This chapter has described the important features of various SA methods, and discussed their strengths and weaknesses.

It is important to note that there is no one "perfect" method which could give a comprehensive indication of sensitivity, and as Helton *et al.* note in Ref. [50] it is not possible to rate one method of SA as superior to another. To decide which method(s) to use will depend on both the model being studied and the type of information desired. If possible one should consider using a number of methods to maximize confidence in the analysis results.

In the two chapters which follow, various SA techniques are applied to three different compartmental models of the global carbon cycle. First, a general background on the carbon cycle, which is useful in helping to understand different aspects of the cycle, is given. That is followed by details about the observed(historic) and

predicted(future) atmospheric CO<sub>2</sub> emissions resulting from fossil fuel burning and deforestation. Then, a description of the model is given. After that, local and global SA techniques are presented. In the last section, the results and the discussions are summarized.

## Chapter 3

# Application of Sensitivity Analysis Techniques to Compartmental Models with Specific Application to Global Carbon Cycle Models

### 3.1 Introduction

Sensitivity Analysis (SA) of the model response to the variations in its input factors is an essential element for improving both the understanding of the model and its performance. As noted by Hora, SA is the first stage of a cycle of investigation [57], and in this chapter various methods of SA for exploring the influence of input factors on the outputs of models will be discussed.

The two test models used in this chapter are linear, time-invariant compartmental models and originate from the context of modelling the global carbon cycle (GCC). These models, compared to some other GCC models available in the scientific literature (for instance, GLOCO model in [40], ANU-BACE model

in [103]), are not very complex. In the next chapter of this thesis, we use one of these more complex models as a test model and investigate effect of model complexity on SA.

Uncertainty analysis (UA), the next stage of investigation cycle, is covered in Chapter 5.

The aims and objectives of this chapter are to (a) compare and contrast SA techniques; (b) extend the techniques to deal with mass balanced GCC models, a special case of constrained systems; and (c) identify which set(s) of model input factors are more influential on the model outputs.

## 3.2 The Global Carbon Cycle

Why is carbon important? The carbon atom is the basic building block of living matter, therefore it is of prime importance to life. Living organisms circulate carbon by simply existing.

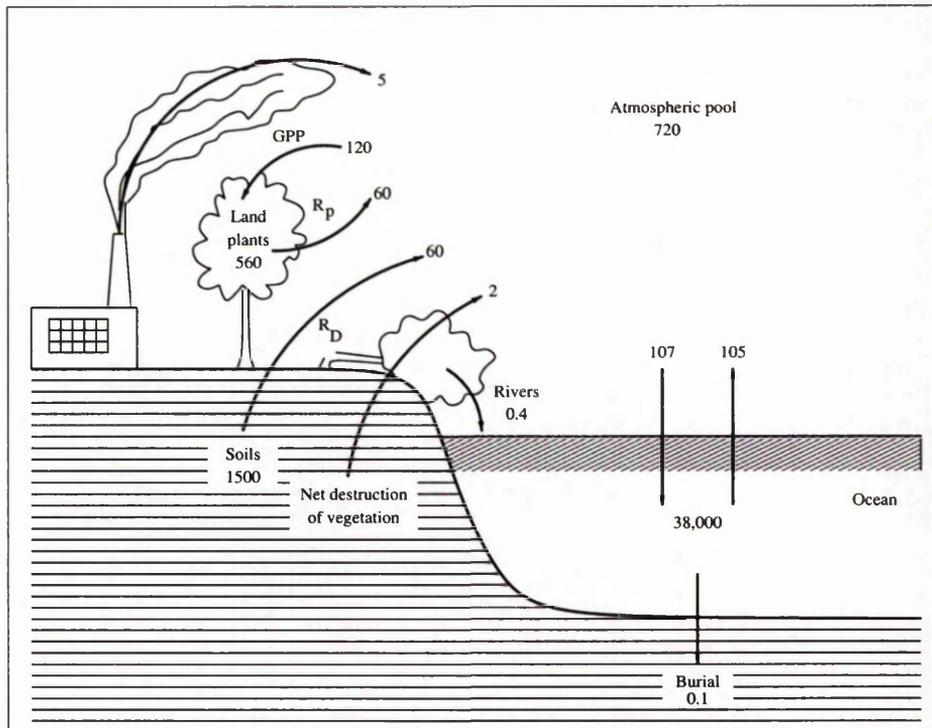
The carbon cycle is the biological circulation of carbon from the atmosphere into living organisms and after their death back again. The carbon cycle is mostly the carbon dioxide cycle. Carbon dioxide, which plays an essential part in metabolism and an intrinsic role in planetary energy, is released by respiration, soil processes, combustion of carbon compounds, oceanic evaporation and volcanic eruptions. It is dissolved in the oceans and consumed via photosynthesis. The two main fluxes of the global carbon cycle, which are nearly equal, are between the atmosphere and terrestrial systems, and the atmosphere and the oceans. The terrestrial and aquatic systems are usually considered virtually independent because production balances consumption of  $\text{CO}_2$  in both [76].

The carbon cycle was historically considered to be in equilibrium. However, human actions and the industrial revolution have resulted in a significant imbalance. According to Neftel *et al.* since 1750 there has been a steady increase in the atmospheric carbon dioxide level, and human activities are largely responsible for the observed increase [89]. Fossil fuel burning is the most important source of

CO<sub>2</sub> and fossil fuels have been injected into the atmosphere at a steadily increasing rate since the beginning of the industrial revolution, around 1860 [68], [79], [81]. Keeling & Whorf (see [70]) have recorded an approximately 25% increase in the atmospheric CO<sub>2</sub> levels from 1800 to 1985 mainly due to human influences first from deforestation and now primarily from fossil fuel burning. They have also noted that the huge increase in fuel combustion since 1950 has led the CO<sub>2</sub> content in the atmosphere to increase gradually from 280 parts per million (ppm) to more than 350 ppm. Currently, there are 369.40 parts per million by volume (ppmv) of CO<sub>2</sub> in the atmosphere [71].

Although the carbon cycle is a highly complex cycle it can be summarized in terms of a few major reservoirs of carbon and the carbon fluxes between them. Figure 3.1 shows a typical global carbon cycle adapted from Krebs [76]. The amounts of carbon are given in units of Gigatons of carbon (1 Gt C = 1 billion metric tons of carbon) and all fluxes in units of Gigatons of carbon per year (Gt C/yr). It shows that the oceanic reservoir contains by far the largest amount of carbon, and the atmosphere is the smallest in terms of carbon storage but it plays an important role in the cycle. The atmosphere exchanges CO<sub>2</sub> with the ocean's surface, and much of the carbon in the oceans is in the deeper waters. Oceanographers believe that about 40% of the CO<sub>2</sub> from fossil fuels enters the oceans each year [76]. The amount of CO<sub>2</sub> in the oceans is fifty times that of CO<sub>2</sub> in the atmosphere, suggesting that the oceans can absorb most of the additional CO<sub>2</sub> injected into the atmosphere. However, uptake of CO<sub>2</sub> into the surface waters of the oceans is relatively slow (half-life 1.3 years), and in addition, the surface waters of the ocean (0 to ~100 m depth) mix with the deep waters even more slowly (half-life 35 years) [7].

In 1938, G. S. Callendar presented the first data showing the increasing CO<sub>2</sub> concentration in the atmosphere and suggested that this increase might affect the Earth's climate. In 1956, G. N. Plass outlined theories to explain the relationship between atmospheric CO<sub>2</sub> and climate. Soon after that R. Revelle and H. Suess described the relationship between CO<sub>2</sub> in the atmosphere and in the oceans, and



**Figure 3.1.** A schematic diagram of the Global Carbon Cycle

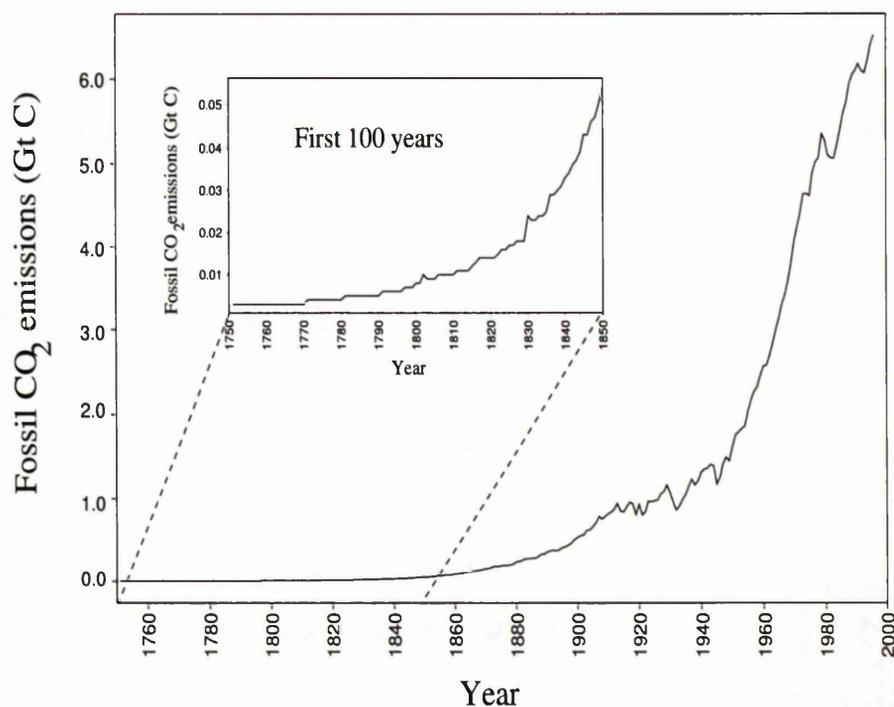
Kaplan enlarged upon the role of  $\text{CO}_2$  in the atmosphere in terms of the global heat balance [106].

In the 1970's, recognition of a growing world population, the rising per capita use of energy, and the accompanying growth in the rates of fossil fuel  $\text{CO}_2$  emissions drew many scientists' attention to the atmospheric  $\text{CO}_2$  increase. Despite international action to control the emission of some greenhouse gases, as we can see in Figure 3.2  $\text{CO}_2$  levels are still rising.  $\text{CO}_2$  emission in 1751 was estimated to be 0.003 Gt C, and in the next 100 years it only increased to 0.054 Gt C. Carbon emissions from fossil fuel burning are estimated to have increased at a rate near 4.3 percent per year from 1860 until 1973 with the exception of brief periods during the great depression and the world wars. Following the 1973 oil embargo and a decline induced by sharp oil price increases in the early 1980's, the amount

of carbon entering the atmosphere began increasing again in the mid-1980's and reached 6.52 Gt C in 1996 [80].

In 1977, a leading group of scientists assembled in Florida to discuss the current understanding of the dynamics of carbon exchanges within the atmosphere, oceans and terrestrial biota that determine the atmospheric CO<sub>2</sub> concentration. Since then extensive research has been carried out by the international scientific communities [106]. The importance of such research activities and measurements obtained has become more widely recognised in the past two decades.

A dramatic increase in greenhouse gases, particularly CO<sub>2</sub>, in the atmosphere is one of the many disasters that the 20<sup>th</sup> century has witnessed. There is growing concern that the resulting increased heat in the atmosphere known as 'greenhouse effect' will affect climates around the world in the 21<sup>st</sup> century and beyond [3],



**Figure 3.2.** CO<sub>2</sub> emissions from fossil fuels (for pre-industrial time see inset of period from 1750 to 1850).

and the resulting increased heat in the atmosphere will have a serious impact on the environment, climate, ocean levels and agriculture.

Projections of future CO<sub>2</sub> concentrations are needed to assess the likelihood of significant global and regional change as a consequence of the continuing use of fossil fuels and to determine if alternative scenarios of future energy use can significantly change this likelihood [73].

Global carbon cycle models are needed to estimate the future change of CO<sub>2</sub> concentrations for specified CO<sub>2</sub> emission scenarios. SA of these models is required to investigate the effects of changes in the inputs on the model outputs.

### 3.2.1 Observed Atmospheric CO<sub>2</sub> since 1744

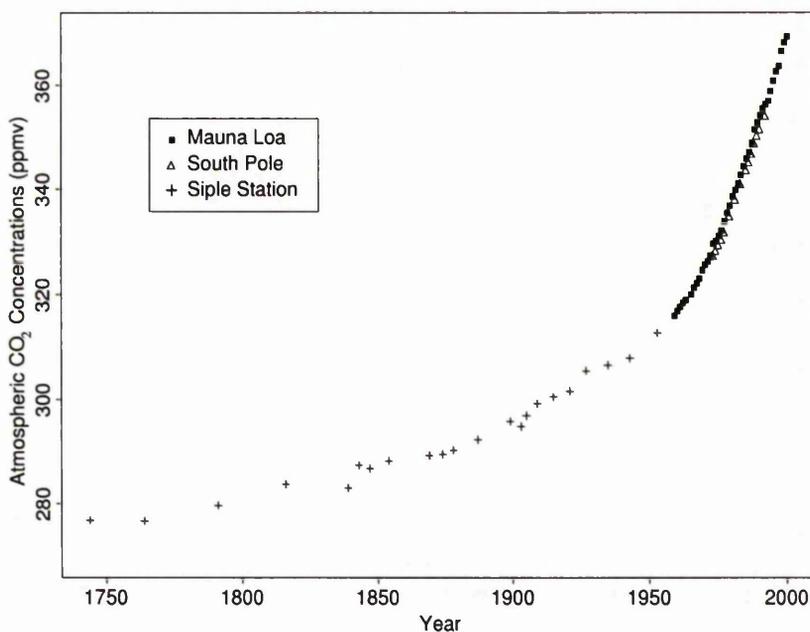
Precise measurements of atmospheric carbon dioxide concentrations have been obtained since March 1958 when Dr. Charles D. Keeling of the Scripps Institution of Oceanography, initiated a monitoring program at Mauna Loa Observatory (MLO) in Hawaii. The site of MLO is one of the most favourable locations for measuring undisturbed air and minimises the possible influences of human activities and vegetation on atmospheric CO<sub>2</sub> concentrations. Air samples have been continuously collected at MLO and analysed by infrared spectroscopy for CO<sub>2</sub> concentrations. Data are averaged to give monthly and annual atmospheric CO<sub>2</sub> concentrations [69]. These measurements, widely recognised as the 'Keeling Curve', constitute the largest, continuous record of atmospheric CO<sub>2</sub> concentrations available in the world. The annual averages of the data collected at MLO have shown a steady rise in annual average concentration from 316 ppmv in 1959 to 369.40 ppmv in 2000 [71].

Also, it has been found that long-term records of atmospheric CO<sub>2</sub> concentrations can be obtained from ice cores. Air is trapped by snow as it is transformed into glacial ice, and by taking ice cores one can sample the atmosphere back in time. For example, in the Soviet Antarctic Expedition at Vostok, Antarctica the scientists collected a 2083 meter long ice core which spans 160,000 years [4].

Observed atmospheric CO<sub>2</sub> concentrations from Mauna Loa, South Pole and

Siple ice cores during the period 1744 through 2000 are presented in Figure 3.3. What we can see in this figure is that concentration level did not change much until 1850, when it was around 280 ppmv, after 1850's the amount of carbon dioxide in the atmosphere can be seen to be increasing year after year. This exponential growth shows that the sources and sinks of atmospheric CO<sub>2</sub> are not exactly in balance, with the rate of increase in the 1980's a little less than 0.5% annually.

Measurements made at other locations all over the world also demonstrate the increase in the atmospheric CO<sub>2</sub> level. South Pole and Siple Station ice core measurements for recent years shows agreement with the Mauna Loa data.



**Figure 3.3.** The annual atmospheric CO<sub>2</sub> concentrations from Mauna Loa Observatory, Hawaii: 1958-2000 (Keeling & Whorf, 2001 [71]); South Pole: 1973-1993 (Keeling & Whorf, 1994 [69]); Siple Station: 1744-1953 (Friedli *et al.*, 1986 [37]).

### 3.2.2 IPCC and Emission Scenarios

The Intergovernmental Panel on Climate Change (IPCC) was jointly established by the World Meteorological Organisation (WMO) and the United Nations Environmental Programme (UNEP) in 1988. The primary objectives of IPCC's three Working Groups are:

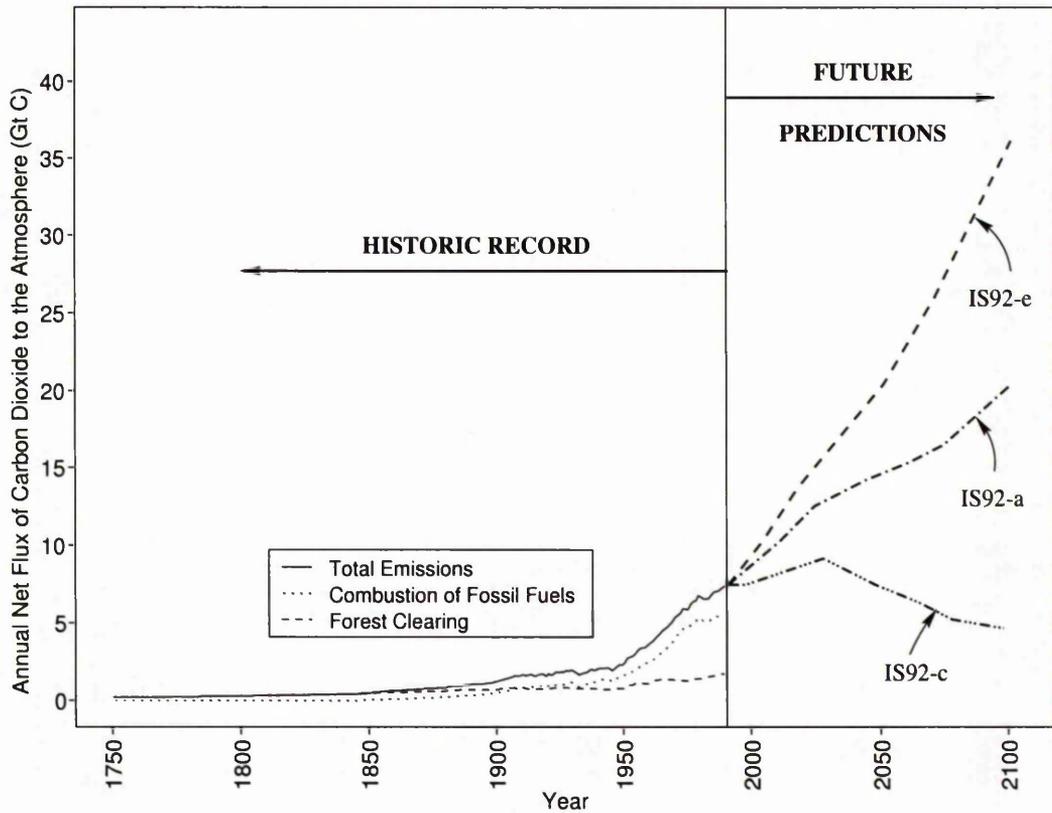
- to assess the available information on climate change,
- to assess the environmental and socio-economic impacts of climate change,
- and to formulate response strategies.

Despite trends which favour considerable increase in net greenhouse gas emissions over the next century, the IPCC says that "significant reductions ... are technically possible and can be economically feasible". In 1992, they developed six alternative emission scenarios based on different assumptions regarding the factors which could have major influences on future levels of CO<sub>2</sub> emissions. Population and economic growth, structural changes in economies, energy prices, technological advance and fossil fuel supplies are among factors considered [58]. These scenarios are referred to as the IS92a to f scenarios, they extend to the year 2100, and they include emissions of other greenhouse-related gases as well as CO<sub>2</sub>. The projection of future CO<sub>2</sub> emissions associated with these scenarios are given in Table 3.1, and three of these six scenarios used as input scenarios in this thesis are illustrated in Figure 3.4 with the historic record of CO<sub>2</sub> emissions.

Scenario IS92a represents a middle-of-the range scenario due to modest and largely offsetting changes in the underlying assumptions. Scenario IS92c has a CO<sub>2</sub> emissions path that eventually falls below its starting value. For this scenario a decline in the population by the middle of the next century, a low economic growth, and severe constraints on fossil fuel supply are assumed. Scenario IS92e has the highest greenhouse gas emissions. It assumes moderate population growth, high economic growth, high fossil fuel availability and eventually hypothetical phase-out of nuclear power [58].

**Table 3.1. IPCC 1992 CO<sub>2</sub> Emission Scenario Results**

Year	Scenario					
	IS92a	IS92b	IS92c	IS92d	IS92e	IS92f
1990	7.40	7.40	7.41	7.33	7.40	7.41
1995	7.93	7.93	7.23	7.27	8.16	7.97
2000	8.44	8.24	7.46	7.49	9.10	8.77
2005	9.16	8.82	7.75	7.81	10.17	9.70
2010	9.89	9.45	8.05	8.17	11.40	10.77
2015	10.64	10.19	8.31	8.54	12.63	11.95
2020	11.38	10.95	8.49	8.78	13.74	13.13
2025	12.23	11.81	8.79	9.29	15.08	14.37
2050	14.52	13.80	7.51	9.02	20.10	17.25
2075	16.31	15.41	5.58	9.27	26.96	21.19
2100	20.28	19.11	4.61	10.33	35.84	26.59



**Figure 3.4. Historic and future CO<sub>2</sub> emissions.**

### 3.2.3 CO<sub>2</sub> Inputs

Global annual CO<sub>2</sub> emissions from the combination of fossil fuels combustion and deforestation from Enting *et al.* [33] are used as input to the models. These data are for the period 1765-1990. Pre-1765 values for total carbon flux were set to 0.2 Gt C, the total 1765 values from deforestation and fossil fuel emissions. As future emissions, the IPCC's three IS92 scenario estimates of CO<sub>2</sub> emissions from fossil fuel combustion and forest clearing are used for the period 1990-2100.

## 3.3 Modelling the Global Carbon Cycle

Modelling complex processes, such as the global carbon cycle (GCC), is not an easy task. The difficulty lies not only in having limited observational data but also in a lack of understanding of the complex characteristics of the system. To investigate the rate of atmospheric CO<sub>2</sub> increase it is essential to understand the GCC linking the atmosphere, oceans and terrestrial systems, and to describe it by mathematical models on the basis of ecology and engineering [39].

Many studies have been carried out to estimate current and future patterns of atmospheric, oceanic, and terrestrial carbon storage. For that purpose usually dynamic linear compartmental models are used.

There is a wide range of mathematical models describing the local and global carbon cycle. The former usually describes the transport of carbon in a specific area or for a particular species. The latter models however usually focus on analysing the GCC. Many GCC models, ranging from a single compartment model to a 25-compartment model have been developed and are available in the scientific literature. As noted by Iman and Conover [59], although the model is the most important link in the study of a physical systems, the proper development and verification of the model is the responsibility of geologists, physicists, engineers and other experts. From a statistician's point of view the model is viewed as a 'black-box' with many inputs and one or more output variables. Taking this into account, we treat the GCC models used in this thesis as black-box models.

Now, in the following section we give a description of two of the three GCC models (both are 8-compartment models) utilised in this thesis (the third GCC model which is a larger and more complex model is described and used in Chapter 4). Then, the results and discussions on the application of various SA methods to one of the 8-compartment models is given in detail. In order to construct a base for the UA covered in Chapter 5, the same SA methods have been applied to the second 8-compartment model and the results are summarized in Appendix B.

### 3.3.1 The Two 8-Compartment GCC Models

The CO<sub>2</sub> distribution between atmosphere and oceans, atmosphere and terrestrial systems, and the responses of these reservoirs to the input resulting from fossil fuel burning and deforestation have been quantitatively described by using two different compartmental models each consisting of eight compartments but different structures. Here, we assume that each of these 8-compartment models constitutes an adequate representation of the GCC.

The release of fossil fuel CO<sub>2</sub> and forest clearing are viewed in the context of this model as perturbations to an initial steady-state condition, and all other inputs to the model are assumed to be zero.

#### 3.3.1.1 Description of Model I

The compartmental diagram of the first model adapted from Emanuel *et al.* (1984) is given in Figure 3.5. The model consists of eight well-mixed compartments and 15 transfer coefficients.

Two compartments represent carbon in the 'surface ocean' and 'deep ocean'. Carbon in living plants is divided between 'tree' and 'ground vegetation' compartments. The 'tree' compartment is separated into two separate compartments, namely 'nonwoody parts of trees' and 'woody parts of trees'. To represent carbon in dead parts of the terrestrial systems and their decomposers, two compartments are used. The 'detritus/decomposers' compartment corresponds to litter and its

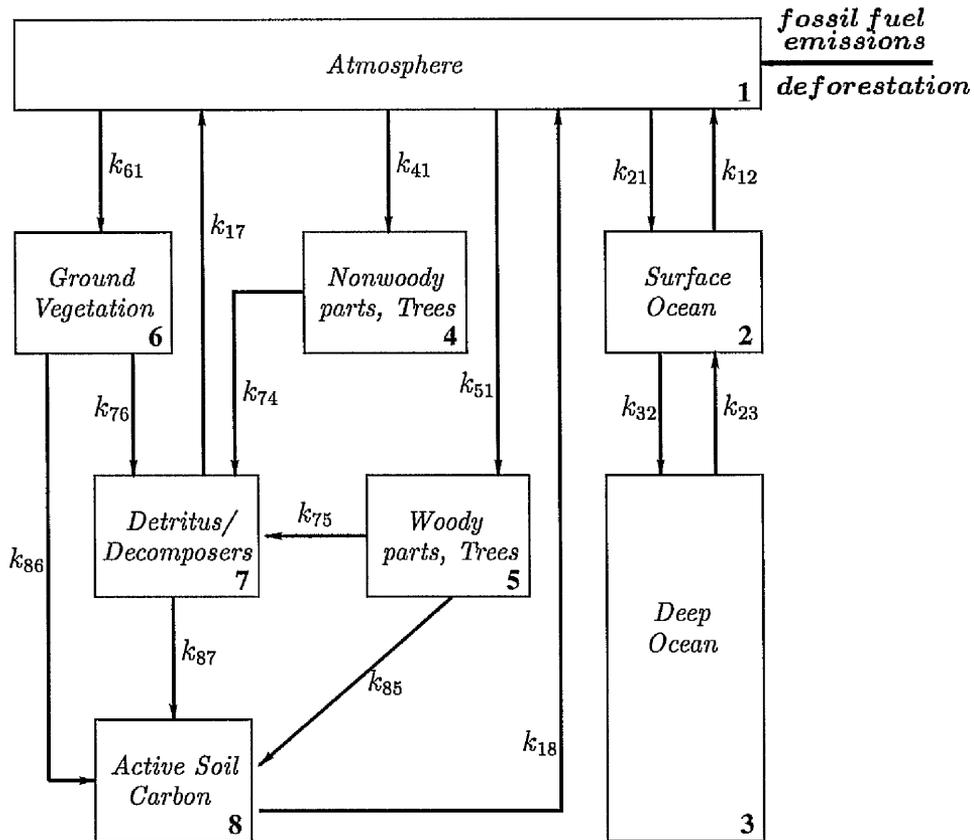


Figure 3.5. Compartment diagram of Global Carbon Cycle Model I (adapted from Emanuel *et al.* (1984) [31]).

decomposers at the soil surface. Carbon input to this reservoir comes from death of aboveground parts of vegetation. The ‘active soil carbon’ compartment consists of carbon in soils and its decomposers. Carbon from death and initial decomposition of below-ground parts of vegetation and transport of decomposed material from the actively decaying litter layer is transferred into this compartment [30].

The flow of  $\text{CO}_2$  between the compartments is described by a set of eight first-order, linear differential equations which contain 23 uncertain model input factors (15 transfer coefficients and 8 initial conditions). To define the model sensitivity to variations in these uncertain input factors, we treat them as random variables.

The nominal values of the initial compartment contents, and the transfer coefficients which satisfy the assumed initial steady-state condition are given in

Table 3.2 and Table 3.3, respectively.

The CO<sub>2</sub> emissions due to fossil fuel burning and forest clearing enter the system through the atmosphere compartment.

The following are the model state equations of Model I:

$$\begin{aligned}
 \dot{x}_1 &= k_{12}x_2 + k_{17}x_7 + k_{18}x_8 - (k_{21} + k_{41} + k_{51} + k_{61})x_1 + u_1(t) \\
 \dot{x}_2 &= k_{21}x_1 + k_{23}x_3 - (k_{12} + k_{32})x_2 \\
 \dot{x}_3 &= k_{32}x_2 - k_{23}x_3 \\
 \dot{x}_4 &= k_{41}x_1 - k_{74}x_4 \\
 \dot{x}_5 &= k_{51}x_1 - (k_{75} + k_{85})x_5 \\
 \dot{x}_6 &= k_{61}x_1 - (k_{76} + k_{86})x_6 \\
 \dot{x}_7 &= k_{74}x_4 + k_{75}x_5 + k_{76}x_6 - (k_{17} + k_{87})x_7 \\
 \dot{x}_8 &= k_{85}x_5 + k_{86}x_6 + k_{87}x_7 - k_{18}x_8
 \end{aligned} \tag{3.1}$$

with  $t$ (time) in years from 1750 to 2100, and  
the initial states  $x_i(t = 1750) = x_i^0, i = 1, 2, \dots, 8$

and the model outputs:

$$\begin{aligned}
 y_1(t) &= x_1(t), y_2(t) = x_2(t), y_3(t) = x_3(t), y_4(t) = x_4(t), \\
 y_5(t) &= x_5(t), y_6(t) = x_6(t), y_7(t) = x_7(t), y_8(t) = x_8(t).
 \end{aligned} \tag{3.2}$$

**Table 3.2.** Model I reference case initial compartment contents (*in units of Gt C*).

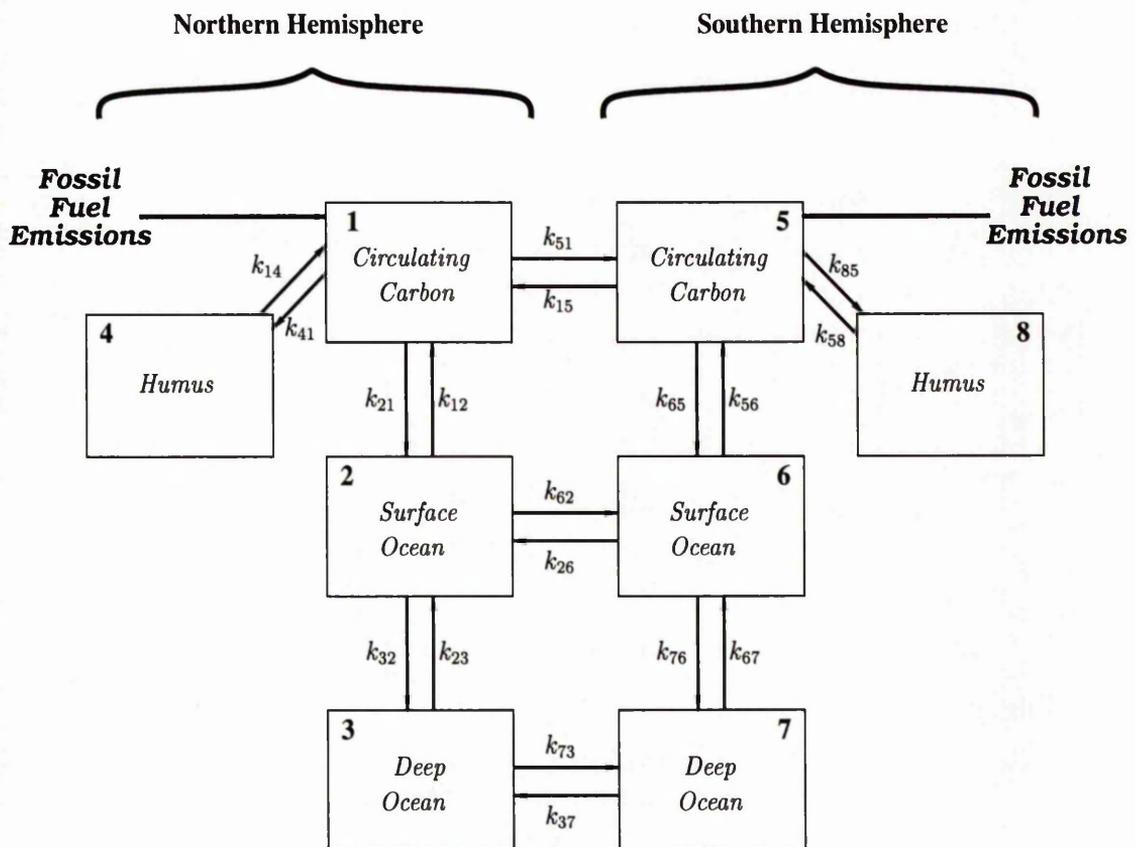
Description (Initial Conditions)	Input Factor	Nominal Value	Range
Atmosphere	$x_1^0$	622.40	497.92 - 746.88
Surface ocean	$x_2^0$	667.37	533.90 - 800.84
Deep ocean	$x_3^0$	37542.00	30033.60 - 45050.40
Nonwoody parts of trees	$x_4^0$	38.21	30.57 - 45.85
Woody parts of trees	$x_5^0$	634.47	507.58 - 761.36
Ground vegetation	$x_6^0$	59.32	47.46 - 71.18
Detritus / decomposers	$x_7^0$	108.22	86.58 - 129.86
Active soil carbon	$x_8^0$	1131.39	905.11 - 1357.67

**Table 3.3.** Model I reference case transfer coefficients (*in units of yr<sup>-1</sup>*) for carbon transfer among compartments.

Description (Transfer Coefficients)	Input Factor	Nominal Value	Range
Atmosphere → Surface Ocean	$k_{21}$	0.1582	0.1266 - 0.1898
Atmosphere → Nonwoody parts of trees	$k_{41}$	0.0354	0.0283 - 0.0425
Atmosphere → Woody parts of trees	$k_{51}$	0.0408	0.0326 - 0.0490
Atmosphere → Ground vegetation	$k_{61}$	0.0241	0.0193 - 0.0289
Surface ocean → Atmosphere	$k_{12}$	0.1476	0.1181 - 0.1771
Surface Ocean → Deep ocean	$k_{32}$	0.0473	0.0378 - 0.0568
Deep ocean → Surface ocean	$k_{23}$	0.0008	0.0006 - 0.0010
Nonwoody parts of trees → Detritus/decomposers	$k_{74}$	0.5758	0.4606 - 0.6910
Woody parts of trees → Detritus/decomposers	$k_{75}$	0.0353	0.0282 - 0.0424
Woody parts of trees → Active soil carbon	$k_{85}$	0.0047	0.0038 - 0.0056
Ground vegetation → Detritus/decomposers	$k_{76}$	0.1667	0.1334 - 0.2000
Ground vegetation → Active soil carbon	$k_{86}$	0.0862	0.0690 - 0.1034
Detritus/decomposers → Atmosphere	$k_{17}$	0.4688	0.3750 - 0.5626
Detritus/decomposers → Active soil carbon	$k_{87}$	0.0328	0.0262 - 0.0394
Active soil carbon → Atmosphere	$k_{18}$	0.0103	0.0082 - 0.0124

## 3.3.1.2 Description of Model II

This model originally developed by Kelly *et al.* (1975) is identical to that utilised by McCartney [78] and Bush *et al.* [8]. It is based on the global carbon cycle, and was mostly used to predict the possible effects of  $^{14}\text{C}$  discharges from the nuclear fuel cycle. The structure of the model, consisting of 8 compartments and 18 transfer coefficients, is given in Figure 3.6. The model is basically that of a four-compartment model separated into northern and southern hemispheres. The main 4 reservoirs are the 'circulating carbon', 'surface ocean', 'deep ocean', and the 'humus' compartments.



**Figure 3.6.** Compartment diagram of Global Carbon Cycle Model II (adopted from McCartney (1987) [78]).

Since  $^{14}\text{C}$ 's primary atmospheric form is gaseous as  $\text{CO}_2$  and it is closely involved in many physical and biological processes notably photosynthesis and exchange within the carbon cycle, we consider the same nominal values of parameters as those used by the model developers. The initial content of each compartment and the transfer coefficients are provided in Table 3.4 and Table 3.5, respectively. The input into the circulating carbon compartments is assumed to be distributed in the same proportion between Northern and Southern hemispheres as the present-day population distribution, i.e. 80% in the North and 20% in the South [78].

The set of differential equations representing Model II are as follows:

$$\begin{aligned}
 \dot{x}_1 &= k_{12}x_2 + k_{14}x_4 + k_{15}x_5 - (k_{21} + k_{41} + k_{51})x_1 + u_1(t) \\
 \dot{x}_2 &= k_{21}x_1 + k_{23}x_3 + k_{26}x_6 - (k_{12} + k_{32} + k_{62})x_2 \\
 \dot{x}_3 &= k_{32}x_2 + k_{37}x_7 - (k_{23} + k_{73})x_3 \\
 \dot{x}_4 &= k_{41}x_1 - k_{14}x_4 \\
 \dot{x}_5 &= k_{51}x_1 + k_{56}x_6 + k_{58}x_8 - (k_{15} + k_{65} + k_{85})x_5 + u_5(t) \\
 \dot{x}_6 &= k_{62}x_2 + k_{65}x_5 + k_{67}x_7 - (k_{26} + k_{56} + k_{76})x_6 \\
 \dot{x}_7 &= k_{73}x_3 + k_{76}x_6 - (k_{37} + k_{67})x_7 \\
 \dot{x}_8 &= k_{85}x_5 - k_{58}x_8
 \end{aligned} \tag{3.3}$$

with  $t$ (time) in years from 1750 to 2100, and

the initial states  $x_i(t = 1750) = x_i^0, i = 1, 2, \dots, 8$

and the model outputs:

$$\begin{aligned}
 y_1(t) &= x_1(t), y_2(t) = x_2(t), y_3(t) = x_3(t), y_4(t) = x_4(t), \\
 y_5(t) &= x_5(t), y_6(t) = x_6(t), y_7(t) = x_7(t), y_8(t) = x_8(t).
 \end{aligned} \tag{3.4}$$

Table 3.4. Model II reference case initial compartment contents (*in units of Gt C*).

Description (Initial Conditions)	Input Factor	Nominal Value	Range
Circulating carbon-(NH) <sup>a</sup>	$x_1^o$	325.21	260.17 - 390.25
Surface ocean-(NH)	$x_2^o$	448.31	358.65 - 537.97
Deep ocean-(NH)	$x_3^o$	12,426.00	9940.80 - 14911.20
Humus-(NH)	$x_4^o$	1042.30	833.84 - 1250.76
Circulating carbon-(SH) <sup>b</sup>	$x_5^o$	291.59	233.27 - 349.91
Surface ocean-(SH)	$x_6^o$	677.54	542.03 - 813.05
Deep ocean-(SH)	$x_7^o$	21,983.00	17,586.40 - 26,379.60
Humus-(SH)	$x_8^o$	356.21	284.97 - 427.45

<sup>a</sup>Northern Hemisphere<sup>b</sup>Southern HemisphereTable 3.5. Model II reference case transfer coefficients (*in units of yr<sup>-1</sup>*) for carbon transfer among compartments.

Description (Transfer Coefficients)	Input Factor	Nominal Value	Range
Circulating carbon (NH) → Surface Ocean (NH)	$k_{21}$	0.1400	0.1120 - 0.1680
Circulating carbon (NH) → Humus (NH)	$k_{41}$	0.0160	0.0128 - 0.0192
Circulating carbon (NH) → Circulating carbon (SH)	$k_{51}$	0.5000	0.4000 - 0.6000
Surface ocean (NH) → Circulating carbon (NH)	$k_{12}$	0.1000	0.0800 - 0.1200
Surface ocean (NH) → Deep ocean (NH)	$k_{32}$	0.0900	0.0720 - 0.1080
Surface ocean (NH) → Surface ocean (SH)	$k_{62}$	0.1000	0.0800 - 0.1200
Deep ocean (NH) → Surface ocean (NH)	$k_{23}$	0.0032	0.0026 - 0.0038
Deep ocean (NH) → Deep ocean (SH)	$k_{73}$	0.0050	0.0040 - 0.0060
Humus (NH) → Circulating carbon (NH)	$k_{14}$	0.0050	0.0040 - 0.0060
Circulating carbon (SH) → Circulating carbon (NH)	$k_{15}$	0.5600	0.4480 - 0.6720
Circulating carbon (SH) → Surface ocean (SH)	$k_{65}$	0.2300	0.1840 - 0.2760
Circulating carbon (SH) → Humus (SH)	$k_{85}$	0.0061	0.0049 - 0.0073
Surface ocean (SH) → Surface ocean (NH)	$k_{26}$	0.0660	0.0528 - 0.0792
Surface ocean (SH) → Circulating carbon (SH)	$k_{56}$	0.1000	0.0800 - 0.1200
Surface ocean (SH) → Deep ocean (SH)	$k_{76}$	0.0900	0.0720 - 0.1080
Deep ocean (SH) → Deep ocean (NH)	$k_{37}$	0.0028	0.0022 - 0.0034
Deep ocean (SH) → Surface ocean (SH)	$k_{67}$	0.0028	0.0022 - 0.0034
Humus (SH) → Circulating carbon (SH)	$k_{58}$	0.0050	0.0040 - 0.0060

### 3.4 Extension and Application of SA to GCC Model I

In this study, various SA methods grouped under three main settings are used. First, local SA which is based on the derivation of the differential equations describing a complete sensitivity matrix for each output variable with respect to each input factor is applied to the model.

Second, individually randomised one-factor-at-a-time (OAT) design is discussed, and data analysis based on the resulting random sample of observed elementary effects is presented. This standard OAT design and Morris's design described in Section 2.5.5 are used for screening purpose.

Third, global SA methods which are based on Monte Carlo simulations are carried out by following these three steps:

1. selection of random, independent sets of values for model input factors,
2. initialisation of the carbon cycle model, and
3. simulation of GCC dynamics between years 1750 and 2100.

Initialisation of the carbon cycle model involves a calibration step in which parameter values and initial conditions are calculated, consistent with an assumed steady-state for atmosphere, ocean and terrestrial biota  $\text{CO}_2$ .

To assess the effect of any one input on the model output one can either ignore the other inputs, i.e. fix them at their nominal values, or adjust them using some assumed condition. Since it is known that long before the industrial revolution the GCC was in balance, the latter situation must be taken into consideration when the two 8-compartment models are considered. Hence, as the first step of the analysis, models are set up such that steady-state is maintained, i.e. the flux of  $\text{CO}_2$  leaving compartment  $i$  is equal to the flux of  $\text{CO}_2$  coming into that compartment, before we start perturbing the system with any input. In other words, we assume that  $dx_i/dt = 0$  in year 1750 when model simulations are started.

Because the model codes of the two 8-compartment models are not supplemented with a calibration routine we have to find a computational way to ensure initial steady-state which is an important modelling assumption. Two different procedures are developed to maintain steady-state condition when we are concerned with the sensitivities of compartment contents to the uncertainties about the initial conditions, and about the transfer coefficients. In both procedures the main idea is to adjust the initial compartment contents and the transfer coefficients so that steady-state condition is satisfied.

For the initial conditions, when the model code is run for a long period of time without any perturbations to the cycle, the model eventually reaches a steady-state at which point fossil-fuel and forest clearing emissions are introduced into the model through the atmosphere compartment and the model output is calculated. This procedure was followed for every model run. Although, for the base-line case, the compartmental contents did not vary much from their original values, this process puts the system in balance. Thus, we can assume that preindustrial emissions were low and relatively constant over a long time period prior to the initial simulation date.

The procedure followed when the transfer coefficients are concerned is based on solving the homogeneous system of equations using the method of Gauss-Jordan elimination. Considering Model I, we have 8 linear equations in 15 unknowns (transfer coefficients), and for Model II we have 8 linear equations in 18 unknowns. Before we go further, we shall introduce notation that makes it easier to explain the steps in the procedure. The nominal values of the compartmental contents are considered as entries of a matrix  $\mathbf{X}$  which will be referred to as the coefficient matrix of the system. The vector of the transfer coefficients denoted by  $\mathbf{k}$  is the vector of unknowns. Hence, the model equations in matrix notation can be written as  $\mathbf{Xk} = \mathbf{0}$ .



To find the solution of  $\mathbf{Xk} = \mathbf{0}$ , first the elementary row operations<sup>c</sup> are applied to the rows of  $\mathbf{X}$  until the matrix is in reduced row echelon form. In a matrix in reduced row echelon form, the non-zero rows come first and the first non-zero entry in those rows are the pivots. The unknowns (transfer coefficients), which are the elements of  $\mathbf{k}$ , are separated into two groups. One group is made up of the *basic variables*, those that correspond to columns with pivots. The other group is made up of the *free variables*, corresponding to columns without pivots. After reaching reduced row echelon form and identifying the basic and the free variables, the next step involves solving the simplified system of equations for the basic variables in terms of free variables. In this process the system is solved in reverse order, from the last equation to the first, by substituting each newly computed value into the previous equation. This process called back substitution continues until all basic variables are computed. As an illustration, the solution of the homogeneous system of equations of Model I is given below. Note that in the matrix notation of the system of equations the symbols of the elements of the coefficient matrix  $\mathbf{X}$  are used instead of their actual numerical values to make the row operations easier.

After applying elementary row operations to the coefficient matrix given in Equation 3.5, the matrix becomes

$$\begin{pmatrix} \boxed{x_1^o} & x_1^o & x_1^o & x_1^o & -x_2^o & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -x_7^o & 0 & -x_8^o \\ 0 & \boxed{x_1^o} & 0 & 0 & 0 & 0 & 0 & -x_4^o & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \boxed{x_1^o} & 0 & 0 & 0 & 0 & 0 & -x_5^o & -x_5^o & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \boxed{x_1^o} & 0 & 0 & 0 & 0 & 0 & 0 & -x_6^o & -x_6^o & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \boxed{x_2^o} & -x_3^o & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \boxed{x_4^o} & x_5^o & 0 & x_6^o & 0 & -x_7^o & -x_7^o & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \boxed{x_5^o} & 0 & x_6^o & 0 & x_7^o & -x_8^o \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

<sup>c</sup>Elementary row operations are arithmetic operations applied to a matrix representation of a system of equations: 1. Multiplying (or dividing) one row by a non-zero number, 2. Adding a multiple of one row to another row, 3. Interchanging two rows.

In this echelon matrix, the first, second, third, fourth, sixth, eighth and tenth columns contain the pivots (given in boxes), so the corresponding elements of  $\mathbf{k}$  vector, i.e. the transfer coefficients  $k_{21}$ ,  $k_{41}$ ,  $k_{51}$ ,  $k_{61}$ ,  $k_{32}$ ,  $k_{74}$ ,  $k_{85}$  are the basic variables. The elements of the free variables group are the transfer coefficients corresponding to the columns of the above matrix without pivots, i.e.  $k_{12}$ ,  $k_{23}$ ,  $k_{75}$ ,  $k_{76}$ ,  $k_{86}$ ,  $k_{17}$ ,  $k_{87}$ ,  $k_{18}$  are the free variables. Having identified the free and the basic variables, now we can determine the basic variables in terms of the free variables by back-substitution. Proceeding upward,

$$x_5^{\circ} k_{85} + x_6^{\circ} k_{86} + x_7^{\circ} k_{87} - x_8^{\circ} k_{18} = 0$$

$$\text{yields } k_{85} = [-x_6^{\circ} k_{86} - x_7^{\circ} k_{87} + x_8^{\circ} k_{18}] / x_5^{\circ}$$

$$x_4^{\circ} k_{74} + x_5^{\circ} k_{75} + x_6^{\circ} k_{76} - x_7^{\circ} (k_{17} + k_{87}) = 0$$

$$\text{yields } k_{74} = [-x_5^{\circ} k_{75} - x_6^{\circ} k_{76} + x_7^{\circ} (k_{17} + k_{87})] / x_4^{\circ}$$

$$x_2^{\circ} k_{32} - x_3^{\circ} k_{23} = 0$$

$$\text{yields } k_{32} = [x_3^{\circ} k_{23}] / x_2^{\circ}$$

$$x_1^{\circ} k_{61} - x_6^{\circ} (k_{76} + k_{86}) = 0$$

$$\text{yields } k_{61} = [x_6^{\circ} (k_{76} + k_{86})] / x_1^{\circ} \quad (3.6)$$

$$x_1^{\circ} k_{51} - x_5^{\circ} (k_{75} + k_{85}) = 0$$

$$\text{yields } k_{51} = [x_5^{\circ} k_{75} - x_6^{\circ} k_{86} - x_7^{\circ} k_{87} + x_8^{\circ} k_{18}] / x_1^{\circ}$$

$$x_1^{\circ} k_{41} - x_4^{\circ} k_{74} = 0$$

$$\text{yields } k_{41} = [-x_5^{\circ} k_{75} - x_6^{\circ} k_{76} + x_7^{\circ} (k_{17} + k_{87})] / x_1^{\circ}$$

$$x_1^{\circ} (k_{21} + k_{41} + k_{51} + k_{61}) - x_2^{\circ} k_{12} - x_7^{\circ} k_{17} - x_8^{\circ} k_{18} = 0$$

$$\text{yields } k_{21} = [x_2^{\circ} k_{12}] / x_1^{\circ}$$

replacing  $x_i^o$ , ( $i = 1, 2, \dots, 8$ )'s with their actual values gives

$$\begin{aligned}
 k_{85} &= [-59.32 k_{86} - 108.22 k_{87} + 1131.39 k_{18}]/634.47 \\
 k_{74} &= [-634.47 k_{75} - 59.32 k_{76} + 108.22 (k_{17} + k_{87})]/38.21 \\
 k_{32} &= [37542.00 k_{23}]/667.37 \\
 k_{61} &= [59.32 (k_{76} + k_{86})]/622.40 \\
 k_{51} &= [634.47 k_{75} - 59.32 k_{86} - 108.22 k_{87} + 1131.39 k_{18}]/622.40 \\
 k_{41} &= [-634.47 k_{75} - 59.32 k_{76} + 108.22 (k_{17} + k_{87})]/622.40 \\
 k_{21} &= [667.37 k_{12}]/622.40.
 \end{aligned}$$

Thus the solution to  $\mathbf{Xk} = \mathbf{0}$ , after carrying out the arithmetic in the above equations, can now be written as

$$\begin{pmatrix} k_{21} \\ k_{41} \\ k_{51} \\ k_{61} \\ k_{12} \\ k_{32} \\ k_{23} \\ k_{74} \\ k_{75} \\ k_{85} \\ k_{76} \\ k_{86} \\ k_{17} \\ k_{87} \\ k_{18} \end{pmatrix} = \begin{pmatrix} 1.07 k_{12} \\ -1.02 k_{75} - 0.10 k_{76} + 0.17 [k_{17} + k_{87}] \\ 1.02 k_{75} - 0.10 k_{86} - 0.17 k_{87} + 1.82 k_{18} \\ 0.10 [k_{76} + k_{86}] \\ k_{12} \\ 56.25 k_{23} \\ k_{23} \\ 16.60 k_{75} - 1.55 k_{76} + 2.83 [k_{17} + k_{87}] \\ k_{75} \\ -0.09 k_{86} - 0.17 k_{87} + 1.78 k_{18} \\ k_{76} \\ k_{86} \\ k_{17} \\ k_{87} \\ k_{18} \end{pmatrix}. \quad (3.7)$$

When the transfer coefficients of Model II is considered in the initialisation process, the solution to  $\mathbf{Xk} = \mathbf{0}$ , by following the same procedure above, is found as

$$\begin{pmatrix} k_{21} \\ k_{41} \\ k_{51} \\ k_{12} \\ k_{32} \\ k_{62} \\ k_{23} \\ k_{73} \\ k_{14} \\ k_{15} \\ k_{65} \\ k_{85} \\ k_{26} \\ k_{56} \\ k_{76} \\ k_{37} \\ k_{67} \\ k_{58} \end{pmatrix} = \begin{pmatrix} 1.38 k_{12} - 0.90 k_{65} + 2.08 k_{56} \\ 3.21 k_{14} \\ 0.90 (k_{15} + k_{65}) - 2.08 k_{56} \\ k_{12} \\ 27.72 k_{23} - 1.51 k_{76} + 49.04 k_{67} \\ -0.65 k_{65} + 1.51 (k_{26} + k_{56} + k_{76}) - 49.04 k_{67} \\ k_{23} \\ -0.05 k_{76} + 1.77 (k_{37} + k_{67}) \\ k_{14} \\ k_{15} \\ k_{65} \\ 1.22 k_{58} \\ k_{26} \\ k_{56} \\ k_{76} \\ k_{37} \\ k_{67} \\ k_{58} \end{pmatrix} \quad (3.8)$$

with the basic variables being  $k_{21}$ ,  $k_{41}$ ,  $k_{51}$ ,  $k_{32}$ ,  $k_{62}$ ,  $k_{73}$ ,  $k_{85}$ , and the free variables being  $k_{12}$ ,  $k_{23}$ ,  $k_{14}$ ,  $k_{15}$ ,  $k_{65}$ ,  $k_{26}$ ,  $k_{56}$ ,  $k_{76}$ ,  $k_{37}$ ,  $k_{67}$  and  $k_{58}$ .

In the beginning of each model run, first the free variables are randomly generated from uniform distribution over the specified ranges given in Table 3.3 (for Model I) and in Table 3.5 (for Model II). Then, using Equations (3.7) when

Model I and Equations (3.8) when Model II is under consideration, the values for the basic variables are calculated.

After the initialisation process which puts the system in steady-state, the anthropogenic release of CO<sub>2</sub> from the combustion of fossil fuels and changes in the land-use mainly from deforestation are introduced into the model through the atmosphere compartment and then global carbon cycle dynamics are simulated using the two linear, time-invariant compartmental models introduced in Section 3.3.1. Each of the model runs is initialised in 1750 and integrated for 350 years to the year 2100. The quantities of carbon stored in each compartment are calculated at annual time intervals. Simple random sampling was used to generate the input matrix. To assess if the sample size considered in the model simulations has any impact on the sensitivities of model outputs we consider two different sample sizes  $N=100$  and  $N=5,000$  in the simulations. In the sensitivity analysis of the model outputs, historical input data (see Section 3.2.1) and IPCC's future emission scenario IS92a, which is based on plausible assumptions about future population and economic growths and energy supplies (see Section 3.2.2), are considered as the model input.

## 3.5 Results and Discussion on 8-Compartment GCC Model I

### 3.5.1 Screening Methods

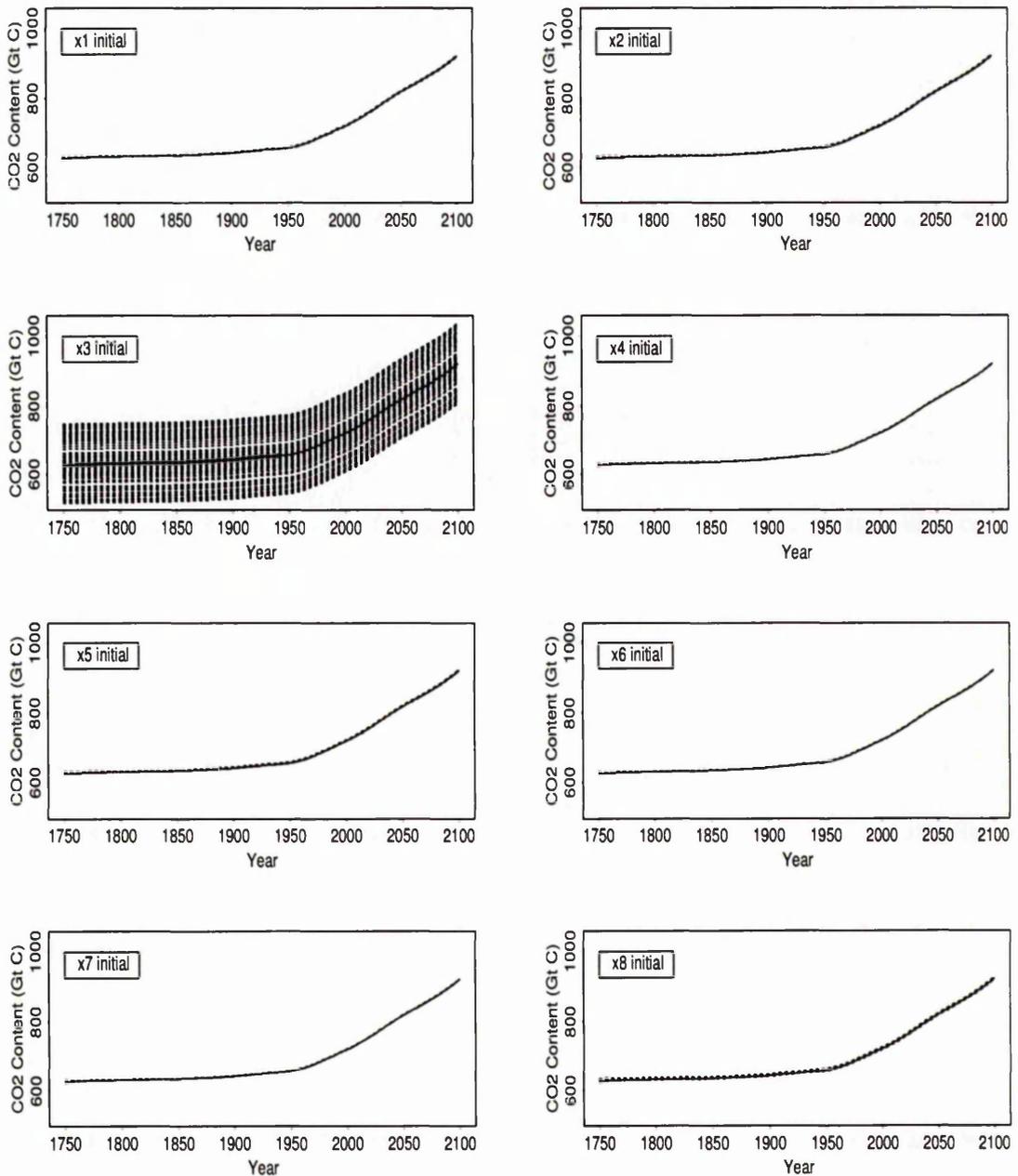
#### 3.5.1.1 Standard OAT Design on Initial Conditions

First, the initial condition of each compartment ( $x_1^o, x_2^o, \dots, x_8^o$ ) is considered in the analysis. These model input factors are listed in Table 3.2 together with their nominal values and ranges. Given the lack of information about the distributions of these initial conditions the sensitivity analysis experiments have been carried out assuming all the initial conditions follow a uniform distribution.

Figure 3.7 illustrates the variation in time dependent behaviour of the atmosphere compartment due to the variation about each initial condition. The variation in the time dependent behaviour of compartmental CO<sub>2</sub> contents of the other seven compartments are presented in Appendix A (see Figures A.1 - A.7). Each dashed-line curve in these graphs corresponds to the prediction associated with one of the 100 Monte Carlo sample vectors, and the solid line curve in each of these graphs is based on the nominal values of the input factors.

Each graph of Figure 3.7 shows the predictions resulting from varying the initial CO<sub>2</sub> content of each compartment ( $x_1^o, \dots, x_8^o$ ) one-at-a-time while holding the others at their nominal values, and the initial condition which is varied is specified at the top-left corner of each graph. It is clear from this figure that the variability in the initial condition of deep ocean compartment  $x_3^o$  has a dramatic effect on the output of the atmosphere compartment during the whole time period. Even though it is not as obvious as the effect of  $x_3^o$ , the variation in  $x_8^o$ ,  $x_2^o$  and  $x_1^o$  also seem to be affecting the output. The atmospheric output does not seem to be very sensitive to the rest of the input factors, all 100 prediction curves lie roughly on the base-line curve.

The output of the other seven compartments also appear to be influenced by the variation in  $x_3^o$  the most. Because there is much more carbon in the deep ocean than in any other compartment, it is apparent that the amount of CO<sub>2</sub> in all compartments is being controlled to a considerable extent by the initial condition of the deep ocean compartment. The output of both ocean compartments especially the deep ocean compartment are least affected by the variation in  $x_4^o$ ,  $x_6^o$  and  $x_7^o$ . These ocean compartments also seem to be sensitive to the variation in  $x_8^o$ ,  $x_5^o$ ,  $x_2^o$  and  $x_1^o$ . The sensitivity of the active soil carbon compartment output to these four initial conditions is more obvious. We can also see that except in the deep ocean and active soil carbon compartments CO<sub>2</sub> content of each compartment is increasing rapidly starting around 1980s. Because these two carbon reservoirs are distinguished from those of other carbon reservoirs in their functions as long-term carbon sinks, the change in the carbon content of



**Figure 3.7.** Atmospheric CO<sub>2</sub> predictions resulting from varying initial compartmental content  $x_i^0$  of compartment  $i$  ( $i = 1, 2, \dots, 8$ ) OAT (given at top-left corner of each graph - see Table 3.2 for description of these input factors).  $N=100$  model simulations, IS92a emission scenario is considered. In each graph solid line represents the base-line case and dashed lines represent the predictions.

these compartments is not as rapid, especially in the deep ocean compartment since transport and mixing processes in this reservoir is very slow. The initial conditions of the terrestrial compartments, except the active soil carbon compartment, do not influence CO<sub>2</sub> levels in any of the compartments much.

Very qualitatively, such pictures allow the analyst to investigate the effect of each input factor (varied OAT) on a given output. They also show which outputs affected by which inputs in the sense of variation in the outputs.

**Sensitivity Index** The sensitivity index (SI) was introduced by Hoffman & Gardner [54] and it is simply calculated by substituting the minimum and maximum (from Table 3.2) for the nominal value of the  $j$ th input factor (the  $j$ th initial condition, in this case), while holding all other initial conditions at their nominal values, to produce a maximum and minimum value of the compartment  $i$  ( $y_{ij}^{max}, y_{ij}^{min}$ ).  $SI_{ij}$  accounts for all possible values when determining input factor sensitivities.

Because the compartmental outputs are time dependent functions of initial conditions, to determine whether their importance changes through time we examine the results from years 1900, 2000 and 2100. Each  $SI_{ij}$  is derived using

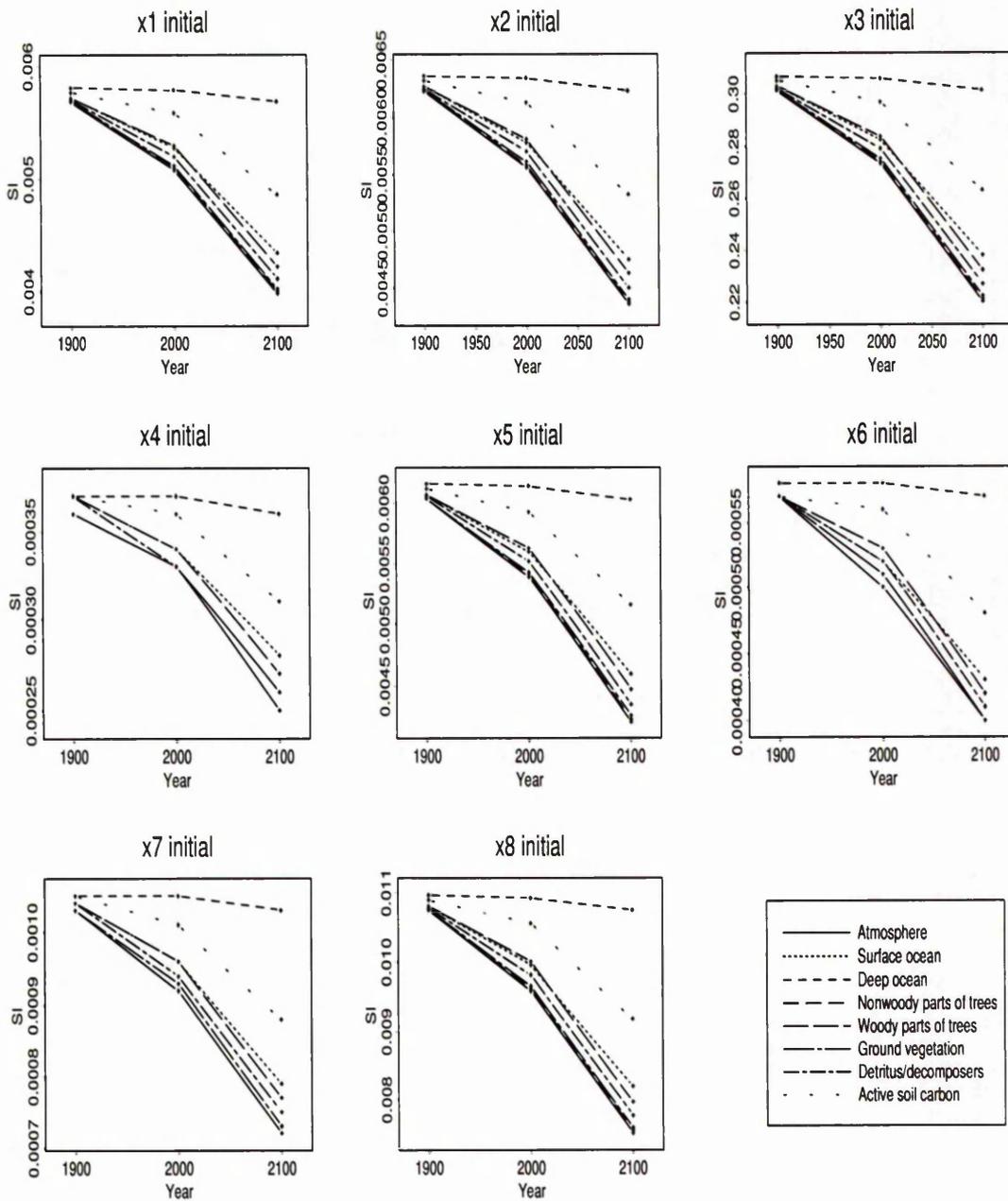
$$SI_{ij} = 1 - [y_{ij}^{min}(t)/y_{ij}^{max}(t)].$$

For example, raising the nominal value of  $x_1^o$  from 622.40 to its maximum of 746.88 produces  $y_{21}^{max}(t = 2100) = 895.37$  Gt C. Lowering the value of  $x_1^o$  to its minimum value 497.92 produces  $y_{21}^{min}(t = 2100) = 891.30$  Gt C. The surface ocean  $SI_{21}$  for initial condition  $x_1^o$  is therefore  $SI_{21} = 1 - (891.30/895.37) = 0.0045$ .

A sensitivity index can take values between 0 and 1. A  $SI_{ij}$  of 1.0 indicates that the  $i$ th model output has maximal variation to changes in values of the  $j$ th input factor, whereas a  $SI_{ij}$  of less than 0.01 on the other hand indicates that the  $i$ th output is not very sensitive to the changes in the  $j$ th input factor. The calculated SIs are presented graphically in Figure 3.8. The sensitivity of the eight

compartments' output to the range of each initial condition at three years are plotted in each graph, with lines joining the SI of each compartmental output. It is clear from these line plots that the sensitivity of each compartment's content to changes in the range of all initial conditions decreases with time. The decrease in the sensitivities of the outputs of the deep ocean and the active soil carbon compartments is less compared to the decrease for the other compartments. It is evident in this figure that for all eight compartmental contents in all three years, uncertainties in model predictions will mainly be dominated by changes in the initial condition  $x_3^0$  (see top-right graph in Figure 3.8) as it is the only initial condition that gives SI of greater than 0.01.

To determine the order of importance of the change in the initial conditions on each compartment in years 1900, 2000 and 2100, the SIs are ranked and given in Table 3.6. The ranking within each compartment in years 1900 and 2100 does not change, whereas for all compartmental outputs the ranking in year 2000 changes slightly. All model outputs, i.e. all compartmental contents in three chosen years, seem to be influenced by the variation in  $x_3^0$  followed by  $x_8^0$  and  $x_2^0$ .



**Figure 3.8.** Sensitivity indices of compartmental outputs to the range of initial conditions. (The results given are calculated using Model I, and IS92a emission scenario).

**Table 3.6.** Rankings of Model I compartmental output sensitivities to the range of the initial conditions based on SI calculated in years 1900, 2000 and 2100.

Compartmental Output	Input Factor	Sensitivity Index (SI)		
		Yr 1900	Yr 2000	Yr 2100
Atmosphere	W <sub>1</sub> <sup>0</sup>	5	6	5
	W <sub>2</sub> <sup>0</sup>	3	3	3
	W <sub>3</sub> <sup>0</sup>	1	1	1
	W <sub>4</sub> <sup>0</sup>	8	8	8
	W <sub>5</sub> <sup>0</sup>	4	4.5	4
	W <sub>6</sub> <sup>0</sup>	7	7	7
	W <sub>7</sub> <sup>0</sup>	6	4.5	6
	W <sub>8</sub> <sup>0</sup>	2	2	2
Surface Ocean	W <sub>1</sub> <sup>0</sup>	5	6	5
	W <sub>2</sub> <sup>0</sup>	3	3	3
	W <sub>3</sub> <sup>0</sup>	1	1	1
	W <sub>4</sub> <sup>0</sup>	8	8	8
	W <sub>5</sub> <sup>0</sup>	4	4.5	4
	W <sub>6</sub> <sup>0</sup>	7	7	7
	W <sub>7</sub> <sup>0</sup>	6	4.5	6
	W <sub>8</sub> <sup>0</sup>	2	2	2
Deep Ocean	W <sub>1</sub> <sup>0</sup>	5	6	5
	W <sub>2</sub> <sup>0</sup>	3	3	3
	W <sub>3</sub> <sup>0</sup>	1	1	1
	W <sub>4</sub> <sup>0</sup>	8	8	8
	W <sub>5</sub> <sup>0</sup>	4	4.5	4
	W <sub>6</sub> <sup>0</sup>	7	7	7
	W <sub>7</sub> <sup>0</sup>	6	4.5	6
	W <sub>8</sub> <sup>0</sup>	2	2	2
Nonwoody parts of Trees	W <sub>1</sub> <sup>0</sup>	5	6	5
	W <sub>2</sub> <sup>0</sup>	3	3	3
	W <sub>3</sub> <sup>0</sup>	1	1	1
	W <sub>4</sub> <sup>0</sup>	8	8	8
	W <sub>5</sub> <sup>0</sup>	4	4.5	4
	W <sub>6</sub> <sup>0</sup>	7	7	7
	W <sub>7</sub> <sup>0</sup>	6	4.5	6
	W <sub>8</sub> <sup>0</sup>	2	2	2
Woody parts of Trees	W <sub>1</sub> <sup>0</sup>	5	6	5
	W <sub>2</sub> <sup>0</sup>	3	3	3
	W <sub>3</sub> <sup>0</sup>	1	1	1
	W <sub>4</sub> <sup>0</sup>	8	8	8
	W <sub>5</sub> <sup>0</sup>	4	4.5	4
	W <sub>6</sub> <sup>0</sup>	7	7	7
	W <sub>7</sub> <sup>0</sup>	6	4.5	6
	W <sub>8</sub> <sup>0</sup>	2	2	2
Ground Vegetation	W <sub>1</sub> <sup>0</sup>	5	6	5
	W <sub>2</sub> <sup>0</sup>	3	3	3
	W <sub>3</sub> <sup>0</sup>	1	1	1
	W <sub>4</sub> <sup>0</sup>	8	8	8
	W <sub>5</sub> <sup>0</sup>	4	4.5	4
	W <sub>6</sub> <sup>0</sup>	7	7	7
	W <sub>7</sub> <sup>0</sup>	6	4.5	6
	W <sub>8</sub> <sup>0</sup>	2	2	2
Detritus/Decomposers	W <sub>1</sub> <sup>0</sup>	5	6	5
	W <sub>2</sub> <sup>0</sup>	3	3	3
	W <sub>3</sub> <sup>0</sup>	1	1	1
	W <sub>4</sub> <sup>0</sup>	8	8	8
	W <sub>5</sub> <sup>0</sup>	4	4.5	4
	W <sub>6</sub> <sup>0</sup>	7	7	7
	W <sub>7</sub> <sup>0</sup>	6	4.5	6
	W <sub>8</sub> <sup>0</sup>	2	2	2
Active Soil Carbon	W <sub>1</sub> <sup>0</sup>	5	6	5
	W <sub>2</sub> <sup>0</sup>	3	3	3
	W <sub>3</sub> <sup>0</sup>	1	1	1
	W <sub>4</sub> <sup>0</sup>	8	8	8
	W <sub>5</sub> <sup>0</sup>	4	4.5	4
	W <sub>6</sub> <sup>0</sup>	7	7	7
	W <sub>7</sub> <sup>0</sup>	6	4.5	6
	W <sub>8</sub> <sup>0</sup>	2	2	2

The SI only addresses input factor sensitivity relative to the point estimates of an input factor. Next, using another sensitivity measure we examine sensitivities with regard to the entire input factor distributions.

**Standardised Range** This statistic was calculated using predictions from the standard OAT design. First, the range of each compartmental content (say, compartment  $i$ ) resulting from varying each input factor (say, initial condition  $j$ ) over its range while leaving all other initial conditions at their nominal values, i.e.  $y_{ij}^{max} - y_{ij}^{min}$  is calculated at year  $t$ . Then, this output range of compartment  $i$  is standardised by the nominal value of the corresponding compartment's initial content. To assess the effect of the number of model runs on the results  $N=100$  and  $N=5,000$  model simulations are considered.

The standardised ranges (SRs) of compartments due to varying each initial condition OAT are demonstrated as dot charts in Figures 3.9 and 3.10. These figures allow us to judge how compartmental contents differ in terms of SRs within the same year and also between the years. In each frame of these figures, we have superimposed the SRs computed on the predictions of the model outputs based on  $N=100$  and  $N=5,000$  model runs, so that the influence of the number of model runs on the SRs can easily be visualized.

Even though the range of SR values are not the same, the degree of sensitivity of each compartment to the uncertainty in the initial conditions of atmosphere ( $x_1^o$ ) and both ocean compartments ( $x_2^o$  and  $x_3^o$ ) does not appear to be changing, that is, the variation in all these three initial conditions influences the ground vegetation compartment the most, the nonwoody parts of trees compartment the second most, and these compartments are followed by the active soil carbon, atmosphere and the detritus/decomposers compartments (see Figure 3.9). The surface ocean, deep ocean and the woody parts of trees compartments are the least influenced by these three initial conditions.

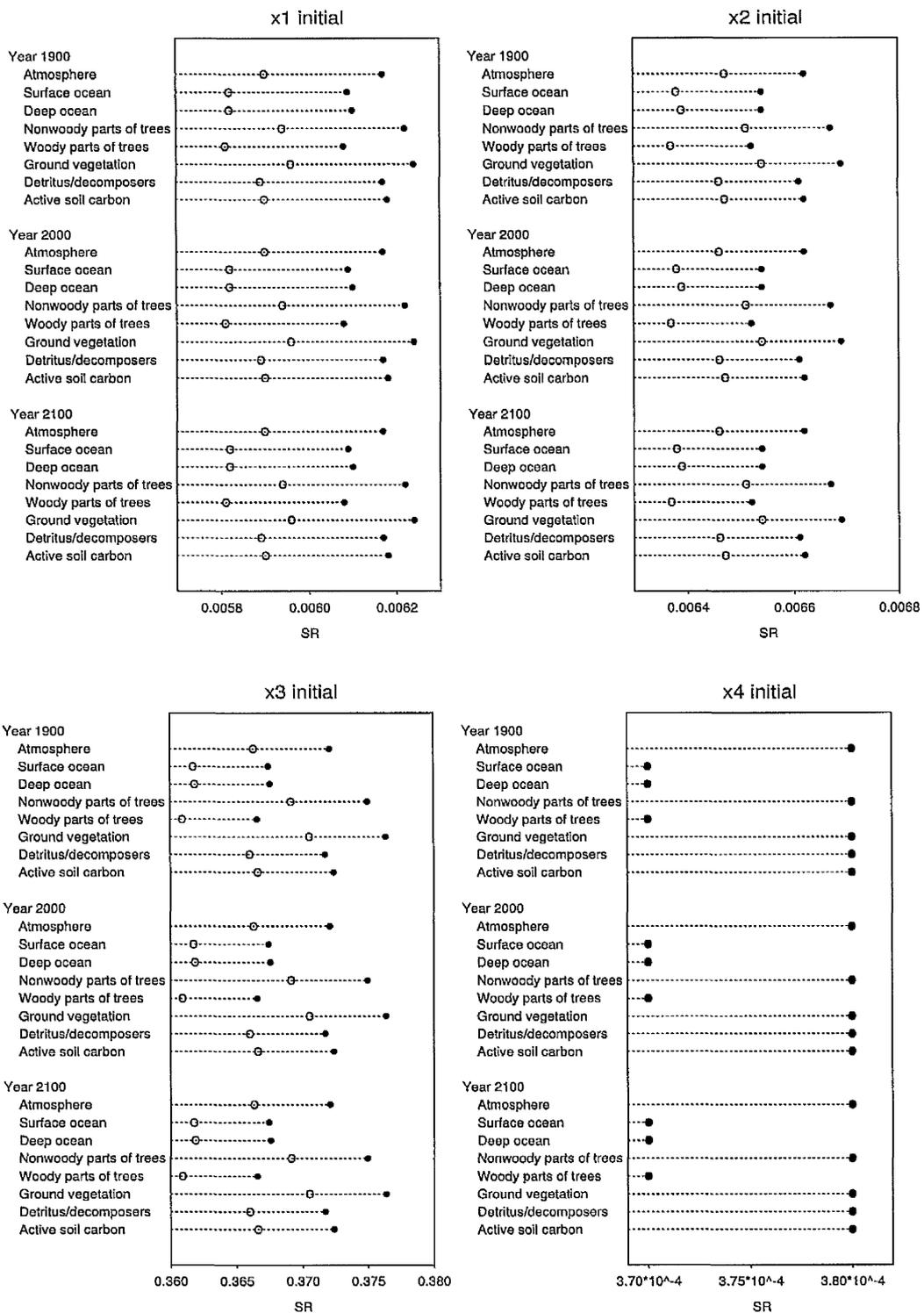
Compared to the SR results on  $x_3^o$ , the SRs of all the compartments due to variation in the other seven initial conditions are much smaller (lower than 0.1), especially the results due to uncertainty about  $x_4^o$  and  $x_6^o$  (see the lower-right frame in Figure 3.9 and the upper-right frame in Figure 3.10). Considering how small the ranges of SRs are it is difficult to say that change in  $x_4^o$  and  $x_6^o$  are effective on any of the compartments at all.

As Figures 3.9 and 3.10 show the SRs hardly change from year to year no matter which  $N$  is used. The highest SR values result from the OAT design conducted on  $x_3^o$ , which are in the range of 0.361 and 0.3708 when  $N=100$ , and between 0.367 and 0.3765 when  $N=5,000$  (see lower-left frame in Figure 3.9).

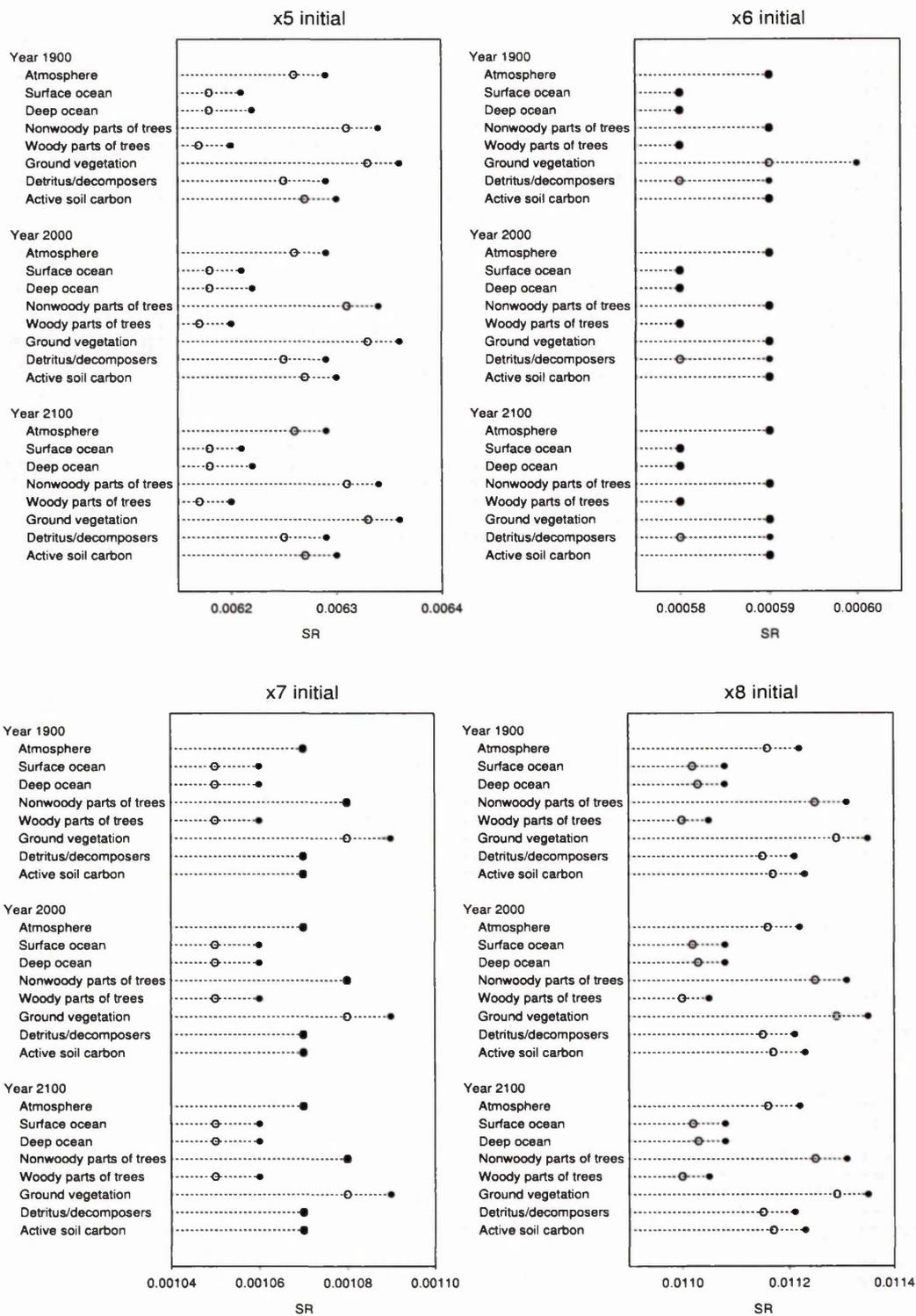
As with  $x_1^o$ ,  $x_2^o$  and  $x_3^o$ , the variation in the initial conditions of the woody parts of trees ( $x_5^o$ ) and active soil carbon ( $x_8^o$ ) compartments appear to influence the ground vegetation, nonwoody parts of trees, active soil carbon, atmosphere and detritus/decomposers compartments more (in descending order of sensitivity in terms of SRs) (see Figure 3.10). The effect of  $x_5^o$  and  $x_8^o$  on the two ocean and the woody parts of trees is relatively low.

In the dotplot showing the SRs obtained from the analysis where only  $x_7^o$  was varied over its entire range, we see that with the results based on  $N=100$  model runs the variation in  $x_7^o$  is equally and most effective on non-woody parts of trees and ground vegetation compartments followed closely by atmosphere, detritus/decomposers and active soil carbon compartments; and the ocean and woody parts of trees compartments are the least influenced by  $x_7^o$ . When 5,000 model runs are considered, there is no change in the SR of four of the compartments, namely nonwoody parts of trees, atmosphere, detritus/decomposers and active soil carbon compartments. For the other three compartments the SRs increase but only by 0.00001.

Again using the SRs of compartmental contents at three chosen years, we can also assess how the order of importance of the initial conditions change within the same compartment, i.e. which input factors are more influential on a compartmental content than others. It is clear from the dotcharts of the SRs given in Figures 3.9 and 3.10 that  $x_3^o$  is the most influential initial condition on all compartmental outputs and it is followed by  $x_8^o$ ,  $x_2^o$ ,  $x_5^o$ ,  $x_1^o$ ,  $x_7^o$ ,  $x_6^o$  and  $x_4^o$  in this given order. This order of relative importance based on the SR values does not change with the number of model runs and time.



**Figure 3.9.** Dotcharts showing how each compartmental output of Model I in years 1900, 2000 and 2100 is effected by the variation in the x1-x4 initial conditions in terms of Standardised Ranges. In each frame (○) show results from N=100 and (●) from N=5,000 model runs.



**Figure 3.10.** Dotcharts showing how each compartmental output of Model I in years 1900, 2000 and 2100 is effected by the variation in the x5-x8 initial conditions in terms of Standardised Ranges. In each frame (○) show results from N=100 and (●) from N=5,000 model runs.

### 3.5.1.2 Morris Design on Initial Conditions

Now, we continue our screening analysis by applying the Morris design to Model I. Here we are first taking into consideration the initial conditions as model input factors. Due to our lack of knowledge regarding the distribution of each input factor, in the analyses we assume that the distribution followed by each factor is uniform. In the application of Morris method, such an assumption is beneficial, i.e. since the levels of the experiment are obtained by dividing in equal parts the interval in which each factor varies, the statistical information contained in the distribution functions will not be lost. As noted by Campolongo *et al.* [15], all the examples of application of this screening design available in the literature are based on the assumption that the input factors follow a uniform distribution (examples can be found in Refs. [9], [12] and [87]). They also note that sensitivity measures provided by the Morris method are only qualitative, i.e. measures capable of ranking the input factors in order of importance.

As described in Section 2.5.5, in this screening test data analysis is based on examination of the finite distributions of elementary effects. In order to estimate the mean and standard deviation of the distribution, from each  $F_i$  we collected a random sample of size  $r = 10$  by using ten independently generated orientation matrices. These orientation matrices were generated using  $p = 4$ . In comparison to the standard OAT screening the Morris method is less expensive and laborious. Based on the  $n = 90$  computed values of each output variable, a random sample of ten elementary effects was observed for each of the 8 input factors. Then, the sample mean and standard deviation which are unbiased estimators of the mean and standard deviation of the distribution of  $F_i$  for input factor  $i$  are calculated.

The analysis was carried out on all eight compartmental outputs in years 1900, 2000 and 2100. The estimated Morris mean  $\mu$  and standard deviation  $\sigma$  values for the 8 initial conditions are reported in Table A.1 (see Appendix A). In order to establish a general order of relative importance for the initial conditions within each compartment at a specific year we used a sensitivity measure, the Euclidean distance from the origin, introduced by Campolongo & Gabric [10].

The higher the distance, the more important the initial condition is. Here we give the results of the Morris screening exercise on the atmosphere compartment at three chosen years in Figure 3.11. In this figure the eight initial conditions  $x_1^0, \dots, x_8^0$  are labeled as x10, ..., x80. Similarly, figures displaying the results of Morris design on the other seven compartments have been produced but since the pattern in those figures is almost the same as the one shown in Figure 3.11, we have not included all these figures here but the results on all eight model outputs are summarised in Table 3.7 in terms of importance ranking.

As seen in Figure 3.11, input factor  $x_3^0$  is clearly separated from the cluster of the remaining initial conditions. Hence, considering both means and standard deviations together, we can conclude that the deep ocean compartment being the largest compartment of the model its initial condition  $x_3^0$  dominate the results and  $x_3^0$  appear to be the most important initial condition for the atmosphere compartment at all three years. This is also the case with the other compartmental outputs. In Figure 3.11, we first draw the plots with all initial conditions, and because of effect of the high mean values for  $x_3^0$ , it is difficult to visualize the relative importance of the other seven factors. Therefore, we redraw the plots in the absence of  $x_3^0$ .

The ranking of the initial conditions according to the Morris mean, which is identical to the ranking of the Euclidean distance from the origin, for each compartmental output is given in Table 3.7. Note that because the ranking of the estimated mean values for the same compartment do not change from year to year, in the table we do not give the rankings from all three chosen years. Table 3.7 shows that the order of importance between the initial conditions is the same for all compartments.

As seen in Figure 3.11 for the atmosphere compartment and summarized in Table 3.7, in terms of ranks, for all compartments after  $x_3^0$  the second most important initial condition appears to be  $x_8^0$  which is followed by  $x_2^0$ ,  $x_5^0$  and  $x_1^0$ . None of the compartments appears to be sensitive to  $x_4^0$ ,  $x_6^0$  and  $x_7^0$  which have estimated morris mean and standard deviations close to zero.

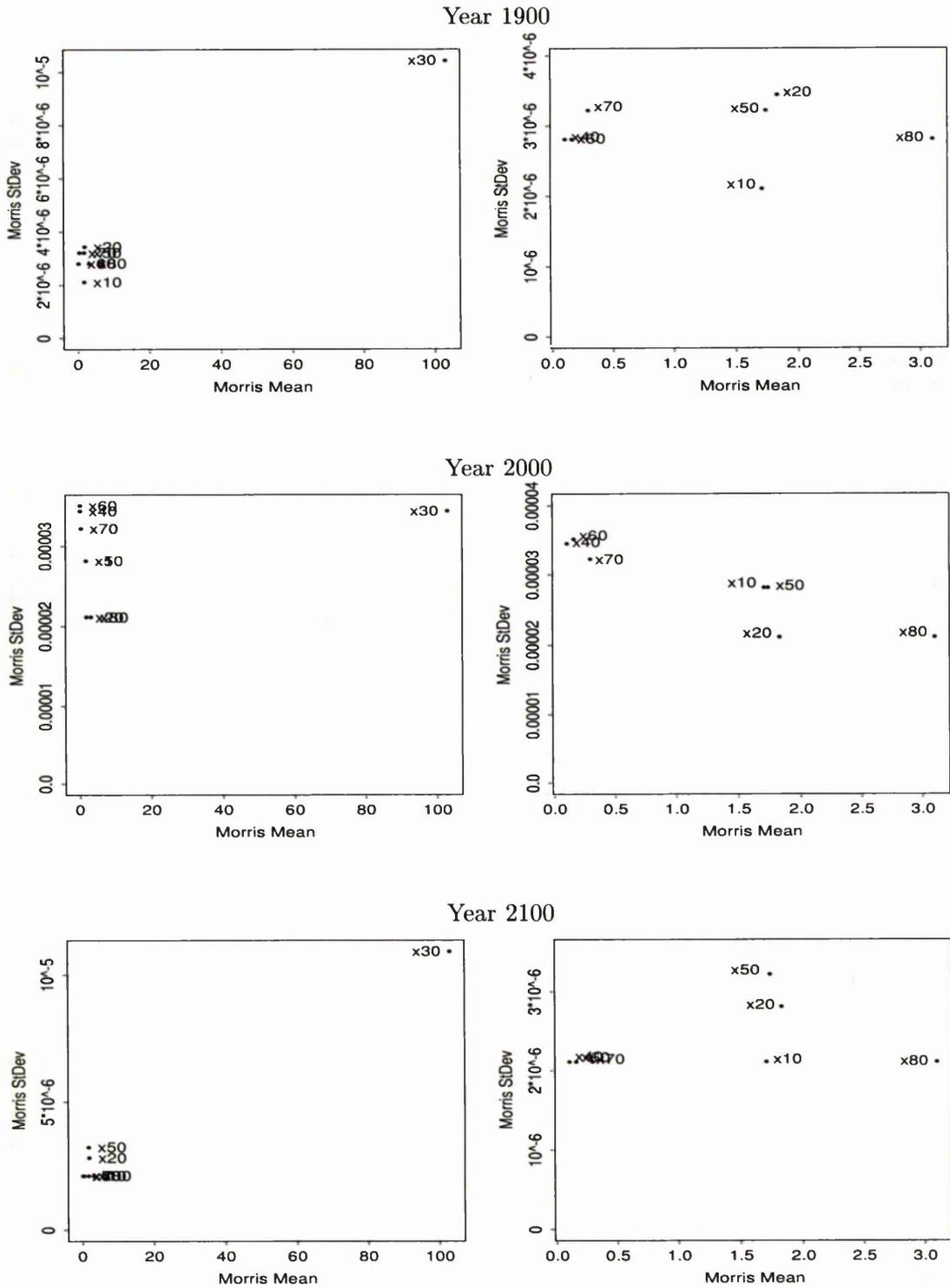


Figure 3.11. Morris screening results on Atmosphere compartment of Model I in years 1900, 2000 and 2100. Mean and standard deviations are associated with the initial conditions considered in the analysis. The panels on the right of the figure display the results excluding the point corresponding to  $x_3^0$ .

**Table 3.7.** Results of Morris experiment on Model I. Initial conditions are ranked in order of importance according to the SA measures of Morris mean  $\mu$ .

Compartmental Output	Input Factor	Morris Rank	Compartmental Output	Input Factor	Morris Rank
Atmosphere	$x_1$	5	Woody parts of Trees	$x_1$	5
	$x_3$	3		$x_3$	3
	$x_2$	1		$x_2$	1
	$x_4$	8		$x_4$	8
	$x_7$	4		$x_7$	4
	$x_6$	7		$x_6$	7
Surface Ocean	$x_5$	6	Ground Vegetation	$x_5$	6
	$x_8$	2		$x_8$	2
	$x_3$	5		$x_3$	5
	$x_1$	3		$x_1$	3
	$x_2$	1		$x_2$	1
	$x_4$	8		$x_4$	8
Deep Ocean	$x_7$	4	Detritus/Decomposers	$x_7$	4
	$x_6$	7		$x_6$	7
	$x_5$	6		$x_5$	6
	$x_8$	2		$x_8$	2
	$x_3$	5		$x_3$	5
	$x_1$	3		$x_1$	3
Nonwoody parts of Trees	$x_2$	1	Active Soil Carbon	$x_2$	1
	$x_4$	8		$x_4$	8
	$x_7$	4		$x_7$	4
	$x_6$	7		$x_6$	7
	$x_5$	6		$x_5$	6
	$x_8$	2		$x_8$	2

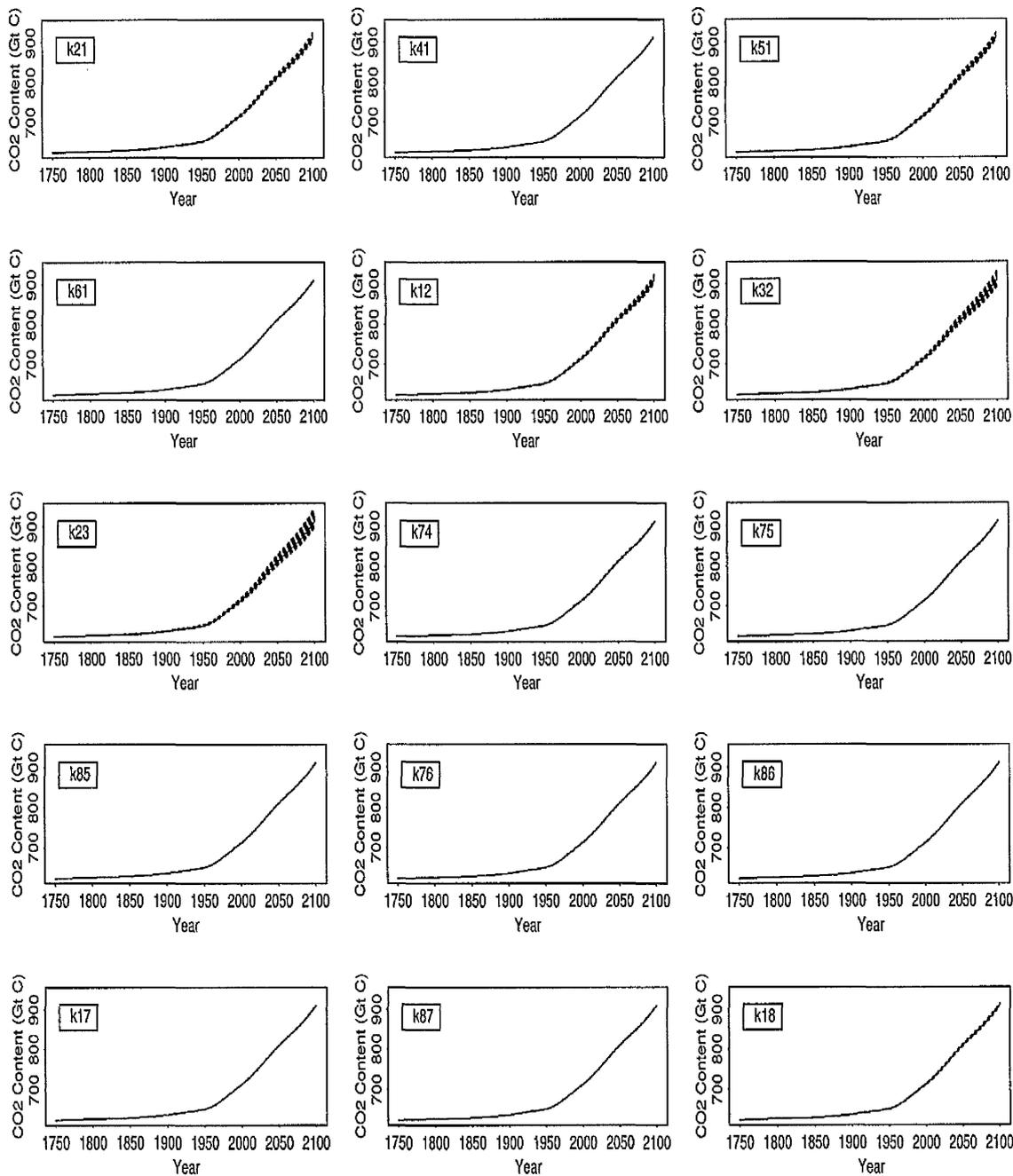
### 3.5.1.3 Standard OAT Design on Transfer Coefficients

The specific form of Model I we consider now is based on fixing the initial conditions of each compartment at their nominal values and treating the transfer coefficients as the uncertain model input factors. Using standard OAT design on the model transfer coefficients, we now want to assess the degree of influence they have on the compartmental outputs. Because of the lack of knowledge about the distribution the transfer coefficients follow, we assume that each of these fifteen transfer coefficients follow uniform distribution over their ranges given in Table 3.3 with their description.

The simple random sampling regime and two different number of model runs,  $N=100$  and  $N=5,000$ , are considered in the analysis. Having performed the required number of model runs using the model and the input data, which is obtained by randomly varying each transfer coefficients OAT over its entire range while keeping the rest at their nominal values, we then determine the influence of each transfer coefficient on the model outputs from the eight compartments both quantitatively and qualitatively. To study if the sensitivity of each compartment to the transfer coefficient changes with time we analyze the outputs from years 1900, 2000 and 2100. Here we shall use the notation  $k_{ij}$  when we refer to the transfer coefficients.

Figure 3.12 shows the simulated time dependent behaviour of the atmosphere compartment over the period 1750 - 2100 obtained with  $N=100$  Monte Carlo simulations by varying each transfer coefficient OAT. At the top-left corner of each graph the varied input factor is specified. In each graph, the 100 prediction curves are presented along with the base-line curves. In order to see the effect of variation in each  $k_{ij}$  the scale on the y-axis kept the same. It is clear from this figure that the variation in  $k_{23}$  and  $k_{32}$  influence the atmospheric content the most.  $k_{21}$ ,  $k_{12}$  and  $k_{51}$  also appear to be potentially important transfer coefficients for the atmosphere compartment. The sensitivity of atmospheric output to these transfer coefficients seem to increase with time.

The time dependent behaviour of the other seven compartments due to varying the transfer coefficients OAT also simulated and the predictions are plotted in Figures A.8 - A.14 (see Appendix A for these figures). As with the atmosphere compartment, all the other compartments except the active soil carbon compartment are influenced by the variation in  $k_{32}$  and  $k_{23}$  the most (Figures A.8 - A.13), the effect of the changes in  $k_{21}$ ,  $k_{51}$  and  $k_{12}$  on these seven compartments is also evident. While it does not show any significant effect on the other compartments,  $k_{17}$  appears to be having some influence on the detritus/decomposers compartment. Figure A.14 shows that the active soil compartment is sensitive to the variation in all  $k_{ij}$ s but with different degrees. The most influential transfer



**Figure 3.12.** Atmospheric CO<sub>2</sub> predictions resulting from varying transfer coefficients  $k_{ij}$  OAT (given in top-left corner of each graph - see Table 3.3 for description of these input factors).  $N=100$  model simulations, IS92a emission scenario is considered. In each graph solid line represents the base-line case and dashed lines represent the predictions.

coefficient for this compartment is  $k_{51}$  which is followed by  $k_{18}$ ,  $k_{61}$ ,  $k_{32}$ ,  $k_{23}$ ,  $k_{75}$  and  $k_{12}$ . The other transfer coefficients do not appear to have significant effect on the output of the active soil carbon compartment.

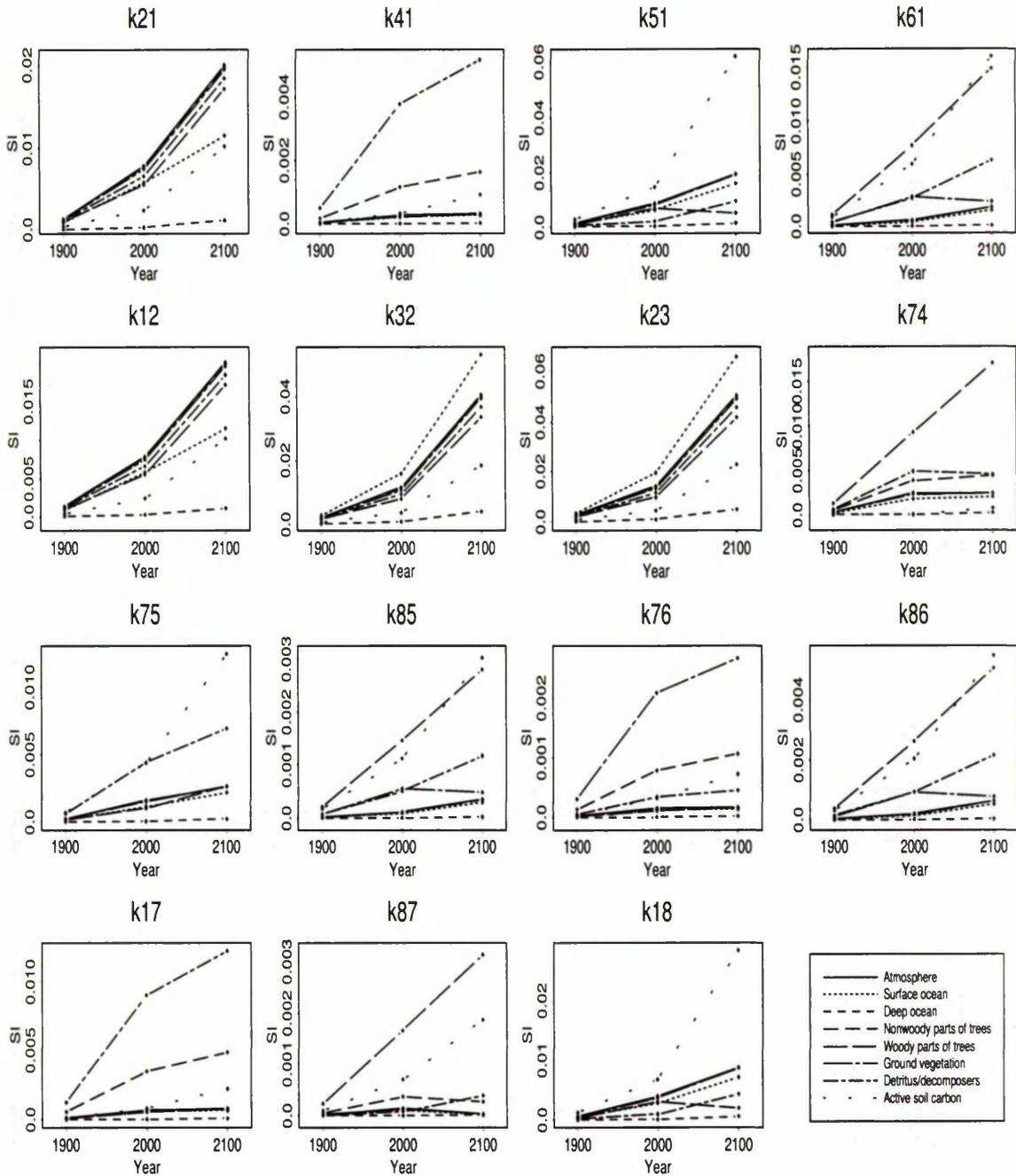
Even though the sensitivity of all compartments to the variation in the transfer coefficients appear to be changing with time, for each compartment the order of importance of the transfer coefficients does not seem to be changing through time. The change of input factor influence on each compartment with time will be examined further with some numerical sensitivity measures.

**Sensitivity Index** Considering all compartmental contents in 1900, 2000 and 2100 the results of SIs performed on the ranges of the 15 transfer coefficients when each of them was varied OAT are presented in Figure 3.13. In each graph of this figure, we give the sensitivities of all compartmental contents of the three years to one of the  $k_{ij}$ s. The SI of each compartment from the three years are connected using a different line type each representing one compartment. Unlike the sensitivities of the compartments to the initial conditions, here the sensitivities tend to increase with time for most of the compartments.

In 1900, no compartment is sensitive to the change in any of the  $k_{ij}$ s. In 2000, even though all the SIs show some degree of increase, sensitivity of most of the compartments to any of the transfer coefficients is still not very large, but SI of some compartments in this year is  $> 0.01$ . The sensitivity of the active soil carbon compartment to  $k_{51}$  is 0.015 and increases to 0.062 in 2100. The SIs of atmosphere, surface ocean, non-woody parts of trees and ground vegetation compartments due to the change in  $k_{51}$  in 2100 are also above 0.01.

In 2000, the atmosphere, surface ocean, non-woody parts of trees, ground vegetation and detritus/decomposers compartments have SI values for  $k_{23}$  and  $k_{32}$  greater than 0.01, and they increase in 2100. The active soil carbon compartment has a low SI in 2000 but it becomes sensitive to both  $k_{23}$  and  $k_{32}$  in 2100.

The results of the OAT designs on  $k_{21}$  and  $k_{12}$  show that in 2100 all the compartments except deep ocean are sensitive to the ranges of these two transfer



**Figure 3.13.** Sensitivity indices of compartmental outputs to each transfer coefficient. (The results given are calculated using Model I, and IS92a emission scenario).

coefficients. The change in  $k_{61}$  shows an effect on active soil carbon and woody parts of trees compartments in 2100. Only woody parts of trees compartment content in 2100 is sensitive to the change in  $k_{74}$ . The changes in  $k_{75}$  and  $k_{18}$  seem to be having an effect on only active soil carbon compartment in 2100. The results due to change in  $k_{17}$  show that only ground vegetation compartment is sensitive to this transfer coefficient and this sensitivity is significant only in 2100.

None of the compartmental contents is sensitive to the ranges of the transfer coefficients  $k_{41}$ ,  $k_{85}$ ,  $k_{76}$ ,  $k_{86}$  and  $k_{87}$  at any of the three chosen years.

In Figure 3.13, we have shown how the range of each  $k_{ij}$  is influencing all 8 compartments. Now, by using the same SI calculations we want to summarize the results in a different way to show if the order of importance between the  $k_{ij}$ s changes for each compartment at different years. Considering all SI values which are  $< 0.07$  as indicating low sensitivity and in the interest of shortening the discussion, we list the most important three  $k_{ij}$ s ranked as first(1), second(2) and third(3) in Table 3.8.

Except for the active soil carbon compartment at all three years and woody parts of trees compartment in 2000, the most influential transfer coefficients on the other seven compartments are  $k_{23}$  and  $k_{32}$ , respectively.  $k_{51}$  appears to be influencing the active soil carbon compartment the most in all three years whereas it is ranked as the third most influential transfer coefficient on atmosphere compartment (in 1900 and 2000), on both ocean compartments (in all three years) on non-woody parts of trees and ground vegetation compartments (in 1900 and 2000). In year 2100, the influence of  $k_{21}$  and  $k_{12}$  ranked as third for atmosphere compartment exceeds the influence of  $k_{51}$  on this compartment, and this is the case with non-woody parts of trees compartment as well. For ground vegetation compartment in 2100  $k_{12}$  is ranked as the third important transfer coefficient. The sensitivity of woody parts of trees compartment content to the change in  $k_{74}$  is ranked as third in 1900 but it became more sensitive to this transfer coefficient in 2000 where  $k_{74}$  was ranked as the second and  $k_{32}$  as the third most important factor. In 2100, however, the influence of  $k_{74}$  on this compartment became less

**Table 3.8.** Three most effective transfer coefficients from SI - Transfer Coefficients of Model I are varied OAT.

Compartmental Output	Input Factor	Sensitivity Index (SI)		
		Year 1900	Year 2000	Year 2100
Atmosphere	1	$k_{23}$	$k_{23}$	$k_{23}$
	2	$k_{32}$	$k_{32}$	$k_{32}$
	3	$k_{51}$	$k_{51}$	$k_{21}, k_{12}$
Surface Ocean	1	$k_{23}$	$k_{23}$	$k_{23}$
	2	$k_{32}$	$k_{32}$	$k_{32}$
	3	$k_{51}$	$k_{51}$	$k_{51}$
Deep Ocean	1	$k_{23}$	$k_{23}$	$k_{23}$
	2	$k_{32}$	$k_{32}$	$k_{32}$
	3	$k_{51}$	$k_{51}$	$k_{51}$
Nonwoody parts of Trees	1	$k_{23}$	$k_{23}$	$k_{23}$
	2	$k_{32}$	$k_{32}$	$k_{32}$
	3	$k_{51}$	$k_{51}$	$k_{21}, k_{12}$
Woody parts of Trees	1	$k_{23}$	$k_{23}$	$k_{23}$
	2	$k_{32}$	$k_{74}$	$k_{32}$
	3	$k_{74}$	$k_{32}$	$k_{21}, k_{12}$
Ground Vegetation	1	$k_{23}$	$k_{23}$	$k_{23}$
	2	$k_{32}$	$k_{32}$	$k_{32}$
	3	$k_{51}$	$k_{51}$	$k_{12}$
Detritus/Decomposers	1	$k_{23}$	$k_{23}$	$k_{23}$
	2	$k_{32}$	$k_{32}$	$k_{32}$
	3	$k_{17}$	$k_{17}$	$k_{12}$
Active Soil Carbon	1	$k_{51}$	$k_{51}$	$k_{51}$
	2	$k_{18}$	$k_{18}$	$k_{18}$
	3	$k_{61}$	$k_{61}$	$k_{23}$

significant than the influence of  $k_{23}$ ,  $k_{32}$ ,  $k_{21}$  and  $k_{12}$ .

After  $k_{23}$  and  $k_{32}$ , the third most important transfer coefficient for detritus/decomposers compartment is  $k_{17}$  (in 1900 and 2000) but  $k_{12}$  in 2100.

The rankings from years 1900 and 2000 for active soil carbon compartment is the same:  $k_{51}$  being the first,  $k_{18}$  the second and  $k_{61}$  the third most important transfer coefficients, but in 2100,  $k_{23}$  becomes the third most influential factor after  $k_{51}$  and  $k_{18}$ .

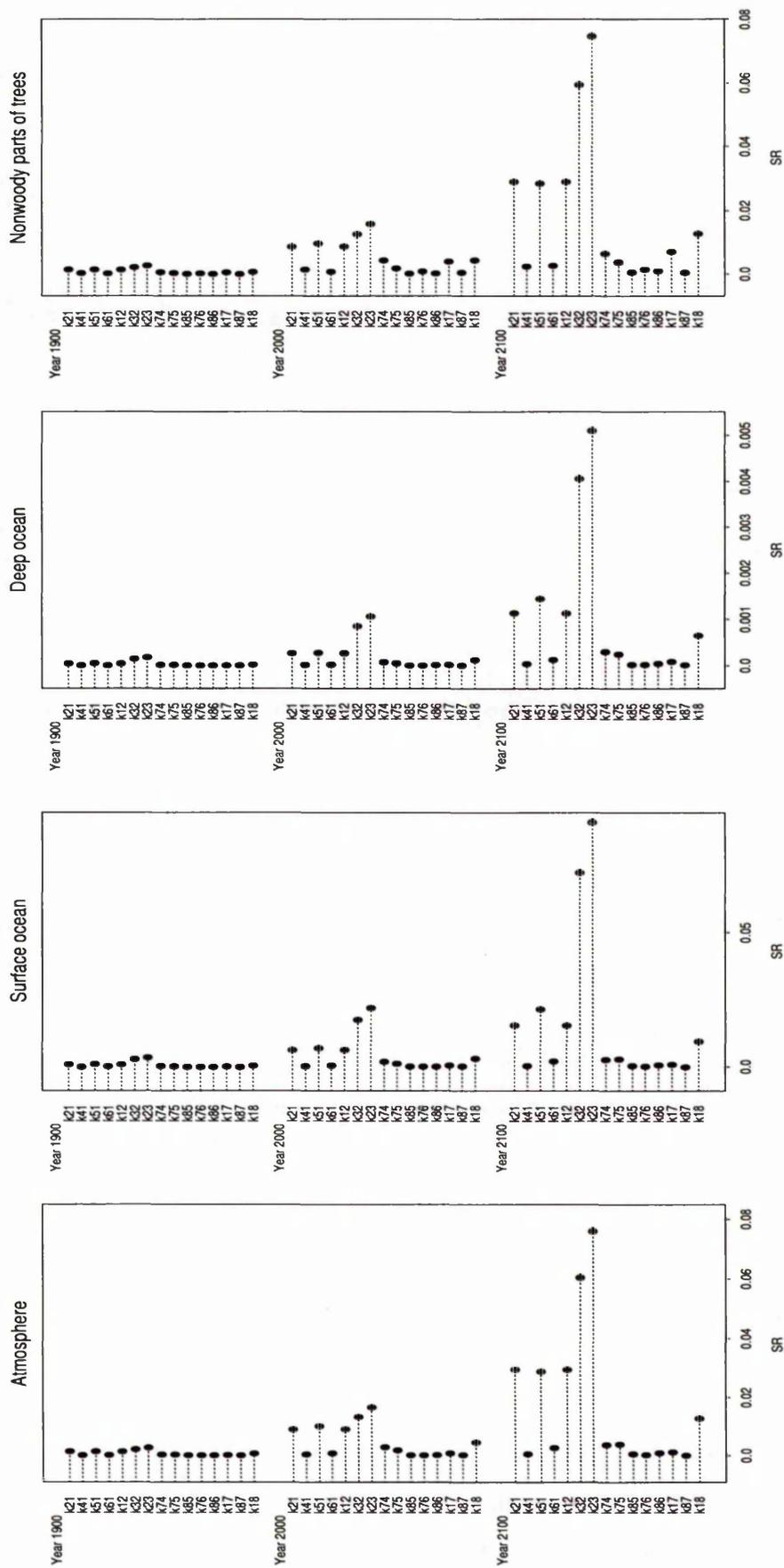
**Standardised Range** We use the output obtained from performing standard OAT design on the 15 transfer coefficients with 100 and 5,000 model evaluations

to calculate standardised rankings which provide another sensitivity measure. The results given in Figures 3.14 - 3.15 show how the SR of each compartment is affected by the change in the 15  $k_{ij}$ s at the three years and by the sample sizes considered in the simulations. In each dotchart of these figures we focus on one compartment. As the (o) represent the SRs from  $N=100$  and (●) from  $N=5,000$  model runs overlap, we point out that there is no noticeable change in the results when we consider different number of model iterations in the analysis, however, the results show some change with time for some of the  $k_{ij}$ s.

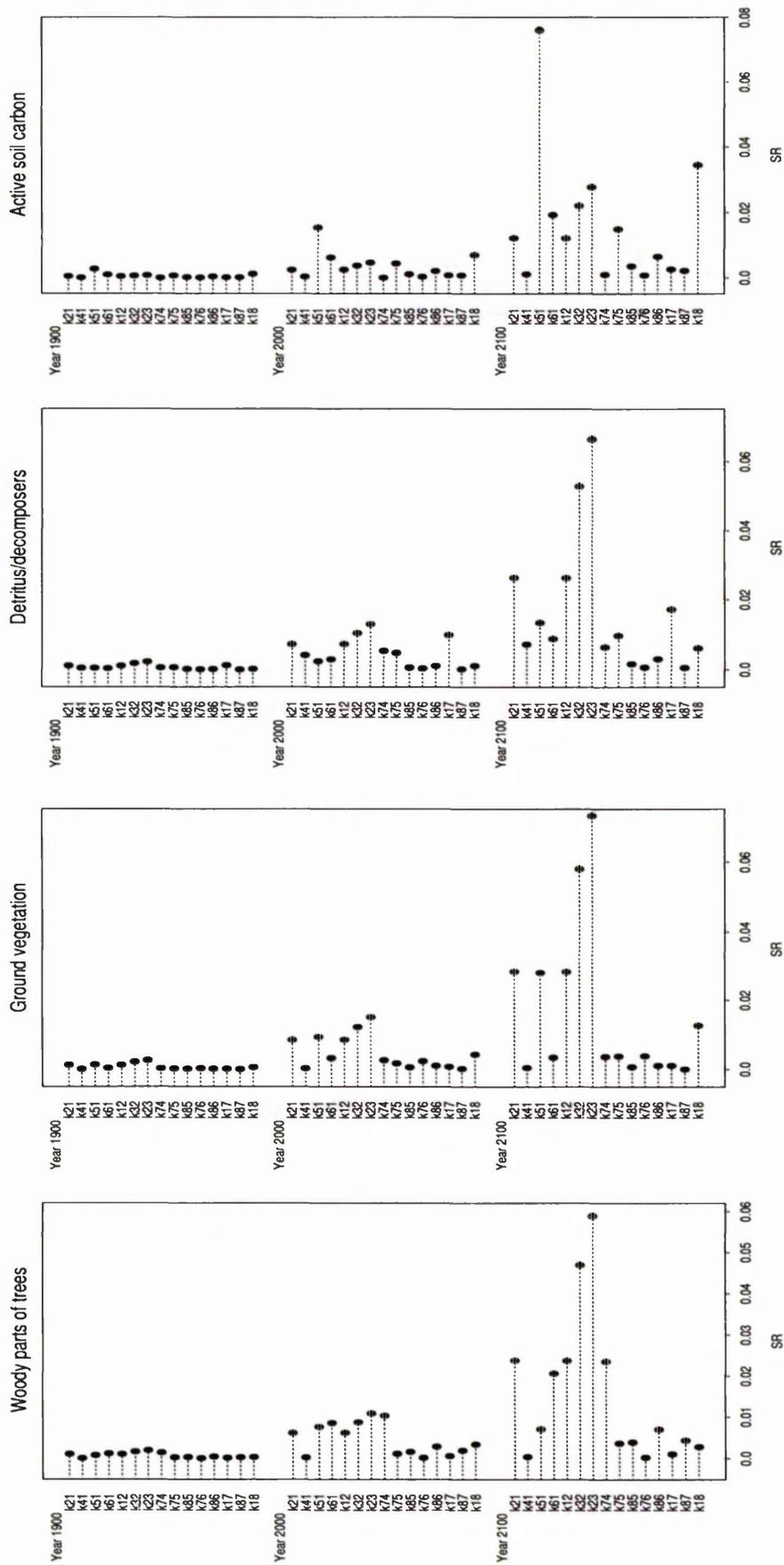
The SRs of each compartment in 1900 due to the changes in all the transfer coefficients hardly differs one from the other. They are very close or equal to zero, and this seems to be the case with all compartments. Year 2000 results for atmosphere and nonwoody parts of trees compartments (see the corresponding dotcharts in Figure 3.14) show a small increase in SR values due to  $k_{23}$ ,  $k_{32}$ ,  $k_{51}$ ,  $k_{21}$  and  $k_{12}$ . The increase with the same transfer coefficients is even higher in 2100.  $k_{18}$  also has some degree of influence on these two compartments in 2100. The effect of the other  $k_{ij}$ s have on the SR values in 2000 and 2100 is not significant.

As examination of the surface ocean and deep ocean dotcharts given in Figure 3.14 show both of these compartments present the same picture with different range of SR values.  $k_{23}$  and  $k_{32}$  are the transfer coefficients effecting the SR of these ocean compartments the most, and their influence on these two compartments is even higher in 2100. The effect of  $k_{51}$ ,  $k_{21}$ ,  $k_{12}$  and  $k_{18}$  on the SR values in 2100 should also be noted.

As seen in the first frame of Figure 3.15, the SR of woody parts of trees compartment is increased a little by the effect of  $k_{23}$ ,  $k_{74}$ ,  $k_{32}$ ,  $k_{61}$ ,  $k_{51}$ ,  $k_{21}$  and  $k_{12}$  (in the given order) in 2000. In 2100, however, the influence of these  $k_{ij}$ s on the results becomes more obvious but in a slightly different importance order;  $k_{23}$  being the most important followed by  $k_{32}$ ,  $k_{21}$ ,  $k_{12}$ ,  $k_{74}$  and  $k_{61}$ . The dotchart of ground vegetation compartment given in Figure 3.15 presents a quite similar picture to what we see in the figures corresponding to the ocean compartments



**Figure 3.14.** Dotcharts showing how output of Atmosphere, Surface ocean, Deep ocean and Nonwoody parts of trees compartments in years 1900, 2000 and 2100 are effected by the variation in the transfer coefficients  $k_{ij}$  in terms of Standardised Ranges. In each frame (○) show results from  $N=100$  and (●) from  $N=5,000$  model runs.



**Figure 3.15.** Dotcharts showing how output of Woody parts of trees, Ground vegetation, Detritus/decomposers and Active soil carbon compartments in years 1900, 2000 and 2100 are effected by the variation in the transfer coefficients  $k_{ij}$  in terms of Standardised Ranges. In each frame (o) show results from N=100 and (●) from N=5,000 model runs.

(see Figure 3.14). The scale for the SR values is different here but the influence of  $k_{23}$ ,  $k_{32}$ ,  $k_{51}$ ,  $k_{21}$ ,  $k_{21}$  and  $k_{18}$  on the SR of this compartment, specially when year 2100 is considered, is clear.

The detritus/decomposers compartment results presented in Figure 3.15 show that  $k_{23}$  and  $k_{32}$  are again the most influential transfer coefficients, and their effect on 2100 results is even higher. For this compartment,  $k_{17}$  is also among the most important input factors along with  $k_{21}$  and  $k_{12}$ . Unlike the other compartments, the SR of active soil carbon compartment is influenced by  $k_{51}$  the most, and its influence on the results increases with time. We should also note the reasonably high effect of  $k_{18}$ ,  $k_{23}$ ,  $k_{32}$  and  $k_{61}$  (in this order) on the SR of this compartment.

From the dotcharts of the SRs, we can easily obtain an importance ranking between the transfer coefficients within each compartment at all three years.

#### 3.5.1.4 Morris Design on Transfer Coefficients

Here, the Morris design has been performed considering eight of the 15 transfer coefficients as input factors and eight compartmental contents in years 1900, 2000 and 2100 as the model output variables of Model I.

The set of 8 transfer coefficients ( $k_{12}$ ,  $k_{23}$ ,  $k_{75}$ ,  $k_{76}$ ,  $k_{86}$ ,  $k_{17}$ ,  $k_{87}$ ,  $k_{18}$ ) taken into consideration here is the set obtained from Gauss-Jordan elimination procedure that we developed to maintain the steady-state condition for the model (see Section 3.4). In the Morris experiment a sample size  $r = 10$  is used. Each of the 8 transfer coefficients is assumed to follow a uniform distribution over its assigned uncertainty range (given in Table 3.3). In the design, each transfer coefficient is varied across  $l = 4$  levels. A total number of  $N = 90$  model evaluations is performed.

The values of the Morris mean and standard deviations, for each of the 8 output variables at the three chosen years, are shown in Table A.2 (given in Appendix A), and these values are displayed in Figures 3.16 - 3.18. Since the plots for the surface and deep ocean compartments show a very similar pattern as the plot of the atmosphere compartment given in Figure 3.16, we have not

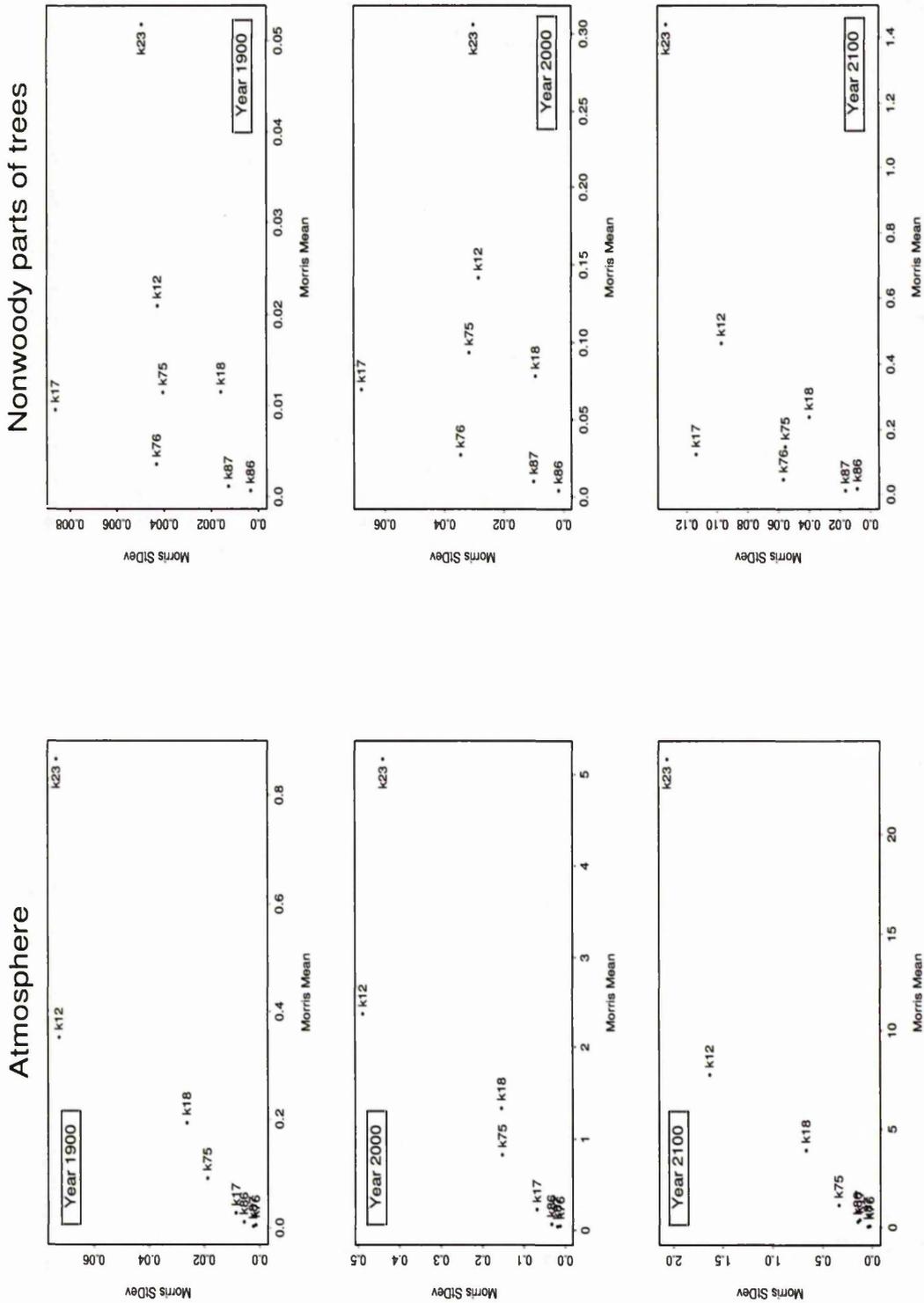


Figure 3.16. Morris screening results on Atmosphere and Nonwoody parts of trees compartments of Model I in 1900, 2000 and 2100. Mean and standard deviations associated with the eight transfer coefficients considered in the analysis.

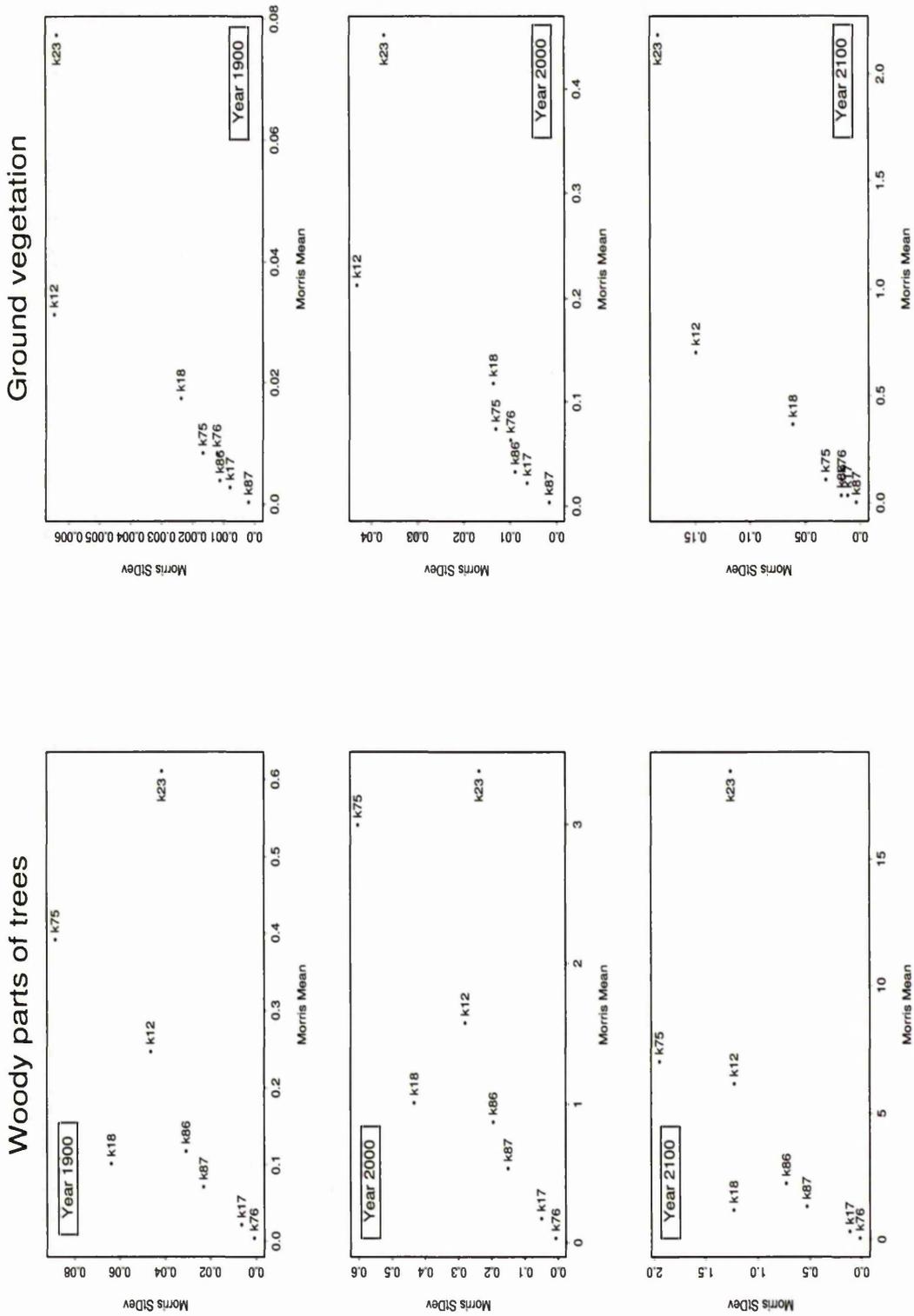


Figure 3.17. Morris screening results on Woody parts of trees and Ground vegetation compartments of Model I in 1900, 2000 and 2100. Mean and standard deviations associated with the eight transfer coefficients considered in the analysis.

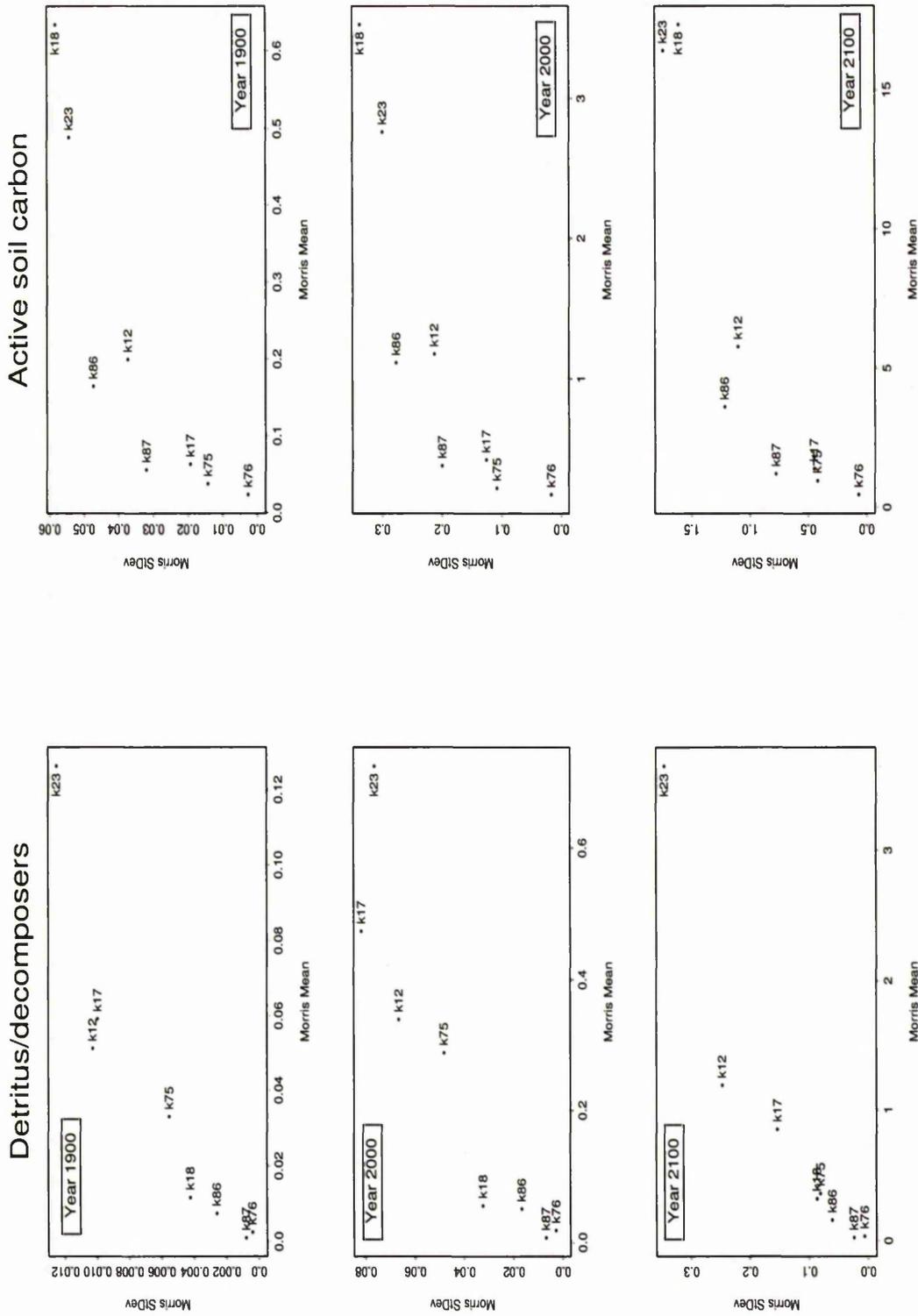


Figure 3.18. Morris screening results on Detritus/decomposers and Active soil carbon compartments of Model I in 1900, 2000 and 2100. Mean and standard deviations associated with the eight transfer coefficients considered in the analysis.

included those plots here. In order to be able to compare the results from the three years for each of the compartments, we present them in the same figure. The plotted values in all these figures are examined relative to each other to see which transfer coefficients are the most important factors for each compartmental output, and if the order of importance changes from year to year. In the examination of these plots, as a criterion we use the Euclidean distance from the origin in the  $(\mu, \sigma)$  plane.

The transfer coefficients are ranked according to the Morris sensitivity measure  $\mu$  (reported in Table A.2). These rankings are given in Table 3.9 for all eight compartments in the three chosen years, and they confirm what we have found by examining the plots.

As Figures 3.16 - 3.18 and Table 3.9 reveal, factor  $k_{23}$  is the most influential on all compartments except the active soil carbon compartment. For this compartment  $k_{18}$  is the most important input factor and  $k_{23}$  is the second most important factor.

The relative importance of the factors for atmosphere, surface and deep ocean compartments in all three years is in complete agreement. The most significant factor  $k_{23}$  is followed by  $k_{12}$ ,  $k_{18}$  and then by  $k_{75}$ . The other four factors have estimated  $\mu$  and  $\sigma$  values around zero which indicates that these factors do not have a large influence on the output of these compartments. For the non-woody parts of trees compartment, after the first two most important factors ( $k_{23}$  followed by  $k_{12}$ ), the order of importance between the other 6 factors changes slightly between the years.

As seen in the left-panels of Figure 3.17, which present the results on woody parts of trees compartment, the top three most important factors are  $k_{23}$ ,  $k_{75}$  and  $k_{12}$  (in this order), and this order does not seem to change with time. As for the rest of the transfer coefficients, the order of importance changes a little from year to year (see Table 3.9). The transfer coefficients  $k_{23}$ ,  $k_{12}$  and  $k_{18}$  appear to be the three most important factors for the ground vegetation compartment in all three years (see the right-panels of Figure 3.17). As Table 3.9 shows, the relative

Table 3.9. Results of Morris experiment on the transfer coefficients of Model I. Input factors (8 transfer coefficients) are ranked in order of importance according to the SA measure of Morris mean  $\mu$ .

Compart. Output	Input Factor	Morris Ranks		
		Yr 1900	Yr 2000	Yr 2100
Atmosphere	$k_{12}$	2	2	2
	$k_{23}$	1	1	1
	$k_{75}$	4	4	4
	$k_{76}$	7	7	7
	$k_{86}$	6	6	6
	$k_{17}$	5	5	5
	$k_{87}$	8	8	8
	$k_{18}$	3	3	3
Surface Ocean	$k_{12}$	2	2	2
	$k_{23}$	1	1	1
	$k_{75}$	4	4	4
	$k_{76}$	7	7	7
	$k_{86}$	6	6	6
	$k_{17}$	5	5	5
	$k_{87}$	8	8	8
	$k_{18}$	3	3	3
Deep Ocean	$k_{12}$	2	2	2
	$k_{23}$	1	1	1
	$k_{75}$	4	4	4
	$k_{76}$	7	7	7
	$k_{86}$	6	6	6
	$k_{17}$	5	5	5
	$k_{87}$	8	8	8
	$k_{18}$	3	3	3
Nonwoody parts of Trees	$k_{12}$	2	2	2
	$k_{23}$	1	1	1
	$k_{75}$	4	3	4
	$k_{76}$	6	6	6
	$k_{86}$	8	8	7
	$k_{17}$	5	5	5
	$k_{87}$	7	7	8
	$k_{18}$	3	4	3
Woody parts of Trees	$k_{12}$	3	3	3
	$k_{23}$	1	1	1
	$k_{75}$	2	2	2
	$k_{76}$	8	8	8
	$k_{86}$	4	5	4
	$k_{17}$	7	7	7
	$k_{87}$	6	6	5
	$k_{18}$	5	4	6
Ground Vegetation	$k_{12}$	2	2	2
	$k_{23}$	1	1	1
	$k_{75}$	4	4	4
	$k_{76}$	5	5	5
	$k_{86}$	6	6	6
	$k_{17}$	7	7	7
	$k_{87}$	8	8	8
	$k_{18}$	3	3	3
Detritus/ Decomposers	$k_{12}$	3	3	2
	$k_{23}$	1	1	1
	$k_{75}$	4	4	4
	$k_{76}$	7	7	7
	$k_{86}$	6	6	6
	$k_{17}$	2	2	3
	$k_{87}$	8	8	8
	$k_{18}$	5	5	5
Active Soil Carbon	$k_{12}$	3	3	3
	$k_{23}$	2	2	2
	$k_{75}$	7	7	7
	$k_{76}$	8	8	8
	$k_{86}$	4	4	4
	$k_{17}$	5	5	5
	$k_{87}$	6	6	6
	$k_{18}$	1	1	1

importance of the transfer coefficients does not change with time.

The influence of the transfer coefficients  $k_{23}$ ,  $k_{17}$ ,  $k_{12}$  and  $k_{75}$  on the detritus/decomposers compartment is higher than the influence of the rest of the factors, but the order of importance changes from year 2000 to 2100 between the second and third most important transfer coefficients. That is,  $k_{17}$  appears to be the second and  $k_{12}$  the third most influential factors on this compartment in years 1900 and 2000, but in 2100  $k_{12}$  becomes more important than  $k_{17}$ . For the other transfer coefficients the order of importance in the three years do not change.

For the active soil carbon compartment the most influential factors are  $k_{18}$  and then  $k_{23}$ . As shown in both Figure 3.18 and Table 3.9 the importance order between the transfer coefficients stays the same all three years.

### 3.5.2 Local SA

As described in Section 2.5.2 of the previous chapter, differential sensitivity analysis is used to provide local information about the behaviour of the model due to small perturbations about a point which is considered to be the base-case scenario, i.e. all input factors set equal to their nominal values.

This SA method based on partial differentiation is computationally efficient [50]. However, depending on complexity of the model equations the implementation of the method can be quite intensive, and in such cases complex numerical procedures are often required [63]. Here we present the results from differential analysis applied to Model I. First, the input going into the atmosphere (denoted by  $u_1(t)$  in Equation 3.1 on page 68) from 1750 to 2100 is estimated by the function

$$u_1(t) = 0.1265 e^{0.01513(t-1750)}$$

where  $u_1(t)$  is the annual emission (Gt C/yr) and  $t$  is the year. Then, using the Laplace transformation approach described in Section 2.3.2 the model equations were solved analytically. Next, a base-case vector consisting of the nominal values of the independent variables was defined and the required first-order partial derivatives of the dependent variables of interest, with respect to the independent variables, were calculated at the corresponding base-case values and the three chosen time points.

3.5.2.1 Local SA on Initial Conditions

The sensitivity coefficient for a particular independent variable, i.e. an initial condition in this case, is calculated from the partial derivative of each dependent variable (i.e. compartmental output,  $y_1(t), \dots, y_8(t)$ ) with respect to the independent variable. Partial derivatives of Model I equations describing the compartmental CO<sub>2</sub> contents were calculated for each initial condition, that is the sensitivity matrix

$$S_{(t)} = \begin{bmatrix} \frac{\partial y_1}{\partial x_1^\circ} & \frac{\partial y_1}{\partial x_2^\circ} & \frac{\partial y_1}{\partial x_3^\circ} & \frac{\partial y_1}{\partial x_4^\circ} & \frac{\partial y_1}{\partial x_5^\circ} & \frac{\partial y_1}{\partial x_6^\circ} & \frac{\partial y_1}{\partial x_7^\circ} & \frac{\partial y_1}{\partial x_8^\circ} \\ \frac{\partial y_2}{\partial x_1^\circ} & \frac{\partial y_2}{\partial x_2^\circ} & \frac{\partial y_2}{\partial x_3^\circ} & \frac{\partial y_2}{\partial x_4^\circ} & \frac{\partial y_2}{\partial x_5^\circ} & \frac{\partial y_2}{\partial x_6^\circ} & \frac{\partial y_2}{\partial x_7^\circ} & \frac{\partial y_2}{\partial x_8^\circ} \\ \dots & \dots \\ \frac{\partial y_8}{\partial x_1^\circ} & \frac{\partial y_8}{\partial x_2^\circ} & \frac{\partial y_8}{\partial x_3^\circ} & \frac{\partial y_8}{\partial x_4^\circ} & \frac{\partial y_8}{\partial x_5^\circ} & \frac{\partial y_8}{\partial x_6^\circ} & \frac{\partial y_8}{\partial x_7^\circ} & \frac{\partial y_8}{\partial x_8^\circ} \end{bmatrix}$$

which is evaluated at the nominal conditions  $x^\circ = (x_1^\circ, \dots, x_8^\circ)$  and at year  $t$  is constructed. Let  $y_p$  represent the output of the  $p$ th compartment ( $p = 1, \dots, 8$ ) and  $x_i^\circ$  the initial content of the  $i$ th compartment ( $i = 1, 2, \dots, 8$ ), and  $t$  the year at which we evaluate the sensitivities. Each row of this sensitivity matrix shows how sensitive a specific compartmental content at a given time is to the initial conditions of the eight compartments.

In order to remove the magnitude of unit effect from the sensitivity coefficients the numerical results are standardised by multiplying the partial derivatives by  $x_i^\circ/y_p^\circ(t)$ , where  $x_i^\circ$  and  $y_p^\circ(t)$  correspond to the base-case result of the model. For example, the partial derivative  $\partial y_1/\partial x_8^\circ$  is standardised as  $(\partial y_1/\partial x_8^\circ) (x_8^\circ/y_1^\circ(t))$ . This coefficient then indicates the effect on the dependent variable ( $y_1$ ) of equivalent fractional change of base-case values for the individual independent variable ( $x_8^\circ$ ). As noted by Iman and Helton [63] such coefficients are often referred to as normalized or standardized sensitivity coefficients. We shall call them standardized sensitivity coefficients.

The rankings of the absolute value of the standardized sensitivity coefficients within each compartment are given in Table 3.10. The ranking starts with the

Table 3.10. Rankings of the initial conditions to Model I based on standardised sensitivity coefficients evaluated in years 1900, 2000 and 2100.

Compartment. Output	Local Sensitivity	Yr 1900	Yr 2000	Yr 2100
Atmosphere (y1)	$\partial y_1 / \partial w_1^0$	4	4	4
	$\partial y_1 / \partial w_2^0$	5	5	5
	$\partial y_1 / \partial w_3^0$	1	1	1
	$\partial y_1 / \partial w_4^0$	8	8	8
	$\partial y_1 / \partial w_5^0$	3	3	3
	$\partial y_1 / \partial w_6^0$	7	7	7
	$\partial y_1 / \partial w_7^0$	6	6	6
	$\partial y_1 / \partial w_8^0$	2	2	2
Surface ocean (y2)	$\partial y_2 / \partial w_1^0$	4	4	4
	$\partial y_2 / \partial w_2^0$	5	5	3
	$\partial y_2 / \partial w_3^0$	1	1	1
	$\partial y_2 / \partial w_4^0$	8	8	8
	$\partial y_2 / \partial w_5^0$	3	3	3
	$\partial y_2 / \partial w_6^0$	7	7	7
	$\partial y_2 / \partial w_7^0$	6	6	6
	$\partial y_2 / \partial w_8^0$	2	2	2
Deep ocean (y3)	$\partial y_3 / \partial w_1^0$	4	4	4
	$\partial y_3 / \partial w_2^0$	2	3	3
	$\partial y_3 / \partial w_3^0$	1	1	1
	$\partial y_3 / \partial w_4^0$	8	8	8
	$\partial y_3 / \partial w_5^0$	5	5	5
	$\partial y_3 / \partial w_6^0$	7	7	7
	$\partial y_3 / \partial w_7^0$	6	6	6
	$\partial y_3 / \partial w_8^0$	3	2	2
Nonwoody parts of trees (y4)	$\partial y_4 / \partial w_1^0$	4	4	4
	$\partial y_4 / \partial w_2^0$	5	5	5
	$\partial y_4 / \partial w_3^0$	1	1	1
	$\partial y_4 / \partial w_4^0$	8	8	8
	$\partial y_4 / \partial w_5^0$	3	3	3
	$\partial y_4 / \partial w_6^0$	7	7	7
	$\partial y_4 / \partial w_7^0$	6	6	6
	$\partial y_4 / \partial w_8^0$	2	2	2
Woody parts of trees (y5)	$\partial y_5 / \partial w_1^0$	4	4	4
	$\partial y_5 / \partial w_2^0$	5	5	5
	$\partial y_5 / \partial w_3^0$	1	1	1
	$\partial y_5 / \partial w_4^0$	8	8	8
	$\partial y_5 / \partial w_5^0$	3	3	3
	$\partial y_5 / \partial w_6^0$	7	7	7
	$\partial y_5 / \partial w_7^0$	6	6	6
	$\partial y_5 / \partial w_8^0$	2	2	2
Ground vegetation (y6)	$\partial y_6 / \partial w_1^0$	4	4	4
	$\partial y_6 / \partial w_2^0$	5	5	5
	$\partial y_6 / \partial w_3^0$	1	1	1
	$\partial y_6 / \partial w_4^0$	8	8	8
	$\partial y_6 / \partial w_5^0$	3	3	3
	$\partial y_6 / \partial w_6^0$	7	7	7
	$\partial y_6 / \partial w_7^0$	6	6	6
	$\partial y_6 / \partial w_8^0$	2	2	2
Detritus/decomposers (y7)	$\partial y_7 / \partial w_1^0$	4	4	4
	$\partial y_7 / \partial w_2^0$	5	5	5
	$\partial y_7 / \partial w_3^0$	1	1	1
	$\partial y_7 / \partial w_4^0$	8	8	8
	$\partial y_7 / \partial w_5^0$	3	3	3
	$\partial y_7 / \partial w_6^0$	7	7	7
	$\partial y_7 / \partial w_7^0$	6	6	6
	$\partial y_7 / \partial w_8^0$	2	2	2
Active soil carbon (y8)	$\partial y_8 / \partial w_1^0$	4	4	4
	$\partial y_8 / \partial w_2^0$	5	5	5
	$\partial y_8 / \partial w_3^0$	1	1	1
	$\partial y_8 / \partial w_4^0$	8	8	8
	$\partial y_8 / \partial w_5^0$	3	3	3
	$\partial y_8 / \partial w_6^0$	7	7	7
	$\partial y_8 / \partial w_7^0$	6	6	6
	$\partial y_8 / \partial w_8^0$	2	2	2

highest coefficient taking rank 1, the second highest taking rank 2, and so on.

Within Table 3.10, except for  $y_3$  the order of all compartmental outputs agree, and this order does not change within the three years. The disagreement between the ranking for  $y_3$  and the ranking for all the other dependent variables is moderate. For all seven compartmental outputs  $w_5^0$  appears to be the third most effective input factor, but it is ranked as the fifth most influential factor for  $y_3$ .  $w_8^0$  is ranked as the second most effective input factor for all compartments including  $y_3$  in years 2000 and 2100 but it is ranked as the third most influential input factor in year 1900. All eight compartmental outputs at all three years

appears to be the least influenced by  $x_4^0$ .

### 3.5.2.2 Local SA on Transfer Coefficients

Again using the analytical solutions of Model I equations, the partial derivatives of  $y_p$ 's with respect to each transfer coefficient are obtained (see Table 3.3 for the description of the transfer coefficients). Then using these partial derivatives, given in the following sensitivity matrix, the sensitivity coefficients are calculated about the vector of base-case values for the transfer coefficients

$k = (k_{21}, k_{41}, k_{51}, k_{61}, k_{12}, k_{32}, k_{23}, k_{74}, k_{75}, k_{85}, k_{76}, k_{86}, k_{17}, k_{87}, k_{18})$  and at  $t = 1900, 2000, 2100$

$$S_{(t)} = \begin{bmatrix} \frac{\partial y_1}{\partial k_{21}} & \frac{\partial y_1}{\partial k_{41}} & \frac{\partial y_1}{\partial k_{51}} & \frac{\partial y_1}{\partial k_{61}} & \frac{\partial y_1}{\partial k_{12}} & \frac{\partial y_1}{\partial k_{32}} & \frac{\partial y_1}{\partial k_{23}} & \frac{\partial y_1}{\partial k_{74}} & \frac{\partial y_1}{\partial k_{75}} & \frac{\partial y_1}{\partial k_{85}} & \frac{\partial y_1}{\partial k_{76}} & \frac{\partial y_1}{\partial k_{86}} & \frac{\partial y_1}{\partial k_{17}} & \frac{\partial y_1}{\partial k_{87}} & \frac{\partial y_1}{\partial k_{18}} \\ \frac{\partial y_2}{\partial k_{21}} & \frac{\partial y_2}{\partial k_{41}} & \frac{\partial y_2}{\partial k_{51}} & \frac{\partial y_2}{\partial k_{61}} & \frac{\partial y_2}{\partial k_{12}} & \frac{\partial y_2}{\partial k_{32}} & \frac{\partial y_2}{\partial k_{23}} & \frac{\partial y_2}{\partial k_{74}} & \frac{\partial y_2}{\partial k_{75}} & \frac{\partial y_2}{\partial k_{85}} & \frac{\partial y_2}{\partial k_{76}} & \frac{\partial y_2}{\partial k_{86}} & \frac{\partial y_2}{\partial k_{17}} & \frac{\partial y_2}{\partial k_{87}} & \frac{\partial y_2}{\partial k_{18}} \\ \dots & \dots \\ \frac{\partial y_8}{\partial k_{21}} & \frac{\partial y_8}{\partial k_{41}} & \frac{\partial y_8}{\partial k_{51}} & \frac{\partial y_8}{\partial k_{61}} & \frac{\partial y_8}{\partial k_{12}} & \frac{\partial y_8}{\partial k_{32}} & \frac{\partial y_8}{\partial k_{23}} & \frac{\partial y_8}{\partial k_{74}} & \frac{\partial y_8}{\partial k_{75}} & \frac{\partial y_8}{\partial k_{85}} & \frac{\partial y_8}{\partial k_{76}} & \frac{\partial y_8}{\partial k_{86}} & \frac{\partial y_8}{\partial k_{17}} & \frac{\partial y_8}{\partial k_{87}} & \frac{\partial y_8}{\partial k_{18}} \end{bmatrix}$$

The rankings of the absolute values of these sensitivity coefficients within each compartment are given in Table 3.11.

The ranking of the sensitivity coefficients for the transfer coefficients (as indicated in Table 3.11) changes slightly from year to year. At all three years,  $k_{23}$  is ranked as the most important input factor for all compartmental outputs. The first three most influential transfer coefficients on the output of the atmosphere ( $y_1$ ) and the surface ocean ( $y_2$ ) compartments appear to be the same,  $k_{23}, k_{32}$ , and  $k_{18}$ . Except for  $y_4$ ,  $k_{74}$  is the least important transfer coefficient for all compartmental outputs in all three years. For  $y_4$ ,  $k_{17}$  is ranked as the least important input factor. The second and third most important transfer coefficients for the deep ocean compartment are  $k_{18}$  and  $k_{32}$ , respectively.  $k_{41}$  is the second most important input factor for  $y_4$ , and it becomes the third most effective transfer coefficient for  $y_7$  by 2100, but for the other compartmental outputs it is ranked as one of the least important input factors.

Table 3.11. Rankings of the transfer coefficients to Model I for each compartmental output based on sensitivity coefficients evaluated in years 1900, 2000 and 2100.

	Atmosphere ( $y_1$ )			Surface ocean ( $y_2$ )			Deep ocean ( $y_3$ )			Nonwoody parts of trees ( $y_4$ )			Woody parts of trees ( $y_5$ )			Ground vegetation ( $y_6$ )			Detritus/ decomposers ( $y_7$ )			Active soil carbon ( $y_8$ )					
	year	year	year	year	year	year	year	year	year	year	year	year	year	year	year	year	year	year	year	year	year	year	year	year			
	1900	2000	2100	1900	2000	2100	1900	2000	2100	1900	2000	2100	1900	2000	2100	1900	2000	2100	1900	2000	2100	1900	2000	2100			
$\partial y_i / \partial k_{21}$	5	5	4	9	9	8	9	9	9	6	6	6	8	8	8	6	6	6	8	8	8	9	9	9	9	9	9
$\partial y_i / \partial k_{41}$	11	11	11	11	11	11	11	11	11	2	2	2	11	11	11	13	13	13	6	6	6	4	4	4	11	11	11
$\partial y_i / \partial k_{51}$	8	8	6	6	4	4	5	5	5	9	9	7	4	4	3	10	10	9	7	7	7	6	5	5	7	7	7
$\partial y_i / \partial k_{61}$	7	7	7	5	5	6	6	6	6	8	8	8	9	9	9	2	2	2	9	9	9	7	7	7	4	4	4
$\partial y_i / \partial k_{12}$	4	4	5	8	8	9	8	8	8	5	5	5	7	7	7	5	5	5	5	5	5	2	2	2	8	8	8
$\partial y_i / \partial k_{32}$	2	2	2	2	2	2	3	3	3	3	3	3	6	6	6	3	3	3	3	3	3	5	5	5	2	2	2
$\partial y_i / \partial k_{23}$	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
$\partial y_i / \partial k_{74}$	15	15	15	15	15	15	15	15	15	11	10	10	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15
$\partial y_i / \partial k_{75}$	9	9	9	7	7	7	7	7	7	10	11	11	3	3	3	4	4	4	11	11	11	10	10	10	10	10	10
$\partial y_i / \partial k_{85}$	6	6	8	4	4	5	4	4	4	7	7	7	2	2	2	2	2	2	7	7	7	9	9	9	3	3	3
$\partial y_i / \partial k_{76}$	13	13	13	13	13	13	13	13	13	14	14	14	13	13	13	13	13	13	9	8	8	14	14	14	13	13	13
$\partial y_i / \partial k_{86}$	12	12	12	12	12	12	12	12	12	13	13	13	12	12	12	8	7	7	8	7	7	13	13	13	10	10	10
$\partial y_i / \partial k_{17}$	14	14	14	14	14	14	14	14	14	15	15	15	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14
$\partial y_i / \partial k_{87}$	10	10	10	10	10	10	10	10	10	12	12	12	10	10	10	10	10	10	12	12	12	11	11	11	6	6	6
$\partial y_i / \partial k_{18}$	3	3	3	3	3	3	2	2	2	4	4	4	5	6	6	4	4	4	4	4	4	4	5	6	2	2	2

We now move from local SA methods, used to evaluate the relative importance of model input factors in a qualitative manner, to global SA methods which provide a more quantitative assessment of the relative influence of the input factors.

### 3.5.3 Global SA Methods Applied to Initial Conditions

In order to identify the model input factors to which model behaviour is most sensitive, we now consider various global SA techniques and analyses, including correlation coefficients, standardized regression coefficients, partial correlation coefficients.

In the Monte Carlo model simulations the steady-state restriction is preserved by following the procedures explained in Section 3.4. In the analyses,  $N=100$  and  $N=5,000$  model runs are considered.

Because, in global methods, we vary the input factors simultaneously, the Monte Carlo iterations might not provide estimates that are compatible with the historic records of atmospheric  $\text{CO}_2$ . With the 8-compartment model this situation arises when we vary the initial conditions simultaneously. In such cases, '*windowing analysis*' can be conducted on the model predictions to obtain a sample of model predictions that are in broad agreement with the observed pattern. Now, we shall describe briefly how this analysis is performed.

#### 3.5.3.1 Windowing Analysis

This analysis has been used by King & Sale (1990) of CDIAC (Carbon Dioxide Information Analysis Center) in a technical report for the U. S. Department of Energy (see [73]). In this report, they introduced windowing analysis as one of the steps in a procedure for uncertainty analysis of atmospheric  $\text{CO}_2$ , in which sensitivity analysis is considered as another step. Because according to them, sensitivity analysis is based on the estimation of changes in model output in response to very small (1% coefficient of variation on all model parameters) or local changes in model parameters, windowing analysis, on the other hand, is

based on model parameters with wide uncertainty ranges ( $\pm 50\%$  of their nominal values) and assuming that they all follow uniform distribution, they consider windowing and sensitivity analysis as two separate analyses.

In more recent work done by Grieb *et al.* [40] windowing is used in the analysis of a global carbon cycle model (called GLOCO) in order to adjust the values of model input parameters to achieve an acceptable match between observed and predicted model conditions. In this paper, after they obtain the GLOCO parameter space using windowing, they then demonstrate a new tree-structured density estimation technique to explore parameter interaction in the terrestrial ecosystem module of the model. The terminology “*windowing*” we adopted from King & Sale is not used by Grieb *et al.* but the procedure followed is essentially the same. Grieb *et al.* define a set of performance criteria by giving acceptable ranges of values that three model output variables take in year 1970 (the end of the historical period considered), and compared the model output, at the conclusion of each historical simulation, with these criteria. Then, the model predictions that meet the performance criteria referred to as “passes” or “behaviours” are taken into account in the analysis.

Windowing analysis can help to reduce the concerns about the validity of predictions made by simulation models, especially global carbon cycle models for which incompleteness of scientific understanding always exists.

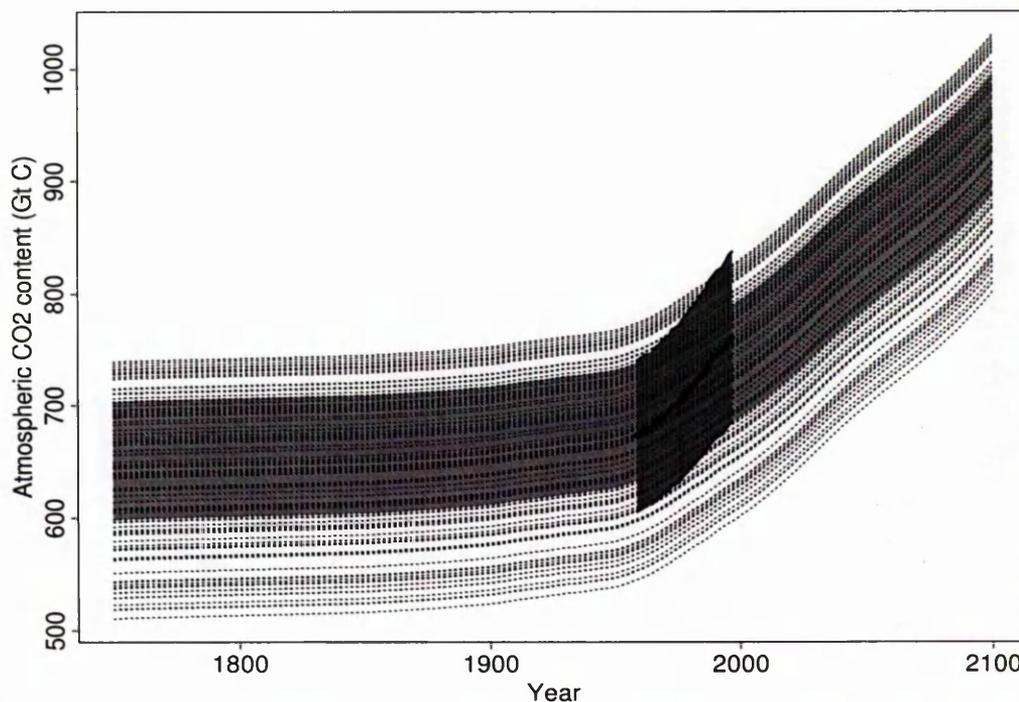
Windowing also helps to improve our understanding about the model input factors, and a better understanding of model inputs will also lead to better model predictions, hence the credibility of the model can be improved.

In this thesis, we consider windowing as a step to sensitivity analysis; first by conducting windowing analysis we filter model predictions through a window defined by the variance around the Mauna Loa records of historic atmospheric  $\text{CO}_2$  from the period 1958-2000; then by considering only the predictions which pass wholly through the filter, i.e. lie within the defined window, we apply sensitivity analysis to the resulting input and output sets. The input sets that result in model predictions which meet the windowing criteria are referred to

as 'good' and the others as 'bad' sets. Windowing can then be described as a screening or filtering algorithm.

**Windowing Analysis with the Initial Conditions of Model I** The specific input factors subject to windowing analysis are the 8 initial conditions. In the figure below (Figure 3.19) an illustration of windowing analysis on  $N=100$  Monte Carlo simulations of atmospheric  $\text{CO}_2$  is presented. Out of 100 model predictions 54 runs that met the window criterion described above were obtained.

Any point within the eight dimensional input space can be identified as leading



**Figure 3.19.** Illustration of windowing  $N=100$  atmospheric  $\text{CO}_2$  predictions resulted from varying Model I initial conditions ( $x_1^0, \dots, x_8^0$ ) simultaneously. The dotted curves are the model predictions for atmosphere compartment. The window obtained using the observed historical data (from Mauna Loa observatory; 1958-2000) is indicated by a dark grey shaded region. The set of the prediction curves that met the windowing criteria (54 out of 100 model simulations) are highlighted with a light grey shaded area.

to a good or bad simulation by running the model with the corresponding input vector and applying the windowing criteria to the output. The good sets and their corresponding model predictions are then used in the global SA methods applied. But before proceeding with the application of these methods, we want to take a step back and find out what is happening in the input space as a result of windowing analysis. In other words, we wish to identify specific regions of the input space where the likelihood of obtaining good model runs is high, and also investigate how input factors contribute to the separation of good and bad runs. For this purpose, discriminant analysis and classification is used.

Using discriminant analysis we aim to find out how well it is possible to separate good and bad model runs. Several methods for discriminant analysis have been developed and these methods can be found in many multivariate analysis books (see [67], [35]). Differences between methods arise because of the variety of the assumptions made about the variables describing each observation to be classified. Here we first consider Fisher's linear discriminant analysis method. This method is one of the most widely used discrimination methods, and it is based on the assumption that the within-group covariances are the same. Because this method involves knowledge of the population covariance matrices and of course we do not have such knowledge, we can either assume that our groups have the same covariance matrix and carry out Fisher's linear method or use quadratic discriminant analysis method which does not require this assumption.

The performance of both Fisher's linear method and quadratic method applied on the original and cross validated data are presented in Tables 3.12(a) and (b) respectively. In both methods, we assume that the prior probabilities for the two groups are equal.

The classification results in Table 3.12(a) show that 73% of good and 63% of bad cases are correctly classified, and the overall classification rate is nearly 68%. The cross-validation routine, which takes approximately four times longer in computation time but provides more realistic misclassification rate, estimates the overall correct classification rate to be 63%.

**Table 3.12.** Classification performance on (a) Fisher's discriminant method and (b) Quadratic discriminant method assessed on the original and cross-validated sample where 100 good and 116 bad sets of model initial conditions are involved.

**(a) Fisher's discriminant method:**

Actual Group		Predicted Group		Total N	N Correct	Proportion
		<i>good</i>	<i>bad</i>			
Original	<i>good</i>	73	27	100	73	0.730
	<i>bad</i>	43	73	116	73	0.629
N = 216 ; N Correct = 146 ; Proportion Correct = 0.676						
Cross Validated	<i>good</i>	65	35	100	65	0.650
	<i>bad</i>	45	71	116	71	0.612
N = 216 ; N Correct = 136 ; Proportion Correct = 0.630						

**(b) Quadratic discriminant method:**

Actual Group		Predicted Group		Total N	N Correct	Proportion
		<i>good</i>	<i>bad</i>			
Original	<i>good</i>	89	11	100	89	0.890
	<i>bad</i>	5	111	116	111	0.957
N = 216 ; N Correct = 200 ; Proportion Correct = 0.926						
Cross Validated	<i>good</i>	71	29	100	71	0.710
	<i>bad</i>	17	99	116	99	0.853
N = 216 ; N Correct = 170 ; Proportion Correct = 0.787						

Table 3.12(b) shows the number of observations in the original and cross-validated samples correctly and incorrectly classified using a quadratic discrimination method. Compared to the linear method this method gives higher overall classification rates, about 93% with original data and about 79% with cross-validation. The quadratic approach proves to be deriving a better decision rule for classifying good and bad model runs.

So far we have interpreted the results of standard discrimination analysis in terms of the linear and quadratic combinations of all eight input factors which separate two groups from each other. Next, we carry out discriminant analysis in a stepwise manner to find out if there is a subset of input factors which maximizes

the discriminating power. With this approach variables (initial conditions  $x_i^0$ ) are added to the discriminant function one by one until it is found that adding extra variables does not provide significantly better discrimination. There are many different criteria which can be used for entering and removing variables. Stepwise method is available in SPSS, and different criteria used for this method are explained in SPSS User's Guide (see [90]). Application of stepwise analysis to our data show that except for  $x_3^0$ , all initial conditions are dropped and the analysis has only one step. The classification results of this analysis based on Fisher's linear method (see Table 3.13(a)) indicates that with only  $x_3^0$  considered as predictor 69% of the data is correctly classified using either original or cross-validated grouped cases. As the results in Table 3.13(b) show, the quadratic method leads to higher proportion of cases to be correctly classified, 91.2% when original and 90.7% when cross-validated data were used.

The stepwise analysis results shows that all initial conditions except  $x_3^0$  are not important as predictors considering the data in hand.

Next, we consider a completely different approach called classification trees. This new method, which is an exploratory technique for revealing structure in data, developed in 1980s is gaining widespread popularity [16]. Using this method we aim to obtain more accurate classifiers. Given a set of variables which might be useful discriminators, this method picks out the best variables and then a binary split is performed using this variable which provides the smallest number of misclassifications for the data. By repeating this process, further binary splits are made in order to reduce the misclassification error on the data. The splitting process continues until we reach a suitable stopping point. This procedure is diagrammed as a tree referred to as classification tree. S-PLUS has built-in software for this method (see Chapter 9 of Ref. [16] for a detailed description of the method). The branches of such trees correspond to divisions in the sample space. Nodes which are labelled by class labels are represented by ellipses (interior nodes) and rectangles (terminal nodes). The misclassification error rates are given under each terminal node.

**Table 3.13.** Classification performance on (a) Fisher's linear discriminant method and (b) Quadratic discriminant method used for stepwise discriminant function analysis where only  $x_3^o$  found to be a significant discriminator.

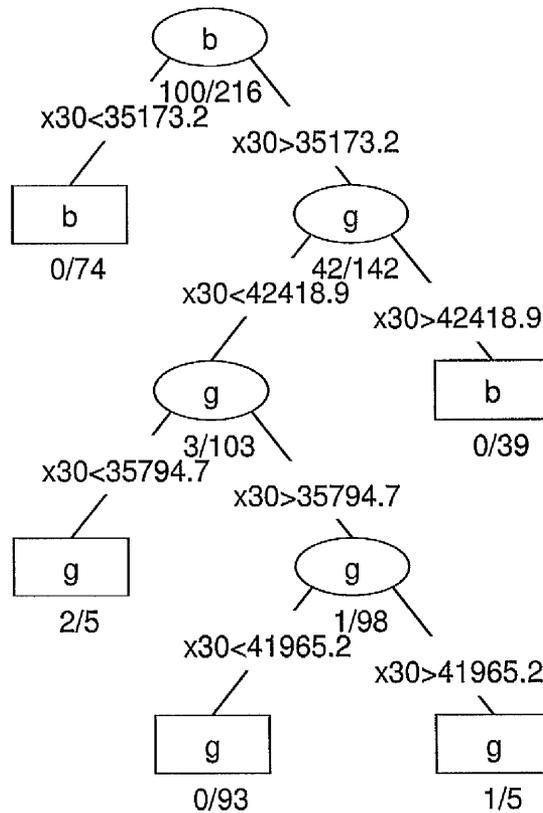
(a) Fisher's linear discriminant method:

Actual Group		Predicted Group		Total N	N Correct	Proportion
		<i>good</i>	<i>bad</i>			
Original	<i>good</i>	73	27	100	73	0.730
	<i>bad</i>	40	76	116	76	0.655
N = 216 ; N Correct = 149 ; Proportion Correct = 0.690						
Cross Validated	<i>good</i>	73	27	100	73	0.730
	<i>bad</i>	40	76	116	76	0.655
N = 216 ; N Correct = 149 ; Proportion Correct = 0.690						

(b) Quadratic discriminant method:

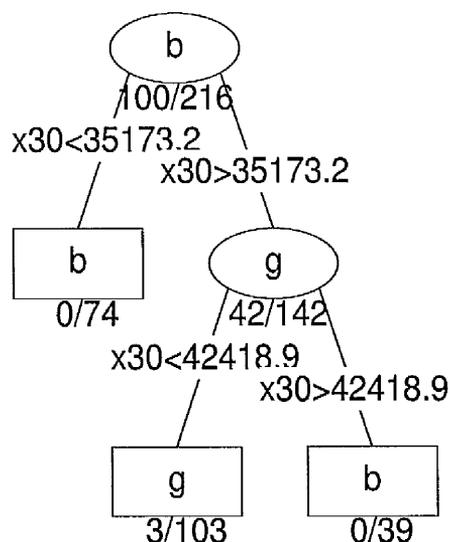
Actual Group		Predicted Group		Total N	N Correct	Proportion
		<i>good</i>	<i>bad</i>			
Original	<i>good</i>	88	12	100	88	0.880
	<i>bad</i>	7	109	116	109	0.940
N = 216 ; N Correct = 197 ; Proportion Correct = 0.912						
Cross Validated	<i>good</i>	87	13	100	87	0.870
	<i>bad</i>	7	109	116	109	0.940
N = 216 ; N Correct = 196 ; Proportion Correct = 0.907						

Figure 3.20 presents a display of a classification tree grown to the windowing data. In this tree only  $x_3^o$  is used. At the first step all observations are classified as bad predictors with 100 misclassifications. The first split is on  $x_3^o$ , the model runs resulted from input sets which have  $x_3^o$  value of less than 35173.2 are classified as 'bad' with an error rate of 0/74, and those with an  $x_3^o$  value of more than 35173.2 are classified as 'good' model runs with an error rate of 42/142. The 'good' node undergoes further splits using again  $x_3^o$ . When  $x_3^o$  takes values of less than 42418.9 the corresponding model predictions are classified as 'good' with an error rate of 3/103 and when the values of  $x_3^o$  are greater than 42418.9 the predictions are classified as 'bad' with an error rate of 0/39.



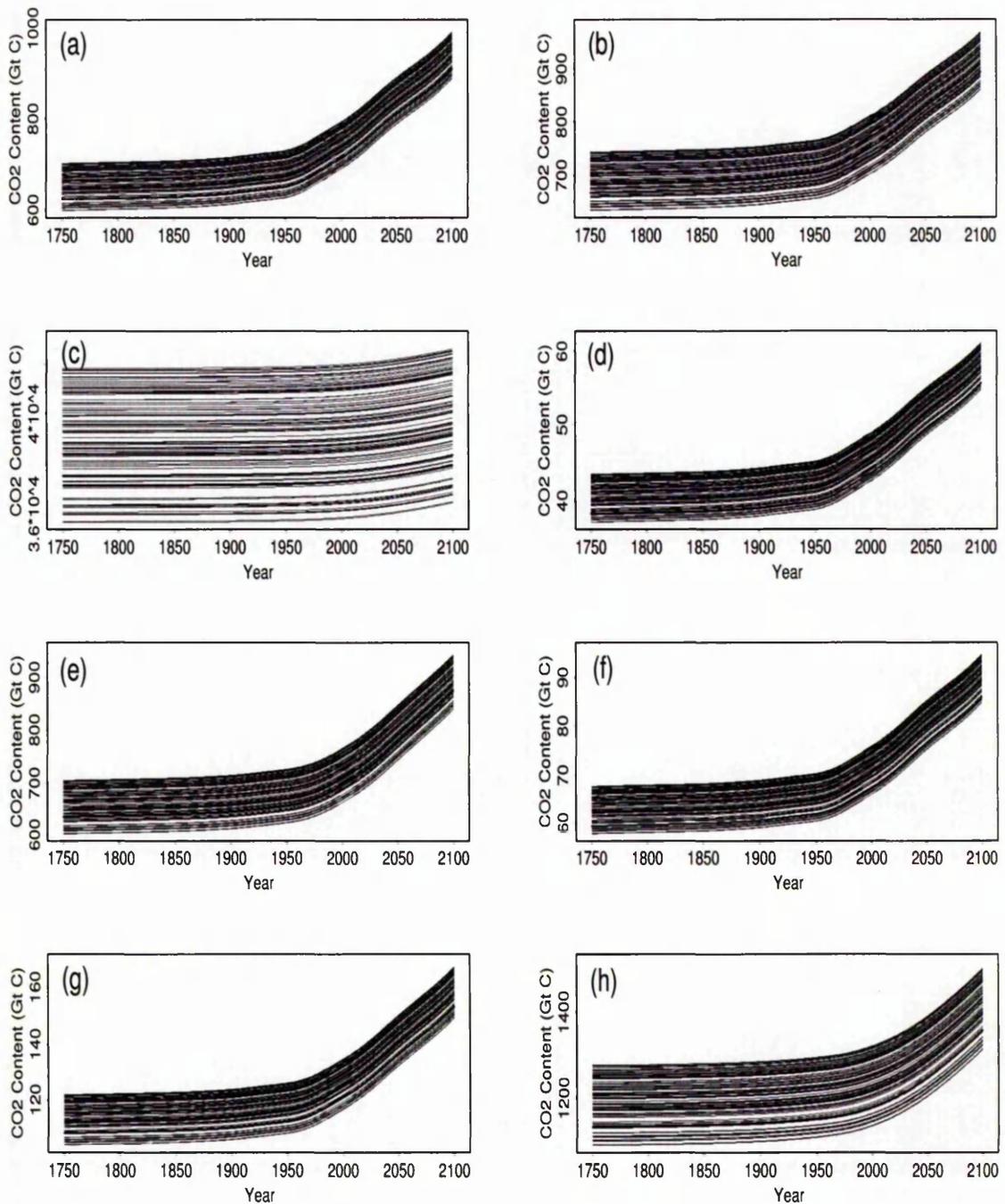
**Figure 3.20.** Classification tree showing the performance of the windowing data. The class of the predicted response variable (g-for 'good', b-for 'bad') is centered in the node. The number underneath each terminal node is the misclassification error rate.

The third split is on a 'good' node but as we can see from the tree an overfitting has occurred here. In such cases, a pruning procedure can be used. Now using ten-fold cross-validation within S-PLUS we prune the classification tree shown in Figure 3.20. The pruned tree is displayed in Figure 3.21. Applying cross-validation suggests that a tree with 3 terminal nodes is suitable and that a more realistic estimate of the likely misclassification rate with new input sets is  $54/216 = 25\%$ .



**Figure 3.21.** A pruned version of the classification tree given in Figure 3.20.

Having found  $N=100$  input sets which we know result in atmospheric  $\text{CO}_2$  predictions that match with the Mauna Loa observations reasonably well, we then perform the model calculation for all eight compartments. The same procedure was repeated using  $N=5,000$  good input sets obtained through windowing analysis. The time-dependent behavior of each compartmental output following windowing analysis are presented in Figure 3.22 for  $N=100$ .



**Figure 3.22.** Dependent variables predicted by Model I following windowing analysis: CO<sub>2</sub> content of (a) Atmosphere, (b) Surface ocean, (c) Deep ocean, (d) Non-woody parts of trees, (e) Woody parts of trees, (f) Ground vegetation, (g) Detritus/decomposers, and (h) Active soil carbon compartments as a result of varying all input factors ( $x_i$ 's) simultaneously. Emission scenario IS92a is considered in the model calculations.

As we have done in screening procedures (Section 3.5.1) and local SA (Section 3.5.2), in the application of global SA methods we consider  $N=100$  and  $N=5,000$  samples, and study the compartmental contents in years 1900, 2000 and 2100.

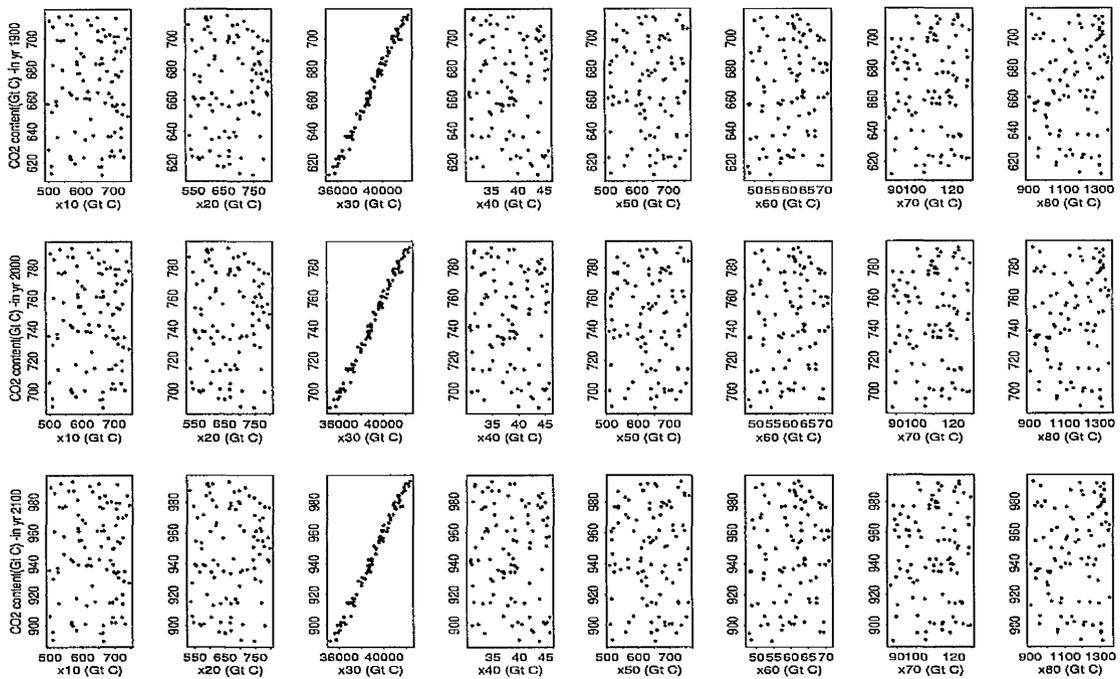
We now start the analyses with the generation of scatterplots which are known to be a good starting point in a sensitivity study.

### 3.5.3.2 Examination of Scatterplots

A scatterplot of each response variable (i.e. the predicted  $\text{CO}_2$  content of each compartment at a chosen year) against each predictor (i.e. initial condition of each compartment) enables us to investigate the relationship between the model outcome and each of the model inputs.

The scatterplots for the atmosphere compartment are given in Figure 3.23 for  $N=100$ . The plots on each row of this multiple plot represent the results from each of the years considered. In this figure, it is clear that patterns in the scatterplots corresponding to the same response and explanatory variables do not change between the three years. Within the same compartment the scatterplots for  $x_3^0$  display a clear linear pattern, but for the other initial conditions the points are widely scattered. There is no evidence to suggest any linear or non-linear relationship between these explanatory variables and any of the response variables. Since we assume that the input factors are independent in this sample-based study and generate an independent sample for the model inputs, the explanatory variables are not correlated. The scatterplots for the other seven compartments with  $N=100$  and for all eight compartments with  $N=5,000$  are also produced but due to space limitations these scatterplots are not given in this thesis. However, these scatterplots present patterns very similar to the patterns we see in Figure 3.23; a very strong linear association between  $x_3^0$  and the individual compartmental outputs and hardly any relationship between the other predictors and the dependent variables.

Examination of these scatterplots has provided us with a good indication



**Figure 3.23.** Scatterplots of predicted Atmospheric CO<sub>2</sub> content in years 1900, 2000 and 2100 versus each compartment's initial condition ( $x_i^0$ ). To calculate these N=100 model predictions Model I with IS92a emission scenario is used.

about  $x_3^0$  being the only initial condition dominating variation in all compartmental outputs. Hence, we do not expect the further analysis to reveal different results, but to see if the degree of influence the other inputs have becomes more/less apparent with different methods we carry out some other SA procedures on the data.

Next, some numerical measures are calculated to distinguish between the input factors that appear to have a significant effect on a predicted compartmental output and the input factors which appear to have little or no effect. Since there is an apparent linear relationship between  $x_3^0$  and all  $y_i$ s (i.e. all compartmental outputs) we have calculated the Pearson correlation coefficients on the input and output values. Then on the basis of their correlation coefficients we have ranked the input factors. We note that the calculated Pearson correlation coefficients hardly change between the three chosen years, that is, the contribution of the

input factors to prediction uncertainty is the same in 1900, 2000 and 2100.

The calculations based on both  $N=100$  and  $N=5,000$  show that  $x_3^0$  with a correlation coefficient of about 0.994 (ranked 1) has a substantial effect on all predicted compartmental contents at all three years. With  $N=100$  simulations  $x_3^0$  (with correlation coefficient of around 0.153, ranked 2) and  $x_8^0$  (with correlation coefficient of 0.132, ranked 3) seem to have some small effect on the dependent variables, with  $N=5,000$  however except for  $x_3^0$  none of the other initial conditions, which have correlation coefficients varying between 0.027 and 0.00007, seem to have any effect on the model predictions of any compartment.

### 3.5.3.3 Regression Methods

The multiple regression of each model prediction ( $y_i, i = 1, \dots, 8$ ) on all the input factors ( $x_i^0, i = 1, \dots, 8$ ) simultaneously estimates all the initial condition sensitivities. However, based on our diagnostics of the scatter plots and the correlation coefficients given in the previous section, it is clear that only  $x_3^0$  (the initial condition of the deep ocean compartment) is a useful explanatory variable in explaining the variability of each of the response variables, and because the relationship between each response and  $x_3^0$  appear to be linear, we use linear regression to explore the dependence of each compartmental output at chosen times on  $x_3^0$ .

Considering each model prediction in 1900, 2000 and 2100 as the dependent variables and  $x_3^0$  as the independent variable, the construction of regression models with  $N=100$  model iterations give us  $R^2 = 98.9\%$  when the predictions  $y_i(t = 1900)$  and  $y_i(t = 2000)$ ; and  $R^2 = 98.7\%$  when  $y_i(t = 2100)$  are considered. When  $N=5,000$  the coefficient of determination  $R^2 = 99.2\%$  is the same for all compartmental predictions at all three times. In short, these very high  $R^2$  values indicate that the regression models of each  $y_i$  on  $x_3^0$  are accounting for most of the uncertainty in the corresponding  $y_i$ . The contribution of the remaining initial conditions to  $R^2$  is about 1% only.

In SA based on regression procedure the standardized regression coefficients

(SRCs) are used as a measure of variable importance. As an example, the regression model, including all eight explanatories in it, for  $N=100$  atmospheric predictions in 2100 with the SRCs is

$$y_1(t = 2100) = 0.0364 x_1^{\circ} + 0.0391 x_2^{\circ} + 1.0018 x_3^{\circ} + 0.0022 x_4^{\circ} + 0.0369 x_5^{\circ} \\ + 0.0035 x_6^{\circ} + 0.0063 x_7^{\circ} + 0.0659 x_8^{\circ}$$

where  $y_1(t = 2100)$  and  $x_i^{\circ}$ , ( $i = 1, \dots, 8$ ) have been standardized to mean zero and standard deviation one. The SRCs in this equation provide a characterization of input factor importance. For instance, for perturbations equal to a fixed fraction of their standard deviation, the impact of  $x_3^{\circ}$  is approximately 1420% larger than the impact of  $x_8^{\circ}$  (i.e.,  $(1.0018 - 0.0659)/0.0659 = 14.20$ ). The situation with the other response variables at all three times and with different  $N$  is not any different.

For the purpose of comparing the sensitivity rankings of the input factors we obtain from various SA methods, here we also provide the absolute ranking of the SRCs:  $x_1^{\circ}$  (rank: 5),  $x_2^{\circ}$  (rank: 3),  $x_3^{\circ}$  (rank: 1),  $x_4^{\circ}$  (rank: 8),  $x_5^{\circ}$  (rank: 4),  $x_6^{\circ}$  (rank: 7),  $x_7^{\circ}$  (rank: 6) and  $x_8^{\circ}$  (rank: 2). We shall note that this ranking does not change with the sample size  $N$  and the model prediction  $y_i$ .

#### 3.5.3.4 Stepwise Regression

In this analysis, a variable was required to be significant at an  $\alpha$ -value of 0.01 to enter a regression model and to remain significant at an  $\alpha$ -value of 0.05 to be kept in a regression model, although no variable was entered and then dropped from a model.

As expected, based on the results we presented in the two previous subsections, in the first step of the stepwise regression procedure  $x_3^{\circ}$  is entered into the least squares multiple regression models for all response variables calculated at three years with  $N=100$  and  $N=5,000$  model runs (see Table 3.14). The analyses with  $x_3^{\circ}$  in the model yields a very high  $R^2$ -value of 98.94% in 1900 and 2000; and

**Table 3.14.** Stepwise regression analyses for output variable  $y_1$  of Model I in years 1900, 2000 and 2100; based on  $N=100$  and  $N=5,000$  model runs; and IS92a emission scenario.

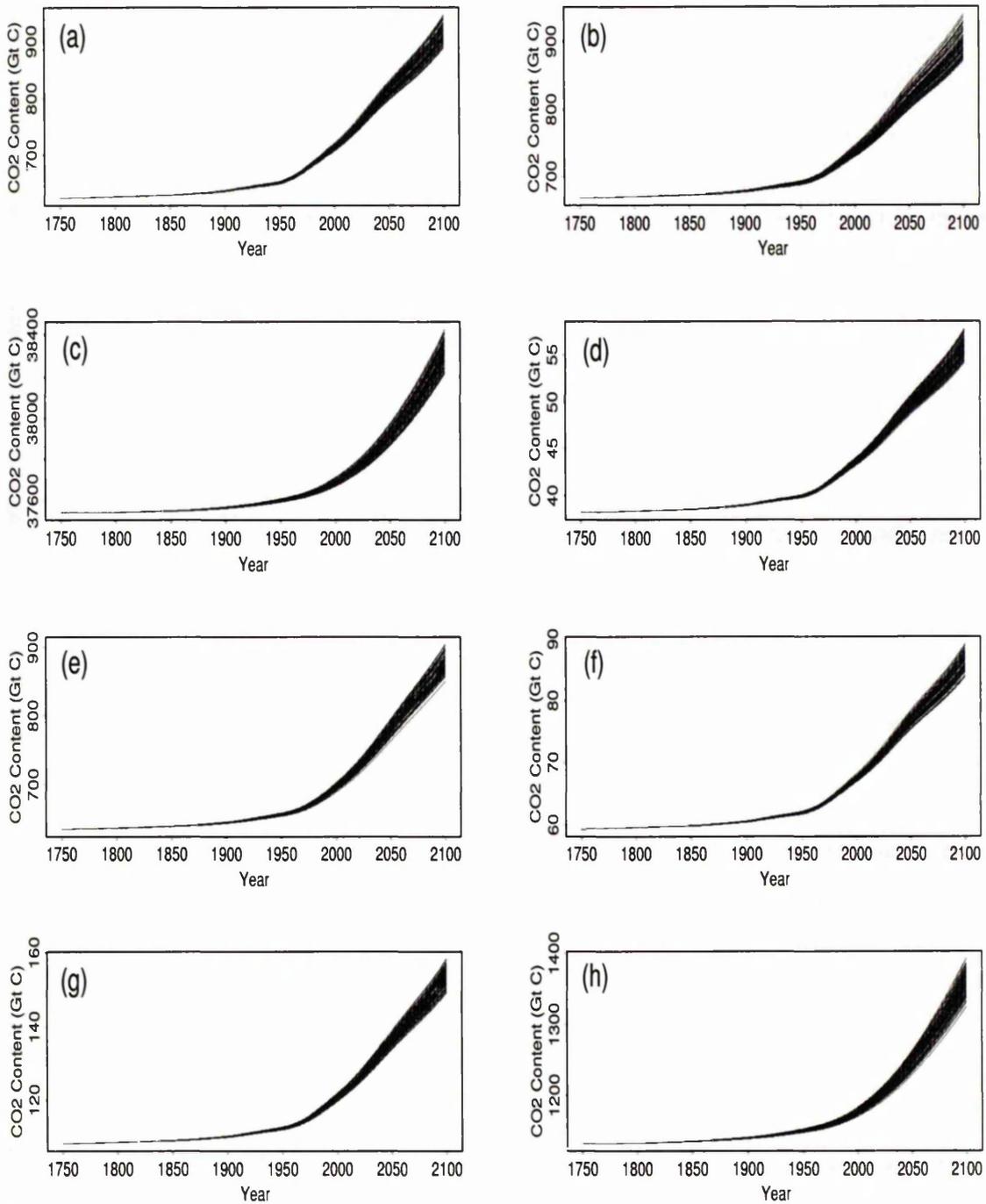
	$y_1(t = 1900)$			$y_1(t = 2000)$		$y_1(t = 2100)$	
	Step	Variable	$R^2$	Variable	$R^2$	Variable	$R^2$
N = 100	1	$x_3^0$	0.9894	$x_3^0$	0.9894	$x_3^0$	0.9868
	2	$x_8^0$	0.9954	$x_8^0$	0.9954	$x_8^0$	0.9934
	3	$x_5^0$	0.9972	$x_5^0$	0.9972	$x_5^0$	0.9954
	4	$x_2^0$	0.9986	$x_2^0$	0.9986	$x_1^0$	0.9966
	5	$x_1^0$	0.9999	$x_1^0$	0.9999	$x_2^0$	0.9976
	6	$x_7^0$	1.0000	$x_7^0$	1.0000		
	7	$x_6^0$	1.0000	$x_6^0$	1.0000		
N = 5,000	1	$x_3^0$	0.9915	$x_3^0$	0.9915	$x_3^0$	0.9915
	2	$x_8^0$	0.9957	$x_8^0$	0.9957	$x_8^0$	0.9957
	3	$x_2^0$	0.9972	$x_2^0$	0.9972	$x_2^0$	0.9972
	4	$x_5^0$	0.9986	$x_5^0$	0.9986	$x_5^0$	0.9986
	5	$x_1^0$	0.9999	$x_1^0$	0.9999	$x_1^0$	0.9999
	6	$x_7^0$	1.0000	$x_7^0$	1.0000	$x_7^0$	1.0000
	7	$x_6^0$	1.0000	$x_6^0$	1.0000	$x_6^0$	1.0000

98.68% when year 2100 predictions are used. The coefficient of multiple determination calculated for all  $y_i$  with  $N=5,000$  model iterations is slightly higher (i.e., 99.15% with all three years' predictions). In the following steps of the analyses, other initial conditions with significant  $p$ -values were also added to the models, but after allowing for the effect of  $x_3^0$  in the model their contribution to  $R^2$  is very small. The results for each compartmental prediction ( $y_i$ ) are very similar in the sense that the same variables were selected with  $R^2$ -values that are quite similar and the order of variable selection did not change. Here we include the results for only the atmosphere compartment.

As mentioned in Section 2.5.6 when there is non-linear monotonic relationship between the variables the rank transformation often is an effective way of improving the resolution of regression based SA, however, in this case with the initial conditions where we have no monotonic relationships applying rank transformation on the data is not appropriate.

### 3.5.4 Global SA Methods Applied to Transfer Coefficients

In this section, we wish to explore the mapping from the uncertain transfer coefficients  $k_{ij}$  listed in Table 3.3 to the corresponding uncertain model outcomes  $y_i(t)$ ,  $i = 1, \dots, 8$ ;  $t = 1900, 2000, 2100$ . First, we illustrate the time dependent behaviour of the outcome of all eight compartments resulting from  $N=100$  model runs in Figure 3.24. Each curve in these plots is calculated conditional on the steady-state restriction explained in Section 3.4. The variability in all compartmental predictions increases with time. Now, using a number of global SA procedures we wish to determine the effects of individual input factors on the model outcomes. As in the SA involving the initial conditions we again start the analysis by examining the scatterplots.



**Figure 3.24.** Dependent variables predicted using Model I: CO<sub>2</sub> content of (a) Atmosphere, (b) Surface ocean, (c) Deep ocean, (d) Nonwoody parts of trees, (e) Woody parts of trees, (f) Ground vegetation, (g) Detritus/decomposers, and (h) Active soil carbon compartments as a result of varying selected 8 transfer coefficients ( $k_{ij}$ s) simultaneously. Emission scenario IS92a is considered in the model calculations.

### 3.5.4.1 Examination of Scatterplots

The scatterplots that show the relationships between the sampled transfer coefficients (see Table 3.3 for the description of these  $k_{ij}$ ) and the predicted CO<sub>2</sub> content of the atmosphere compartment (i.e.,  $y_1$ ) in 1900, 2000 and 2100 are given in Figure 3.25. The plots of  $y_1$  in the three years versus each transfer coefficient present quite similar pictures. This is also true for all the other compartments. At this point we shall note that to preserve space the scatterplots for the rest of the responses ( $y_2, \dots, y_8$ ) at the three chosen years (1900, 2000 and 2100) with both  $N=100$  and  $N=5,000$  model runs are not included in this thesis.

Here, we give the scatterplot matrix of  $y_1(t = 2100)$  and the  $k_{ij}$  (see Figure 3.26) considering the results from  $N=100$  model runs. In this scatterplot matrix, we can examine the relationships between the output variable and each of the input factors as well as the relationships among the input factors.

The correlation structure between the sets of free and basic transfer coefficients, which are obtained using Gauss-Jordan approach to maintain the steady-state condition (see Section 3.4, Equation 3.7), is revealed in this matrix plot (see Figure 3.26). As given in Equation 3.7,  $k_{21}$  is directly proportional to  $k_{12}$ ,  $k_{32}$  to  $k_{23}$  and  $k_{41}$  to  $k_{74}$ . So each of these pairs of transfer coefficients are perfectly correlated. The associations between  $k_{41}$  and  $k_{17}$ ;  $k_{51}$  and  $k_{75}$ ;  $k_{61}$  and  $k_{76}$ ;  $k_{74}$  and  $k_{17}$ ;  $k_{85}$  and  $k_{18}$  are very strong. Since each of the basic transfer coefficients ( $k_{21}$ ,  $k_{41}$ ,  $k_{51}$ ,  $k_{61}$ ,  $k_{32}$ ,  $k_{74}$  and  $k_{85}$ ) is highly correlated with at least one of the free variables ( $k_{12}$ ,  $k_{23}$ ,  $k_{75}$ ,  $k_{76}$ ,  $k_{86}$ ,  $k_{17}$ ,  $k_{87}$ ,  $k_{18}$ ) that are independent in the further analysis we take into consideration only the eight free  $k_{ij}$  as model input factors.

If we scan across the first row of the scatterplot matrix given in Figure 3.26, it is no surprise to see the same nature of a relationship between  $y_1(t = 2100)$  and both  $k_{32}$  and  $k_{23}$ ; and also with both  $k_{21}$  and  $k_{12}$ . It is clear that the high  $y_1(t = 2100)$  values are those with low  $k_{23}$  (or  $k_{32}$ ). It is less obvious (with  $N=100$ ) but also clear (with  $N=5,000$ ) that as  $k_{12}$  (or  $k_{21}$ ) increases, less CO<sub>2</sub> is transferred into the atmosphere compartment in year 2100 and  $y_1(t = 2100)$  falls.

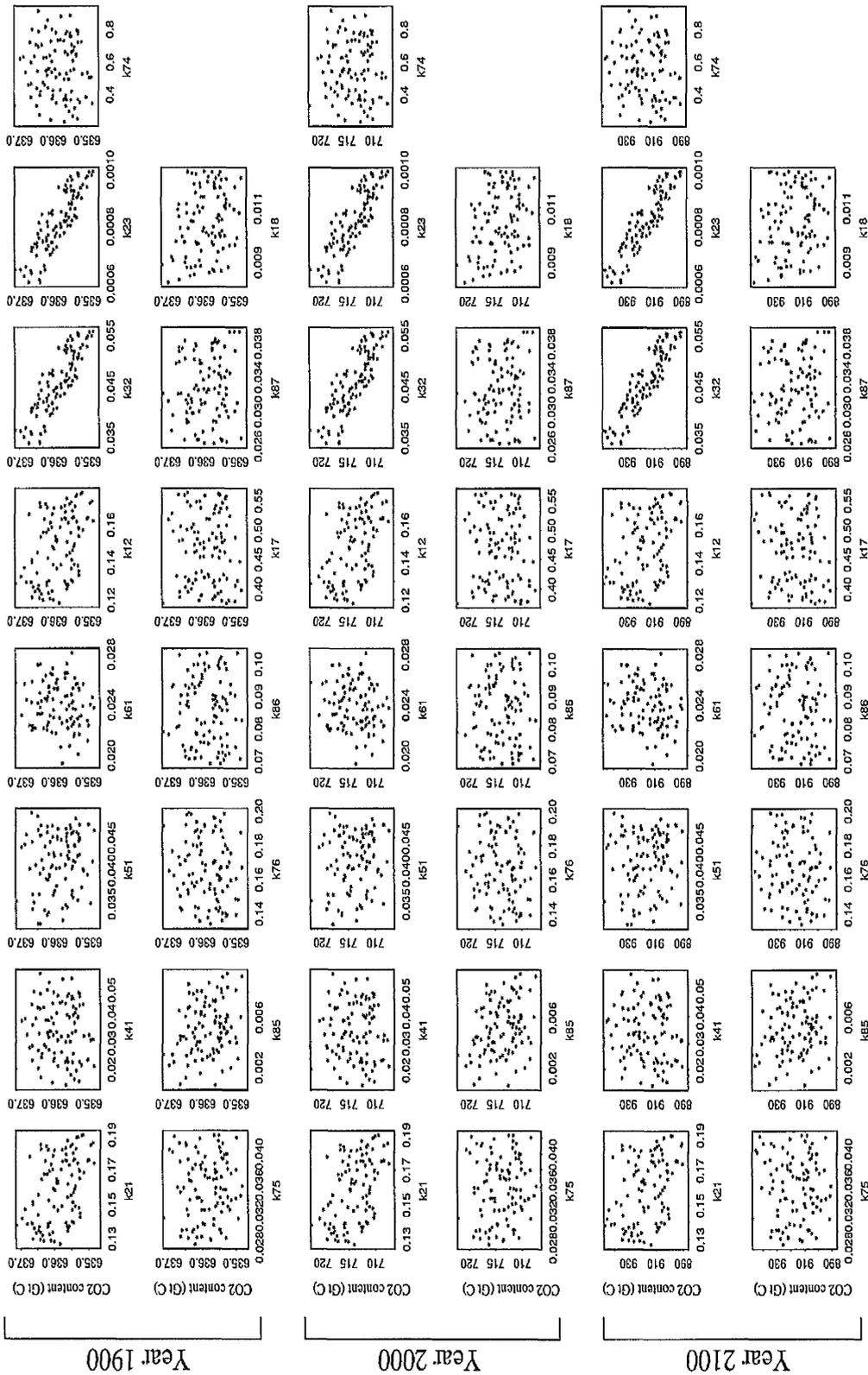


Figure 3.25. Scatterplots of predicted Atmospheric CO<sub>2</sub> content in years 1900, 2000 and 2100 versus each transfer coefficient  $k_{ij}$  (given in Table 3.3). To calculate these N=100 model predictions, Model I with IS92a emission scenario is used.

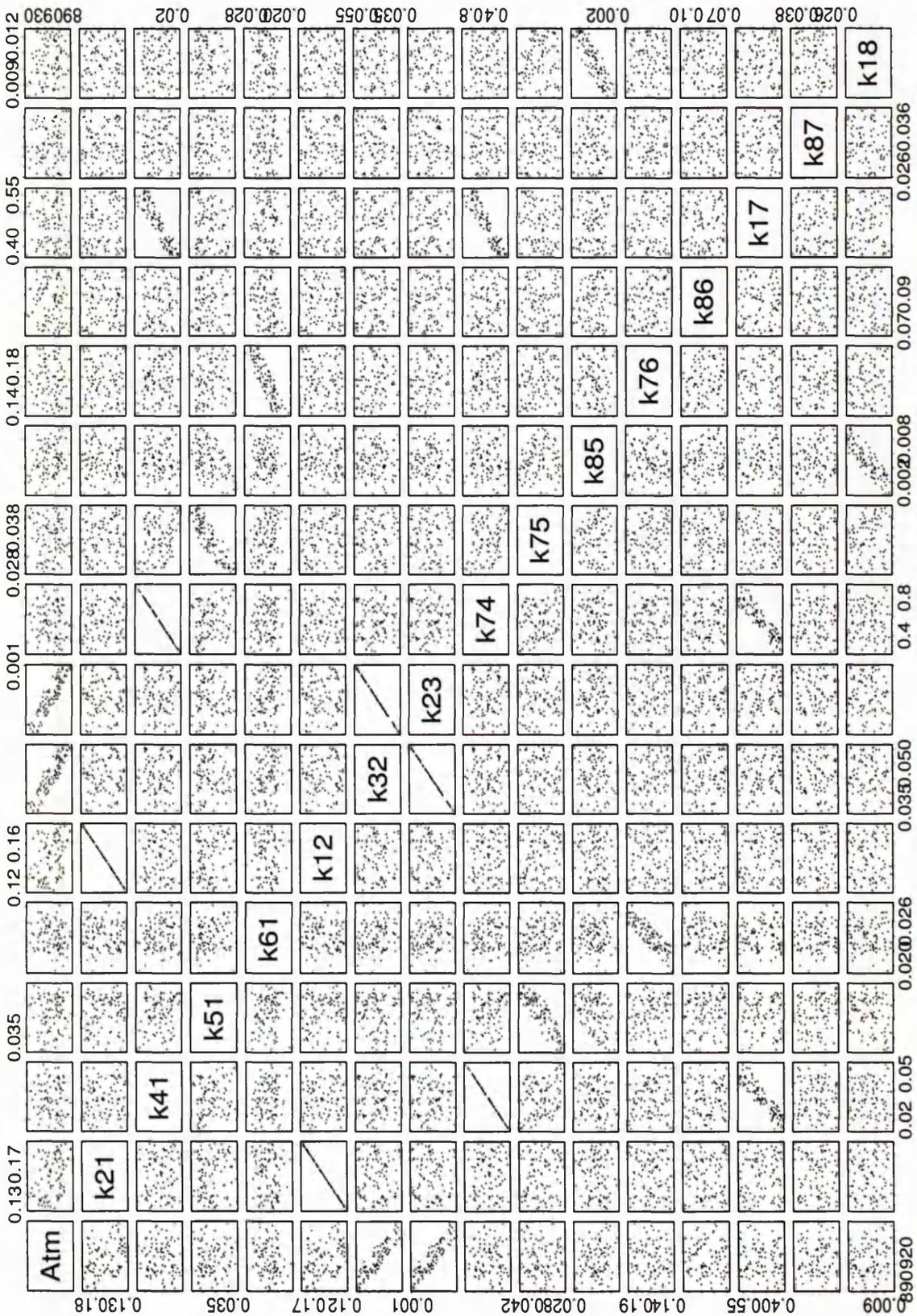


Figure 3.26. Scatterplot matrix of predicted Atmospheric CO<sub>2</sub> content in 2100 and each transfer coefficient  $k_{ij}$  (see Table 3.3). To calculate these  $N=100$  model predictions, Model I with IS92a emission scenario is used.

Overall, there is a strong linear association between  $y_1(t)$  and both  $k_{23}$  and  $k_{12}$ , but no particular association with the rest of the free input factors.

The rankings of the absolute correlations (i.e., the largest CC is given rank 1, the next largest CC is given rank 2, and so on) of the free transfer coefficients with all the response variables are given in Table 3.15. As the table of the absolute CC values (given in Table A.3 in Appendix A) reveals most of the CC values based on  $N=100$  model runs are slightly higher than the CCs obtained from  $N=5,000$  model runs, and also the order of importance between the transfer coefficients for the same compartment at a certain year changes with  $N$  (see Table 3.15). Because of the high dimension of the input space (eight, in this case), and because compared to  $N=100$  sample size,  $N=5,000$  provides a better coverage of the sample space of the input factors, and hence more reliable SA results, we have decided to present and discuss the results from the global SA methods based on  $N=5,000$  model evaluations in the remainder of this chapter.

The importance rankings in Table 3.15 indicates that the top three most influential transfer coefficients on the atmosphere, surface and deep ocean, nonwoody parts of trees, and ground vegetation compartments are  $k_{23}$ ,  $k_{12}$  and  $k_{18}$  in this order. Even though the corresponding CC values change with time, the order of importance stays the same.  $k_{23}$  appears to be the most influential transfer coefficient on all compartmental outputs except the active soil carbon compartment. The calculated CCs between  $k_{23}$  and all seven compartmental outputs at all three chosen years are quite high; in absolute value they vary between 0.69 and 0.97.  $k_{23}$  is particularly highly correlated with the ocean compartments; CC with the surface ocean compartment is around -0.96 in year 1900, decreases to about -0.94 in year 2000 and increases to around -0.98 in year 2100. As for the CCs between the output of the deep ocean compartment and  $k_{23}$ , it is about 0.97 in all three years (see Table A.3 in Appendix A).

For the woody parts of trees compartment, following  $k_{23}$  the second and the third most important transfer coefficients appear to be  $k_{75}$  and  $k_{12}$ , respectively, and this importance order does not change from year to year. The CC with  $k_{75}$

Table 3.15. Rankings of absolute Pearson correlation coefficients (CC) for the outputs of Model I. The outputs from years 1900, 2000 and 2100 based on N=100 and N=5,000 model runs are considered. The eight free transfer coefficients are ranked in order of importance, and the rankings based on N=5,000 model evaluations are highlighted.

Compartmental Output	Input Factor	CC Ranks					
		Yr 1900		Yr 2000		Yr 2100	
		N=100	N=5,000	N=100	N=5,000	N=100	N=5,000
Atmosphere (y1)	k <sub>12</sub>	2	2	2	2	2	2
	k <sub>23</sub>	1	1	1	1	1	1
	k <sub>75</sub>	4	4	4	4	4	4
	k <sub>76</sub>	8	7	8	7	8	7
	k <sub>86</sub>	7	8	6	8	7	8
	k <sub>17</sub>	6	5	7	5	6	5
	k <sub>87</sub>	5	6	5	6	5	6
	k <sub>18</sub>	3	3	3	3	3	3
Surface ocean (y2)	k <sub>12</sub>	3	2	3	2	3	2
	k <sub>23</sub>	1	1	1	1	1	1
	k <sub>75</sub>	5	4	5	4	5	4
	k <sub>76</sub>	7	7	6	7	8	7
	k <sub>86</sub>	8	6	8	6	7	6
	k <sub>17</sub>	6	5	7	5	6	5
	k <sub>87</sub>	4	8	4	8	4	8
	k <sub>18</sub>	2	3	2	3	2	3
Deep ocean (y3)	k <sub>12</sub>	2	2	2	2	2	2
	k <sub>23</sub>	1	1	1	1	1	1
	k <sub>75</sub>	6	4	6	4	6	4
	k <sub>76</sub>	8	7	7	7	8	7
	k <sub>86</sub>	5	5	5	5	5	5
	k <sub>17</sub>	4	8	4	8	4	8
	k <sub>87</sub>	3	6	3	6	3	6
	k <sub>18</sub>	7	3	8	3	7	3
Nonwoody parts of trees (y4)	k <sub>12</sub>	2	2	2	2	2	2
	k <sub>23</sub>	1	1	1	1	1	1
	k <sub>75</sub>	3	4	3	4	4	4
	k <sub>76</sub>	7	7	7	7	8	8
	k <sub>86</sub>	8	8	8	8	7	7
	k <sub>17</sub>	5	5	5	5	5	5
	k <sub>87</sub>	6	6	6	6	6	6
	k <sub>18</sub>	4	3	4	3	3	3
Woody parts of trees (y5)	k <sub>12</sub>	3	3	3	3	2	3
	k <sub>23</sub>	1	1	1	1	1	1
	k <sub>75</sub>	2	2	2	2	3	2
	k <sub>76</sub>	6	8	6	8	6	8
	k <sub>86</sub>	4	5	5	5	4	4
	k <sub>17</sub>	7	7	7	7	5	7
	k <sub>87</sub>	8	6	8	6	7	6
	k <sub>18</sub>	5	4	4	4	8	5
Ground vegetation (y6)	k <sub>12</sub>	2	2	2	2	2	2
	k <sub>23</sub>	1	1	1	1	1	1
	k <sub>75</sub>	4	5	4	4	4	5
	k <sub>76</sub>	6	4	6	5	7	4
	k <sub>86</sub>	8	6	7	6	8	6
	k <sub>17</sub>	7	7	8	7	6	7
	k <sub>87</sub>	5	8	5	8	5	8
	k <sub>18</sub>	3	3	3	3	3	3
Detritus/ decomposers (y7)	k <sub>12</sub>	3	3	3	3	2	2
	k <sub>23</sub>	1	1	1	1	1	1
	k <sub>75</sub>	4	4	4	4	4	5
	k <sub>76</sub>	8	8	8	8	8	8
	k <sub>86</sub>	7	6	5	6	7	6
	k <sub>17</sub>	2	2	2	2	3	3
	k <sub>87</sub>	6	7	6	7	6	7
	k <sub>18</sub>	5	5	7	5	5	4
Active soil carbon (y8)	k <sub>12</sub>	3	3	3	3	3	3
	k <sub>23</sub>	2	2	2	2	2	2
	k <sub>75</sub>	8	7	8	8	8	6
	k <sub>76</sub>	6	8	7	7	6	8
	k <sub>86</sub>	7	4	6	4	7	4
	k <sub>17</sub>	4	6	4	6	4	7
	k <sub>87</sub>	5	5	5	5	5	5
	k <sub>18</sub>	1	1	1	1	1	1

is high; about 0.53 in year 1900, 0.63 in 2000 and 0.36 in 2100. The CC with  $k_{12}$  given rank 3 is about -0.35 in all three years.

The second most important transfer coefficient for the detritus/decomposers compartment, after  $k_{23}$ , is  $k_{17}$  in years 1900 and 2000, but  $k_{12}$  in year 2100. The CC with  $k_{17}$  is moderately high, around 0.41 in year 1900 and about 0.52 in 2100.  $k_{12}$  which takes rank 2 in 2100 has an associated CC around -0.35. The correlation between  $k_{12}$  and the output of this compartment in years 1900 and 2000 (ranked as third) is around -0.38 which is still high compared to the CCs with the other transfer coefficients.

For the active soil carbon compartment the most important transfer coefficient appears to be  $k_{18}$  having CC around 0.78 in year 1900, and decreasing slightly with time.  $k_{23}$  is the second most important transfer coefficient for this compartment with a CC value of around -0.52 in years 1900 and 2000, and about -0.60 in year 2100.  $k_{12}$  is the third most influential transfer coefficient on this compartment with the CC value of around -0.25 at all three time points.

#### 3.5.4.2 Regression Methods

Considering the CO<sub>2</sub> content of each compartment in years 1900, 2000 and 2100 as the response variables, and the free transfer coefficients (i.e., independent input factors) as the explanatory variables, we now apply multiple regression method on the results of N=5,000 Monte Carlo iterations to investigate the effect of the transfer coefficients on the compartmental outputs. In a SA without any priori knowledge about the degree of influence the input factors have on the outputs, we have to construct a regression model with all explanatories. For illustration purpose a summary of the regression model involving the predicted 5,000 atmospheric CO<sub>2</sub> content in year 2100 (i.e.,  $y_1(t = 2100)$ ) as the response variable is given in Table 3.16. Seven of the eight transfer coefficients (i.e.,  $k_{12}$ ,  $k_{23}$ ,  $k_{75}$ ,  $k_{76}$ ,  $k_{86}$ ,  $k_{17}$  and  $k_{18}$ ) with  $p$ -values less than 0.01 appear to effect  $y_1(t = 2100)$ .

Since we know from the matrix plot and the correlation measures that there is either a linear or no association between the variables, we have used linear

**Table 3.16.** Summary of regression analysis with Model I output variable  $y_1(t = 2100)$  (Atmospheric CO<sub>2</sub> content at year 2100) and input factors  $k_{12}, k_{23}, k_{75}, k_{76}, k_{86}, k_{17}, k_{87}$  and  $k_{18}$ .

Variable	Regression Coefficient	Std. Error of Coeff.	T-test value	<i>p</i> -value	
$k_{12}$	-299.199	0.959	-311.95	0.000	
$k_{23}$	-117528.000	140.000	-836.69	0.000	
$k_{75}$	-167.267	3.998	-41.84	0.000	
$k_{76}$	-3.200	0.855	-3.74	0.000	
$k_{86}$	-16.564	1.635	-10.13	0.000	
$k_{17}$	-4.091	0.300	-13.62	0.000	
$k_{87}$	6.571	4.311	1.52	0.128	
$k_{18}$	-1982.95	13.380	-148.25	0.000	
R-Squared = 99.4%		Intercept = 1085.00			
Source	DF	Sum of Squares	Mean Sum of Squares	F-statistic	<i>p</i> -value
Regression	8	1095506	136938	103231.78	0.000
Residual	4991	6621	1		
Total	4999	1102126			

regression approach. However, it is important to note that a regression analysis, specially if it is based on identifying a linear relationship when there are other types of relationships between the variables, can fail to show that a variable has an effect on the response.

As we have done in Table 3.16, to present regression analyses results for all response variables evaluated at three different years is rather cumbersome and such tables also involve variables that appear to have no significant effect on the response. Stepwise regression analysis is a more informative and less cumbersome way of constructing and displaying regression models, and we will be presenting the results from this procedure in Section 3.5.4.4.

Because of the effects of units and distributional assumptions, it is difficult to obtain input factor importance from the regression coefficients. As we shall shortly

see in the following section, input factor importance is more clearly assessed by using standardized regression coefficients.

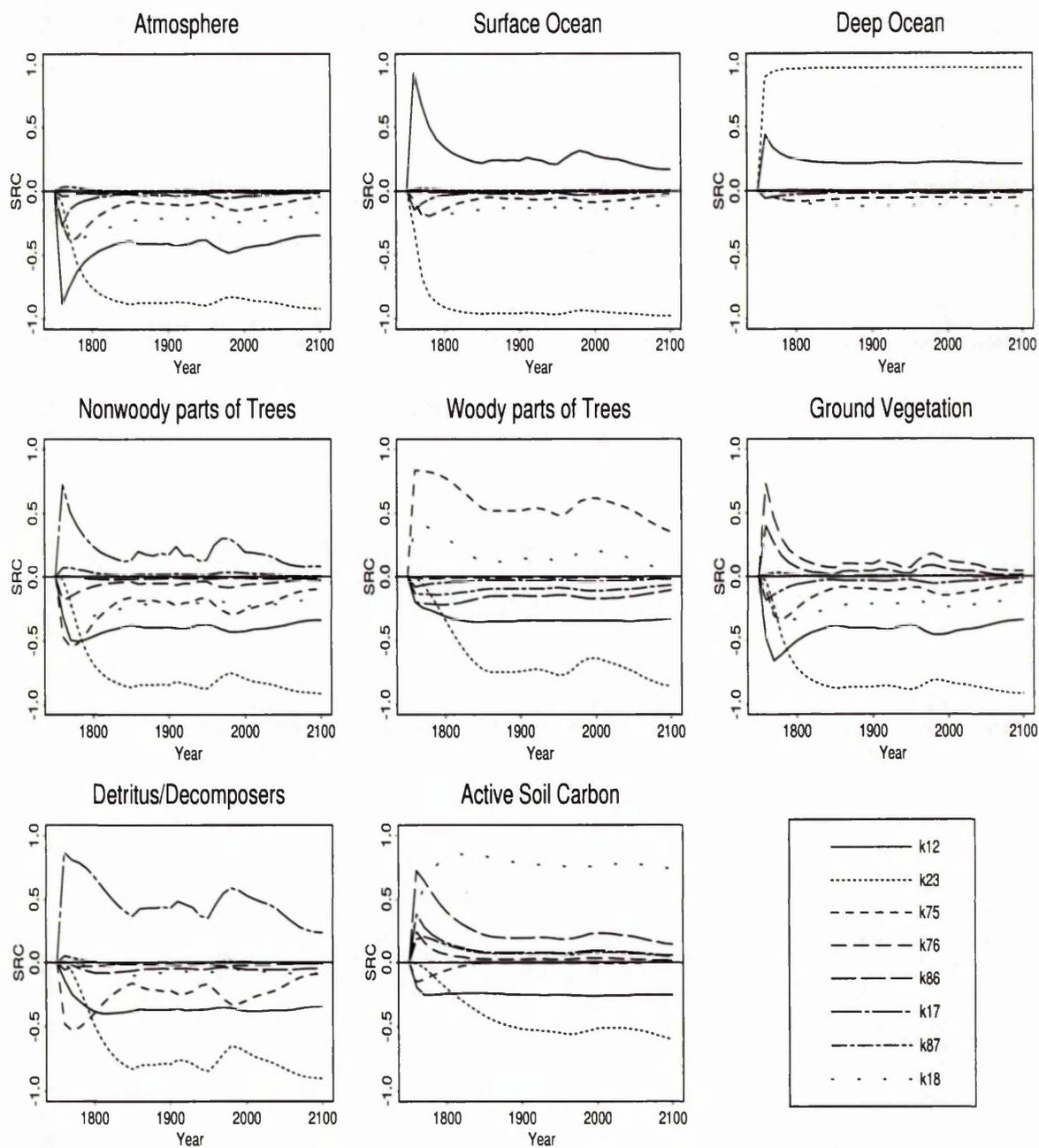
### 3.5.4.3 SRC and PCC

From the SA point of view, to obtain a measure of relative importance of the transfer coefficients we have calculated the standardized regression coefficients (SRCs). As another measure of input factor importance, we have also computed the partial correlation coefficients (PCCs) for all eight compartmental output variables.

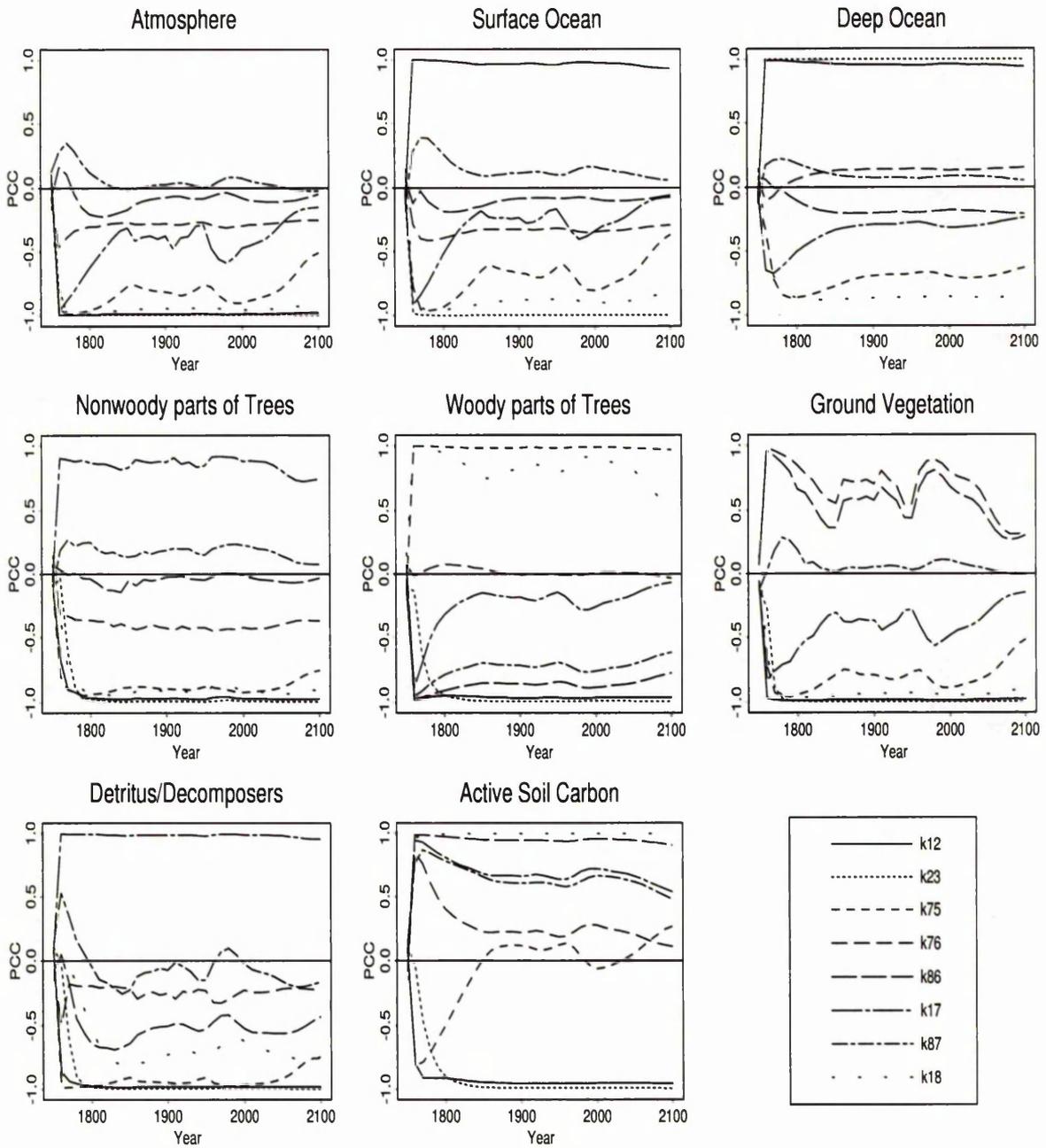
When using SRC, it is also important to consider the coefficient of determination  $R^2$  of the linear regression models fitted to the data. The  $R^2$  values for all regression models constructed from the output of the  $i$ -th compartment at time  $t$  ( $y_i(t)$ ,  $i=1,\dots,8$ ) and the eight independent transfer coefficients are all very high, they lie between 98.4% and 99.6%. Note that the calculations here are based on  $N=5,000$  model evaluations.

We do not list the SRC and the PCC values here, but instead we present the time dependent behaviour of these coefficient estimates for the whole time period from 1750 to 2100 graphically (see Figure 3.27 for SRC, and Figure 3.28 for PCC plots). This is a useful way of presenting sensitivity results for output variables that are functions of time, like in our case. The SRC and PCC values associated with each transfer coefficient for a certain compartment and at a certain year can easily be read from these figures.

Each frame in Figures 3.27 and 3.28 shows results for each output variable, and each curve in these frames displays, respectively, the SRC & PCC values relating the corresponding output variable to one input factor as a function of time. According to what we see in both of these figures, based on the SRCs and PCCs, the sensitivity of some compartments to some of the transfer coefficients changes over time, but within the same compartment the order of importance between the input factors hardly changes with time. The most influential factor is  $k_{23}$  for all compartments except for active soil carbon compartment for which



**Figure 3.27.** Standardized regression coefficients (SRCs) for the eight independent transfer coefficients, with the predicted  $\text{CO}_2$  concentrations in each compartment (i.e., with the dependent variables  $y_1(t), \dots, y_8(t)$ ),  $N=5,000$ .



**Figure 3.28.** Partial correlation coefficients (PCCs) for the eight independent transfer coefficients, with the predicted  $\text{CO}_2$  concentrations in each compartment (i.e., with the dependent variables  $y_1(t), \dots, y_8(t)$ ),  $N=5,000$ .

$k_{18}$  appears to be the first and  $k_{23}$  the second most important input factors. Except for  $k_{23}$ , the influence of all transfer coefficients on all compartmental outputs tend to decrease over time. This decrease with  $k_{12}$  for all outputs and with  $k_{18}$  for the output of active soil carbon compartment is much slower than with the other transfer coefficients.

In Table 3.17, we give the ranking of the absolute value of the SRCs and PCCs for each model output evaluated at the three years. With both SRCs and PCCs, a comparison of the rankings within the same compartment but at different years reveal that -for the surface ocean, nonwoody parts of trees and the woody parts of trees compartments when SRCs are concerned, and the woody parts of trees compartment when the PCCs are concerned- the importance ordering of the input factors changes from year to year, but not dramatically. For instance, ranking with the PCCs for the woody parts of trees compartment shows that  $k_{86}$  is the fourth and  $k_{18}$  the fifth most important factors in 1900, then in 2000  $k_{86}$  becomes the fifth and  $k_{18}$  the fourth. Considering the most important three transfer coefficients (highlighted in Table 3.17) we can see that the same transfer coefficients identified as the three most influential ones by using CCs (based on the data from  $N=5,000$  model evaluations) are picked out by this method of SA as well: in descending order of importance, the top three most important input factors for the atmosphere, surface and deep ocean, nonwoody parts of trees and ground vegetation compartments are  $k_{23}$ ,  $k_{12}$  and  $k_{18}$ ; for the woody parts of trees compartment  $k_{23}$ ,  $k_{12}$  and  $k_{75}$ ; for the detritus decomposers compartment  $k_{23}$ ,  $k_{17}$  and  $k_{12}$ ; and lastly for the active soil carbon compartment  $k_{18}$ ,  $k_{23}$  and  $k_{12}$ .

In comparison to the SRCs (given in Figure 3.27), the PCC estimates (given in Figure 3.28) are higher, but the rankings of transfer coefficient importance obtained from the PCCs are almost identical to the rankings from the SRCs. In ideal circumstances, when the sampled input factor values are independent, the use of CCs, SRCs and PCCs will produce identical rankings of input factor importance [49], but if even small correlations exist between the input factors in the sample this can result in the importance rankings not being identical. In our

Table 3.17. Rankings of absolute Standardized Regression Coefficients (SRC) and Partial Correlation Coefficients (PCC) for the outputs of Model I. The outputs from years 1900, 2000 and 2100 based on N=5,000 model runs are considered. The eight free transfer coefficients are ranked in order of importance, and the 3 most important factors are highlighted.

Compart. Output	Input Factor	SRC Ranks		
		Yr 1900	Yr 2000	Yr 2100
Atmosphere (y1)	k12	2	2	2
	k23	1	1	1
	k75	4	4	4
	k76	7	7	7
	k86	6	6	6
	k17	5	5	5
	k87	8	8	8
	k18	3	3	3
Surface ocean (y2)	k12	2	2	2
	k23	1	1	1
	k75	4	4	4
	k76	8	8	8
	k86	6	7	6
	k17	5	5	5
	k87	7	6	7
	k18	3	3	3
Deep ocean (y3)	k12	2	2	2
	k23	1	1	1
	k75	4	4	4
	k76	7	7	7
	k86	6	6	6
	k17	5	5	5
	k87	8	8	8
	k18	3	3	3
Nonwoody parts of trees (y4)	k12	2	2	2
	k23	1	1	1
	k75	4	4	4
	k76	6	6	6
	k86	8	8	7
	k17	5	5	5
	k87	7	7	8
	k18	3	3	3
Woody parts of trees (y5)	k12	3	3	3
	k23	1	1	1
	k75	2	2	2
	k76	8	8	8
	k86	4	5	4
	k17	7	7	7
	k87	6	6	5
	k18	5	4	6
Ground vegetation (y6)	k12	2	2	2
	k23	1	1	1
	k75	4	4	4
	k76	5	5	5
	k86	6	6	6
	k17	7	7	7
	k87	8	8	8
	k18	3	3	3
Detritus/ decomposers (y7)	k12	3	3	3
	k23	1	1	1
	k75	4	4	4
	k76	7	7	7
	k86	6	6	6
	k17	2	2	2
	k87	8	8	8
	k18	5	5	5
Active soil carbon (y8)	k12	3	3	3
	k23	2	2	2
	k75	8	8	8
	k76	7	7	7
	k86	4	4	4
	k17	5	5	5
	k87	6	6	6
	k18	1	1	1

Compart. Output	Input Factor	PCC Ranks		
		Yr 1900	Yr 2000	Yr 2100
Atmosphere (y1)	k12	2	2	2
	k23	1	1	1
	k75	4	4	4
	k76	6	6	6
	k86	7	7	7
	k17	5	5	5
	k87	8	8	8
	k18	3	3	3
Surface ocean (y2)	k12	2	2	2
	k23	1	1	1
	k75	4	4	4
	k76	5	5	5
	k86	8	8	8
	k17	6	6	6
	k87	7	7	7
	k18	3	3	3
Deep ocean (y3)	k12	2	2	2
	k23	1	1	1
	k75	4	4	4
	k76	7	7	7
	k86	6	6	6
	k17	5	5	5
	k87	8	8	8
	k18	3	3	3
Nonwoody parts of trees (y4)	k12	2	2	2
	k23	1	1	1
	k75	4	4	4
	k76	6	6	6
	k86	8	8	8
	k17	5	5	5
	k87	7	7	7
	k18	3	3	3
Woody parts of trees (y5)	k12	3	3	3
	k23	1	1	1
	k75	2	2	2
	k76	8	8	8
	k86	4	5	4
	k17	7	7	7
	k87	6	6	5
	k18	5	4	6
Ground vegetation (y6)	k12	2	2	2
	k23	1	1	1
	k75	4	4	4
	k76	5	5	5
	k86	6	6	6
	k17	7	7	7
	k87	8	8	8
	k18	3	3	3
Detritus/ decomposers (y7)	k12	3	3	3
	k23	1	1	1
	k75	4	4	4
	k76	7	7	7
	k86	6	6	6
	k17	2	2	2
	k87	8	8	8
	k18	5	5	5
Active soil carbon (y8)	k12	3	3	3
	k23	2	2	2
	k75	8	8	8
	k76	7	7	7
	k86	4	4	4
	k17	5	5	5
	k87	6	6	6
	k18	1	1	1

case, as shown in Tables 3.17 (for SRCs and PCCs) and 3.15 (for CCs) there are slight changes in the importance order of the input factors but this is the case with the least important transfer coefficients given ranks between 4 and 8. For the most important three input factors, all three methods provide the same rankings.

#### 3.5.4.4 Stepwise Regression

Considering an  $\alpha$ -value of 0.01 to add an input factor to a regression model and an  $\alpha$ -value of 0.05 to drop a factor from the model, we have carried out stepwise procedure on the data. The order in which the inputs are added to the model at each step provides the order of importance between the transfer coefficients that are included in the final model.

In Table 3.18, the results of the stepwise regression for the output of atmosphere compartment in 1900, 2000 and 2100 are given. All analyses results seem effective with very high  $R^2$ -values. The results, based on  $N=5,000$  model runs, show that seven of the eight transfer coefficients (i.e.,  $k_{23}$ ,  $k_{12}$ ,  $k_{18}$ ,  $k_{75}$ ,  $k_{17}$ ,  $k_{86}$  and  $k_{76}$ , in this order) are included in the final model, no matter the output in which year is considered. The  $R^2$ -values over 99% indicate that these regressions are successful in accounting for the observed variability in  $y_1$  at the selected years.

Considering the changes in the  $R^2$ -values that occur as additional inputs are added to the regression models, we see that having included  $k_{23}$ ,  $k_{12}$  and  $k_{18}$  in all of these three models, the contribution to the % variation explained in the output variable due to the other input factors is very small (less than 5%). For example, let us consider the model for  $y_1(t = 1900)$ . The  $R^2$ -value of 77% indicates that  $k_{23}$  accounts for 77% of the variability in the output, while  $k_{23}$  and  $k_{12}$  taken together account for 93.60%, and  $k_{23}$ ,  $k_{12}$  and  $k_{18}$  account for 98.28% of the variability. The rest of the input factors all together account for  $99.44\% - 98.28\% = 1.14\%$  of the variability in  $y_1(t = 1900)$ . One might even argue that only  $k_{23}$  and  $k_{12}$  have large impact on the output.

Stepwise regression is a valuable tool for selecting the most important input

**Table 3.18.** Summary of stepwise regression analyses for Atmospheric CO<sub>2</sub> content (i.e., output variable  $y_1$ ) of Model I in years 1900, 2000 and 2100; based on N=5,000 model runs; and IS92a emission scenario.

Step	$y_1(t = 1900)$		$y_1(t = 2000)$		$y_1(t = 2100)$	
	Variable	$R^2$	Variable	$R^2$	Variable	$R^2$
1	$k_{23}$	0.7700	$k_{23}$	0.7137	$k_{23}$	0.8462
2	$k_{12}$	0.9360	$k_{12}$	0.9105	$k_{12}$	0.9646
3	$k_{18}$	0.9828	$k_{18}$	0.9704	$k_{18}$	0.9915
4	$k_{75}$	0.9932	$k_{75}$	0.9925	$k_{75}$	0.9936
5	$k_{17}$	0.9942	$k_{17}$	0.9941	$k_{17}$	0.9939
6	$k_{86}$	0.9943	$k_{86}$	0.9943	$k_{86}$	0.9940
7	$k_{76}$	0.9944	$k_{76}$	0.9944	$k_{76}$	0.9940

factors and providing an order of importance among these factors, but often with models involving large number of explanatory variables there is possibility of overfitting the data. So, it is important to check if the fitted regression model is a reasonable one. For this purpose, as suggested by Helton & Davis (see [49]), predicted error sum of squares (PRESS) can be used. To calculate PRESS values for a regression model with  $p$  variables (i.e. transfer coefficients in this case), the following procedure is used. For  $n = 1, 2, \dots, N$ , the  $n$ th observation is deleted from the original set of  $N$  observations and then a regression model containing the original  $p$  variables is obtained from the remaining  $N-1$  observations. Based on this new regression model, the value  $\hat{y}_p(n)$  is estimated for the deleted observation  $y_n$ . PRESS value is then defined using the preceding predictions and the  $N$  original observations as

$$PRESS_p = \sum_{n=1}^N (y_n - \hat{y}_p(n))^2.$$

Most statistical packages, such as Minitab, can be used to easily compute PRESS values. Table 3.19 reports the PRESS values for the regression models summarized in Table 3.18. As shown by the decreasing PRESS values in Table 3.19, the regression models in these analyses are probably not overfitting the

**Table 3.19.** Predicted error sum of squares (PRESS) values for the regression models summarized in Table 3.18.

Step	$y_1(t = 1900)$		$y_1(t = 2000)$		$y_1(t = 2100)$	
	Variable	PRESS	Variable	PRESS	Variable	PRESS
1	$k_{23}$	369.598	$k_{23}$	17636.700	$k_{23}$	169629.000
2	$k_{12}$	102.839	$k_{12}$	5512.000	$k_{12}$	39040.900
3	$k_{18}$	27.615	$k_{18}$	1824.960	$k_{18}$	9393.180
4	$k_{75}$	10.958	$k_{75}$	462.853	$k_{75}$	7032.160
5	$k_{17}$	9.346	$k_{17}$	361.149	$k_{17}$	6792.710
6	$k_{86}$	9.133	$k_{86}$	353.058	$k_{86}$	6662.710
7	$k_{76}$	9.042	$k_{76}$	346.872	$k_{76}$	6646.860

data from which they were constructed.

Resulting from stepwise regression analyses that are found to be successful in the sense that they have quite high  $R^2$ -values, the order of importance between the top three transfer coefficients (i.e., the order in which they are added to the corresponding regression model) and the  $R^2$ -values obtained for successive models are shown in Table 3.20. As examination of this table shows the top three input factors account for most of the variation in all the outputs considered in the analyses. The contribution to the variability in the outputs due to the other input factors which were also added to the regression models is very small. The complete order of importance for all output variables and the corresponding  $R^2$ -values are given in Table A.4, Appendix A.

Except for the active soil carbon compartment in all three years,  $k_{23}$  appears to be having the most significant influence on all compartments in all three years. The same three most influential input factors  $k_{23}$ ,  $k_{12}$  and  $k_{18}$  (in descending order of importance) are selected in the analyses of atmosphere, both ocean and ground vegetation compartments. This order of input factor selection is almost the same for nonwoody parts of trees compartment with a minor change in 2000 ( $k_{75}$  is added to the regression model before  $k_{18}$ ).

The analyses for the woody parts of trees compartment show that following  $k_{23}$ ,  $k_{75}$  is the second and  $k_{12}$  the third most effective inputs on this compartment

**Table 3.20.** Three most important transfer coefficients identified by stepwise regression procedure on Model I. The importance ranking based on the order at which the transfer coefficients were added to the corresponding model and the  $R^2$ -values for the regression models at each step are given. In the analysis, the outputs in years 1900, 2000 and 2100 are considered.

Compartmental Output	Step	Yr 1900		Yr 2000		Yr 2100	
		Input	$R^2$	Input	$R^2$	Input	$R^2$
Atmosphere	1	$k_{23}$	0.770	$k_{23}$	0.714	$k_{23}$	0.846
	2	$k_{12}$	0.936	$k_{12}$	0.911	$k_{12}$	0.965
	3	$k_{18}$	0.983	$k_{18}$	0.970	$k_{18}$	0.992
Surface Ocean	1	$k_{23}$	0.919	$k_{23}$	0.892	$k_{23}$	0.955
	2	$k_{12}$	0.974	$k_{12}$	0.967	$k_{12}$	0.982
	3	$k_{18}$	0.991	$k_{18}$	0.987	$k_{18}$	0.994
Deep Ocean	1	$k_{23}$	0.932	$k_{23}$	0.931	$k_{23}$	0.935
	2	$k_{12}$	0.977	$k_{12}$	0.979	$k_{12}$	0.976
	3	$k_{18}$	0.992	$k_{18}$	0.992	$k_{18}$	0.992
Nonwoody parts of Trees	1	$k_{23}$	0.715	$k_{23}$	0.636	$k_{23}$	0.830
	2	$k_{12}$	0.875	$k_{12}$	0.822	$k_{12}$	0.948
	3	$k_{18}$	0.920	$k_{75}$	0.889	$k_{18}$	0.976
Woody parts of Trees	1	$k_{23}$	0.547	$k_{23}$	0.399	$k_{23}$	0.735
	2	$k_{75}$	0.823	$k_{75}$	0.790	$k_{75}$	0.864
	3	$k_{12}$	0.941	$k_{12}$	0.904	$k_{12}$	0.973
Ground Vegetation	1	$k_{23}$	0.760	$k_{23}$	0.693	$k_{23}$	0.843
	2	$k_{12}$	0.925	$k_{12}$	0.892	$k_{12}$	0.961
	3	$k_{18}$	0.972	$k_{18}$	0.951	$k_{18}$	0.989
Detritus/ Decomposers	1	$k_{23}$	0.612	$k_{23}$	0.469	$k_{23}$	0.801
	2	$k_{17}$	0.798	$k_{17}$	0.752	$k_{12}$	0.920
	3	$k_{12}$	0.934	$k_{12}$	0.894	$k_{17}$	0.975
Active Soil Carbon	1	$k_{18}$	0.606	$k_{18}$	0.579	$k_{18}$	0.536
	2	$k_{23}$	0.881	$k_{23}$	0.853	$k_{23}$	0.902
	3	$k_{12}$	0.946	$k_{12}$	0.924	$k_{12}$	0.968

at the chosen three years. For the detritus/decomposers compartment, the analyses define  $k_{23}$  as the most important input factor at all three years. This factor is followed by  $k_{17}$  and then  $k_{12}$  in 1900 and 2000, but by  $k_{12}$  first and then  $k_{17}$  in 2100. The analyses with active soil carbon compartment data have identified the top three most important input factors to be  $k_{18}$  as the first,  $k_{23}$  the second and  $k_{12}$  the third.

### 3.5.5 Discussion on the Results

Considering the simplicity of the state equations, i.e. mathematical representation of the model and the magnitude of the uncertainties in the values for the initial conditions, it is expected that the initial conditions with larger uncertainty ranges would dominate the sensitivity results and appear to be the most influential factors on model outputs. A sensitive model input factor with a given large uncertainty range will contribute more to the uncertainty in model output.

Taking into consideration the fact that the mathematical structure of the model has been kept as simple as possible, and the magnitude of the uncertainty allowed about each of the input factors (20% of their literature values), the factors with high nominal values and wider uncertainty ranges, therefore are dominant on output sensitivity results.

Even though OAT sensitivity analysis approach produces some benefits, it is not a very efficient way of performing a sensitivity analysis and it is limited in application to models which are not very expensive to run and have small number of input factors [64].

Since we have relatively small models with relatively small number of input factors and the model codes are very fast running, we have used OAT design to obtain a sensitivity ranking. If the number of input factors considered is not small and the model runs slow, this design is very time consuming and not very practical. Even with the relatively small number of inputs we have in our model (8 initial conditions and 15 transfer coefficients for Model I, and 8 initial conditions and 18 transfer coefficients for Model II) the application of this design was quite impractical. A very important disadvantage of OAT design is that it has an underlying assumption frequently not valid for the models which can result in a confused picture of how input factors affect model behaviour. This is the case with GCC models for which we have to take into account the fact that the system has to be in steady-state before introducing any perturbations to the system. So to retain this condition, when one input factor is varied over its entire range at least one other input factor has to be changed to initialize the GCC model. As

a result, carbon content of each compartment change during a simulation as a function of not only the factor varied but also the one(s) that are calculated, of course in addition to the time-dependent releases of carbon from fossil fuel emissions and forest clearing.

In the local analysis methods based on sampling we considered a small  $N=100$  and a large  $N=5,000$  sample size to investigate the effect of sample size on the results. It was shown that even though there was some variation in the importance ranking of the input factors as a result of different sample size, that was not the case with the most important factors, that is, sample size  $N=100$  was large enough to identify the factors that have large effects on the model outputs. With global SA methods, considering the large dimension of the input space the sample size of  $N=5000$  was considered. As a rule of thumb, about 100 runs for each input factor are usually performed [24]. Based on this argument we believe 5000 runs are sufficiently large to cover the whole input space.

In this chapter, we have found out that the variability associated with a sensitive input factor is transferred through the model resulting in a large contribution to the overall output variability. We also found out that model results can be highly correlated with an input factor so that small changes in the input value result in significant changes in the output.

The techniques used in this chapter, such as correlations, regression coefficients, partial correlation coefficients are optimal choices for the input selection since the output variables appear to behave in a linear fashion. Because, there is no nonlinear relationships present between the output variables and the inputs, the analysis was not performed on the ranks of the data.

Because model outputs are time dependent function of input factors, individual input factors have been examined at various time points. As a result we have found out that their importance changes through time. We also found out that different model input factors are important for different model compartments.

An important question is the extent to which the different techniques agree

in their identification of important input factors. As the results for both 8-compartment models utilised in this chapter reveal (see Appendix B for the results of Model II), even though there are small variations in the input factor importance order with the different SA methods, the overall results are quite similar.

## Chapter 4

# Sensitivity Analysis Techniques Applied to a 25-Compartment GCC Model

### 4.1 Introduction

In the previous chapter, various sensitivity analysis (SA) techniques were applied to two 8-compartment global carbon cycle (GCC) models. In this chapter we consider a more complex, 25-compartment, GCC model. To study the performance of this more complex compartmental model, we analyse the model sensitivity to the input factors by employing various SA techniques defined in Chapter 2 and already applied to the 8-compartment GCC models in Chapter 3, which include sensitivity indices, standardized ranges, Morris screening, regression methods, standardized regression coefficient (SRC), partial correlation coefficient (PCC), SRC and PCC on ranks (i.e. SRRC and PRCC). The Pearson coefficient (CC), Spearman coefficient (RCC) and the Smirnov test are also considered, together with a few other non-parametric tests. By using these SA methods we aim to analyse the relative performance of the different SA techniques employed as well as the model performance.

A description of the linear, time-invariant 25-compartment GCC model under consideration here is given in Section 4.2, together with the characteristics of the model input factors selected for SA. In Section 4.3 we report the results of standard OAT and Morris designs. Then in Section 4.4 analyses results from global SA techniques are presented. Discussion and conclusions are given in Section 4.5.

## 4.2 The Model

Compared to the two 8-compartment GCC models, the 25-compartment model has a more detailed formulation of the dynamics of the carbon cycle.

The model and its computer implementation are adapted from a technical report (see [31]) prepared for the United States Department of Energy by Emanuel *et al.* (1984). It was originally developed for the purpose of predicting the future extent of the greenhouse effect. The model represents three major components -the atmosphere, oceans, and terrestrial systems- of the cycle with 25 compartments. The atmosphere is represented by a single compartment and the oceans by 19 globally averaged layers with depth. The terrestrial systems component of the model is described by 5 compartments. The Figure 4.1 illustrates the model structure. The ocean component takes into account the dependence of the ocean's horizontal cross-sectional area and the carbon concentration on depth. In this component, 'surface ocean' corresponds to waters above 75 m., and the 'deep ocean', which is divided into 18 horizontal layers, corresponds to waters in 75-4500 m. depth. The surface ocean compartment exchanges carbon with the atmosphere and the deep ocean exchanges carbon only with the surface ocean. The terrestrial component of the model allocates carbon among 5 compartments, namely 'nonwoody parts of trees', 'woody parts of trees', 'ground vegetation', 'detritus/decomposers' and 'active soil carbon'. A detailed description of the model is contained in [31]. In this model, CO<sub>2</sub> is released to the atmosphere by fossil fuel

combustion. Deforestation also results in a direct transfer of carbon to the atmosphere from 'tree' compartments as well as a transfer to 'detritus/decomposers'. The relative magnitudes of transfers from the atmosphere to 'tree' and 'ground vegetation' are altered as a result of land-use change.

Carbon in the atmosphere, the inorganic carbon in the oceans, and carbon storage in the terrestrial systems are calculated. The primary dynamic variables (state variables) of this model are the masses of total carbon in each compartment. Beginning from a preindustrial steady-state, model results at annual time scales are considered. Substantial alterations had been made to the computer code of the model in order to implement it for sensitivity and uncertainty analysis. The historical and the future fossil fuel combustion and land-use change data given in the code have been updated. For historical fossil fuel CO<sub>2</sub> emissions, estimates given by Marland *et al.* (1999) (see [80]); for historical land-use change, data given by Enting *et al.* (1994) (see [32]); and for the future predictions of fossil fuel combustion and deforestation IPCC scenarios (see [58]) are used. The masses of CO<sub>2</sub> are expressed in Gigatons (=10<sup>15</sup> g), time in years, and the concentration of CO<sub>2</sub> in the atmosphere in parts per million by volume (ppmv).

Inspection of the relationships between the model input factors and outputs led to the selection of 30 independent input factors that are subject to uncertainty. The description of the model input factors, their nominal values, variability ranges and units are given in Table 4.1. The first six input factors describe the initial conditions of the atmosphere and the terrestrial biota compartments. The next seven describe land-use practices. These are followed by the inputs related to the chemical and physical parameters of the oceans. The remaining input factors are used to calculate the coefficients that control the fluxes between the terrestrial components of the model. Due to the lack of information about the distribution of these input factors, we assume that they all follow uniform distributions over their assigned ranges. For the dynamic equations of the model considered by Emanuel *et al.* see Table 4.2.

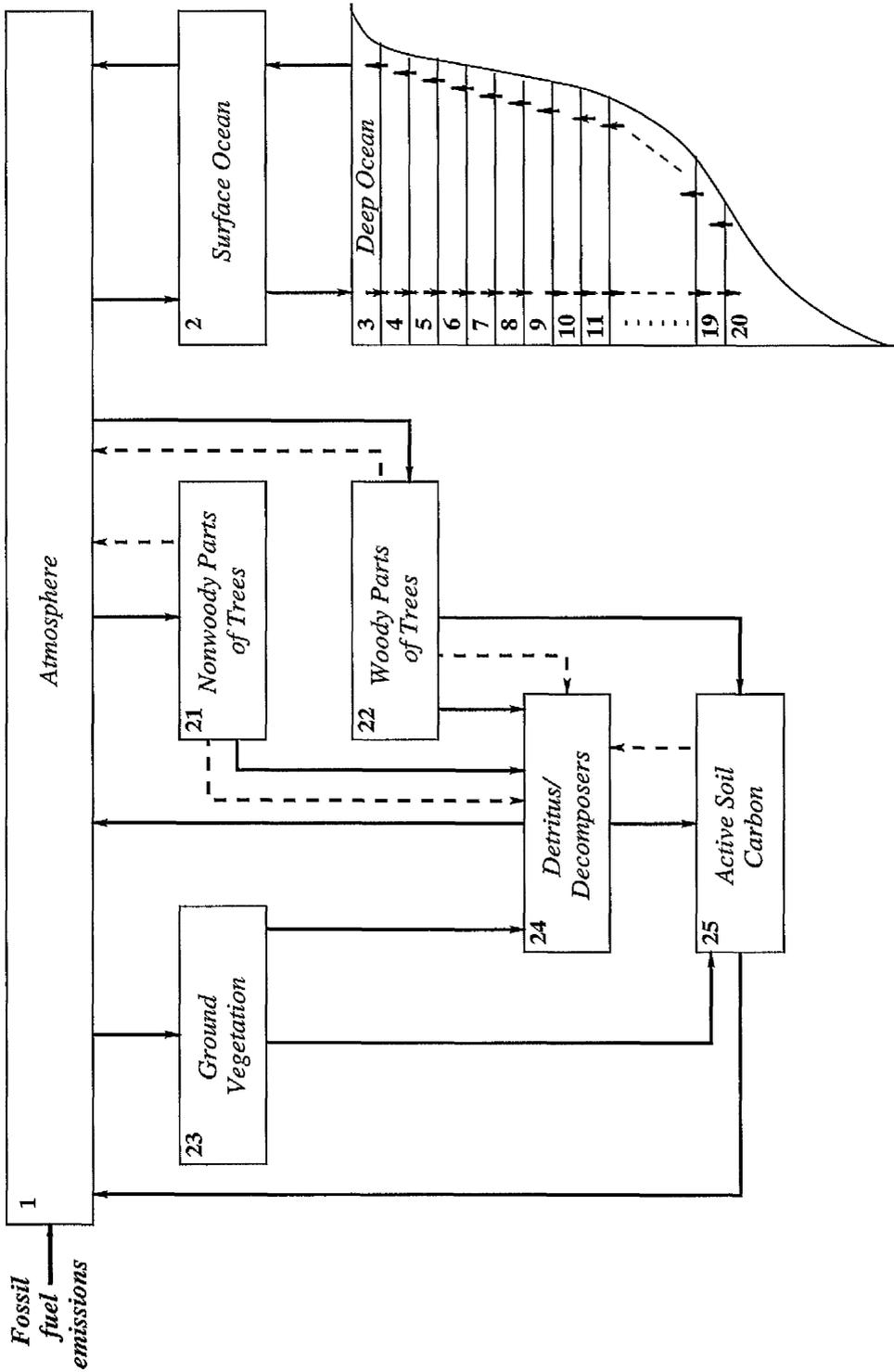


Figure 4.1. Diagram of 25-Compartment Global Carbon Cycle Model (adopted from Emanuel *et al.* (1984) [31]). Solid arrows indicate naturally occurring carbon fluxes, while dashed arrows are associated with forest clearing.

Table 4.1. Model input factors selected for sensitivity analysis

Description	Factor <sup>a</sup>	Value	Range	Unit
<b>Initial conditions</b>				
Atmosphere	$c_1^o$ (CA0)	548.80	510.7 - 596.0	Gt C
Nonwoody parts of trees	$c_{21}^o$ (CF0)	38.20	30.0 - 46.0	Gt C
Woody parts of trees	$c_{22}^o$ (CW0)	634.50	507.0 - 762.0	Gt C
Ground vegetation	$c_{23}^o$ (CG0)	59.30	47.0 - 72.0	Gt C
Detritus/decomposers	$c_{24}^o$ (CD0)	108.20	86.0 - 130.0	Gt C
Active soil carbon	$c_{25}^o$ (CSL0)	1131.00	905.0 - 1348.0	Gt C
<b>Forest clearing</b>				
Fraction of forest clearing carbon transferred to atmosphere	$\varphi_A$ (PHIA)	0.5	0.4 - 0.6	—
Fraction of forest clearing carbon transferred to detrit./decomp.	$\varphi_D$ (PHID)	0.5	0.4 - 0.6	—
Ratio of soil to detrit./decomp. flux to forest clearing flux	$\psi_s$ (PSIS)	0.1	0.08 - 0.12	—
Fraction of forest clearing release that serves to decrease capacity for carbon storage in trees	$\xi_r$ (SXIT)	0.5	0.4 - 0.6	—
<b>Reforestation</b>				
Rate of re-establishment of tree compartments	$\sigma_r$ (SIG)	1.0E-6	0.8E-6 - 1.2E-6	yr <sup>-1</sup>
Rate coefficient controlling the time required for trees to dominate ground vegetation	$\kappa_s$ (SS)	0.2	0.16 - 0.24	yr <sup>-1</sup>
Fraction of the change in capacity for carbon storage in trees that causes a change in capacity for storage in ground vegetation	$\epsilon$ (EPS)	0.5	0.4 - 0.6	—
<b>Chemical ocean</b>				
Total boron concentration in surface ocean	$\Sigma B$ (SIGB)	4.1E-4	3.27E-4 - 4.90E-4	mol/L
Initial temperature of surface ocean	$T_0$ (TEMPO)	292.75	290.75 - 294.75	K
Chlorinity of surface water	$Cl$ (CL)	19.24	15.0 - 23.0	mL <sup>-1</sup>
Relative humidity in atmosphere	$RH$ (RELHUM)	0.75	0.6 - 0.9	—

<sup>a</sup>letters in parentheses indicate the FORTRAN names of the input factors and they will be used in the text.

Table 4.1. *cont.*

Description	Factor	Value	Range	Unit
<b>Physical ocean</b>				
Depth of surface ocean	$HM$ (HM)	75.0	60.0 - 90.0	m
Area of surface ocean	$AREA$ (AREA)	3.61E+14	2.88E+14 - 4.33E+14	m <sup>2</sup>
Temperature change in surface ocean as a result of doubling atmospheric carbon content	$DT$ (DELTP)	3.0	1.5 - 4.5	K
<b>Terrestrial turnover times</b>				
Nonwoody parts of trees	$\tau_{21}$ (TF)	1.75	1.4 - 2.1	yr
Woody parts of trees	$\tau_{22}$ (TW)	25.00	20.0 - 30.0	yr
Ground vegetation	$\tau_{23}$ (TG)	4.00	3.2 - 4.8	yr
Detritus/decomposers	$\tau_{24}$ (TD)	2.00	1.6 - 2.4	yr
Active soil carbon	$\tau_{25}$ (TSL)	100.00	80.0 - 120.0	yr
<b>Soil-forming fractions</b>				
Woody parts of trees	$\theta_{22}$ (THW)	0.1180	0.094 - 0.14	—
Ground vegetation	$\theta_{23}$ (THG)	0.3330	0.26 - 0.40	—
Detritus/decomposers	$\theta_{24}$ (THD)	0.0625	0.05 - 0.075	—
<b>Intrinsic recovery times</b>				
Nonwoody parts of trees	$\nu_T$ (TT2)	20.0	16.0 - 24.0	yr
Ground vegetation	$\nu_V$ (TV2)	4.0	3.2 - 4.8	yr

Table 4.2. Dynamic equations of 25-compartment global carbon cycle model<sup>b</sup>

Compartment	
(1)	<p>Atmosphere</p> $\dot{c}_1 = -k_{AS}c_1 + k_{SA}\bar{c}_2(P_S(c_2)/P_S(\bar{c}_2)) - (F_{1,21}^c(c_{21}) + F_{1,22}^c(c_{21}) + F_{1,23}^c(c_{23})) + \alpha_{24,1}c_{24} + \alpha_{25,1}c_{25} + F_F^c(t) + \varphi_A F_B^c(t)$
(2)	<p>Surface ocean</p> $\dot{c}_2 = k_{AS}c_1 - k_{SA}\bar{c}_2(P_S(c_2)/P_S(\bar{c}_2)) - k_{23}c_2 + k_{32}c_3$
(3-20)	<p>Deep ocean</p> $\dot{c}_i = k_{i-1,i}c_{i-1} - (k_{i,i-1} + k_{i,i+1})c_i + k_{i+1,i}c_{i+1}, \quad i = 3, 4, \dots, 19$ $\dot{c}_{20} = k_{19,20}c_{19} - k_{20,19}c_{20}$
(21)	<p>Nonwoody parts of trees</p> $\dot{c}_{21} = F_{1,21}^c(c_{21}) - \alpha_{21,24}c_{21} - F_B^c(t)c_{21}/(c_{21} + c_{22})$ <p>where <math>F_{1,21}^c(c_{21}) = \nu_T c_{21} - \rho_T c_{21}^2</math></p> $\dot{\rho}_T = -\sigma_T \rho_T + \omega_T(\nu_T - \alpha_{21,24}) + [\rho_T^2/(\nu_T - \alpha_{21,24})] \xi_T F_B^c(t)[c_{21}/(c_{21} + c_{22})]$
(22)	<p>Woody parts of trees</p> $\dot{c}_{22} = F_{1,22}^c(c_{21}) - (\alpha_{22,24} + \alpha_{22,25})c_{22} - F_B^c(t)c_{22}/(c_{21} + c_{22})$ <p>where <math>F_{1,22}^c(c_{21}) = (\nu_T c_{21} - \rho_T c_{21}^2)(\bar{F}_{1,22}^c/\bar{F}_{1,21}^c)</math></p>
(23)	<p>Ground vegetation</p> $\dot{c}_{23} = F_{1,23}^c(c_{23}) - (\alpha_{23,24} + \alpha_{23,25})c_{23}$ <p>where <math>F_{1,23}^c(c_{23}) = \nu_V c_{23} - \rho_V c_{23}^2</math></p> $\dot{\rho}_V = \{\epsilon\sigma_T - \epsilon\omega_T(\nu_T - \alpha_{21,24})/\rho_T - \rho_T[(1 + \xi_T(\epsilon - 1))F_B^c(t)c_{21}/(c_{21} + c_{22}) + k_S\eta]/(\nu_T - \alpha_{21,24})\}\rho_V$ $\dot{\eta} = -\kappa_S\eta + (1 - \xi_T)F_B^c(t)c_{21}/(c_{21} + c_{22})$
(24)	<p>Detritus/decomposers</p> $\dot{c}_{24} = \alpha_{21,24}c_{21} + \alpha_{22,24}c_{22} + \alpha_{23,24}c_{23} - (\alpha_{24,25} + \alpha_{24,1})c_{24} + \varphi_D F_b^c(t) + \psi_S F_B^c(t)$
(25)	<p>Active soil carbon</p> $\dot{c}_{25} = \alpha_{22,25}c_{22} + \alpha_{23,25}c_{23} + \alpha_{24,25}c_{24} - \alpha_{25,1}c_{25} - \psi_S F_B^c(t)$

<sup>b</sup>symbols not appearing in Table 4.1 are described in the following page

In the equations given in Table 4.2 :

$c_i$	=	mass of carbon in the $i$ th compartment,
$k_{AS}, k_{SA}$	=	rate coefficients for surface ocean invasion and evasion,
$P_S$	=	partial pressure of dissolved $\text{CO}_2$ in the surface ocean compartment,
$F_{ij}$	=	flux from compartment $i$ to compartment $j$ ,
$\alpha_{ij}$	=	rate coefficient in the terrestrial component of the model corresponding to the flux $F_{ij}$ ,
$k_{ij}$	=	rate coefficient in the ocean component of the model corresponding to the transfer from compartment $i$ to compartment $j$ ,
$F_B(t)$	=	release at time $t$ from 'nonwoody parts of trees' and 'woody parts of trees' due to forest clearing,
$F_F(t)$	=	release at time $t$ from 'nonwoody parts of trees' and 'woody parts of trees' due to fossil fuel burning,
$\omega_T$	=	parameter that controls the ultimate level in the equilibrium value of carbon storage in 'woody parts of trees' that can be forced by re-establishment,
$\eta$	=	dynamic variable to incorporate a delay in the dominance of trees over 'ground vegetation',
—		over a variable indicates steady-state value.

For this model, the steady-state condition is satisfied within the model calibration process. Random sampling has been used, and for each model run, the

CO<sub>2</sub> content of each compartment is calculated as a function of inputs resulting from fossil-fuel combustion and forest clearing. These time-dependent releases of carbon are described as model input for the years 1750 through 2100 consisting of the historical data (1750 - 1990) and the IPCC's IS92a scenario future projections (1990 - 2100) (see Sections 3.2.1 - 3.2.3 in the previous chapter).

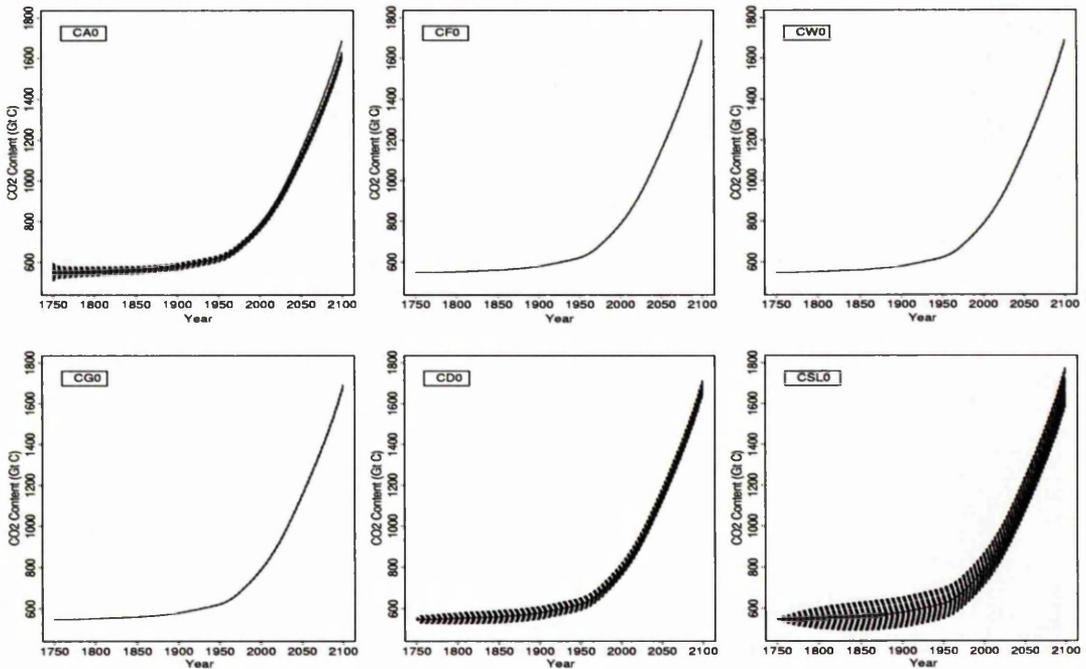
## 4.3 Results from Screening Methods

### 4.3.1 Standard OAT Design

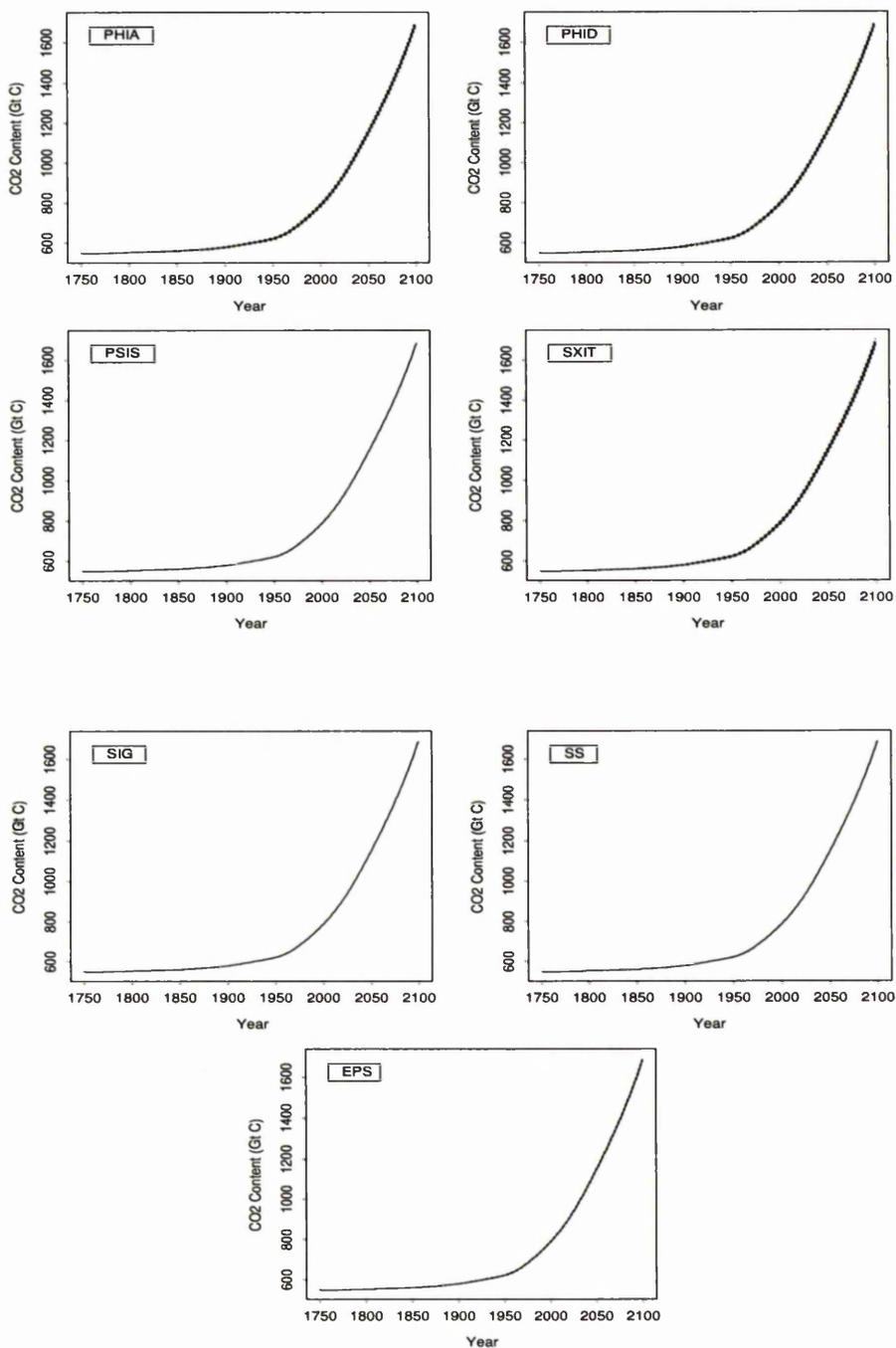
First, we have applied a standard OAT design on the 30 model input factors listed in Table 4.1 and their corresponding outputs. As described in Chapter 3, in this design we vary one input factor at a time over its entire range while keeping the others at their nominal values, and exploring how sensitive the model outputs are to these local changes.

As an illustration, the time-dependent behaviour of the atmosphere compartment resulting from varying some of the input factors OAT is shown in Figures 4.2 - 4.4. Each dotted-line curve in these figures corresponds to the prediction associated with one of the 100 input sample vectors, and solid line curves are the base-line curves.

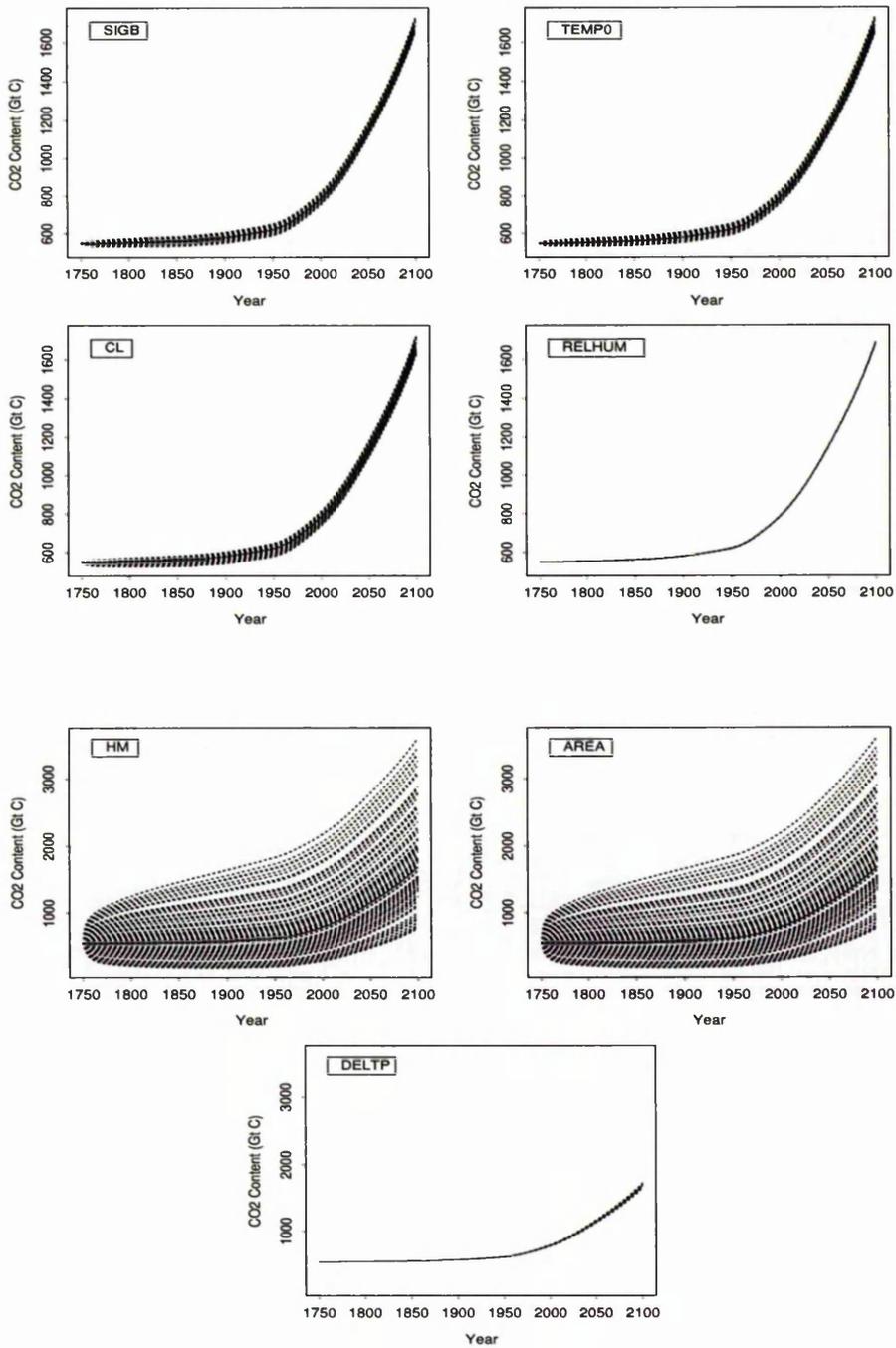
The atmosphere compartment appears to be sensitive to the initial conditions of the atmosphere, detritus/decomposers and active soil carbon compartments, i.e. CA0, CD0 and CSL0 (see Figure 4.2). The forest clearing and reforestation input factors do not seem to be having any significant effect on the atmosphere compartment (see Figure 4.3). Among the ocean related input factors, HM, AREA, SIGB, TEMP0 and CL influence the atmosphere compartment the most (see Figure 4.4). The effect of the variability in the terrestrial input factors (i.e., TF, TW, TG, TD, TSL, THW, THG, THD, TT2 and TV2) on the atmosphere compartment were found to be relatively unimportant.



**Figure 4.2.** Atmospheric CO<sub>2</sub> predictions resulting from varying the initial conditions OAT (given at top-left corner of each graph - see Table 4.1 for description of these input factors). N=100 model simulations, IS92a emission scenario is considered. In each graph the solid line represents the base-line case and dashed lines represent the predictions.



**Figure 4.3.** Atmospheric CO<sub>2</sub> predictions resulting from varying the forest clearing and the reforestation input factors OAT (given at top-left corner of each graph - see Table 4.1 for description of these input factors). N=100 model simulations, IS92a emission scenario is considered. In each graph the solid line represents the base-line case and dashed lines represent the predictions.



**Figure 4.4.** Atmospheric CO<sub>2</sub> predictions resulting from varying the chemical and the physical ocean input factors OAT (given at top-left corner of each graph - see Table 4.1 for description of these input factors). N=100 model simulations, IS92a emission scenario is considered. In each graph the solid line represents the base-line case and dashed lines represent the predictions.

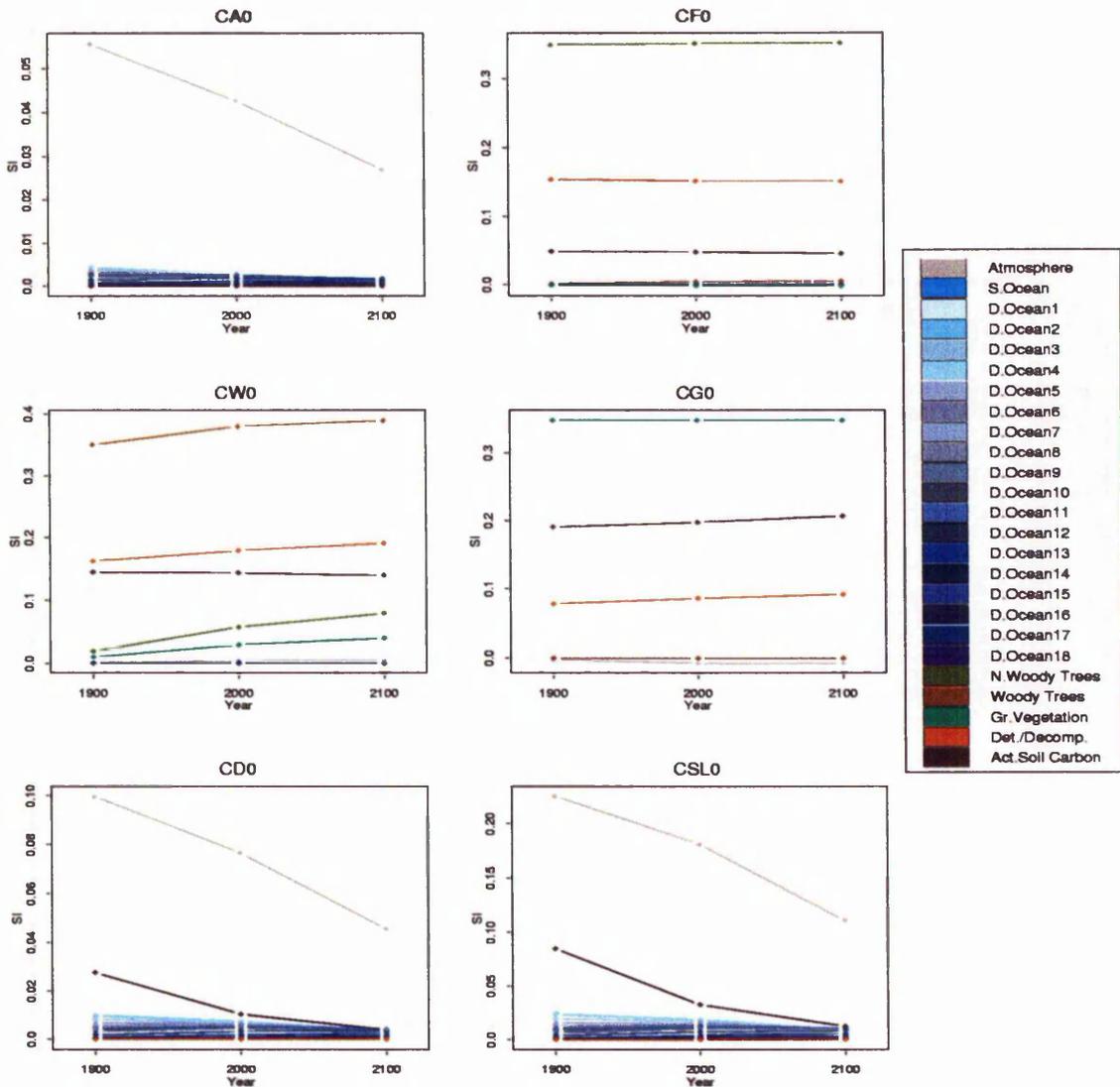
#### 4.3.1.1 Sensitivity Index

In this section, we present the results from a local SA method namely the Sensitivity Index (SI).

The SIs of all 25 compartmental outputs at three chosen years (1900, 2000 and 2100) to the ranges of the 30 input factors are calculated and results presented graphically in Figures 4.5 - 4.8. Each plot in these figures shows the sensitivities of all compartments to one of the model inputs. In Figure 4.5, it is apparent that except for the sensitivity of the atmosphere compartment to CA0, CD0 and CSL0, which decreases over time, the other SIs hardly change with time. Only the atmosphere compartment seems to be effected by the variation in CA0. None of the ocean compartments are sensitive to CF0, CW0 and CG0.

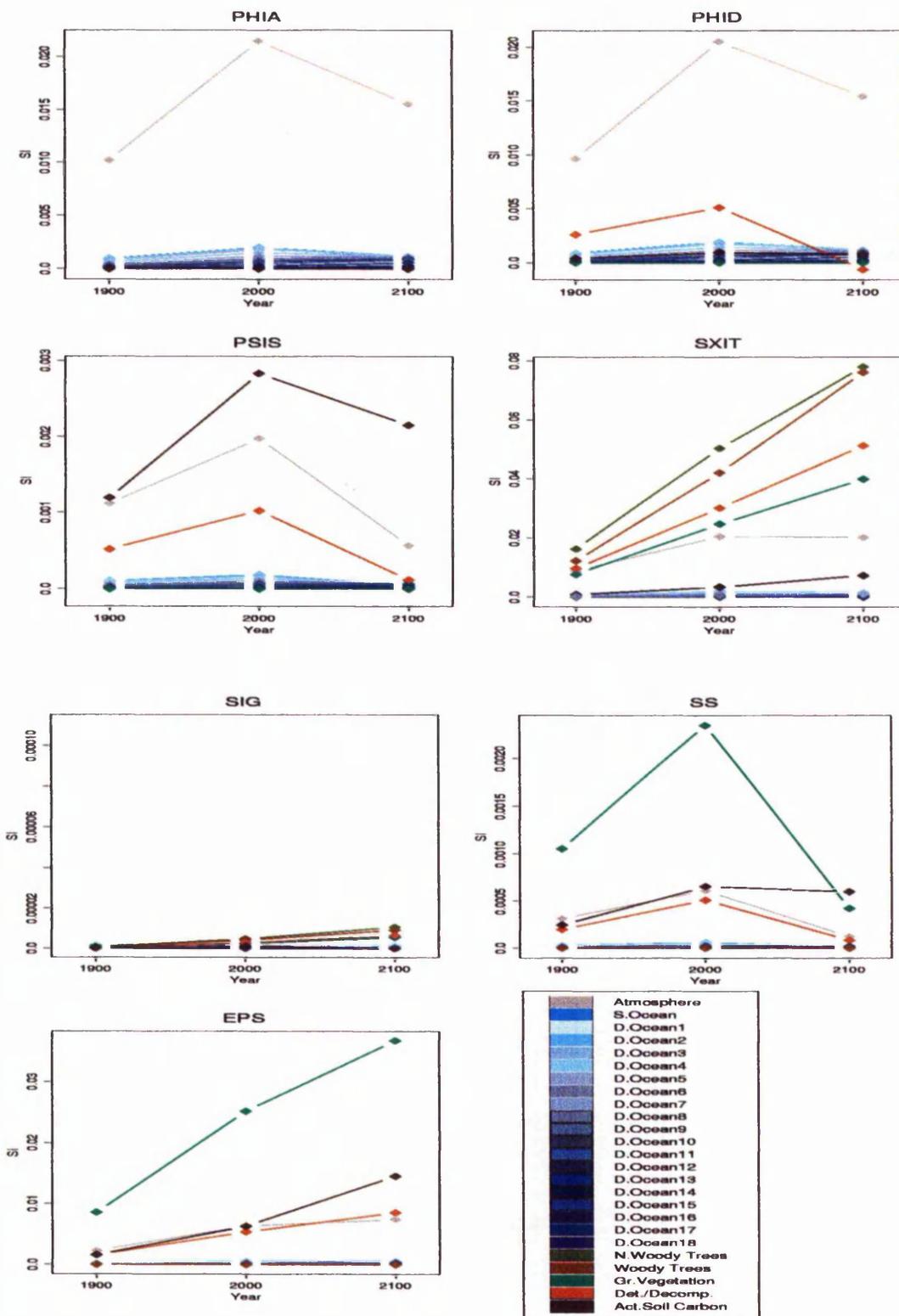
The variation in CD0 and CSL0 seem to be effecting some of the ocean compartments but not much since their corresponding SI values are all less than 0.01 which indicates that the output of these compartments is not very sensitive to the changes in CD0 and CSL0. The variation in CF0 appears to be effecting the nonwoody parts of trees compartment the most. The detritus/decomposers and active soil compartments also show sensitivity to CF0. All five terrestrial compartments show some degree of sensitivity to the input factor CW0; in the order the woody parts of trees, detritus/decomposers, active soil carbon, and finally the nonwoody parts of trees and the ground vegetation compartments. The variation in CG0 appears to be influencing the ground vegetation most, and its influence on the active soil carbon and the detritus/decomposers compartments is also important. The initial conditions CD0 and CSL0 seem to affect the atmosphere compartment only. In early years, the active soil carbon compartment also shows some sensitivity to these two input factors (SI of 0.027 with CD0 and 0.084 with CSL0 in 1900), but later (see the results from 2100) these sensitivities become insignificant with SI values below 0.01.

The SIs of all compartmental outputs due to the range of the forest clearing and the reforestation input factors are shown in Figure 4.6. Except for the atmosphere compartment, none of the other compartments show any sensitivity



**Figure 4.5.** Sensitivity Indices of compartmental outputs due to the range of the initial conditions (CA0, CF0, CW0, CG0, CD0, CSL0).

to the variation in PHIA and PHID. The sensitivity of the atmosphere to these two input factors is low in 1900, then it becomes more significant in 2000 but decreases again in 2100. The detritus/decomposers compartment also shows some sensitivity to PHID but it is very small as the SIs at all three years are very low.



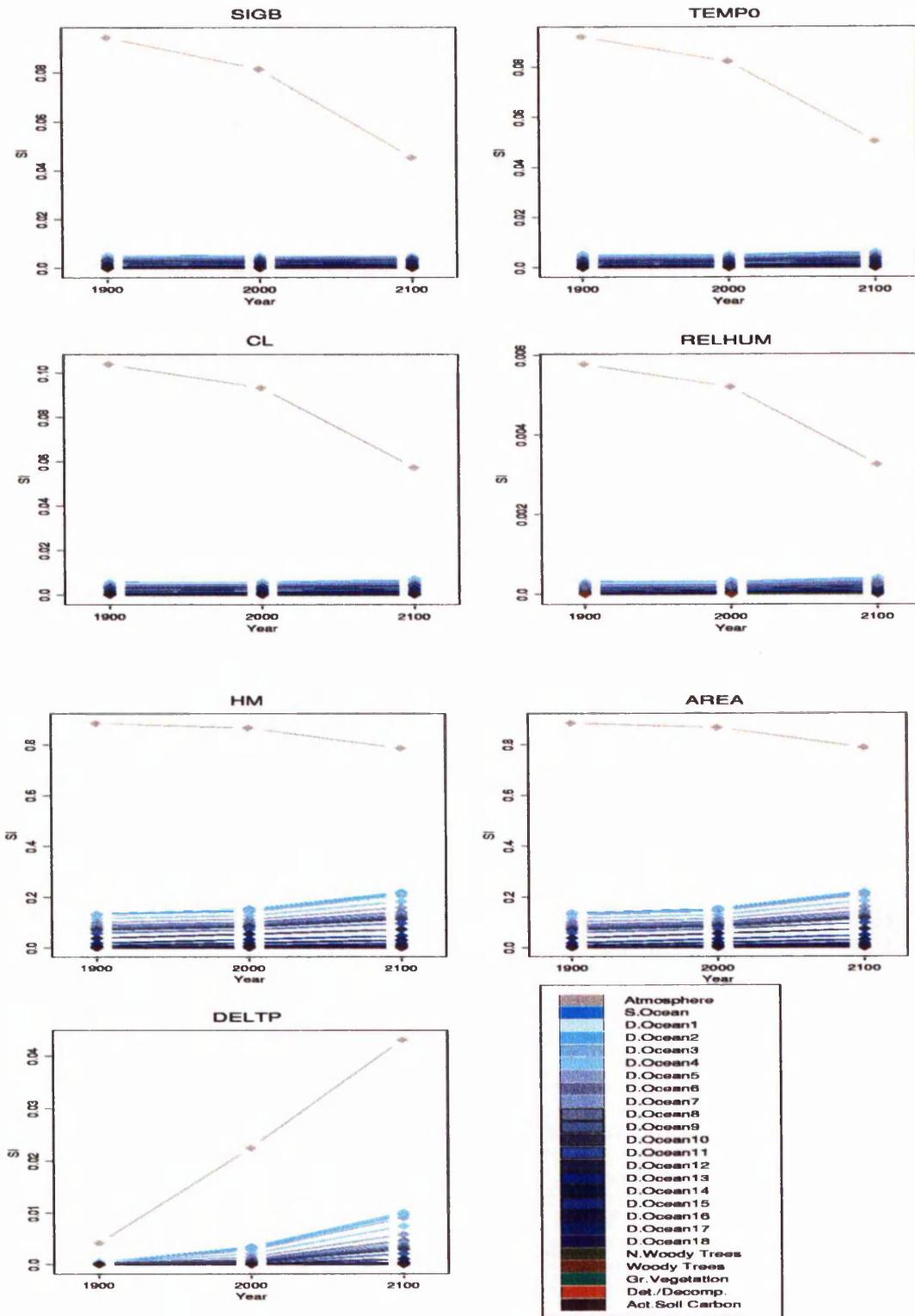
**Figure 4.6.** Sensitivity Indices of compartmental outputs due to the range of the forest clearing input factors (PHIA, PHID, PSIS, SXIT) and the reforestation input factors (SIG, SS, EPS).

Even though the graph for PSIS shows that three of the compartments are sensitive to this input factor, the range of PSIS is not influencing these compartments very much since the SI values are very low, below 0.003. The SXIT appears to be influential on the nonwoody parts of trees, woody parts of trees, detritus/decomposers, ground vegetation and the atmosphere compartments (in this order, from most to least influenced). The SIs corresponding to these compartments tend to increase with time.

None of the ocean compartments is sensitive to the reforestation input factors SIG, SS and EPS. Hardly any of the other compartments show any sensitivity to the range of SIG, i.e. all SIs are very close to zero. With time the SI values corresponding to the terrestrial compartments show some increase but not to a very high value. As for the sensitivities due to the range of SS, compared to the other compartmental outputs the SIs for the ground vegetation, active soil carbon, atmosphere and the detritus/decomposers compartments appear to be slightly higher, especially in year 2000. However, considering the SI values, it is apparent that none of the compartments is very sensitive to SS. The ground vegetation, active soil carbon and detritus/decomposers terrestrial compartments and atmosphere compartment seem to be influenced by the range of EPS, and the sensitivity of these compartments to EPS increases with time.

As shown in Figure 4.7, the ranges of the chemical and physical ocean input factors do not appear to be having any influence on the terrestrial compartments. The atmosphere compartment seems to be the only compartment that is effected by the chemical ocean inputs SIGB, TEMP0, CL and RELHUM. The sensitivity of the atmosphere compartment to these inputs decreases with time. The input factors SIGB, TEMP0, CL and RELHUM also influence the ocean compartments but not very much. The SIs for the atmosphere compartment decreases as the SIs for the ocean compartments increase with time. But the decrease in the SI values of the atmosphere is more rapid.

The sensitivity of the atmosphere compartment to the range of physical ocean input factors HM and AREA is very high (SI values of around 0.89 in 1900, 0.87



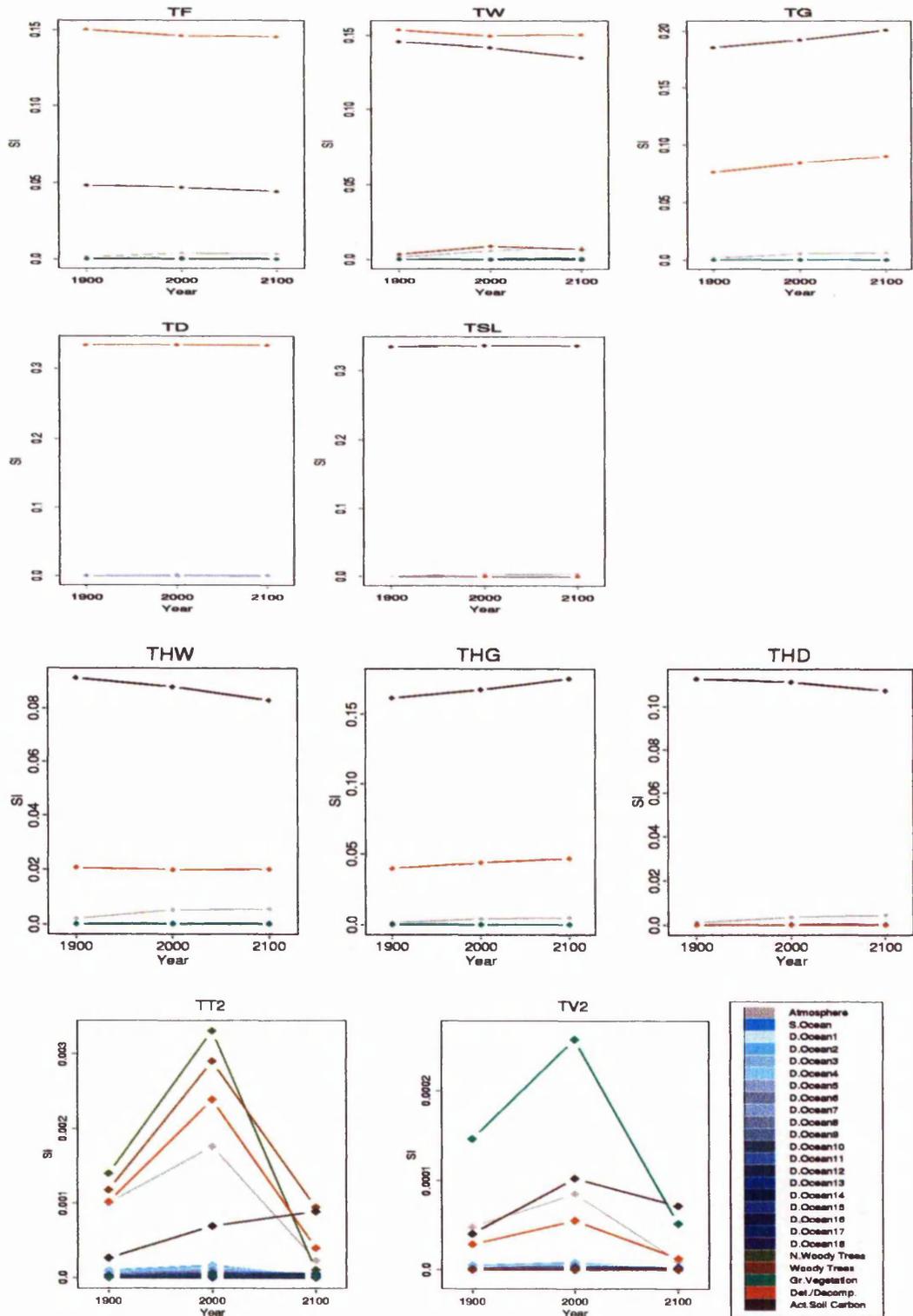
**Figure 4.7.** Sensitivity Indices of compartmental outputs due to the range of the chemical ocean input factors (SIGB, TEMPO, CL, RELHUM) and the physical ocean input factors (HM, AREA, DELTP).

in 2000 and 0.79 in 2100 for both factors). As the plots of HM and AREA show, the SI related to the atmosphere compartment decreases with time, whereas the SIs of the ocean compartments increase with time. After the atmosphere compartment, the most sensitive compartments to the range of HM and AREA are the surface ocean, deep ocean-layer 1, deep ocean-layer 2, and so on, in the given order. The compartmental SIs due to DELTP show that this input factor influences the atmosphere compartment the most. The ocean compartments also show some sensitivity to this input factor but it is not very high. Especially for the atmosphere, the SI values increase quite rapidly with time.

In Figure 4.8, we show the SIs related to the terrestrial input factors TF, TW, TG, TD, TSL (the terrestrial turnover times in, respectively, nonwoody parts of trees, woody parts of trees, ground vegetation, detritus/decomposers and active soil carbon compartments); THW, THG, THD (the soil-forming fractions in woody parts of trees, ground vegetation and detritus/decomposers compartments); and TT2, TV2 (the intrinsic recovery times in nonwoody parts of trees and ground vegetation compartments). The ocean compartments do not seem to be sensitive to any of these input factors. First, considering the terrestrial turnover times input factors, the TF, TW and TG appear to have significant influence only on the detritus/decomposers and the active soil carbon; TD only on the detritus/decomposers; and TSL only on the active soil carbon compartment.

The SIs of the active soil carbon and detritus/decomposers compartments due to the ranges of THW and THG are high. THD, on the other hand, shows an effect only on the active soil carbon compartment. The atmosphere compartment also seems to be effected by these three inputs but not very much.

Except for the ground vegetation compartment, all terrestrial compartments and the atmosphere compartment appear to be sensitive to TT2, but SI values are quite small (below 0.01). The plot for TV2 shows that the ground vegetation compartment is the most sensitive compartment to the range of TV2 in years 1900 and 2000 but becomes the second most sensitive in year 2100 after active soil carbon compartment. Again, very small SI values indicate that the compartments



**Figure 4.8.** Sensitivity Indices of compartmental outputs due to the range of the terrestrial turnover times input factors (TF, TW, TG, TD, TSL), the soil-forming fractions input factors (THW, THG, THD) and the intrinsic recovery times input factors (TT2, TV2).

is in-sensitive to this input factor. Next, by ranking the calculated SI values we obtain the relative order of importance of the input factors for each compartmental output evaluated in years 1900, 2000 and 2100. See Table 4.3 for the rankings of the 10 most important input factors (rank 1 indicating the most influential factor, 2 the second most influential factor, and so on). Since the rankings do not change significantly for the 18 deep ocean compartments, we include only the rankings of the deep ocean layer-5 and layer-13 (i.e. compartments 7 and 15 in Figure 4.1) in the table. Except for the two or three most influential factors, the rankings of the other factors change with time. For example, for the surface ocean compartment, CD0 is the fourth most important input factor in 1900, but in 2100 it becomes the eight most important input. The input factor rankings for the atmosphere, surface ocean, deep ocean-5 and deep ocean-13 compartments, show a quite good agreement. For these four compartments the area of the surface ocean (AREA) appears to be the most, and the depth of the surface ocean (HM) the second most influential input factors.

The nonwoody parts of trees, woody parts of trees and ground vegetation compartments are mostly effected by their initial conditions CF0, CW0 and CG0, respectively. For the nonwoody parts of trees compartment the second most important input seems to be CW0 and it is followed by SXIT. SXIT is the second most influential factor for the woody parts of trees compartment but it is less influential on the ground vegetation compartment (ranked as the 4th). TD, CW0 and CF0 are the three most important input factors (in descending order of importance) for detritus/decomposers compartment. The top three inputs on the active soil carbon compartment are identified as TSL then CG0 and TG.



#### 4.3.1.2 Standardised Range

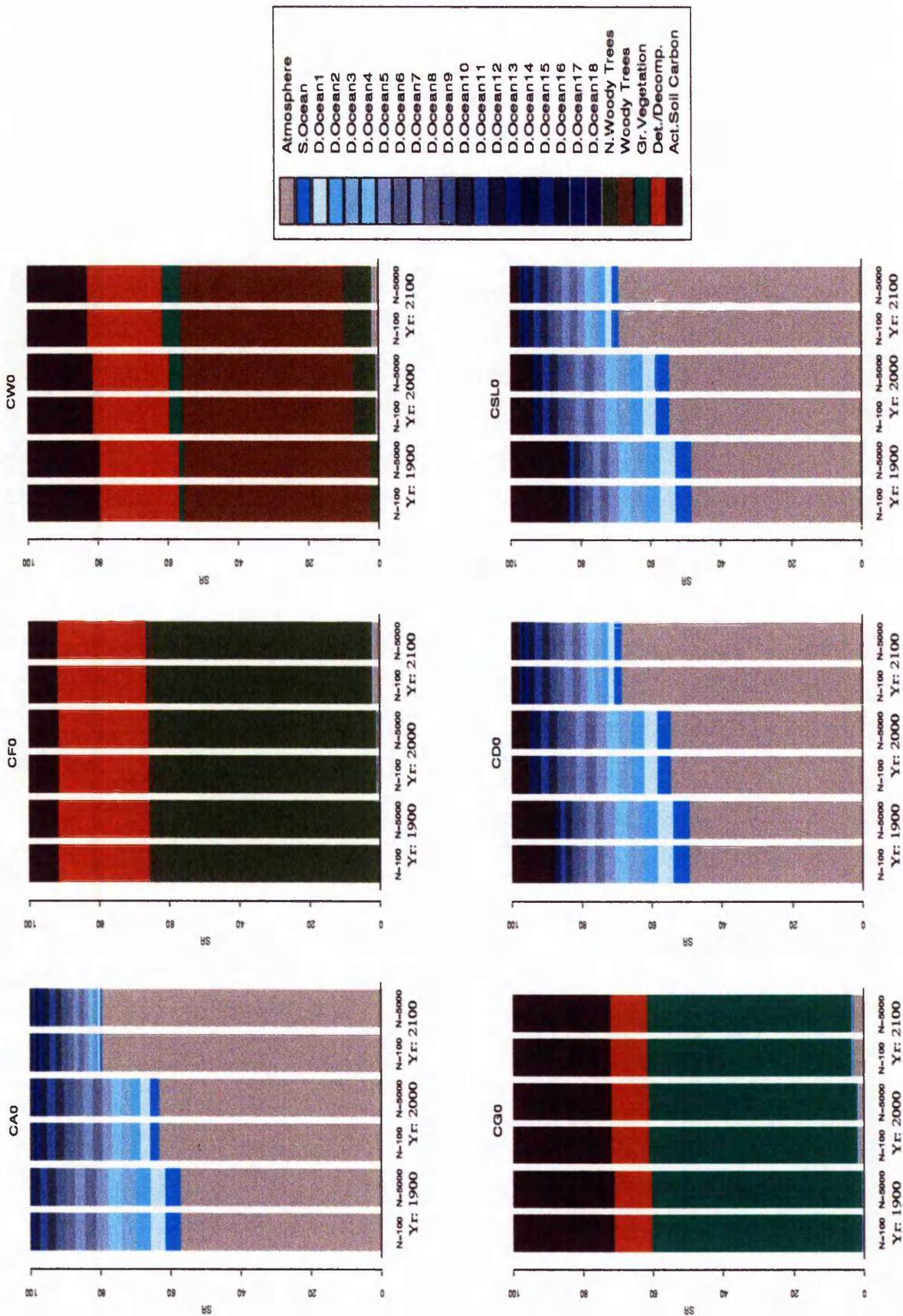
Another OAT sensitivity measure, the standardised ranges (SRs) of each compartmental output resulted from varying each input factor over its range OAT are calculated. Here we consider two different sample sizes,  $N=100$  and  $N=5000$ , to find out if the number of model runs has any effect on the results.

The SRs of all compartments due to varying each input factor are demonstrated in Figures 4.9-4.12 as bar charts. Each graph in these figures shows the influence of one input factor on all compartmental outputs at three chosen years and with two different sample sizes. Examination of these bar plots shows that except for the SRs resulting from varying SIG (see Figure 4.10), there does not seem to be any apparent difference between the SRs based on the two different number of model runs. However, the SR values do change with time.

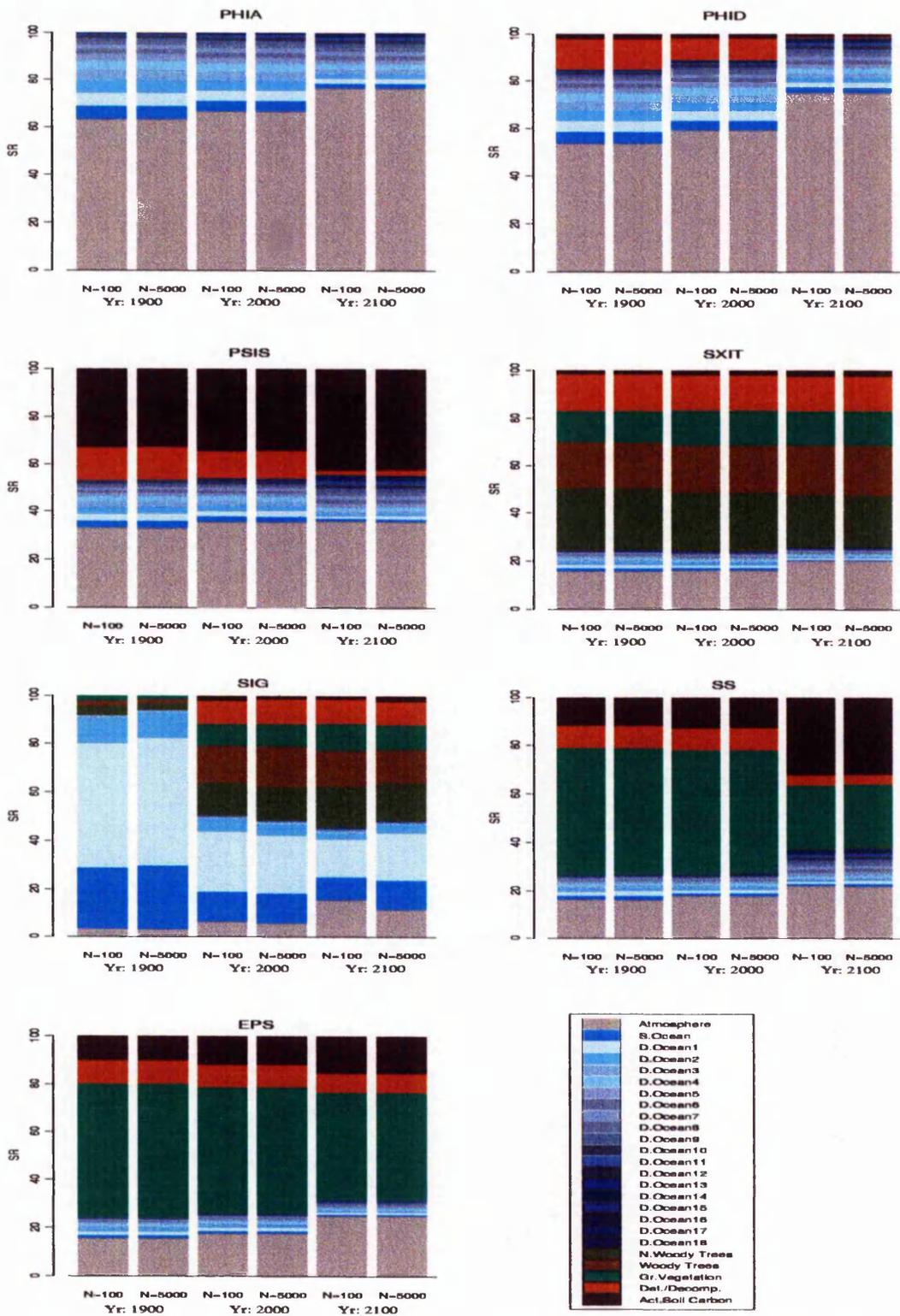
The same compartments, which were identified as the most influenced compartments using the sensitivity indices, are also identified as the most effected outputs by the SRs. The only difference we see between the graphical presentations of the SIs and the SRs is that the sensitivities of the compartments show either a continuing increase or decrease over time when the SRs are considered, but the results of the SIs show that the sensitivity of some compartments can increase from a small SI in 1900 to a larger value in 2000 and then decrease again in 2100. Because our aim in using these two local screening methods is mainly to identify the most important input factors for each compartmental output at a given time, as with the SIs we use the calculated SRs to rank the inputs in order of their importance.

The ranking of the top 10 most important inputs for each model output considered at three different years are given in Table 4.4. The order of importance between the most important 10 factors for the atmosphere, surface ocean, deep ocean layers 5 and 13, and active soil carbon compartments are in complete agreement with the rankings obtained with the SIs. Apart from the detritus/decomposers compartment the top five inputs for the other compartments are also the same as the inputs identified by the SIs (see Table 4.3). Using the

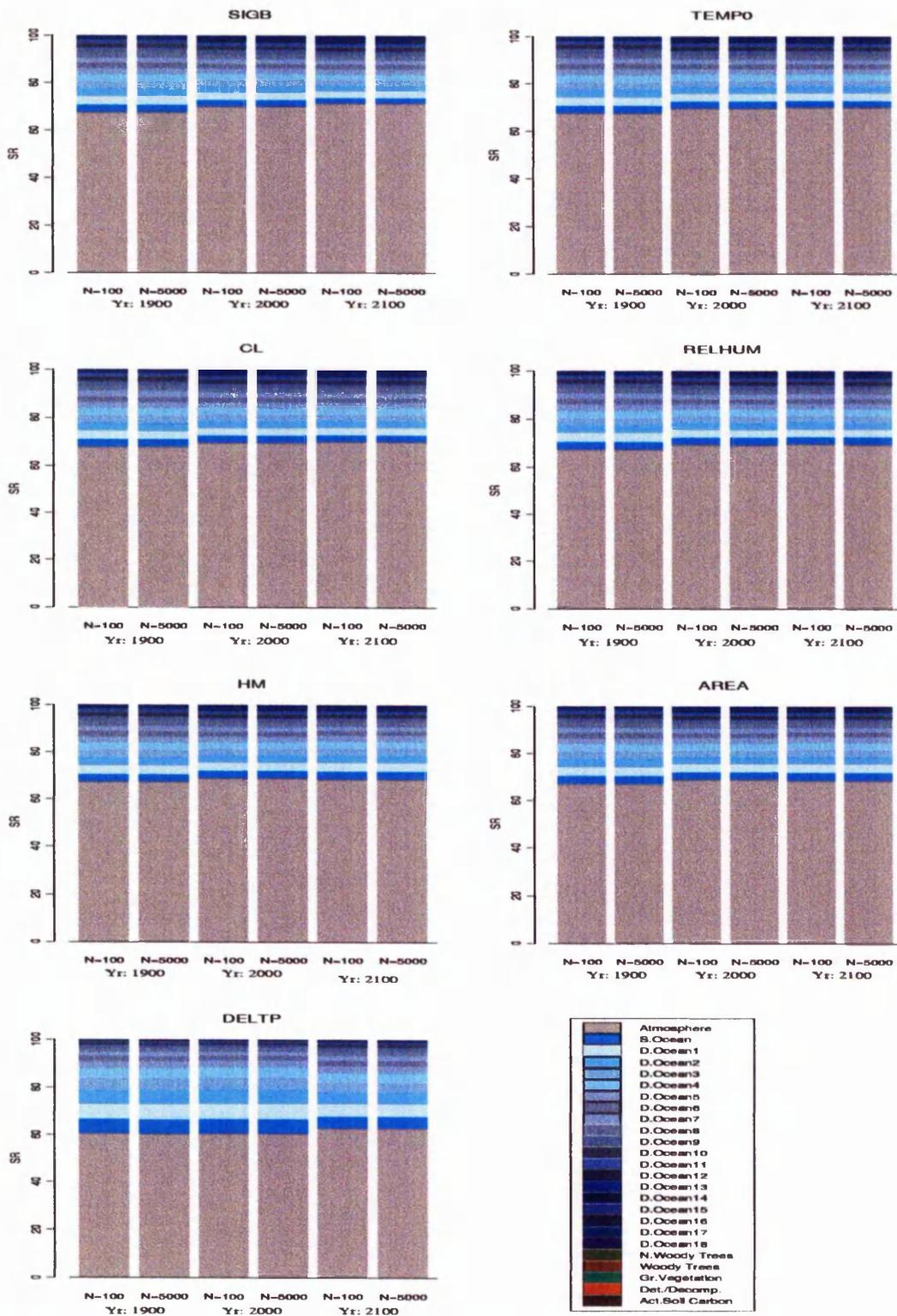
SRs, TW is ranked as the third and CF0 as the fourth influential factors on the detritus/decomposers compartment, while the other local screening method SI identified these factors in reverse order, i.e. CF0 as the third and TW as the fourth important factor. The ranks of the other input factors on this compartment are the same as the ranks obtained from the SIs.



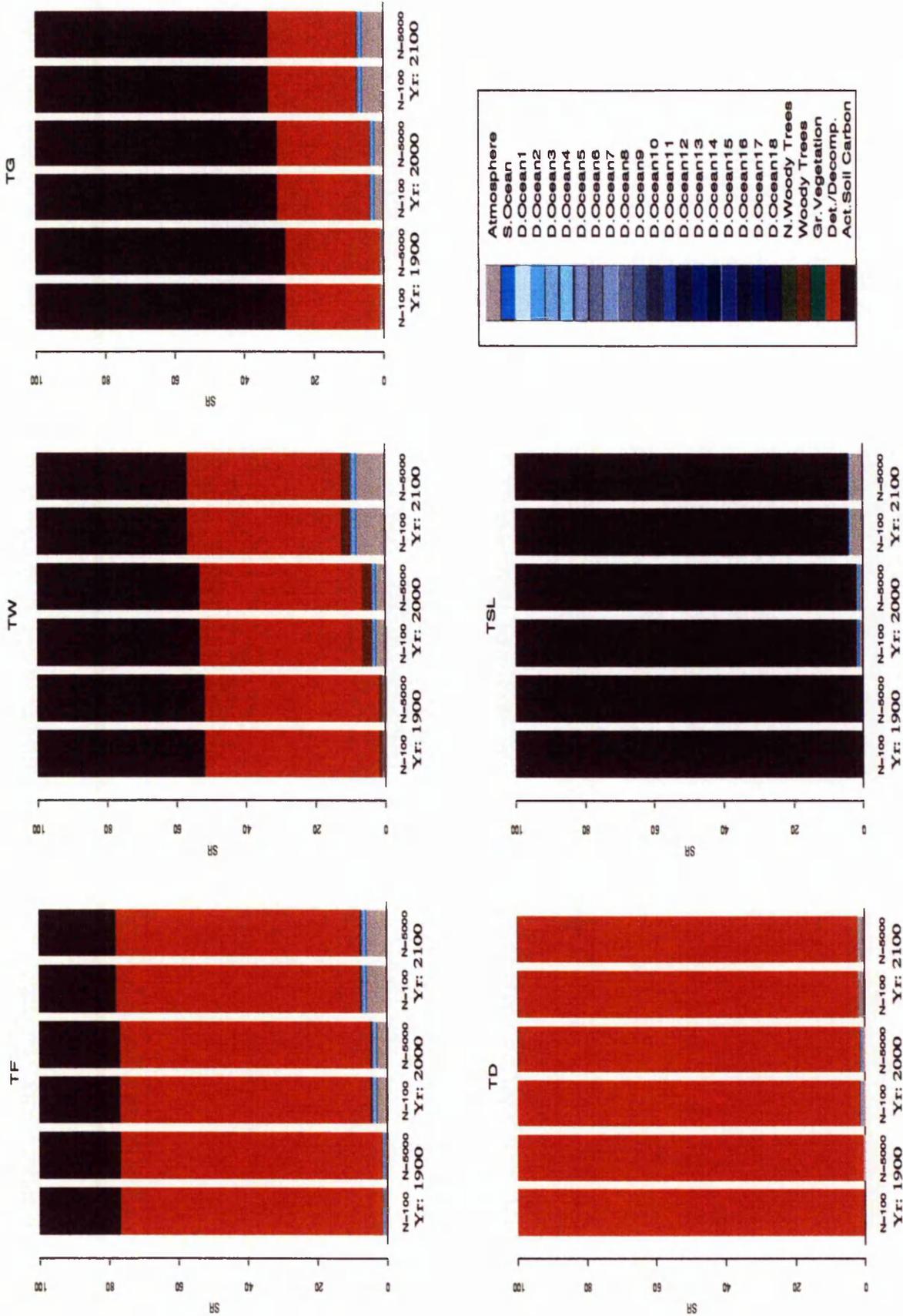
**Figure 4.9.** Bar charts showing how each compartmental output of the model in years 1900, 2000 and 2100 is effected by the variation in the initial conditions (CA0, CF0, CW0, CG0, CD0, CSL0) in terms of standardized ranges. The results from both N=100 and N=5,000 are also compared.



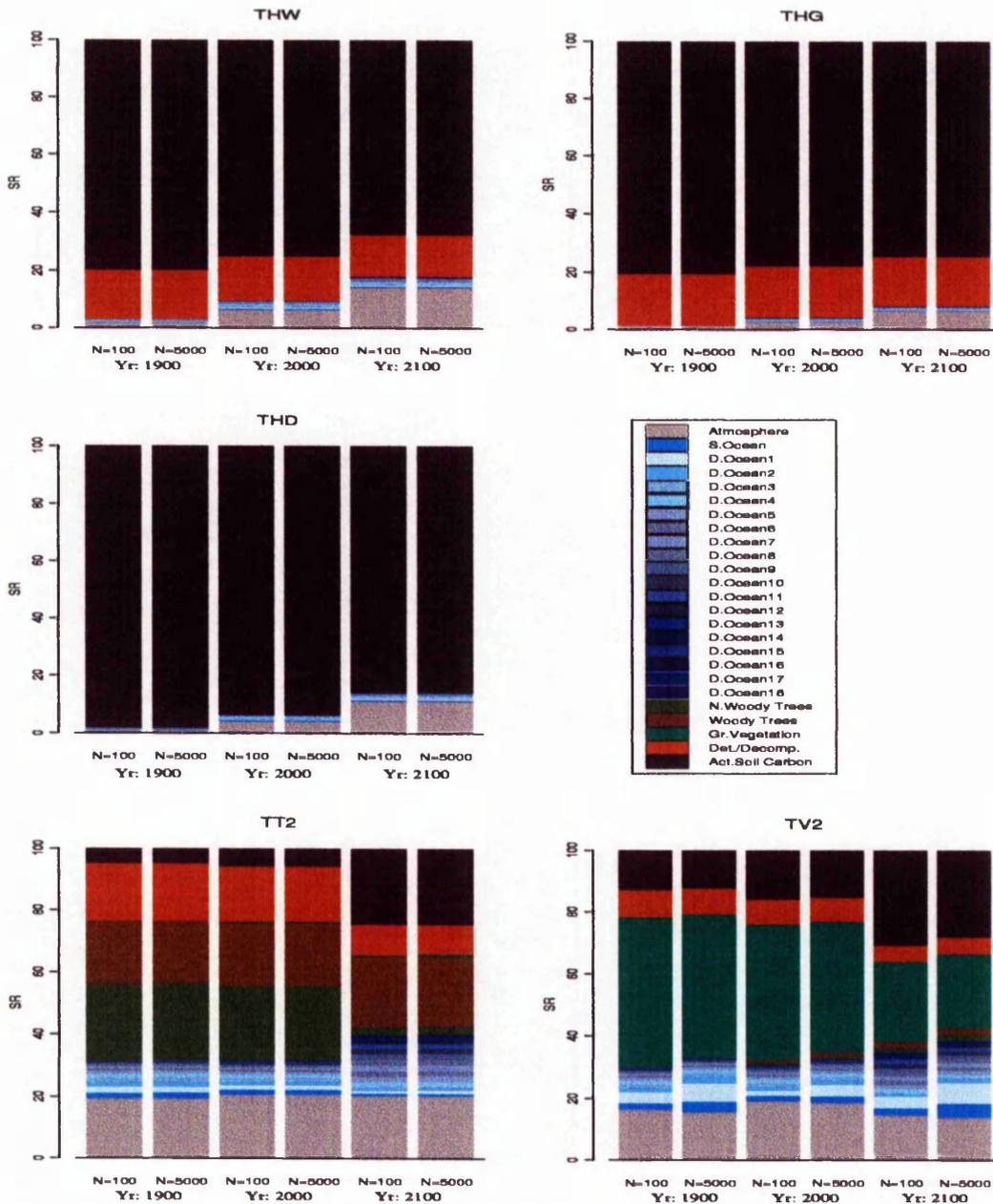
**Figure 4.10.** Bar charts showing how each compartmental output in years 1900, 2000 and 2100 is effected by the variation in the forest clearing input factors (PHIA, PHID, PSIS, SXIT) and the reforestation input factors (SIG, SS, EPS) in terms of standardized ranges. The results from both N=100 and N=5,000 are also compared.



**Figure 4.11.** Bar charts showing how each compartmental output in years 1900, 2000 and 2100 is effected by the variation in the chemical ocean input factors (SIGB, TEMPO, CL, RELHUM) and the physical ocean input factors (HM, AREA, DELTP) in terms of standardised ranges. The results from both N=100 and N=5,000 are also compared.



**Figure 4.12.** Bar charts showing how each compartmental output of the model in years 1900, 2000 and 2100 is effected by the variation in the terrestrial turnover times input factors (TF, TW, TG, TD, TSL) in terms of standardized ranges. The results from both N=100 and N=5,000 are also compared.



**Figure 4.13.** Bar charts showing how each compartmental output in years 1900, 2000 and 2100 is effected by the variation in the soil-forming fractions input factors (THW, THG, THD) and the intrinsic recovery times input factors (TT2, TV2) in terms of standardized ranges. The results from both N=100 and N=5,000 are also compared.



### 4.3.2 Morris Design

We next present the results of SA performed using the Morris method described in Chapter 2 Section 2.5.5.2. For each input factor, the absolute value of the estimated Morris mean and standard deviation, which are considered as two sensitivity measures that respectively indicate the factor's influence on the model output and the influence due to interactions with other input factors and/or nonlinear effects, are calculated. Because of the vast amount of results that our model produces, we decided not to report the estimated means and standard deviations numerically here, instead we display the results graphically in Figures 4.14 - 4.16. In these figures, the input factors are labelled using their FORTRAN names. A general order of importance between the input factors can be obtained considering the Euclidean distance from the origin, the larger the distance the more influential the factor is.

The screening results on each output from the three years are overlaid on the same figure to show how the influence of the inputs change between these three years. For instance, on the atmosphere, surface ocean and the two deep ocean compartments (see Figure 4.14 and the top frame of Figure 4.15), AREA appears to be the most and HM the second most important factor, and this does not change from year to year. For these compartments, the influence of the inputs AREA and HM increases over time.

By ranking the estimated Morris means we establish a relative importance of the inputs in terms of their overall influence on the outputs. The rankings of the most important ten factors for each compartmental output considered at three different years are given in Table 4.5. These rankings confirm what the morris plots show. As Table 4.5 reveals, for each compartment the ranking of the most important 10 input factors do not change considerably in the three years. We discuss the results here by comparing them with the results we have obtained from the two standard OAT methods - the SIs and SRs - since the results from these three screening methods present strong similarities and dissimilarities in terms of the importance rankings they provide.

Unlike the SI and SR methods, the Morris method identifies some of the terrestrial component related inputs (namely TW, TG, TSL, THG and THD) among the most important ten input factors for the atmosphere and the three ocean compartments. However, for the same compartments, the input factors CA0, CD0, PHIA, SIGB and DELTP which are identified among the top 10 most important inputs by the SI and SR methods are not so identified by the Morris method.

For the atmosphere, surface ocean and the two selected deep ocean (layers 5 and 13) compartments the two most influential inputs (first AREA then HM) are the same with all three screening methods. As the Morris plots of the atmosphere and the three ocean compartments (see Figure 4.14 and the top frame of Figure 4.15) show except for the inputs AREA and HM, the other input factors are not as important for these compartments.

CF0, CW0, SXIT, TT2, TW and TF (in this order from most to less important) are the six most influential input factors for the nonwoody parts of trees compartment. This order of importance is the same with the SI, SR and Morris methods. As the plot showing the results from the Morris method (given in the middle frame of Figure 4.15) reveal, except for the most influential six inputs, the other input factors have almost no influence on this compartment.

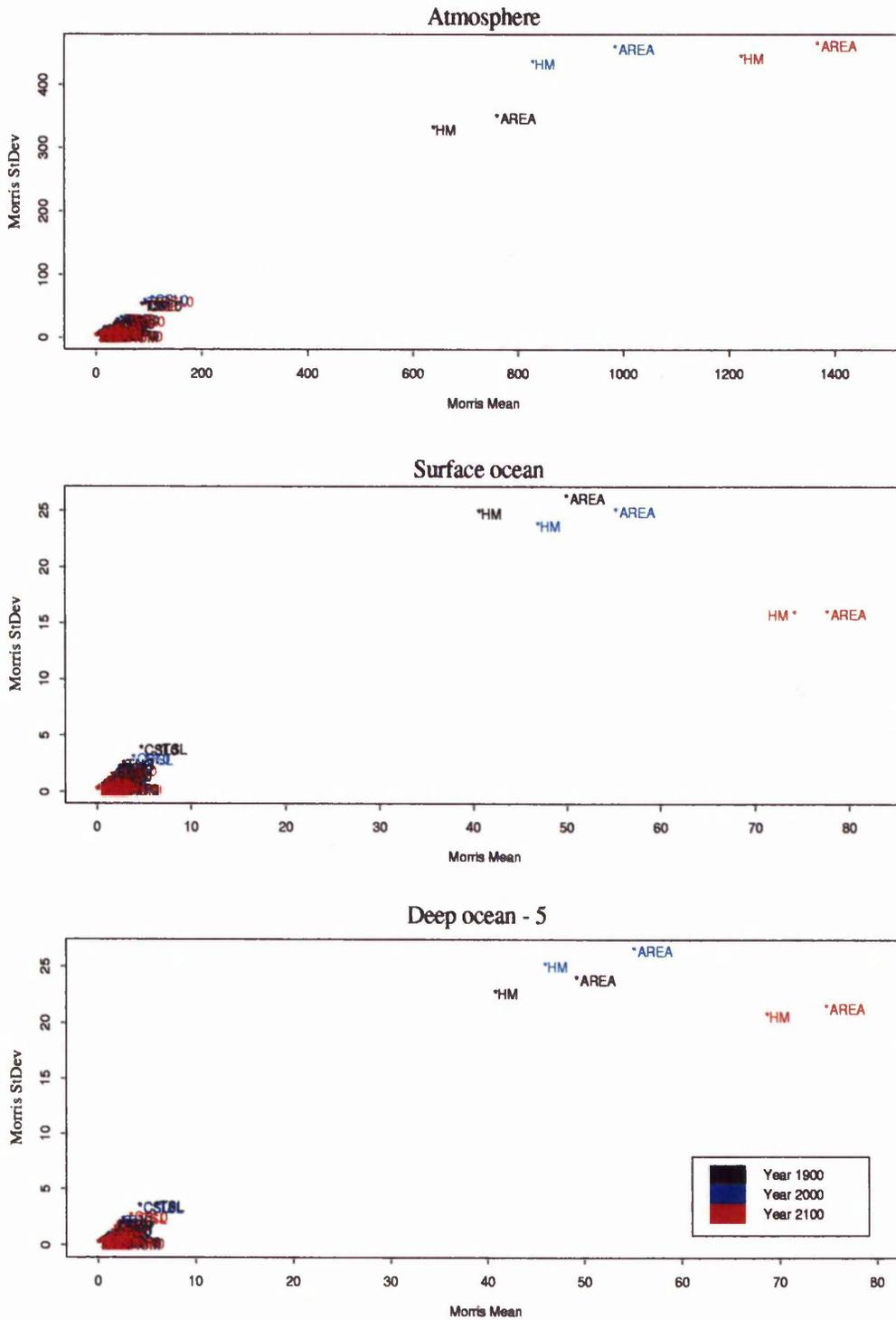
For the woody parts of trees compartment, comparison of the input factors' rankings obtained from the three screening results show that the relative importance between the most important six input factors are in complete agreement: CW0 being the most influential factor followed by SXIT, TW, CF0, TT2 and TF. The rest of the inputs have estimated Morris means and standard deviations around zero which indicates that these inputs do not have a large influence on this compartment (see the bottom frame of Figure 4.15).

Even though the rankings change slightly between the three years, the most important five input factors, namely CG0, CW0, EPS, SXIT and SS, identified by the Morris method are the same inputs identified by the SI and SR measurements, for the ground vegetation compartment. For this compartment, the rankings of

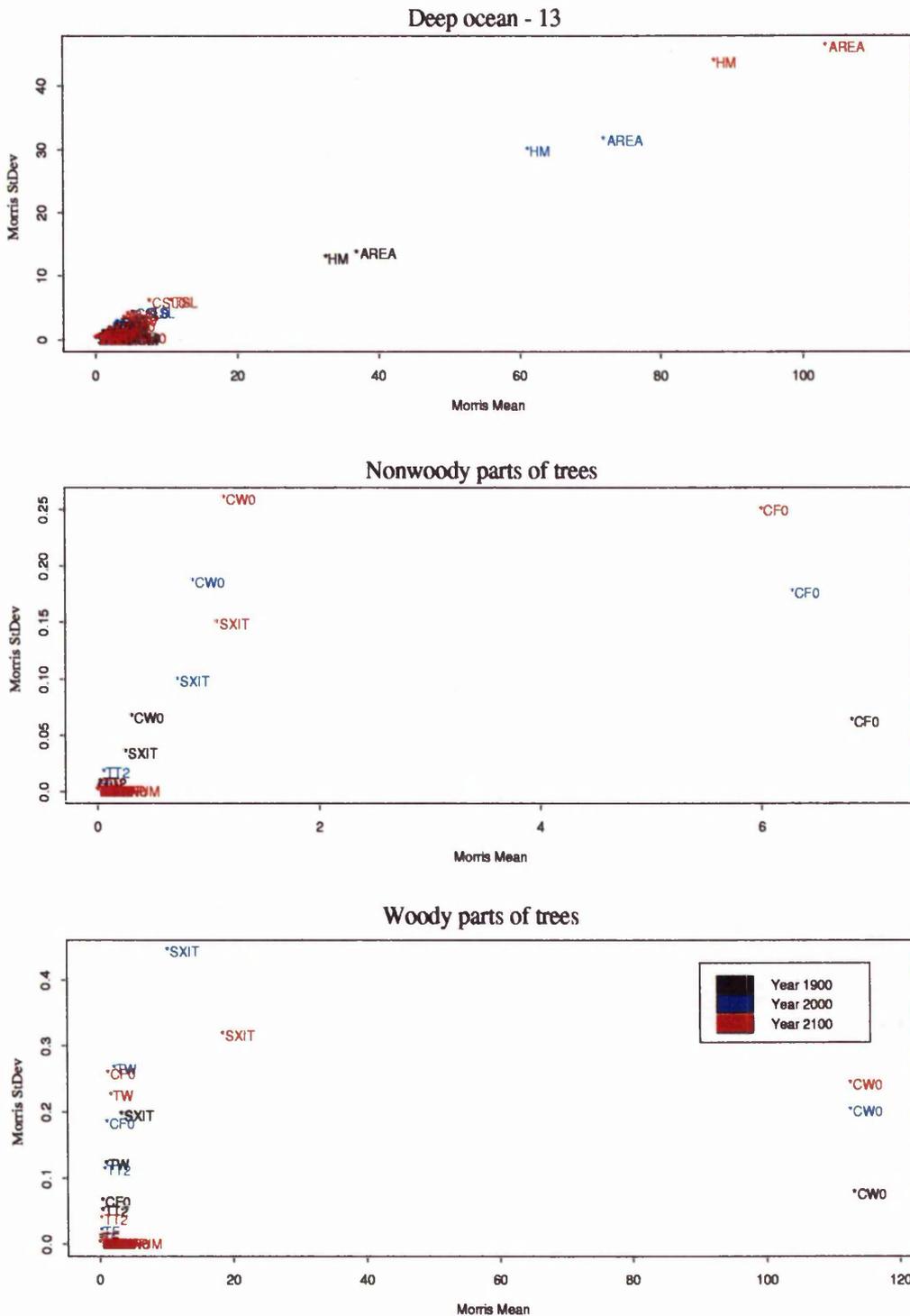
the top ten inputs from the Morris and SR methods appear to be in agreement. According to the Morris method results (presented in the top frame of Figure 4.16) except for the five inputs listed above, the other input factors are not influential on the ground vegetation compartment.

For the detritus/decomposers compartment, the three screening methods identified the same input factors as the most influential ten input factors but the order of importance between these inputs change slightly, for instance, TG is identified as the sixth most important input by the Morris method, but by SI and SR it is ranked as the seventh most important input. It appears that the inputs TD, CF0, CW0, TW, TF, TG, CG0, THG, SXIT and THW have influence on this compartment. See the middle frame of Figure 4.16 for the graphical presentation of the Morris results on this compartment.

The most important ten input factors identified by the Morris method are the same inputs that SI and SR methods picked out as most important ten inputs for the active soil carbon compartment. The order of importance between these inputs is almost identical except with the Morris method THG is ranked 3 and TG ranked 4 where as with the SI and SR measurements this ranking is reversed. Based on the Morris results, the ranking of the inputs, in decreasing importance, is TSL, CG0, THG, TG, TW, CW0, THD, THW, CF0 and TF, and as the Morris plot of this compartment show (see the bottom panel of Figure 4.16) all these inputs have some degree of influence on this compartmental output.



**Figure 4.14.** Morris screening results on Atmosphere, Surface ocean and Deep ocean (layer 5) compartments in years 1900, 2000 and 2100. Mean and standard deviations are associated with the 30 input factors considered in the analysis.



**Figure 4.15.** Morris screening results on Deep ocean (layer 13), Nonwoody parts of trees and Woody parts of trees compartments in years 1900, 2000 and 2100. Mean and standard deviations are associated with the 30 input factors considered in the analysis.





## 4.4 Global SA Methods

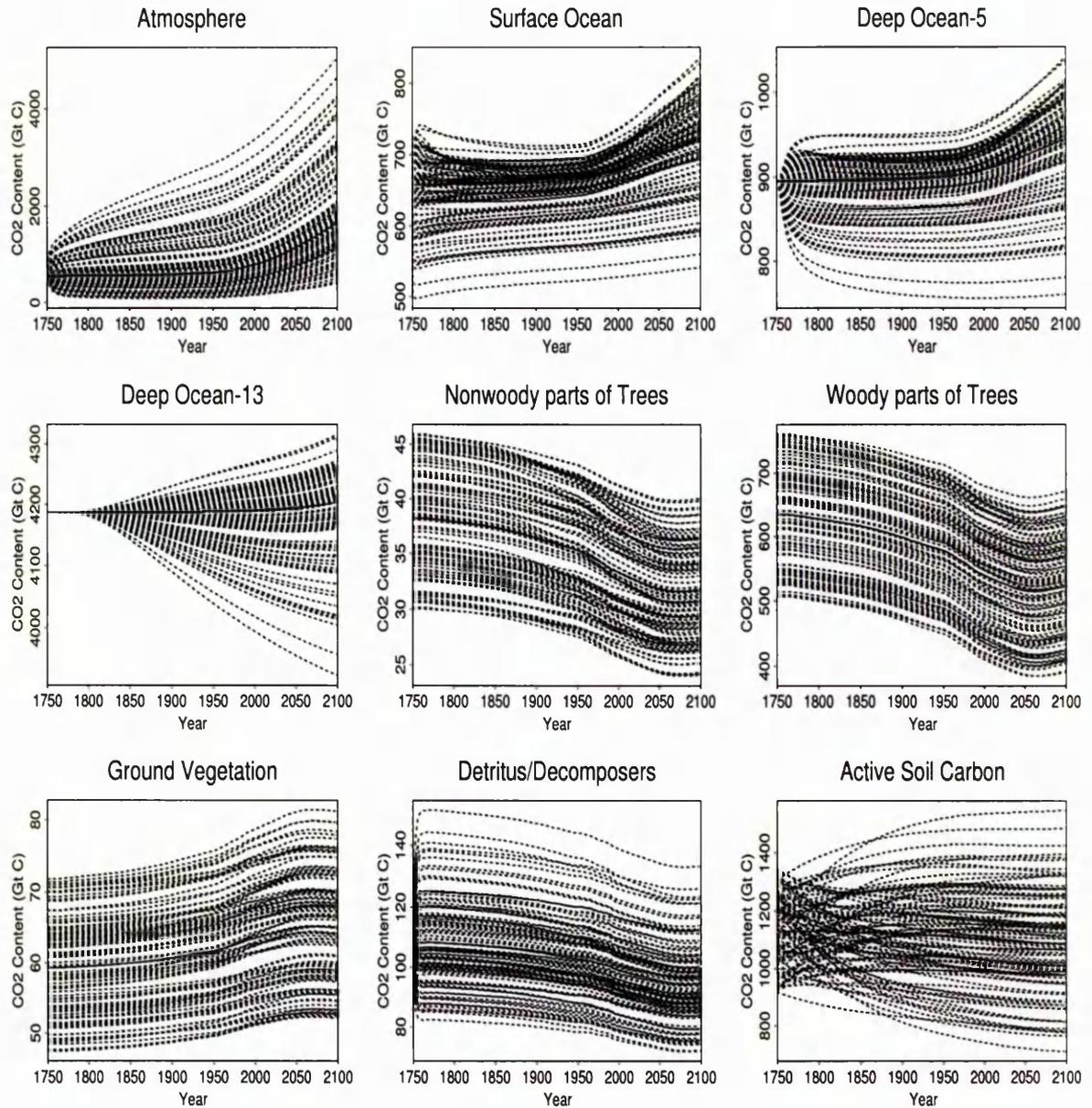
In the previous section (Section 4.3), three different screening exercises have been considered. These methods only provide sensitivity measures that are qualitative and can be used to rank the model inputs in order of their importance. In order to obtain quantitative sensitivity measures, however, global SA methods are required, and in this section we investigate the sensitivity of the model outputs to the model input factors using various global techniques.

All 30 input factors that are subject to uncertainties and are varied simultaneously in the analyses are assumed to follow uniform distributions with upper and lower bounds based on  $\pm 20\%$  of their nominal values (see Table 4.1).

As the emission scenario, the IPCC's IS92a scenario (also known as the Business-as-Usual scenario) is adopted. As sampling technique simple random sampling is considered to obtain a multivariate input sample. The  $N$  simulations of the model were performed with the randomly generated input factors. For each random set of inputs, the steady-state conditions in year 1750 are calculated in the model initialization process. Then the system is run to year 2100 to obtain the time-dependent behaviour of  $\text{CO}_2$  in each model compartment.

First of all, we show in Figure 4.17 the time-dependent behaviour of the atmosphere (model compartment 1), surface ocean (2), layers 5 and 13 of deep ocean (7 and 15), and the terrestrial biota (21-25) compartments as a result of input uncertainties and the anthropogenic inputs. Each plot in this figure corresponds to one of the compartments under consideration and each line curve in each of these plots corresponds to the prediction associated with one of the  $N=100$  sample vectors.

As the plot in Figure 4.17 shows, the  $\text{CO}_2$  content of the atmosphere increases over time. To achieve a steady-state condition for year 1750, once the initial value for the atmospheric  $\text{CO}_2$  content is chosen, the initial condition of the surface ocean compartment has to be adjusted. Thus, this process results in a large initial variability for the surface ocean and the upper layers of the deep ocean. As a result of forest clearing, reforestation and vegetation of cleared land, the



**Figure 4.17.** Time dependent behaviour of the nine dependent variables - CO<sub>2</sub> content of Atmosphere, Surface Ocean, Deep Ocean-5, Deep Ocean-13, Nonwoody Parts of Trees, Woody Parts of Trees, Ground Vegetation, Detritus/Decomposers and Active Soil Carbon compartments - predicted using 25-compartment model as a result of varying all 30 input factors simultaneously.

behaviour of the terrestrial compartments are shown in the same figure. For instance, due to forest clearing the carbon present in the nonwoody and woody parts of trees compartments is released into the atmosphere. On the other hand, the time-dependent pattern of reforestation and vegetation results in an eventual increase in the flux of carbon to the ground vegetation compartment. These land-use changes also effect the detritus/decomposers and the active soil carbon compartments.

In the analyses, the amount of CO<sub>2</sub> stored in the nine chosen compartments in years 1900, 2000 and 2100 are considered as model outputs. To gain as much information as possible from the model runs, input values that completely cover the input space must be selected. Here considering the high dimension of the input space to ensure that the input space is scanned as much as possible we use a sample size of N=5,000.

#### 4.4.1 Examination of Scatterplots

We start the analyses by first producing scatterplots of model inputs versus outputs and also calculating correlation coefficients to examine the relationships between each input factor and each output variable. To conserve space we did not include the scatterplots for all model outputs that were calculated at three chosen years, instead we only present the scatterplots of some model outputs (estimated in 2100) versus each of the 30 input factors (see Figures 4.18 - 4.24). The scatterplots of the model inputs versus model outputs evaluated in the other two years (1900 and 2000) display very similar patterns that we see in the scatterplots present here. It should be noted that the scatterplots given in these figures are produced from the model simulations based on N=1,000 model runs to show a clearer picture of the patterns in the plots, but all the quantitative sensitivity measures (including the correlation coefficients) reported in the remainder of this chapter are based on N=5,000 model evaluations. For convenience, we use notations like  $y_{Atm}(t)$  (for Atmosphere),  $y_{SO}(t)$  (for Surface Ocean) and so on to

denote a model output evaluated at time  $t$ .

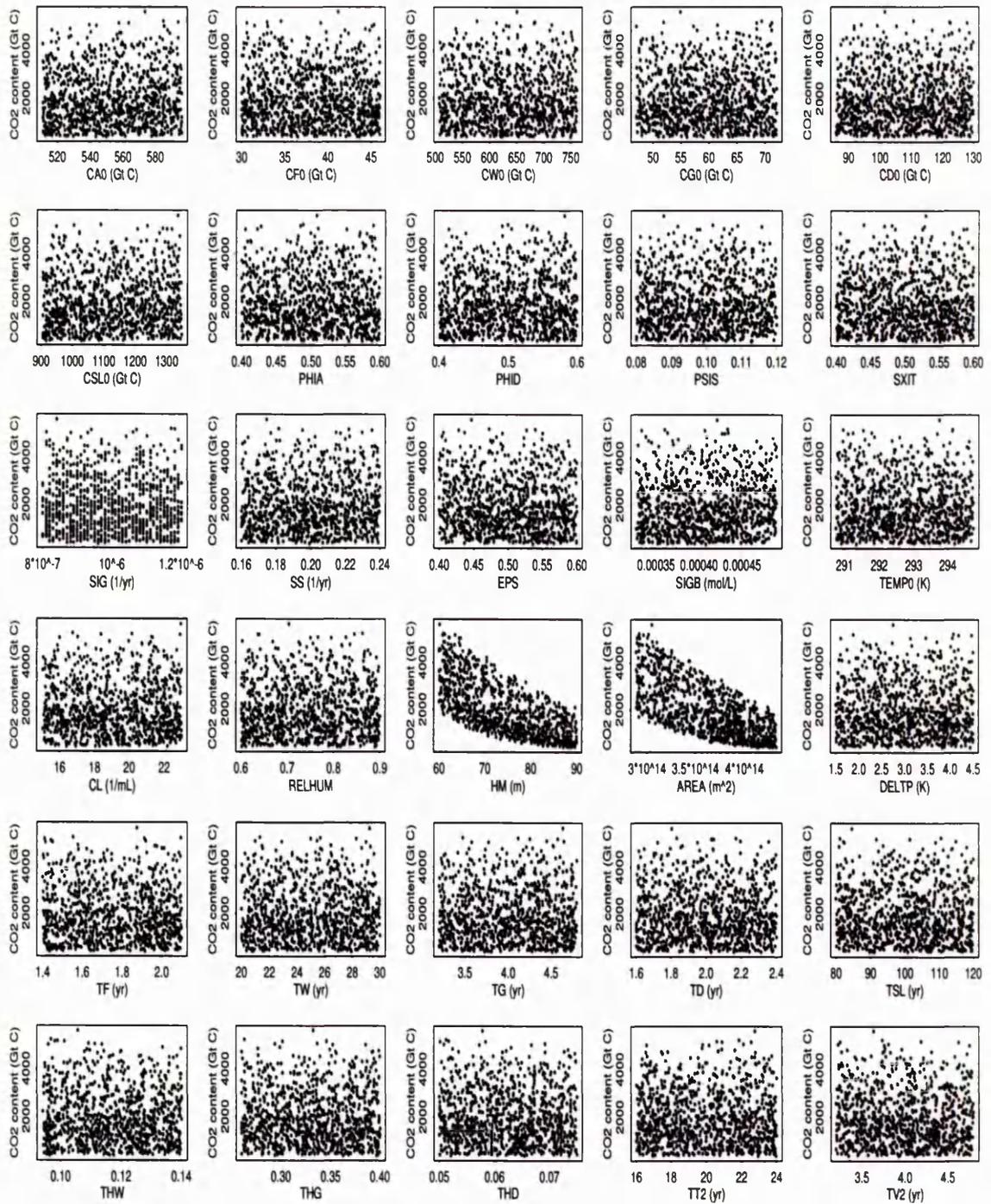
From examination of Figure 4.18 which shows the relationships between  $y_{Atm}(t = 2100)$  and each input factor, it is clear that there is a strong negative association between this model output and both HM and AREA. The relationship with these two inputs appears to be near linear, but one might argue that it is rather a nonlinear but monotonic relationship. In such cases, using rank-transformed data in regression and correlation analyses provide more reliable results. In the other plots in this figure, the vast majority of the points lie close to the abscissa but there is no particularly strong patterns in any of these plots.

Since the scatterplots for the  $y_{SO}(t = 2100)$ ,  $y_{DO_5}(t = 2100)$  and  $y_{DO_{13}}(t = 2100)$  were very similar, here we give the scatterplots of the deep ocean-layer 13 only (see Figure 4.19). Again the HM and AREA plots display well defined patterns. In the other plots corresponding to the other 28 input factors there is no obvious pattern indicating that there may be no relationship between these input factors and the output variable.

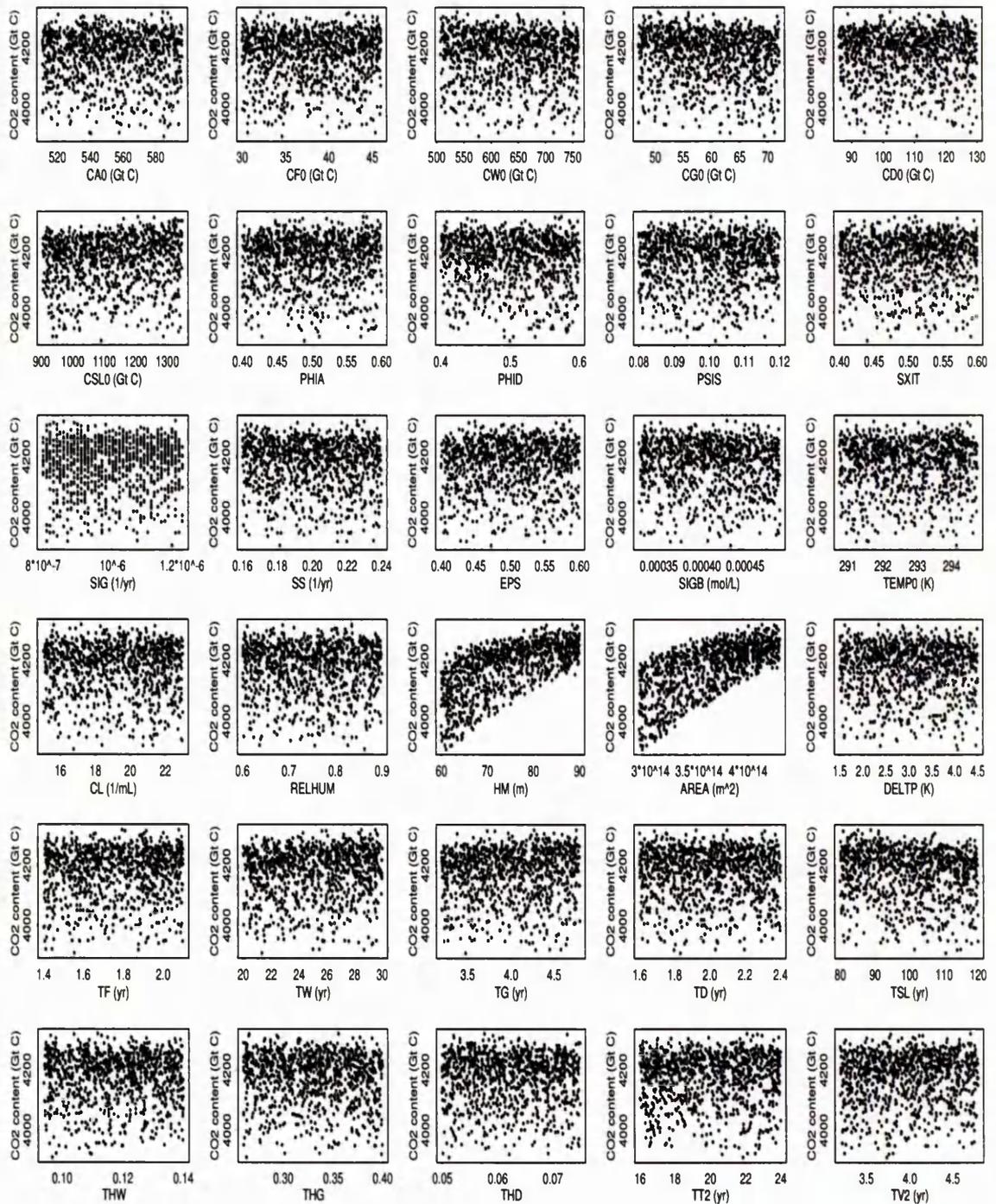
Scatterplots for the  $y_{NWPT}(t = 2100)$  (see Figure 4.20) show that there is a perfect linear association between this output variable and the factor CF0. There also seems to be a detectable linear association with SXIT and CW0, and there is hardly any relationship with the rest of the input factors.

Figure 4.21 shows that there is a positive strong association between the  $y_{WPT}(t = 2100)$  and this compartment's initial condition (i.e., CW0). Among the remaining input factors, SXIT also seems to have a linear relationship with this output variable.

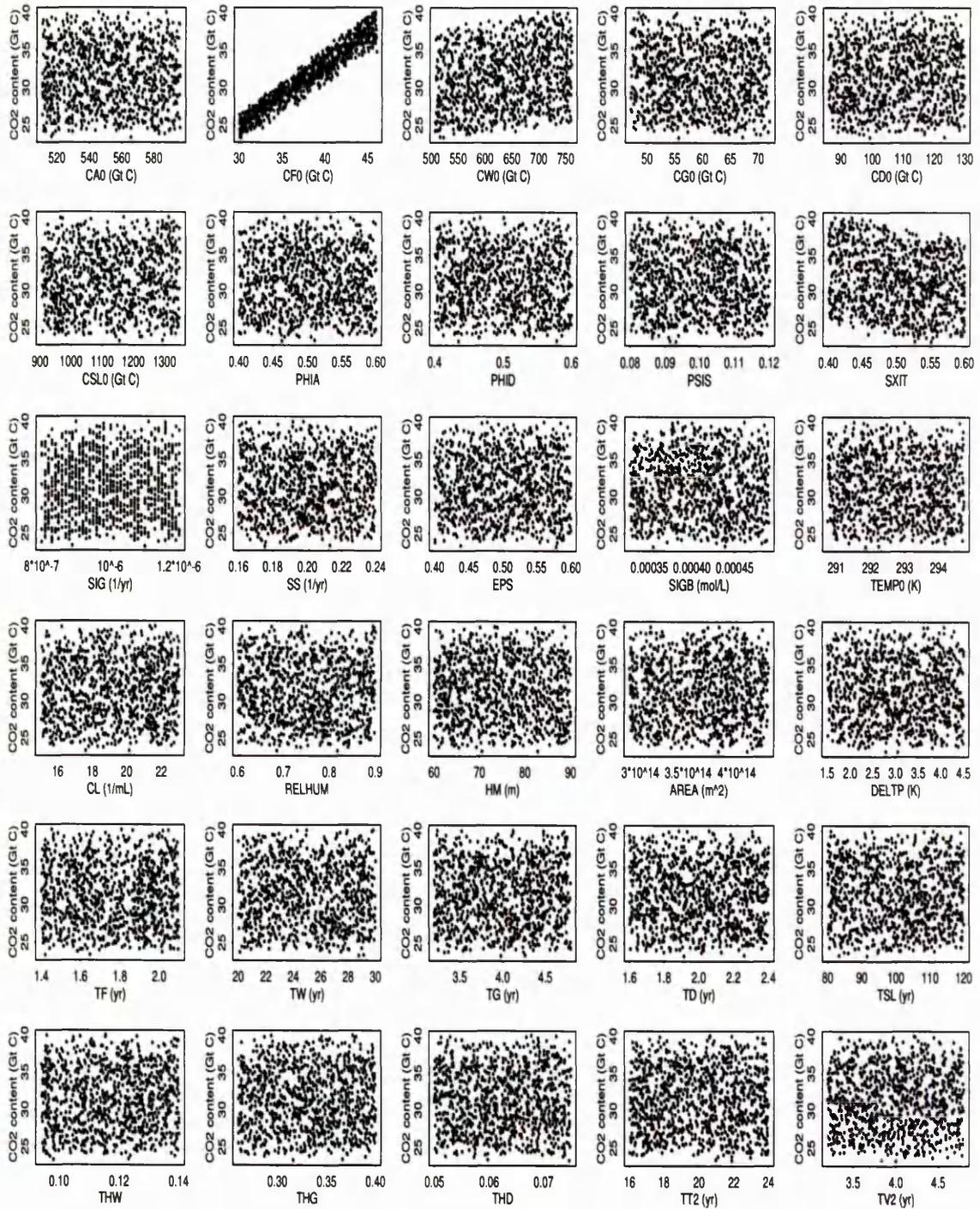
As for the output variable  $y_{GV}(t = 2100)$ , the scatterplot with the input CG0 displays a very strong linear association. The input factors EPS, CW0 and SXIT also appear to have a linear but not very strong relationship with this output variable (see Figure 4.22).



**Figure 4.18.** Scatterplots of predicted Atmosphere CO<sub>2</sub> content in year 2100 (i.e.,  $y_{Atm}(t = 2100)$ ) versus each input factor (listed in Table 4.1).



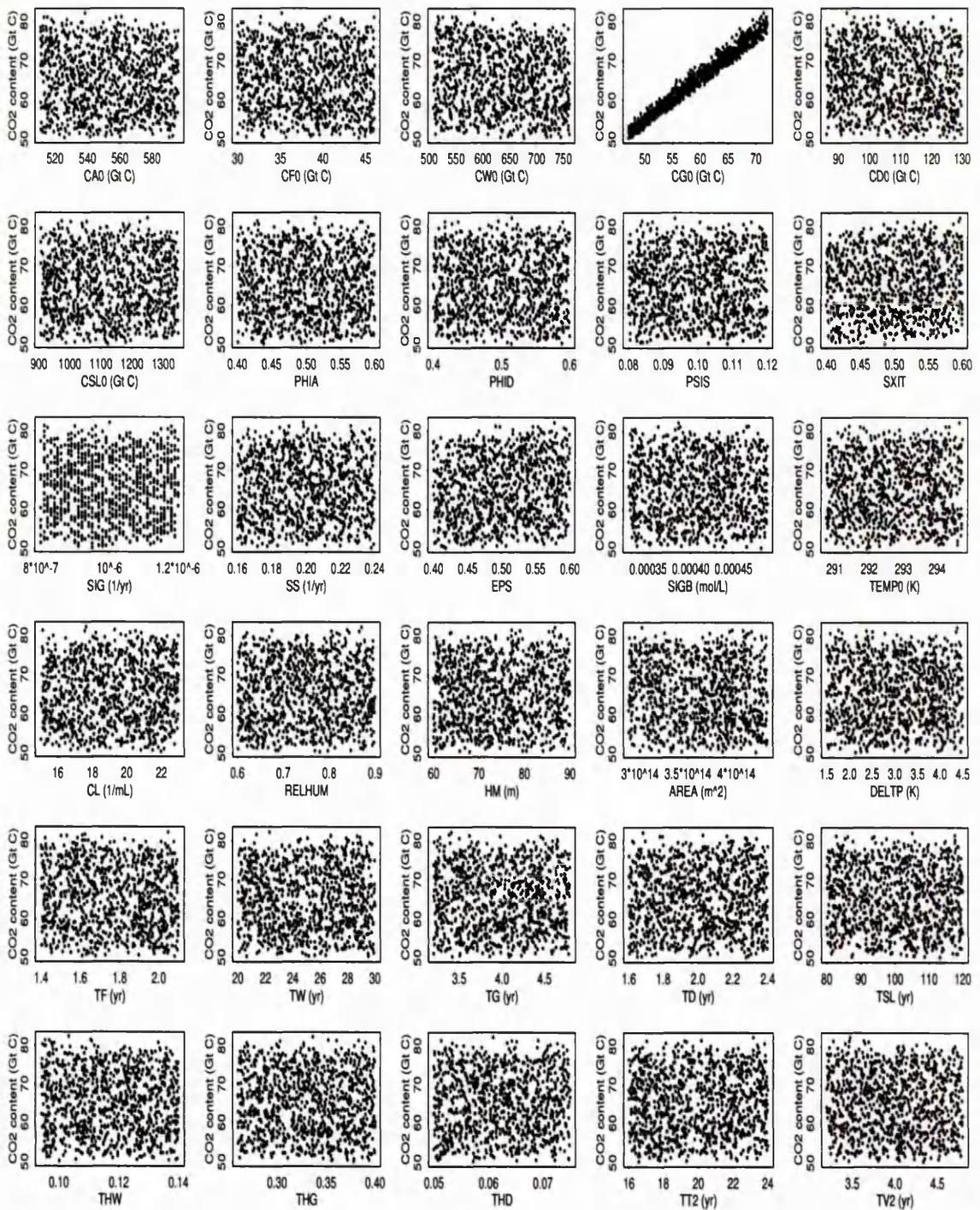
**Figure 4.19.** Scatterplots of predicted Deep Ocean (layer 13) CO<sub>2</sub> content in year 2100 (i.e.,  $y_{D_{O13}}(t = 2100)$ ) versus each input factor (listed in Table 4.1).



**Figure 4.20.** Scatterplots of predicted Nonwoody parts of Trees  $\text{CO}_2$  content in year 2100 (i.e.,  $y_{NWPT}(t = 2100)$ ) versus each input factor (listed in Table 4.1).



Figure 4.21. Scatterplots of predicted Woody parts of Trees CO<sub>2</sub> content in year 2100 (i.e.,  $y_{WPT}(t = 2100)$ ) versus each input factor (listed in Table 4.1).



**Figure 4.22.** Scatterplots of predicted Ground Vegetation CO<sub>2</sub> content in year 2100 (i.e.,  $y_{GV}(t = 2100)$ ) versus each input factor (listed in Table 4.1).

The scatterplots of  $y_{DD}(t = 2100)$  presented in Figure 4.23 show many well-defined patterns involving the input factors like TD, TF, TW, TG, CW0, CF0, CG0 and THG. As seen in Figure 4.24, there are detectable patterns in the plots of  $y_{ASC}(t = 2100)$  versus the input factors TSL, TG, TW, CG0, CW0, THG and THD.

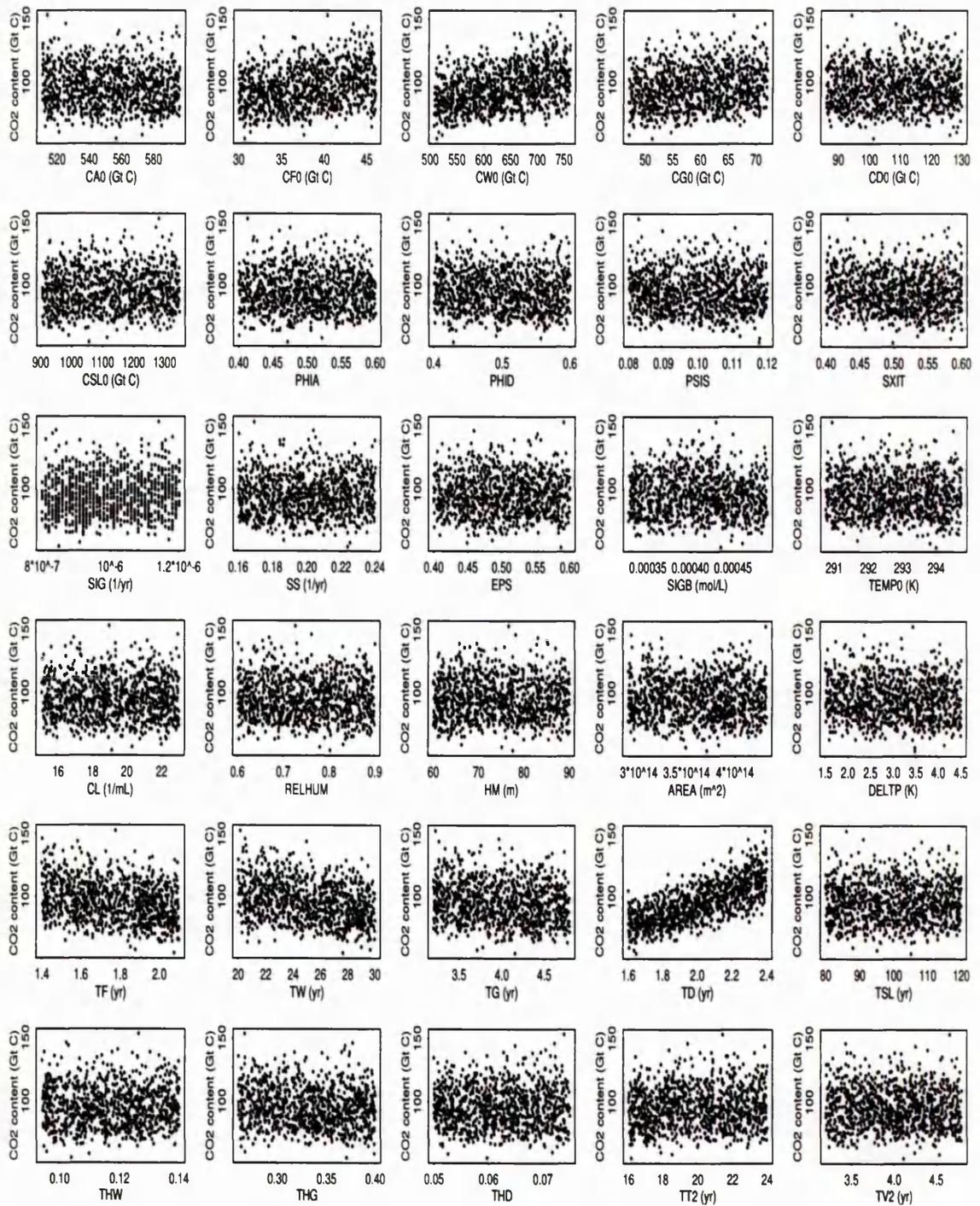
The Pearson correlation coefficients (CCs) are calculated on the model input factors and output variables along with their  $p$ -values for the selected nine compartmental contents evaluated in years 1900, 2000 and 2100. We do not report the CC values here but we obtain importance ranking of the ten most important input factors for each dependent variable and these rankings are given in Table 4.6.

In terms of the strength of the linear relationship with the output variables, some input factors become more(or less) influential over time. The CC between AREA and  $y_{Atm}(t)$ , which appears to be the strongest relationship with rank 1, is around -0.66. The CC with HM is just slightly smaller (about -0.65). All the other CC values corresponding to the rest of the inputs are quite low ( $<0.1$ ) in absolute value.

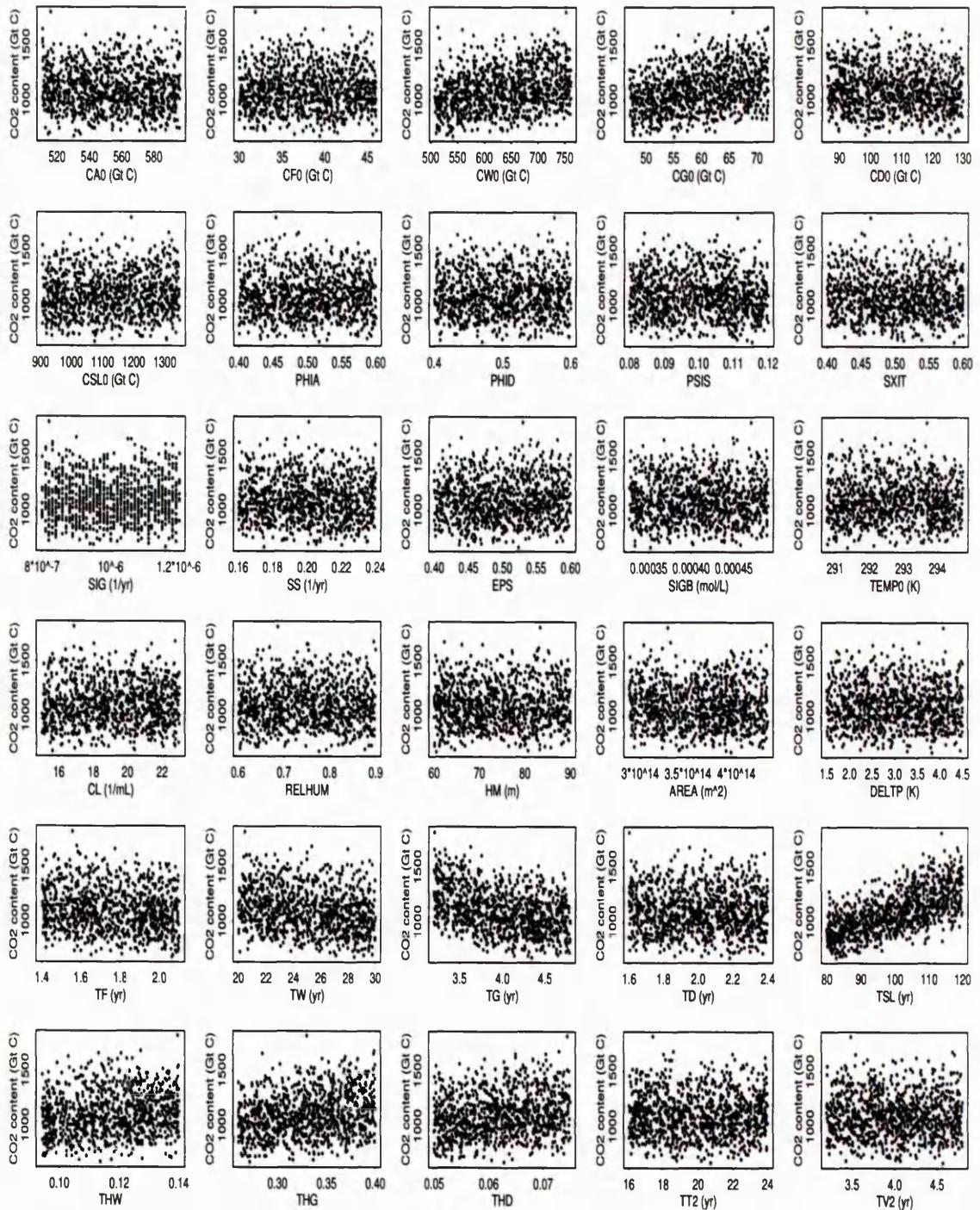
The CC values computed with the input factors and the outputs of all three ocean compartments are fairly similar. As with atmospheric output, AREA and HM are the two factors having the strongest linear relationship with the outputs of the surface and the two deep ocean compartments.

The CC between the output  $y_{NWP}(t)$  and the inputs CF0, CW0, SXIT, CSL0 and SIGB are found to be significant with  $p$ -values less than 0.1. The CC with CF0 is the strongest (0.998 in year 1900, 0.983 in 2000 and 0.965 in 2100), and the CCs with the other four factors mentioned above are less than 0.2.

For the woody parts of trees compartment, the CC with this compartment's initial condition CW0 is the strongest (0.999 in 1900 and decreasing just slightly to 0.987 in 2100). Compared to this CC, the absolute CCs related to the other inputs identified as significant (with  $p$ -values less than 0.1) are quite low, varying between 0.2 (with SXIT in 2100) and 0.02 (with AREA in 2100).



**Figure 4.23.** Scatterplots of predicted Detritus/Decomposers CO<sub>2</sub> content in year 2100 (i.e.,  $y_{DD}(t = 2100)$ ) versus each input factor (listed in Table 4.1).



**Figure 4.24.** Scatterplots of predicted Active Soil Carbon CO<sub>2</sub> content in year 2100 (i.e.,  $y_{ASC}(t = 2100)$ ) versus each input factor (listed in Table 4.1).

As the scatterplot in Figure 4.22 reveals for 2100, the output variable  $y_{GV}(t)$  and the input CG0 are highly correlated at all three chosen years. The CCs between this output variable and the inputs CW0, EPS, SXIT and CF0 are also relatively high. Apart from the CC related to CG0 which shows a small decrease over time, the CCs with the other inputs listed above are increasing over time.

The most highly correlated input factors with the output  $y_{DD}(t)$  are TD (with CC of 0.75 in year 1900), CW0 (0.34), CF0 (0.32), TW (-0.31), TF (-0.29), TG (-0.15), CG0 (0.14), THG (-0.08), CA0 (-0.03). The CC values reported here with the output at year 1900. Over time, the absolute CCs corresponding to the factors TD, CF0, TW and TF are decreasing, whereas the absolute CCs corresponding to the inputs CW0, TG, CG0, THG and CA0 are increasing.

For the output  $y_{ASC}(t)$ , TSL appears to be the most highly correlated input with a CC of about 0.66 in all three years. The correlations between this output variable and the inputs TG, CG0, THG, CW0, TW and THD are also relatively high (varying between 0.20 and 0.36 in absolute value).



### 4.4.2 Regression Methods

For quantifying the relative importance of the input factors to the model outputs, we now consider regression methods. As an illustration, we have constructed a regression model to investigate the effects of all 30 input factors on the dependent variable  $y_{Atm}(t = 2100)$  (i.e., the predicted carbon content of the atmosphere compartment in 2100). The outcome of this regression model is summarized in Table 4.7. According to the  $p$ -values for each of the 30 input factors, it seems that out of the 30 inputs, 17 (given in bold-italic in the table) have  $p$ -values less than 0.02 and hence we can say that these input factors, namely CF0, CW0, CG0, CSL0, PHIA, PHID, PSIS, SIGB, TEMP0, CL, HM, AREA, TW, TG, TSL, THG and THD, appear to influence the response variable.

The regression coefficients in a linear regression model, which can provide rankings of all inputs, depend on their units. If the input factors included in a regression model have different units which is the case with our input factors here, then it is difficult to obtain a meaningful ranking of the input variables based directly on the regression coefficients. Therefore, it is necessary to standardise to remove the unit effects on the final coefficients. These standardised coefficients (SRCs) have been calculated on this model and the results are presented in the following section (Section 4.4.3) along with the partial correlation coefficients (PCCs).

As an alternative to constructing a regression model, which contains all input factors, for each output variable and presenting vast amount of analysis results, we have performed stepwise regression analysis using the predictions from the 5,000 model runs and the results are given in Section 4.4.4.

**Table 4.7.** Summary of regression analysis for  $y_{Atm}(t = 2100)$  (Atmospheric CO<sub>2</sub> content at year 2100) and input factors CA0, CF0, CW0, CG0, CD0, CSL0, PHIA, PHID, PSIS, SXIT, SIG, SS, EPS, SIGB, TEMP0, CL, RELHUM, HM, AREA, DELTP, TF, TW, TG, TD, TSL, THW, THG, THD, TT2 and TV2 (see Table 4.1 for description of these inputs).

Variable	Regression Coefficient	Std. Error of Coeff.	T-test	p-value	
CA0	3.280000e-01	1.610000e-01	2.036700e+00	0.0417	
CF0	-2.082400e+00	8.590000e-01	-2.424200e+00	0.0154	
CW0	-1.872000e-01	5.400000e-02	-3.466200e+00	0.0005	
CG0	-4.376700e+00	5.551000e-01	-7.885200e+00	0.0000	
CD0	6.064000e-01	3.120000e-01	1.944000e+00	0.0520	
CSL0	4.238000e-01	3.090000e-02	1.371680e+01	0.0000	
PHIA	1.729196e+02	6.951790e+01	2.487400e+00	0.0129	
PHID	1.689955e+02	6.913010e+01	2.444600e+00	0.0145	
PSIS	-8.895654e+02	3.452463e+02	-2.576600e+00	0.0100	
SXIT	1.508464e+02	6.920080e+01	2.179800e+00	0.0293	
SIG	4.969968e+06	3.465408e+07	1.434000e-01	0.8860	
SS	-6.150850e+01	1.732762e+02	-3.550000e-01	0.7226	
EPS	-8.934900e+01	6.934510e+01	-1.288500e+00	0.1976	
SIGB	2.882116e+05	8.498279e+04	3.391400e+00	0.0007	
TEMP0	1.395650e+01	3.454400e+00	4.040100e+00	0.0001	
CL	8.120200e+00	1.716300e+00	4.731100e+00	0.0000	
RELHUM	-1.289150e+01	4.579230e+01	-2.815000e-01	0.7783	
HM	-8.647960e+01	4.600000e-01	-1.879872e+02	0.0000	
AREA	0.000000e+00	0.000000e+00	-1.894515e+02	0.0000	
DELTP	-2.464400e+00	4.585500e+00	-5.374000e-01	0.5910	
TF	2.754300e+01	1.959520e+01	1.405600e+00	0.1599	
TW	5.550500e+00	1.380700e+00	4.020000e+00	0.0001	
TG	6.492810e+01	8.541700e+00	7.601300e+00	0.0000	
TD	4.147000e-01	1.723220e+01	2.410000e-02	0.9808	
TSL	-4.555600e+00	3.410000e-01	-1.336040e+01	0.0000	
THW	-5.437331e+02	2.968994e+02	-1.831400e+00	0.0671	
THG	-6.446091e+02	9.867310e+01	-6.532800e+00	0.0000	
THD	-2.048286e+03	5.493671e+02	-3.728400e+00	0.0002	
TT2	-1.767300e+00	1.719200e+00	-1.028000e+00	0.3040	
TV2	-3.449500e+00	8.577800e+00	-4.021000e-01	0.6876	
R-Squared = 93.46%		Intercept = 10696.55			
Source	DF	Sum of Squares	Mean Sum of Squares	F-statistic	p-value
Regression	30	5594310605	186477020	2368.48	0.000
Residual	4969	391223410	78733		
Total	4999	5985534015			

### 4.4.3 SRC and PCC

Using the results of  $N=5,000$  model runs, for each input and output combination of interest, we have calculated the SRCs and PCCs which provide measures of the relative contribution of each of the inputs to the observed output variations.

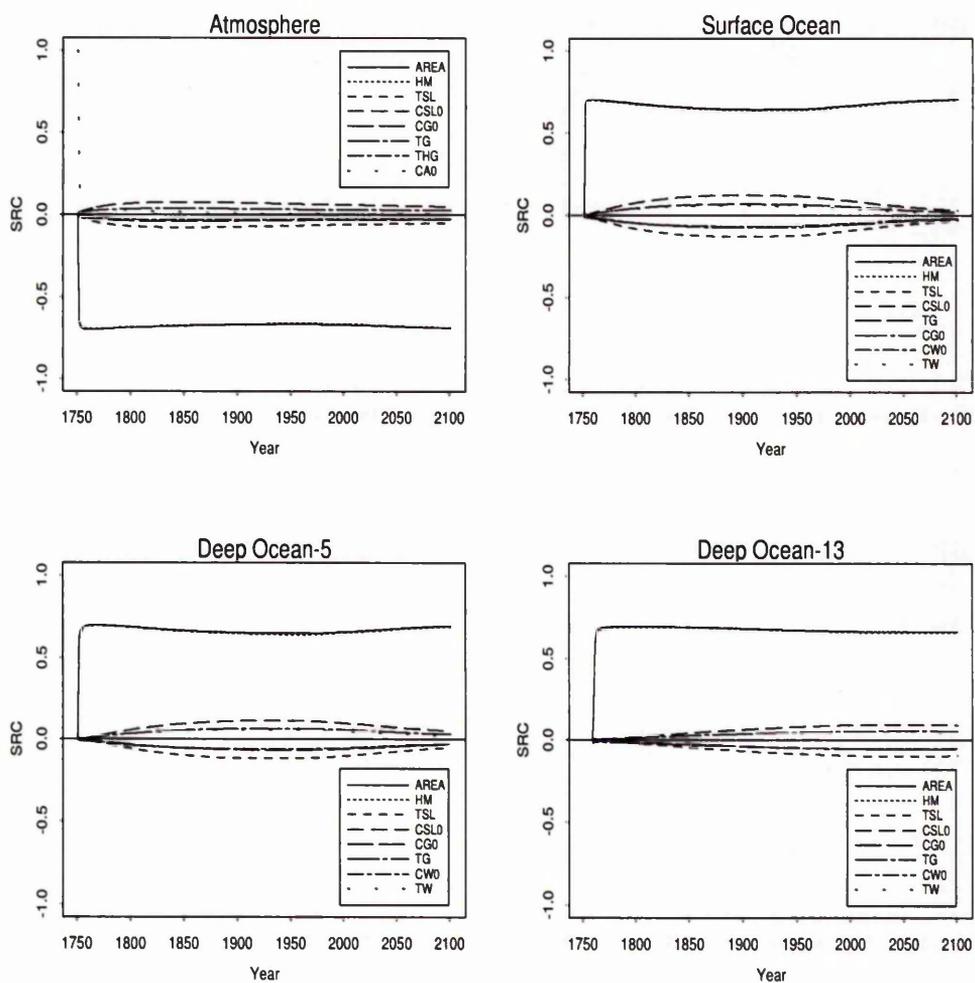
When using the SRCs it is also important to consider the model coefficient of determination  $R^2$ . The  $R^2$ -values we have obtained from the regression models in which the nine compartmental contents in 1900, 2000 and 2100 are the response variables and the 30 input factors are the explanatory variables vary between 85.5% and 100%. These high  $R^2$ -values indicate that the SRCs are valid as a measure of sensitivity.

Since our response variables are functions of time, we have calculated the SRCs and PCCs at each time point and presented these sensitivity results with plots, thus indicating the importance of each input on an output over time. The time-dependent behaviour of these estimated coefficients for eight input factors having the highest values are shown in Figures 4.25 and 4.26 for the SRCs, and Figures 4.27 and 4.28 for the PCCs. For both sets of SRC and PCC curves given in these figures, the dependent variables are the nine compartmental  $\text{CO}_2$  contents and each curve displays the values of the SRCs and PCCs relating these compartmental contents to a single input factor as a function of time. Note that in these figures we show the time-dependent behaviour of the SRCs and PCCs associated with the eight most important input factors only.

Figures 4.25 and 4.27 show that the input factors AREA and HM seem to have a large influence on the atmosphere, surface ocean and the two deep ocean (layer5 and layer13) compartments (i.e.,  $y_{Atm}(t)$ ,  $y_{SO}(t)$ ,  $y_{DO5}(t)$  and  $y_{DO13}(t)$ , respectively), and they are almost equally important. As seen in the top-left frames of these two figures, CA0 has the SRC and PCC values of 1 at the starting year 1750 but this value decreases very rapidly and this input does not appear to have any significant effect on the output variable  $y_{Atm}(t)$ . Although the input factors included in these two figures are the factors that are associated with higher values of SRCs and PCCs (see legends of Figures 4.25 and 4.27), compared to

AREA and HM the other inputs do not seem to be as influential on these four compartments.

The sensitivity of the surface ocean and the deep ocean-layer5 compartments to the other input factors (TSL, CSL0, CG0, TG, CW0 and TW) shown in the figures seems to be increasing for the first 100-150 years and then decreasing gradually over time. As the results for deep ocean-layer13 reveals it takes longer for



**Figure 4.25.** Time-dependent behaviour of the SRCs for the atmosphere and three ocean compartments considered.

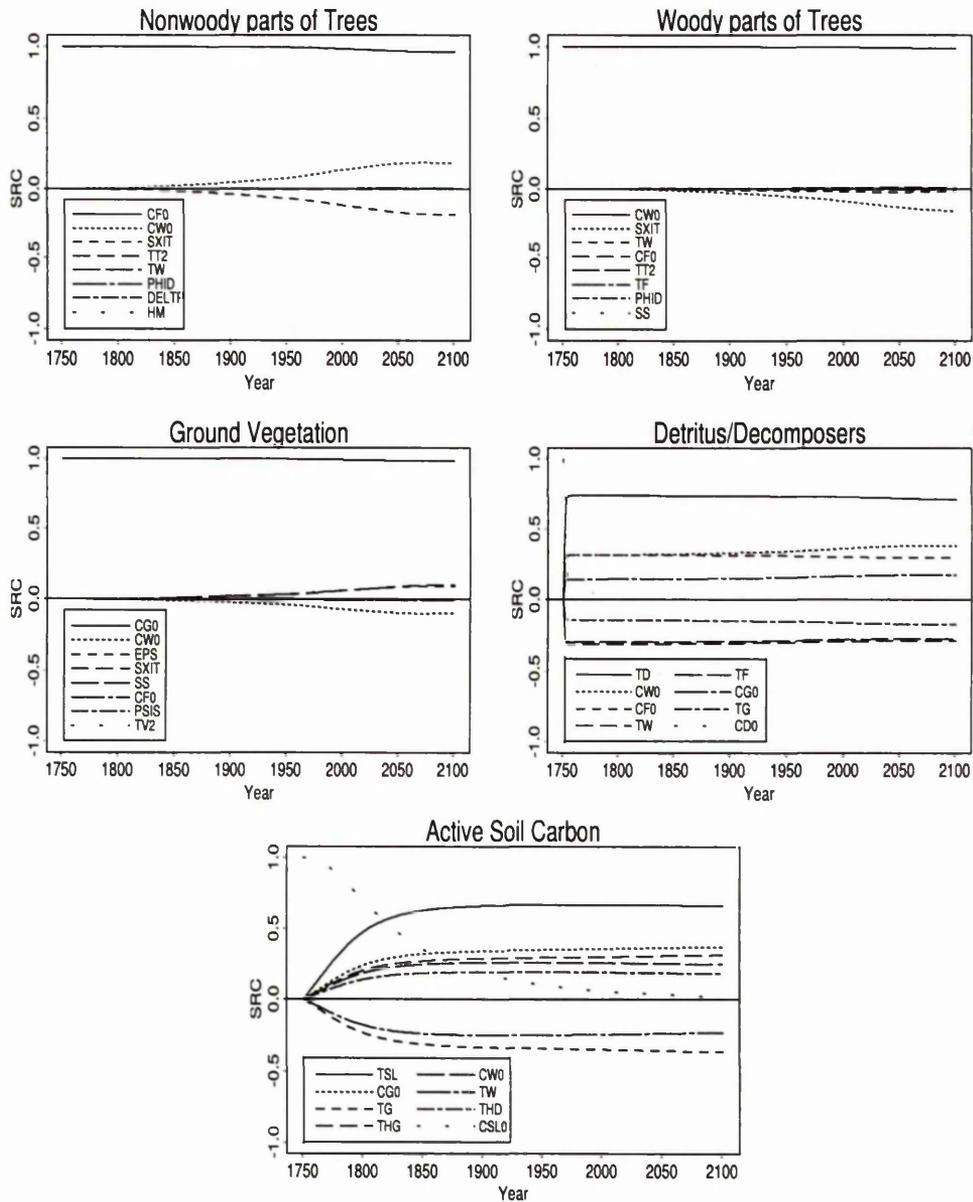
CO<sub>2</sub> to move into the deeper layers of the ocean, so the sensitivities of these compartments to the model input factors in terms of the time-dependent behaviour of their SRC and PCC values show much slower change over 350 years.

The SRC and PCC curves computed on the sampled input factors and the terrestrial compartments are given in Figures 4.26 and 4.28. The nonwoody parts of trees compartment is most sensitive to its initial condition CF0 which has both SRC and PCC of 1 in 1750 and decreases only slightly over the whole time period. By looking at the SRC curves of the most important 8 inputs for  $y_{NWPT}(t)$ , we can see that except for CF0, CW0 and SXIT the rest of the inputs do not seem to be having any influence on this compartment. In terms of the PCC curves, CF0, CW0, SXIT and TT2 appear to be having significant influence on the nonwoody parts of trees compartment. The influence of the other inputs is insignificant. The PCC values associated with CF0 and CW0 are about 1 which indicates that the output variable  $y_{NWPT}(t)$  and these two inputs increase together while the PCC value associated with SXIT, which is about -1, implies that  $y_{NWPT}(t)$  decreases as SXIT increases. The sensitivity of  $y_{NWPT}(t)$  to TT2 decreases with time, and by 2100 the influence of this factor becomes insignificant.

The sensitivity of the output variable  $y_{WPT}(t)$  to CW0 is the highest and this does not seem to change over the years. As shown in the corresponding SRC graph of Figure 4.26, this compartment is also sensitive to SXIT but shows hardly any sensitivity to the rest of the inputs presented in this graph and hence to the rest of the other inputs included in the analysis (since only the top 8 inputs having the highest SRC values are included in the figure).

The graph with the PCC curves for  $y_{WPT}(t)$  show that CW0, SXIT and TW are almost equally influential on this model output. The influence of CF0 is also very high over the whole time period. The sensitivity of  $y_{WPT}(t)$  to TT2 and TF is also relatively high but decreasing over time, especially after year 2050.

As for the ground vegetation compartment, the influence of CG0 on  $y_{GV}(t)$ , in terms of the SRCs, is very high (see the left frame in the second row of Figure 4.26). CG0's SRC value is decreasing over time but only slightly. The



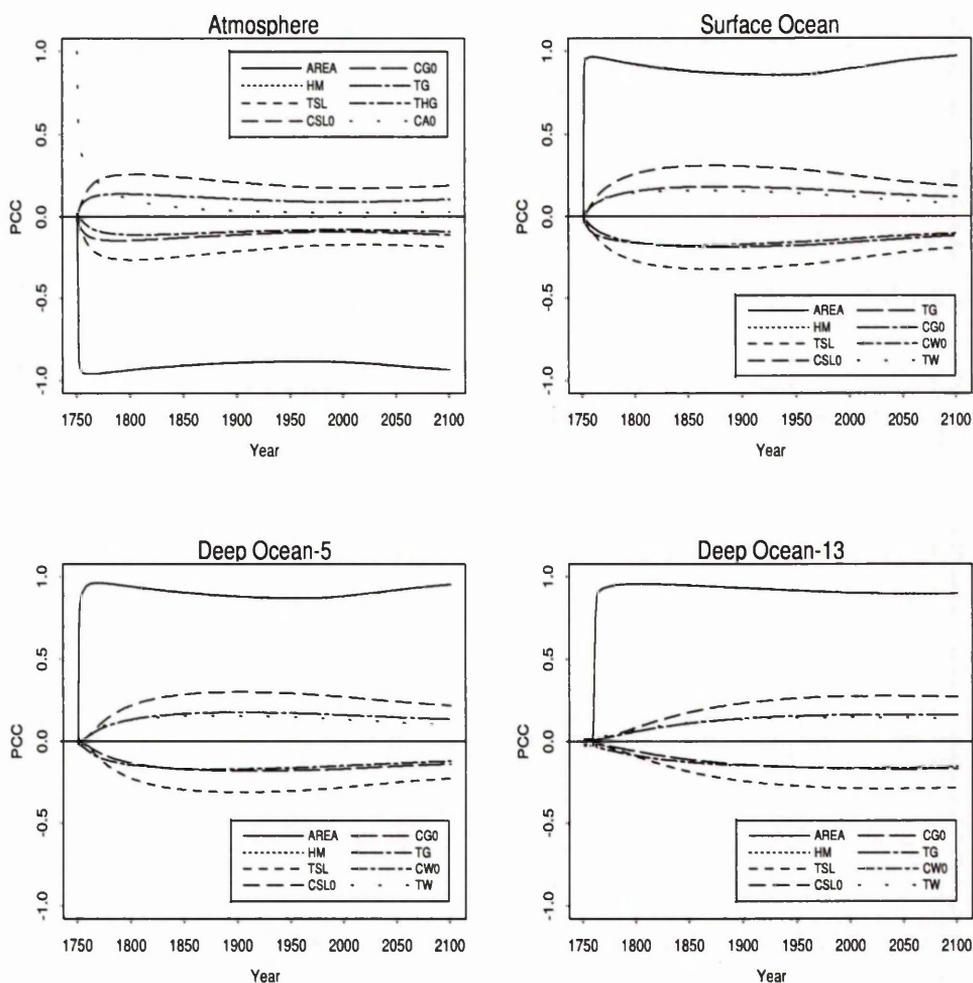
**Figure 4.26.** Time-dependent behaviour of the SRCs for the terrestrial compartments.

sensitivity of  $y_{GV}(t)$  to CW0, EPS and SXIT is also relatively high, and increasing over time. The SRCs related to the other inputs included in this figure are around zero. The PCC curves provided in Figure 4.28 for the ground vegetation

compartment show that CG0 and CW0 are the most influential inputs on this output variable, and they are followed by EPS and SXIT having very high PCC values. The time-dependent behaviour of the PCC of SXIT present an interesting picture. Its PCC value is -0.0157 for the first 15 years, then it increases quite fast in the next three years, reaching to -0.7895 in 1768. Later, it starts decreasing until year 1775 (taking an SRC value of -0.0788) and then increasing rapidly in the following 10-15 years in the positive direction. Its SRC value in 1800 is about 0.91, and continues increasing and reaches an SRC value of 0.951 in 2100. The input factor SS also seems to be important for this compartment especially in the early years, but its influence is diminishing fast and becoming insignificant by 2100. The PCC curve associated with CF0 shows that  $y_{GV}(t)$  is also sensitive to this input. The degree of sensitivity to this input decreases slightly over the years. The PCC curve of TV2 presents a decreasing influence on this output variable whereas the influence of PSIS which is not very high does not change with time.

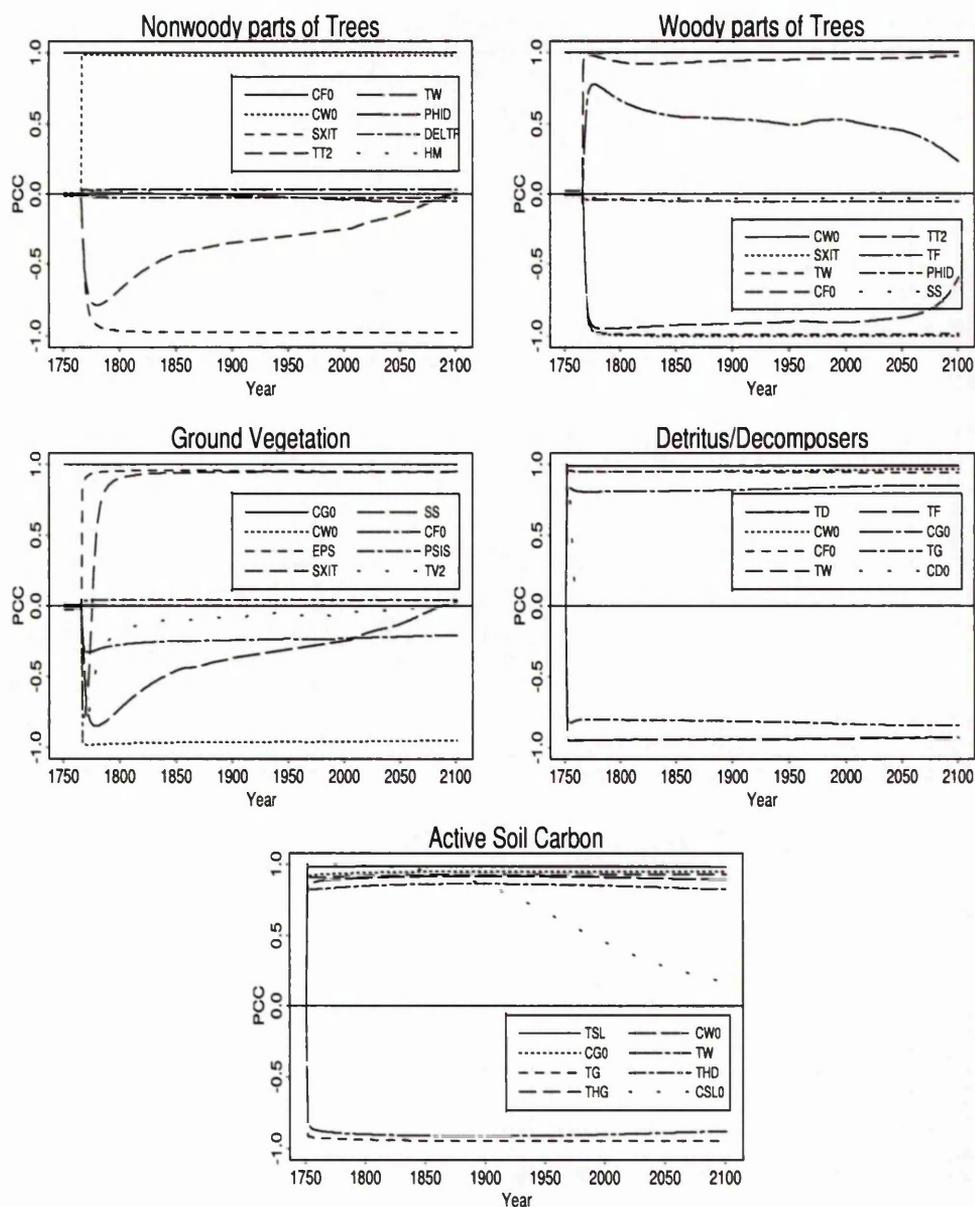
Even though the PCCs are higher than the SRCs, the corresponding graphs for the detritus/decomposers compartment (i.e.,  $y_{DD}(t)$ ), which display two sets of curves, present very similar pictures (see the right frames in the second rows of Figures 4.26 and 4.28). The time-dependent behaviour of the SRCs and PCCs of the eight most important input factors show small changes over time. Both estimated coefficients of the inputs CW0, CG0 and TG increase whereas the SRCs and PCCs of TD, CF0, TW, TF and CD0 decrease over time but very slowly.

In terms of both SRCs and PCCs, the sensitivity of the active soil carbon compartment to CSL0 is highest in the early years but over the years the influence of this input on  $y_{ASC}(t)$  decreases (see the corresponding graphs in Figures 4.26 and 4.28). The SRC curves of the other seven input factors show that the sensitivity of  $y_{ASC}(t)$  to these inputs keeps increasing gradually in the first 100 years, then do not seem to be changing much in the next 250 years. With the PCCs, the time-dependent behaviour of the input factors, except for CSL0 which shows a decrease, does not appear to be changing significantly over the whole time period.



**Figure 4.27.** Time-dependent behaviour of the PCCs for the atmosphere and the three ocean compartments considered.

The SRCs and PCCs provide related but not identical measures of input factor importance. From Figures 4.25 - 4.28, we see that the PCCs tend to be larger than the SRCs, but because the input factors under consideration here are independent, the rankings of input importance based on the absolute SRCs and PCCs are identical. The rankings of the most important ten input factors for the nine model outputs fixed at three years are given in Table 4.8.



**Figure 4.28.** Time-dependent behaviour of the PCCs for the terrestrial compartments.

The rankings for each output variable evaluated at the three chosen years do not appear to differ much from year to year. The input factor rankings for the atmosphere, surface ocean and the two deep ocean compartments are in quite

a good agreement. For the nonwoody and woody parts of trees compartments, almost the same input factors are identified among the ten most influential inputs with slightly different ranks. Most of the ten inputs picked out as important for the tree compartments are also influential on the ground vegetation compartment. In the table, the importance rankings for the detritus/decomposers and the active soil carbon compartments show that most of the ten inputs which are important for one of these compartments are also important for the other but with different order of importance.



#### 4.4.4 Stepwise Regression

In this section, we present results from the stepwise regression analysis. The stepwise regression procedure was performed using the results from the  $N=5,000$  model runs associated with the carbon content of nine selected compartments as the response variables and the 30 input factors as the explanatory variables. As often done in sensitivity studies, an  $\alpha$ -value of 0.01 is used to add a variable to each regression model, and an  $\alpha$ -value of 0.05 is considered to drop a variable from the model. The results are shown in Table 4.9 (for the atmosphere and the three ocean compartments) and in Table 4.10 (for the terrestrial compartments) in terms of the order in which the input factors are selected (the most important input factor is selected first, the next most important factor is selected second, and so on), the  $R^2$  and the PRESS (predicted error sum of squares) values at successive steps of the analysis are reported.

We first examine the adequacy of the regression models by looking at the PRESS values reported in Tables 4.9 and 4.10. The PRESS values computed for each regression model obtained at each step of the analysis appear to decrease in size as additional input variables are added to the corresponding regression model, which indicates that the regression models are not overfitting the data.

In Table 4.9, which presents the analysis results for  $y_{Atm}(t)$ ,  $y_{SO}(t)$ ,  $y_{DO_5}(t)$  and  $y_{DO_{13}}(t)$  (where  $t=1900, 2000, 2100$ ), we see that under each response variable the input factor AREA is selected at the first step of the analysis indicating that this input factor has the greatest impact on these twelve output variables. The  $R^2$ -values computed with only AREA in each regression model vary between 0.4041 and 0.4796. That is, AREA accounts for approximately 40-48% of the uncertainty in the output variables under consideration here.

In the second step of the analysis, HM is identified as the second most influential input on all twelve output variables. Including HM in the regression models, in addition to the input factor AREA, increases  $R^2$  values significantly. Thus, AREA and HM together account for approximately 80-97% of the uncertainty in the output variables. With the addition of 5 more inputs which are identified as

having impact on the response variables the effectiveness of the regression models are further improved. In each of the regression models summarised in Table 4.9, seven out of 30 inputs are selected. The seven inputs identified as influential on a compartment at three chosen years are the same, but the order in which they were added to the model changes slightly from year to year. For instance, for  $y_{SO}(t = 1900)$  and  $y_{SO}(t = 2000)$ , CG0 is added to the regression model on the fifth step, but when  $y_{SO}(t = 2100)$  is considered CG0 is added to the model on the sixth step of the analysis.

The input factors AREA, HM, TSL, CSL0 and TG are all selected in the regression analyses for all response variables  $y_{Atm}(t)$ ,  $y_{SO}(t)$ ,  $y_{DO_5}(t)$  and  $y_{DO_{13}}(t)$ . THG which is among the influential inputs for the atmosphere compartment is not identified as influential for the three ocean compartments. CW0, on the other hand, is selected in the regression analyses on the three ocean compartments, but not in the analyses for  $y_{Atm}(t)$ . Unlike the regression models on  $y_{DO_{13}}(t = 2000)$  and  $y_{DO_{13}}(t = 2100)$ , the regression model for  $y_{DO_{13}}(t = 1900)$  includes TW and excludes CG0.

The analyses in Table 4.10 for the response variables related to the terrestrial compartments appear to be very effective with  $R^2$ -values over 0.91. Under each response variable, a decreasing sequence of PRESS values implies that the regression models are not overfitting the data which they are based on. The regression on  $y_{NWPT}(t)$  selects the input factors CF0, CW0, SXIT and TT2 when  $t=1900$ . In addition to these four input factors, in 2000, TW is also included in the regression model. Later in 2100, TT2 is dropped from the regression model, and the order of the selection of the second and third inputs is reversed. However, including TT2 and TW in the regression models for  $y_{NWPT}(t)$ , in addition to CF0, CW0 and SXIT, does not improve the effectiveness of these models.

For  $y_{WPT}(t)$ , seven of the inputs involved in the analyses (i.e., CW0, SXIT, TW, CF0, TT2, TF and PHID -in this order) are identified in the regression analyses when the outputs from years 1900 and 2100 are considered. The analysis on  $y_{WPT}(t = 2000)$ , however, does not include PHID in the final regression model.

The order of input selection does not change from year to year.

The input factors CG0, CW0, EPS and SXIT are selected by the regression analysis on  $y_{GV}(t)$  in all three years. In addition to these four input factors, SS and TV2 in year 1900, SS and PSIS in 2000, and PSIS and TW in 2100 are included in the regression models for  $y_{GV}(t)$ . The orders in which these inputs are added to the models can be seen in Table 4.10.

Considering the regression analyses results for  $y_{NWPT}(t)$ ,  $y_{WPT}(t)$  and  $y_{GV}(t)$ , we see that even though the analyses identify a number of inputs as influential, including only the first inputs selected at the first step of the analyses in the

**Table 4.9.** Summary of stepwise regression analyses for the CO<sub>2</sub> content of the atmosphere and the three ocean compartments (i.e., output variables  $y_{Atm}$ ,  $y_{SO}$ ,  $y_{DO_5}$  and  $y_{DO_{13}}$ ) in 1900, 2000 and 2100; based on N=5,000 model runs and IS92a emission scenario.

Comp. Output	Step	$y_i(t = 1900)$			$y_i(t = 2000)$			$y_i(t = 2100)$		
		Variable	$R^2$	PRESS	Variable	$R^2$	PRESS	Variable	$R^2$	PRESS
$y_{Atm}$	1	AREA	0.4287	$9.51 \times 10^8$	AREA	0.4298	$15.90 \times 10^8$	AREA	0.4565	$32.60 \times 10^8$
	2	HM	0.8694	$2.18 \times 10^8$	HM	0.8703	$3.63 \times 10^8$	HM	0.9260	$4.44 \times 10^8$
	3	TSL	0.8745	$2.09 \times 10^8$	TSL	0.8737	$3.54 \times 10^8$	CSLO	0.9283	$4.30 \times 10^8$
	4	CSLO	0.8795	$2.01 \times 10^8$	CSLO	0.8772	$3.44 \times 10^8$	TSL	0.9306	$4.16 \times 10^8$
	5	TG	0.8809	$1.99 \times 10^8$	TG	0.8783	$3.41 \times 10^8$	CG0	0.9314	$4.12 \times 10^8$
	6	CG0	0.8823	$1.96 \times 10^8$	CG0	0.8793	$3.38 \times 10^8$	TG	0.9322	$4.07 \times 10^8$
	7	THG	0.8832	$1.95 \times 10^8$	THG	0.8800	$3.36 \times 10^8$	THG	0.9328	$4.04 \times 10^8$
$y_{SO}$	1	AREA	0.4041	$4.25 \times 10^6$	AREA	0.4328	$5.20 \times 10^6$	AREA	0.4796	$11.38 \times 10^6$
	2	HM	0.7982	$1.44 \times 10^6$	HM	0.8608	$1.28 \times 10^6$	HM	0.9662	$0.74 \times 10^6$
	3	TSL	0.8144	$1.32 \times 10^6$	TSL	0.8688	$1.20 \times 10^6$	TSL	0.9674	$0.71 \times 10^6$
	4	CSLO	0.8286	$1.22 \times 10^6$	CSLO	0.8758	$1.14 \times 10^6$	CSLO	0.9684	$0.69 \times 10^6$
	5	CG0	0.8340	$1.19 \times 10^6$	CG0	0.8788	$1.11 \times 10^6$	TG	0.9688	$0.68 \times 10^6$
	6	TG	0.8391	$1.15 \times 10^6$	TG	0.8815	$1.09 \times 10^6$	CG0	0.9692	$0.67 \times 10^6$
	7	CW0	0.8436	$1.12 \times 10^6$	CW0	0.8838	$1.07 \times 10^6$	CW0	0.9696	$0.66 \times 10^6$
$y_{DO_5}$	1	AREA	0.4171	$4.09 \times 10^6$	AREA	0.4229	$5.12 \times 10^6$	AREA	0.4689	$10.03 \times 10^6$
	2	HM	0.8268	$1.22 \times 10^6$	HM	0.8391	$1.43 \times 10^6$	HM	0.9401	$1.13 \times 10^6$
	3	TSL	0.8401	$1.13 \times 10^6$	TSL	0.8496	$1.34 \times 10^6$	TSL	0.9428	$1.08 \times 10^6$
	4	CSLO	0.8518	$1.04 \times 10^6$	CSLO	0.8588	$1.26 \times 10^6$	CSLO	0.9451	$1.04 \times 10^6$
	5	CG0	0.8562	$1.01 \times 10^6$	CG0	0.8625	$1.22 \times 10^6$	CG0	0.9462	$1.02 \times 10^6$
	6	TG	0.8604	$0.98 \times 10^6$	TG	0.8660	$1.19 \times 10^6$	TG	0.9471	$1.00 \times 10^6$
	7	CW0	0.8644	$0.96 \times 10^6$	CW0	0.8690	$1.17 \times 10^6$	CW0	0.9479	$0.98 \times 10^6$
$y_{DO_{13}}$	1	AREA	0.4494	$2.35 \times 10^6$	AREA	0.4328	$8.74 \times 10^6$	AREA	0.4330	$17.98 \times 10^6$
	2	HM	0.8971	$0.44 \times 10^6$	HM	0.8598	$2.16 \times 10^6$	HM	0.8587	$4.49 \times 10^6$
	3	TSL	0.9017	$0.42 \times 10^6$	TSL	0.8691	$2.02 \times 10^6$	TSL	0.8678	$4.20 \times 10^6$
	4	CSLO	0.9057	$0.40 \times 10^6$	CSLO	0.8772	$1.90 \times 10^6$	CSLO	0.8757	$3.95 \times 10^6$
	5	TG	0.9075	$0.39 \times 10^6$	CG0	0.8803	$1.85 \times 10^6$	CG0	0.8789	$3.85 \times 10^6$
	6	TW	0.9092	$0.38 \times 10^6$	CW0	0.8833	$1.80 \times 10^6$	TG	0.8818	$3.76 \times 10^6$
	7	CW0	0.9109	$0.37 \times 10^6$	TG	0.8862	$1.76 \times 10^6$	CW0	0.8847	$3.67 \times 10^6$

**Table 4.10.** Summary of stepwise regression analyses for the CO<sub>2</sub> content of the terrestrial ecosystem compartments (i.e., output variables  $y_{NWPT}$ ,  $y_{WPT}$ ,  $y_{GV}$ ,  $y_{DD}$  and  $y_{ASC}$ ) in years 1900, 2000 and 2100; based on N=5,000 model runs and IS92a emission scenario.

Comp. Output	Step	$y_i(t = 1900)$			$y_i(t = 2000)$			$y_i(t = 2100)$		
		Variable	$R^2$	PRESS	Variable	$R^2$	PRESS	Variable	$R^2$	PRESS
$y_{NWPT}$	1	CF0	0.9964	355.27	CF0	0.9668	2838.00	CF0	0.9307	5600.99
	2	CW0	0.9983	162.37	CW0	0.9845	1322.70	SXIT	0.9648	2844.21
	3	SXIT	0.9999	13.47	SXIT	0.9991	75.58	CW0	0.9983	136.28
	4	TT2	0.9999	12.34	TT2	0.9992	70.61	TW	0.9983	135.98
	5				TW	0.9992	70.53			
$y_{WPT}$	1	CW0	0.9990	26033.20	CW0	0.9912	238331.00	CW0	0.9738	720820.00
	2	SXIT	0.9999	3105.15	SXIT	0.9995	14302.00	SXIT	0.9997	9336.42
	3	TW	1.0000	1024.08	TW	0.9999	4017.40	TW	0.9999	3881.73
	4	CF0	1.0000	697.62	CF0	0.9999	1764.59	CF0	1.0000	716.48
	5	TT2	1.0000	473.43	TT2	1.0000	720.14	TT2	1.0000	621.28
	6	TF	1.0000	455.53	TF	1.0000	640.66	TF	1.0000	610.36
	7	PHID	1.0000	455.09	PHID	1.0000	609.60	PHID	1.0000	609.60
$y_{GV}$	1	CG0	0.9987	364.18	CG0	0.9875	3742.62	CG0	0.9728	8677.02
	2	CW0	0.9992	224.32	CW0	0.9922	2324.29	SXIT	0.9820	5746.57
	3	EPS	0.9996	112.15	EPS	0.9959	1224.79	CW0	0.9909	2910.39
	4	SXIT	0.9999	17.66	SXIT	0.9995	147.28	EPS	0.9989	338.52
	5	SS	0.9999	16.03	SS	0.9995	138.51	CF0	0.9990	324.38
	6	CF0	0.9999	15.28	CF0	0.9996	131.34	PSIS	0.9990	323.85
	7	TV2	0.9999	15.24	PSIS	0.9996	131.16	TW	0.9990	323.39
$y_{DD}$	1	TD	0.5591	$6.17 \times 10^5$	TD	0.5451	$5.81 \times 10^5$	TD	0.5258	$5.71 \times 10^5$
	2	CW0	0.6692	$4.63 \times 10^5$	CW0	0.6768	$4.13 \times 10^5$	CW0	0.6719	$3.95 \times 10^5$
	3	CF0	0.7621	$3.33 \times 10^5$	CF0	0.7633	$3.02 \times 10^5$	CF0	0.7549	$2.96 \times 10^5$
	4	TW	0.8542	$2.04 \times 10^5$	TW	0.8484	$1.94 \times 10^5$	TW	0.8390	$1.94 \times 10^5$
	5	TF	0.9379	$0.87 \times 10^5$	TF	0.9250	$0.96 \times 10^5$	TF	0.9122	$1.06 \times 10^5$
	6	TG	0.9594	$0.57 \times 10^5$	CG0	0.9509	$0.63 \times 10^5$	CG0	0.9410	$0.71 \times 10^5$
	7	CG0	0.9812	$0.26 \times 10^5$	TG	0.9769	$0.29 \times 10^5$	TG	0.9700	$0.36 \times 10^5$
$y_{ASC}$	1	TSL	0.4304	$69.89 \times 10^6$	TSL	0.4381	$92.36 \times 10^6$	TSL	0.4301	$104.00 \times 10^6$
	2	CG0	0.5458	$55.75 \times 10^6$	CG0	0.5655	$71.45 \times 10^6$	CG0	0.5688	$78.58 \times 10^6$
	3	TG	0.6595	$41.81 \times 10^6$	TG	0.6904	$50.93 \times 10^6$	TG	0.7044	$53.90 \times 10^6$
	4	THG	0.7416	$31.74 \times 10^6$	THG	0.7796	$36.27 \times 10^6$	THG	0.8005	$36.38 \times 10^6$
	5	CW0	0.8159	$22.62 \times 10^6$	CW0	0.8511	$24.51 \times 10^6$	CW0	0.8666	$24.34 \times 10^6$
	6	TW	0.8792	$14.85 \times 10^6$	TW	0.9108	$14.69 \times 10^6$	TW	0.9192	$14.75 \times 10^6$
	7	THD	0.9178	$10.11 \times 10^6$	THD	0.9479	$8.59 \times 10^6$	THD	0.9529	$8.61 \times 10^6$

regression models gives very high  $R^2$  values indicating that most of the variabilities (changing between 93% and 99.9%, in this case) in the predictions of these three compartments are explained by their dependence on the first selected input factors (i.e., CF0 for  $y_{NWPT}(t)$ , CW0 for  $y_{WPT}(t)$  and CG0 for  $y_{GV}(t)$ ).

The stepwise procedures for  $y_{DD}(t)$  at three chosen years selected the same seven input factors, namely TD, CW0, CF0, TW, TF, TG and CG0, with only

one minor variation in the order of selection of the last two inputs in the regression model for  $y_{DD}(t = 1900)$ . With the addition of each input factor to the model the  $R^2$ -value increases gradually.

The three regression analyses for  $y_{ASC}(t)$  identified the same input factors TSL, CG0, TG, THG, CW0, TW and THD as significant inputs. The order of input selection is the same at all three years. The addition of each input to these three regression models appears to increase the  $R^2$ -values.

#### 4.4.5 Rank Transformation

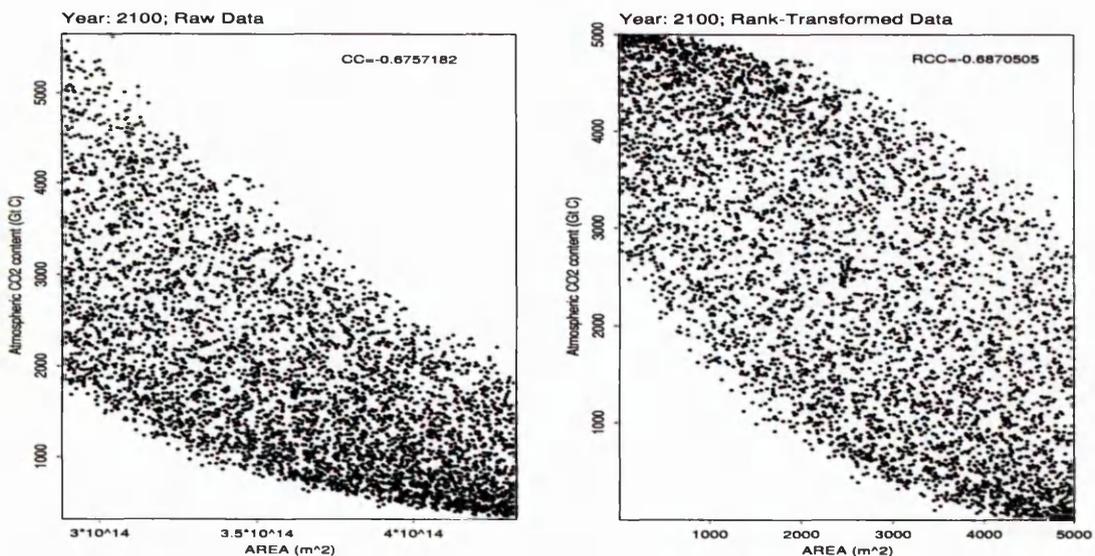
Since regression and correlation analyses are based on developing linear relationships between variables, regression and correlation based sensitivity analyses can perform poorly, if the relationships between the input and the output variables are non-linear but monotonic. In such cases, by using rank transformation the performance of these analyses can be improved and hence more reliable results can be obtained.

As seen in the two dimensional scatterplots, out of 30 input factors most of them have no relationship with the output variables and only a small number of them have a detectable linear association with the outputs. However, for the atmosphere and the three ocean compartments the relationships with AREA and HM may be identified as monotonic rather than linear (see Figures 4.18 and 4.19). As an example, we present in Figure 4.29 the scatterplots of  $y_{Atm}(t = 2100)$  versus AREA with both raw and rank-transformed data. Both AREA and HM show a slightly stronger linear relationship with  $y_{Atm}(t)$  and similarly with  $y_{SO}(t)$ ,  $y_{DO_5}(t)$  and  $y_{DO_{13}}(t)$  after the rank transformation, as it is revealed by both the examination of scatterplots and the computed Spearman's rank correlation coefficients (RCCs).

In Table 4.11, we present the analyses results for  $y_{Atm}(t = 2100)$ ,  $y_{SO}(t = 2100)$ ,  $y_{DO_5}(t = 2100)$  and  $y_{DO_{13}}(t = 2100)$  with CCs, SRCs and PCCs calculated with both raw and rank-transformed data. The five input factors with the largest CC values are included in the table, and the inputs are ordered by  $p$ -values for

CCs. For  $y_{Atm}(t = 2100)$ , the two analyses differ slightly in the importance order assigned to AREA and HM, but this difference appears to be with the CCs (calculated on both raw and rank-transformed data) only. The analysis with rank-transformed data identifies HM as the most important input with a RCC of -0.6887; on the other hand the analysis with raw data identifies HM as the second most important input with a CC of -0.6724. The rank-transformation does not appear to improve the results because a RCC of -0.6887 implies that HM can account for 47.4% of the uncertainty in  $y_{Atm}(t = 2100)$  in the rank-transformed space while a CC of -0.6724 implies that HM can account for 45.2% of the variability in  $y_{Atm}(t = 2100)$  in the non-transformed original space. Considering how close the CC, SRC, PCC, RCC, SRRC and PRCC values of AREA and HM are, it is reasonable to think that this difference in the importance ranking of these two input factors can simply be due to random sampling, and both of these inputs seem to be equally important for the atmosphere compartment.

The results on the use of raw and rank-transformed data in the analyses of



**Figure 4.29.** Scatterplot for the CO<sub>2</sub> content of Atmosphere compartment at year 2100 versus AREA with raw and rank transformed data.

$y_{SO}(t = 2100)$ ,  $y_{DO_5}(t = 2100)$  and  $y_{DO_{13}}(t = 2100)$  are quite similar in terms of estimated coefficients. Furthermore, the use of both raw and rank-transformed data in the analyses appear to lead to the same importance order between the most important input factors (see Table 4.11). As for the other response variables related to the terrestrial compartments, because there is either no relationship between the response variables and the input factors or the relationship is linear, use of raw and rank-transformed data produces almost identical results.

An additional perspective on the use of raw and rank-transformed data can be obtained by examining the results of stepwise regression analyses. Table 4.12 shows the analyses results for the output variables  $y_{Atm}$ ,  $y_{SO}$ ,  $y_{DO_5}$  and  $y_{DO_{13}}$  in 2100. In this table, the input factors are listed in order of selection in the analyses, the SRCs and SRRCs in final regression models, and the cumulative  $R^2$ -values with entry of each input into the regression models are given.

As seen in this table, the use of rank-transformed data leads to a regression model for  $y_{Atm}(t = 2100)$  with 7 input factors and an  $R^2$ -value of 97.02%; in contrast, the use of raw data leads to a regression model also with 7 inputs and an  $R^2$ -value of 93.28%. Thus, the use of rank-transformed data is resulting in an analysis that can account for slightly more of the variation in  $y_{Atm}(t = 2100)$  than can be accounted for in an analysis with raw data, but the difference is trivial. Similarly, for the other three output variables, the use of rank-transformed data does little to improve the quality of the analyses.

Table 4.11. Correlation Coefficients (CCs, RCCs), Standardized Regression Coefficients (SRCs, SRRCs) and Partial Correlation Coefficients (PCCs, PRCCs) with Raw and Rank-Transformed Data for  $y_{Atm}$ ,  $y_{SO}$ ,  $y_{DO5}$  and  $y_{DO13}$  in year 2100.

$y_{Atm}$											
Raw Data											
Input Factor	CC		SRC		PCC						
	p-value	Rank	Value	Rank	Value	Rank	Value				
AREA	0.000	1	-0.6757	1	-0.6894	1	-0.9372				
HM	0.000	2	-0.6724	2	-0.6839	2	-0.9363				
TSL	0.000	3	-0.0594	4	-0.0486	4	-0.1862				
TG	0.001	4	0.0467	6	0.0276	6	0.1072				
CSLO	0.004	5	0.0406	3	0.0498	3	0.1910				
Rank-Transformed Data											
Input Factor	RCC		SRRC		PRCC						
	p-value	Rank	Value	Rank	Value	Rank	Value				
HM	0.000	1	-0.6887	2	-0.7000	2	-0.9734				
AREA	0.000	2	-0.6871	1	-0.7001	1	-0.9733				
TSL	0.000	3	-0.0588	4	-0.0475	4	-0.2766				
TG	0.001	4	0.0473	6	0.0282	6	0.1687				
CSLO	0.006	5	0.0393	3	0.0482	3	0.2805				
$y_{SO}$											
Raw Data											
Input Factor	CC		SRC		PCC						
	p-value	Rank	Value	Rank	Value	Rank	Value				
AREA	0.000	1	0.6926	1	0.7042	1	0.9718				
HM	0.000	2	0.6845	2	0.7000	2	0.9715				
CSLO	0.006	3	0.0386	4	0.0326	4	0.1875				
TSL	0.133	4	-0.0211	3	-0.0333	3	-0.1914				
CGO	0.158	5	-0.0200	6	-0.0202	6	-0.1174				
Rank-Transformed Data											
Input Factor	RCC		SRRC		PRCC						
	p-value	Rank	Value	Rank	Value	Rank	Value				
AREA	0.000	1	0.6916	1	0.7022	1	0.9706				
HM	0.000	2	0.6860	2	0.7005	2	0.9705				
CSLO	0.005	3	0.0399	4	0.0329	4	0.1857				
TSL	0.120	4	-0.0219	3	-0.0339	3	-0.1914				
CGO	0.187	5	-0.0186	6	-0.0205	6	-0.1173				
$y_{DO5}$											
Raw Data											
Input Factor	CC		SRC		PCC						
	p-value	Rank	Value	Rank	Value	Rank	Value				
AREA	0.000	1	0.6847	1	0.6954	1	0.9524				
HM	0.000	2	0.6739	2	0.6901	2	0.9517				
CSLO	0.000	3	0.0552	4	0.0500	4	0.2194				
TSL	0.005	4	-0.0395	3	-0.0517	3	-0.2264				
CGO	0.028	5	-0.0307	5	-0.0309	5	-0.1376				
Rank-Transformed Data											
Input Factor	RCC		SRRC		PRCC						
	p-value	Rank	Value	Rank	Value	Rank	Value				
AREA	0.000	1	0.6881	1	0.6931	1	0.9510				
HM	0.000	2	0.6807	2	0.6909	2	0.9507				
CSLO	0.000	3	0.0615	4	0.0504	4	0.2118				
TSL	0.001	4	-0.0459	3	-0.0524	3	-0.2264				
CGO	0.024	5	-0.0319	5	-0.0313	5	-0.1376				
$y_{DO13}$											
Raw Data											
Input Factor	CC		SRC		PCC						
	p-value	Rank	Value	Rank	Value	Rank	Value				
AREA	0.000	1	0.6576	1	0.6661	1	0.8973				
HM	0.000	2	0.6412	2	0.6589	2	0.8955				
CSLO	0.000	3	0.0952	4	0.0922	4	0.2710				
TSL	0.000	4	-0.0843	3	-0.0963	3	-0.2820				
CGO	0.000	5	-0.0551	5	-0.0551	5	-0.1663				
Rank-Transformed Data											
Input Factor	RCC		SRRC		PRCC						
	p-value	Rank	Value	Rank	Value	Rank	Value				
AREA	0.000	1	0.6661	1	0.6632	1	0.8955				
HM	0.000	2	0.6524	2	0.6602	2	0.8947				
CSLO	0.000	3	0.1218	4	0.0925	4	0.2704				
TSL	0.000	4	-0.1130	3	-0.0969	3	-0.2821				
CGO	0.000	5	-0.0675	5	-0.0555	5	-0.1664				

**Table 4.12.** Comparison of Stepwise Regression Analyses with Raw and Rank-Transformed Data for  $y_{Atm}$ ,  $y_{SO}$ ,  $y_{DO5}$  and  $y_{DO13}$  in year 2100.

Compart.	Step	Raw Data			Rank-Transformed Data		
		Input Factor	SRC	R-squared	Input Factor	SRRC	R-squared
$y_{Atm}$	1	AREA	-0.6892	0.4565	HM	-0.6999	0.4743
	2	HM	-0.6834	0.9260	AREA	-0.6997	0.9634
	3	CSLO	0.0484	0.9283	CSLO	0.0468	0.9656
	4	TSL	-0.0487	0.9306	TSL	-0.0471	0.9677
	5	CGO	-0.0284	0.9314	CGO	-0.0322	0.9687
	6	TG	0.0280	0.9322	TG	0.0288	0.9695
	7	THG	-0.0237	0.9328	CL	0.0266	0.9702
$y_{SO}$	1	AREA	0.7048	0.4796	AREA	0.7033	0.4782
	2	HM	0.6994	0.9662	HM	0.7004	0.9660
	3	TSL	-0.0340	0.9674	TSL	-0.0356	0.9673
	4	CSLO	0.0320	0.9684	CSLO	0.0331	0.9684
	5	TG	0.0212	0.9688	TG	0.0224	0.9686
	6	CGO	-0.0210	0.9692	CL	-0.0216	0.9693
	7	CW0	-0.0186	0.9696	CGO	-0.0196	0.9697
$y_{DO5}$	1	AREA	0.6963	0.4689	AREA	0.6991	0.4733
	2	HM	0.6893	0.9401	HM	0.6958	0.9533
	3	TSL	-0.0520	0.9428	TSL	-0.0588	0.9567
	4	CSLO	0.0493	0.9451	CSLO	0.0555	0.9596
	5	CGO	-0.0324	0.9462	TG	0.0358	0.9609
	6	TG	0.0313	0.9471	CGO	-0.0335	0.9620
	7	CW0	-0.0282	0.9479	CW0	-0.0293	0.9629
$y_{DO13}$	1	AREA	0.6679	0.4330	AREA	0.6755	0.4442
	2	HM	0.6574	0.8587	HM	0.6692	0.8836
	3	TSL	-0.0962	0.8678	TSL	-0.1258	0.8991
	4	CSLO	0.0904	0.8757	CSLO	0.1171	0.9124
	5	CGO	-0.0567	0.8789	TG	0.0718	0.9175
	6	TG	0.0544	0.8818	CGO	-0.0699	0.9224
	7	CW0	-0.0533	0.8847	CW0	-0.0657	0.9267

#### 4.4.6 Two-sample Nonparametric Tests

As described in Section 2.6 of Chapter 2, two-sample test statistics of the Smirnov test and Cramér-von Mises test have been used as measures of sensitivity. In the following subsections we present the SA results based on these two tests.

The test statistics and the sensitivity rankings based on these statistics can vary considerably depending on the choice of the quantile for splitting the samples. In order to fully study whether an input factor has more influence on the

median (indicating a general input importance) or on the 90th quantile (indicating a greater influence on the extremes), we have considered both the median and the 90th quantiles of each output distribution when partitioning each input factor under consideration into two samples.

#### 4.4.6.1 Smirnov Test

Considering the random sample of 5,000 runs, the sensitivities of the nine chosen compartments in years 1900, 2000 and 2100 based on the Smirnov test statistic are calculated and their corresponding ranks are presented in Table 4.13 (where rank 1 is assigned to the most important input factor, rank 2 to the second most important one, and so on). Only the top 10 most important inputs are specified in Table 4.13. For the sake of comparing rankings from both test results (with the two choices of quantiles), we present the rankings in the same table with the ranks for the 90th quantiles given in parentheses. As seen in this table, for each compartment, the order of importance between the input factors show some variation from year to year. For the same output variable, the rankings based on the two partitioned distribution methods also differ considerably for all compartments except for the detritus/decomposers and the active soil carbon compartments for which there is a reasonably good agreement between the rankings resulting from the two partitioned distribution methods.

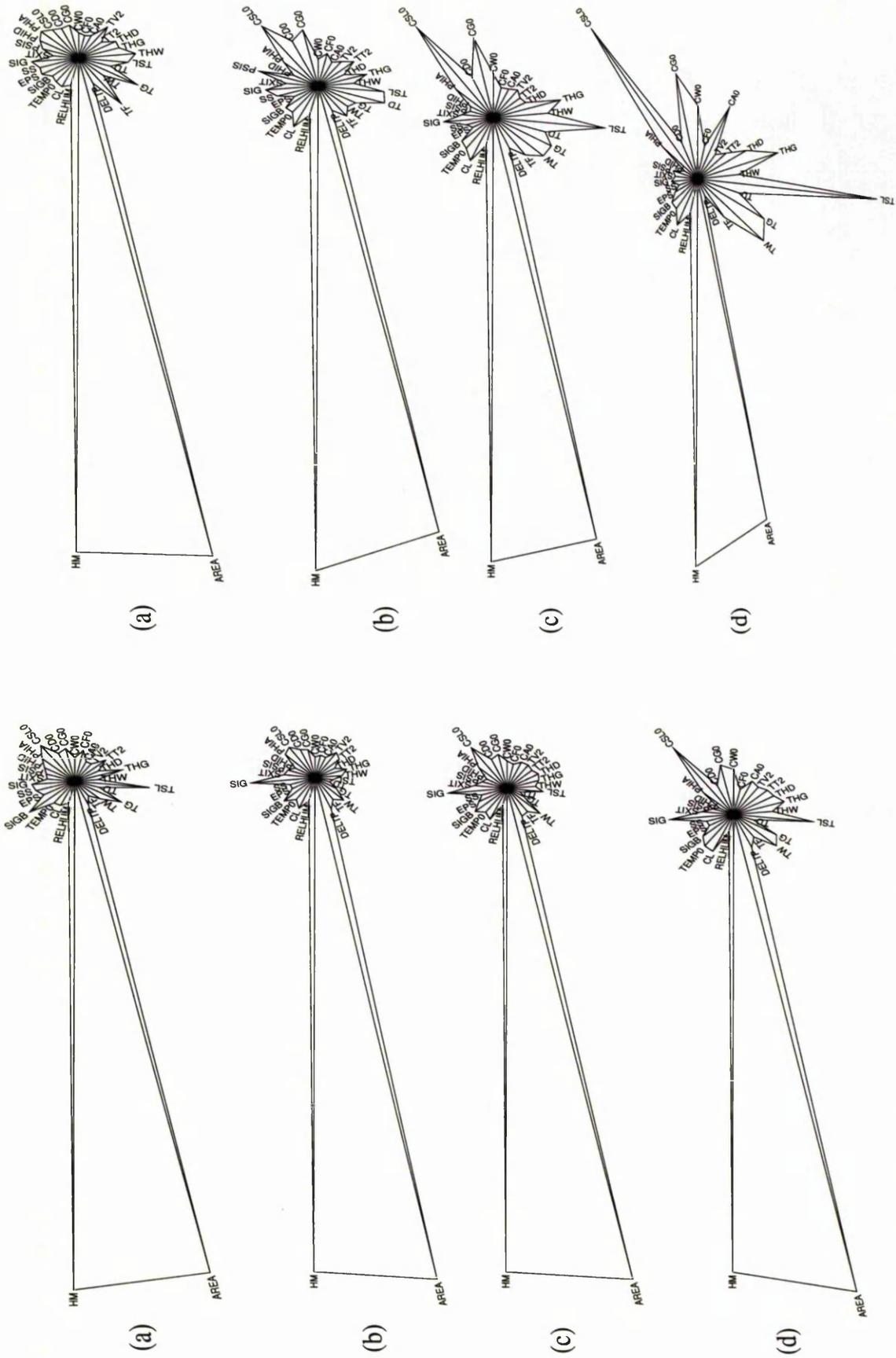


#### 4.4.6.2 Cramér-von Mises Test

Using the same simulation of 5,000 runs the test statistics are calculated. The order of importance obtained using this method is in reasonably good agreement with the results based on the Smirnov test. As noted in Section 2.6 of Chapter 2, the Cramér-von Mises and Smirnov statistics resemble each other very closely; however, the test statistic for the Cramér-von Mises scans the total area enclosed by the two cumulative distributions where as the Smirnov statistic is defined as the maximum vertical distance between the two curves. If the output variable is a non-monotonic function of the input, then the rankings provided by the two methods can be quite different, but in such cases the use of the Cramér-von Mises statistic as a sensitivity measure is more appropriate. In our case both of these methods seem to give reasonably consistent results.

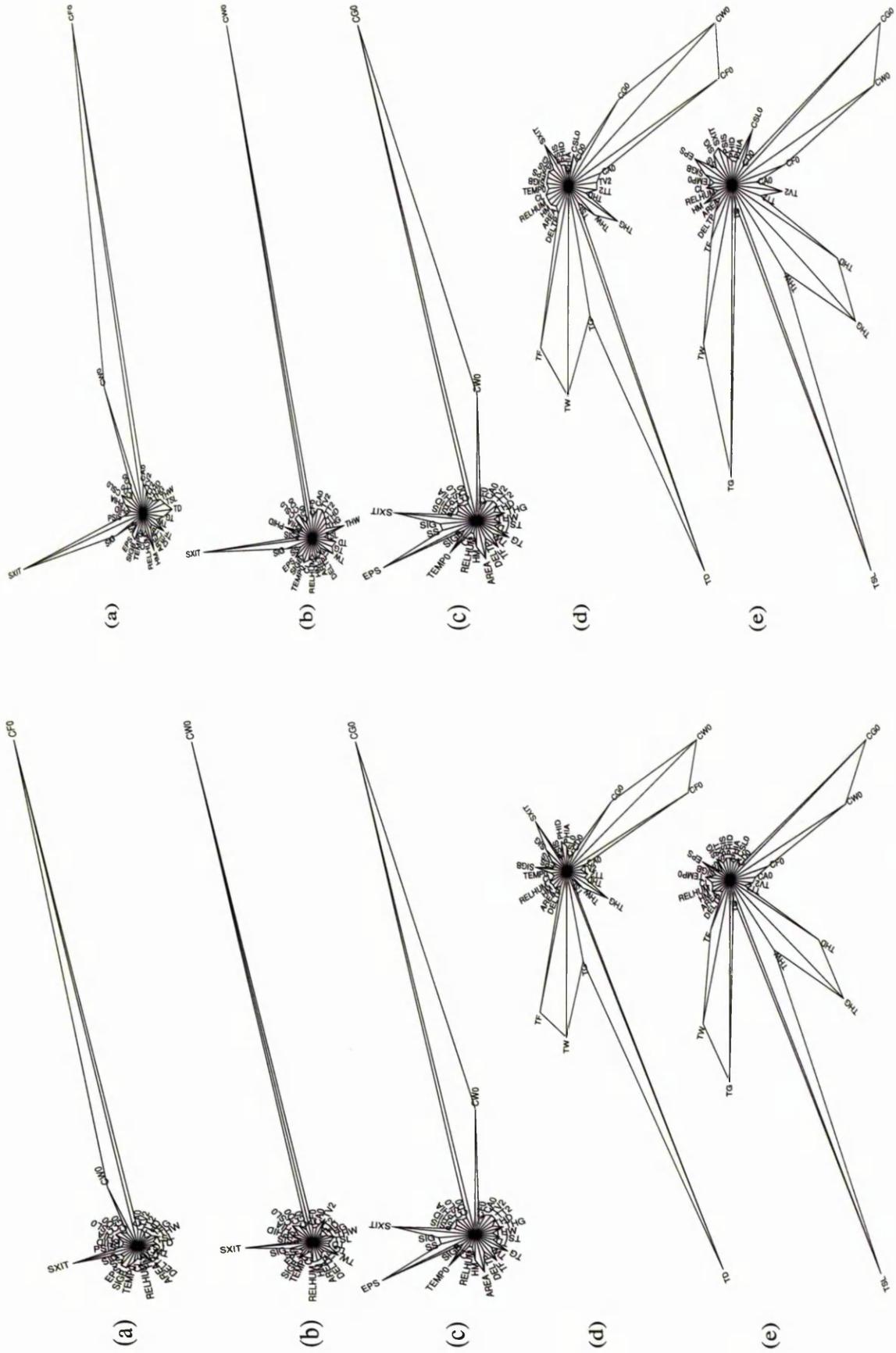
Here, instead of ranking the Cramér-von Mises test statistics, we present the results (from only year 2100) graphically using star plots (Figures 4.30 and 4.31) which allow us to compare the influences of the input factors on each compartmental output. The radius of the stars extending from the centre of a circle represent a test statistic value computed on an input and an output variable. A long radius indicates that the corresponding input factor is important for the output variable under consideration. This graphical display proves to be a good way of picking out the most important input factors at a glance.

In Figure 4.30, which shows the influence of the 30 input factors on the outputs of the atmosphere and the three ocean compartments in 2100 in terms of the Cramér-von Mises test statistic based on both the 50th and the 90th quantile partitioning, it is clear that the input factors AREA and HM have quite a significant influence on these four compartments. Although the degree of importance of these two inputs is very close, the relative importance between these two most important inputs change with the two quantiles considered. For instance, for the surface ocean compartment with the 50th quantile partitioning AREA appears to be the most important and HM the second important input, but with the 90th quantile partitioning this order is the other way around.



**Figure 4.30.** Star plots of Cramér-von Mises test statistics for the output of: (a) Atmosphere; (b) Surface Ocean; (c) Deep Ocean-layer5; and (d) Deep Ocean-layer13 compartments at  $t = 2100$ . The star plots on the left are based on the 90th quantile and the star plots on the right are based on the 50th quantile partitioning.

Compared to AREA and HM, the rest of the input factors do not appear to have much influence on the atmosphere and the three chosen ocean compartments (especially when the partitioning is based on the median). Figure 4.30(a) shows that based on the 50th quantile partitioning the input factors TSL is the third and SIGB the fourth most influential inputs; and based on the 90th quantile partitioning TF is the third and TG the fourth most influential inputs, following AREA and HM. Similarly, an importance ranking of the input factors for the other output variables related to the other compartments can be obtained by examining their star plots.



**Figure 4.31.** Star plots of Cramér-von Mises test statistics for the output of: (a) Nonwoody parts of trees; (b) Woody parts of trees; (c) Ground vegetation; (d) Detritus/Decomposers; and (e) Active soil carbon compartments at  $t = 2100$ . The star plots on the left are based on the 50th quantile and the star plots on the right are based on the 90th quantile partitioning results.

## 4.5 Discussion on the Results

When the models under study are computationally expensive to run and have a large number of input factors, the choice of the SA method is restricted to the techniques that require a relatively small number of model runs, i.e. are computationally not very expensive. As noted by Campolongo *et al.* (see [15]), the computational cost is determined by the number of model evaluations which is a function of the number of input factors considered, the number of model outputs examined and the complexity of the model. In addition, there is the number of input scenarios, and also output times to be taken into account when model outputs are time-dependent. In this chapter, with the model under consideration the number of model outputs (i.e. the number of compartments) is 25, the number of time points is 350 years (from 1750 to 2100), only one of the IPCC emission scenarios (IS92a) is considered, the number of input factors is 30, and  $N=5,000$  runs is considered.

We started the analysis with the simplest parametric approach, standard OAT design, that proceeds by varying only one input factor at a time while holding all other factors fixed at some nominal value. Two different standard OAT methods (SIs and SRs) are applied to the model. Both of these methods are local SA methods and like any local SA technique the information they provide is limited. Neither of these local methods is computationally very cheap. The calculation of the SIs were based on  $2k$  model evaluations where  $k$  is the number of inputs. For the calculation of SRs, the computational cost is more expensive; it requires  $k \times N$  model evaluations, where  $N$  is the sample size considered in the analysis. Both  $N=100$  and  $N=5,000$  is used for the calculation of SRs, and even though the graphical summaries showed that the results hardly change, the ranking of the inputs changed considerably. Therefore, we have presented the results based on  $N=5,000$  model evaluations which assures a better coverage of the input space.

The Morris method, which again changes one factor at a time but is considered as a global method because it explores the entire input factor space, was also used to assess the sensitivity of the model outputs. Compared to the standard OAT

techniques, the information produced by the Morris method is more general. The computational cost with this method is  $r \times (k + 1)$ , where  $k$  again is the number of inputs and  $r$  is the number of orientation matrices which was taken to be 10.

Note that all three screening exercises used here only provide sensitivity measures that are qualitative and were used to rank the inputs in order of their importance. In practice, a screening method allows us to identify a subset of the most important model input factors. Afterwards, we can apply a quantitative method to the subset of preselected inputs. By doing so, we can sometimes reduce the computational cost of the experiment considerably, especially when only a few of the input factors have a significant effect on the model output(s). Since the number of uncertain input factors of this GCC model under consideration in this chapter is relatively small, we have decided not to exclude the input factors, which were identified as unimportant by the three screening designs, in the application of the global SA methods.

The results from the sensitivity measurements based on the three screening designs (SIs, SRs and Morris) show that all these screening methods reveal the same subset of the most important input factors for the same output variable, with minor differences in their importance rankings when the detritus/decomposers and the active soil carbon compartments are concerned.

Although the global SA methods require higher number of model evaluations, they have been used in order to obtain quantitative sensitivity measures, and their use is recommended by researchers like Saltelli, Campolongo, Iman, Helton and Conover (see [47] for a review).

As the first global SA method we produced and examined scatterplots of the model predictions, in order to assess the nature of the relationship between each model output variable and each input factor. These scatterplots did not show any complex patterns. However, because the dimension of the input factor space is quite high (30), we have to keep in mind that these scatterplots can be misleading since the structure present in the original 30-dimensional space is not necessarily

reflected by the individual structures present in the scatterplots of pairs of input-output variables. On the other hand, if a small number of input factors play a relatively dominant role compared to the other inputs (like in our case especially with the atmosphere and the ocean compartments) the corresponding scatterplots can be of value.

Other global SA measures such as CCs, SRCs, PCCs; and also non-parametric measures such as SRRCs, PRCCs (which are based upon the ranks of the input and output values) were calculated, and the absolute value of these estimates were used to rank the factors in order of influence on the outputs. Using stepwise regression approach for each compartment in three chosen years, the most important input factor sets were obtained and model coefficient of determination ( $R^2$ ) values were calculated. These high  $R^2$ -values suggested that the output variables under consideration are reproduced by the linear regression models very well.

We shall emphasize that less than 10 out of the 30 input factors were found to have sizeable influence on the output variables under consideration. In this discussion section, we summarize the results by focussing on the inputs that were identified as important by the stepwise regression model, but when we presented the results earlier in the chapter - although it is an arbitrary choice - we assessed the influence of the top 10 input factors on the outputs in order not to fail to identify a factor which is important.

The subset of the most important input factors, which account for at least 5% of the variability in the output variable, consists of

- AREA and HM for the atmosphere and the three ocean compartments in the three chosen years;
- CF0 for the nonwoody parts of trees compartment in all three years;
- CW0 for the woody parts of trees compartment in all three years;
- CG0 for the ground vegetation compartment in all three years;

- TD, CW0, CF0, TW and TF for the detritus/decomposers compartment in all three years;
- TSL, CG0, THG, TG, TW and CW0 for the active soil carbon compartment in all three chosen year.

Note that the order of importance between these important input factors changes slightly from year to year and from method to method; however, all SA techniques considered here have identified these input factors. There is some variation in the importance order of the rest of the input factors under different methods, but these factors appear to have small effects on the total uncertainty in the output variables under consideration.

Because of monotonic but slightly non-linear relationship detected in the scatterplots of AREA versus each of the output variables related to the atmosphere and the three ocean compartments (similar relationship also found between the same output variables and HM), rank transformation is considered. The results of the analysis with rank-transformed data indicated that since the use of rank transformation improved the performance of the analysis very little, this approach is not really worth being pursued for the model under consideration here.

Input factor importance for each compartment found to be time dependent.

Because of the random sampling, different rankings can be obtained from different simulations. In this chapter, we have used the same set of input and output values in the calculations of different global SA measures so that the rankings provided by these methods can be comparable. For a future study, different simulations (the seed used for the random number generation is changed each time) can be taken into account to compare the variances of the SA estimator prediction over the various simulations.

# Chapter 5

## Uncertainty Analysis

### 5.1 Introduction

The real-world system is precise but complex; a model, which is a simplified representation of this system, may be imprecise but simple. This trade-off between precision and simplicity is the main essence of the modelling process. The effectiveness of a model lies in its simplicity of use, as well as an understanding of the level of imprecision [92]. Because of the imprecision in the model, uncertainties exist in the conclusions derived from the model.

Perincherry *et al.* [92] point out that “Handling uncertainty is perhaps the most pervasive and the most difficult aspect in the analysis of systems”. Some investigators may have a tendency to ignore the uncertainties since it simplifies the decision making process, but ignoring uncertainties may be disastrous. It is a duty of the researchers to highlight the uncertainties associated with the inferences that they make.

In practice there is a wide variety of sources of uncertainty involved in the modelling process, and some of the important ones arising in climate models are shown in Figure 5.1. Uncertainty analysis (UA) plays an important role in estimating the reliability of climate model projections [43]. In this chapter, focusing

on the three main sources of uncertainty - namely input factor, model and scenario uncertainty - and using computer simulation we apply this framework to global carbon cycle (GCC) models. The application of UA to GCC models can make significant contribution to the understanding of the GCC and its role in determining future atmospheric CO<sub>2</sub> concentrations.

Another important source of uncertainty is the modeller. Different modellers can draw different conclusions from the same output, depending on the context of the problem and the attitude of the modeller. Therefore, the resulting model can depend largely upon a modeller. We refer to this type of uncertainty as '*modeller uncertainty*'. This aspect of uncertainty is also illustrated in this chapter with a case study.

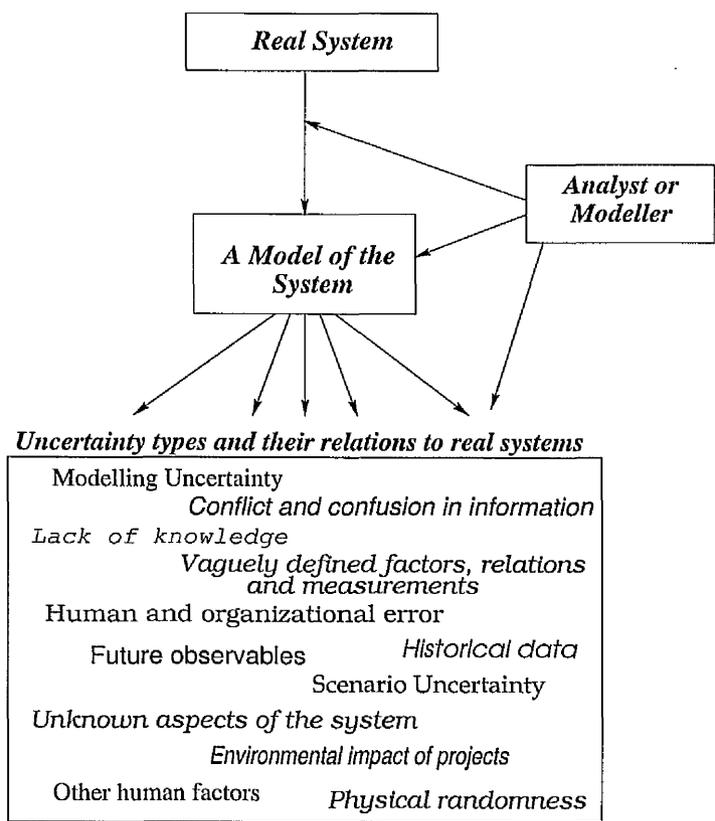


Figure 5.1. A diagram of sources of uncertainty.

To determine the effect the sources of uncertainty have on model results, sensitivity analysis (SA) is used. In the previous two chapters (see Chapters 3 and 4), to assess how sensitive the output of a model can be to variation in one or a set of model input factors, various SA methods have been applied to the three GCC models under specific settings. Those settings being: within each model each input factor is assumed to follow a uniform distribution; IPCC's IS92a emission scenario is used for forecasting CO<sub>2</sub> content of each compartment; as sampling method simple random sampling is used.

At the SA stage of this study, for the three GCC models under consideration, we have determined which of the input factors are important and hence which cannot be ignored in future investigations. It should be noted that in the SA chapters we focused only on uncertainties in the model input factors and assessed their influence on model outputs within each model.

As pointed out by McKay, it is necessary to validate a model before prediction uncertainty can be evaluated sensibly [83]. Hence, in this chapter, we first in Section 5.2 do model intercomparison and model validation, that is compare the historical atmospheric CO<sub>2</sub> predictions of the three GCC models with each other and with the historical CO<sub>2</sub> record, to determine the uncertainty in the results. Then in Section 5.3, we talk about type of uncertainties involved in GCC models. In Section 5.4 three main sources of uncertainty we focus on in this chapter are explained. These sources of uncertainties: input factor uncertainty, scenario uncertainty and model structure uncertainty are investigated in Sections 5.5 through 5.7, respectively. In Section 5.8, partitioning the uncertainty in model predictions between the sources of uncertainty is discussed. Section 5.9 introduces modeller uncertainty and presents results from a case study in environmental radioactivity.

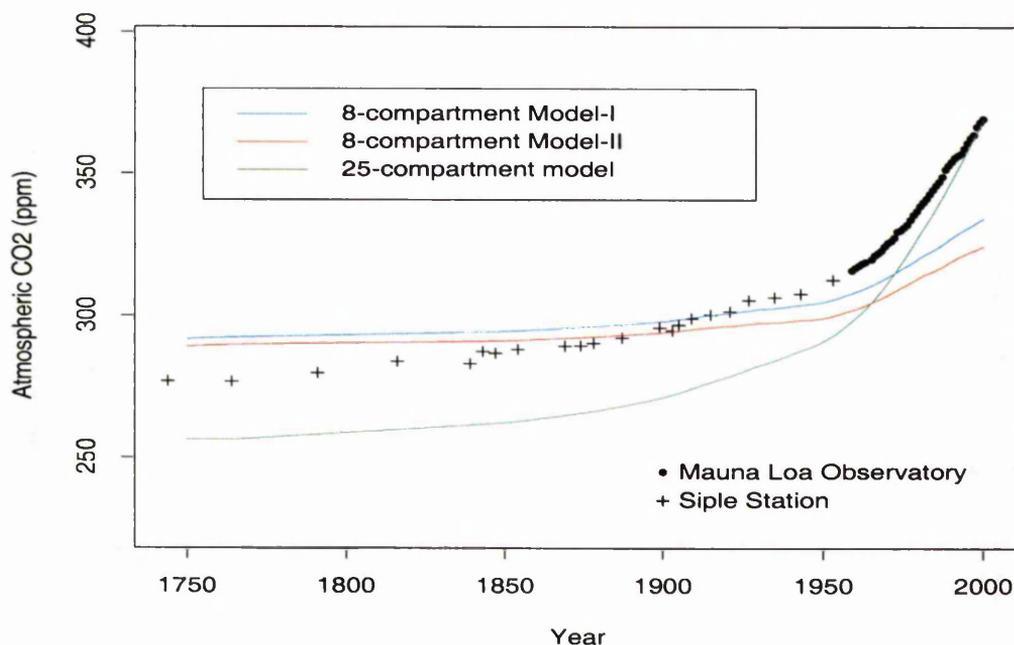
## 5.2 Model Intercomparison and Model Validation

Model intercomparison and model validation are essential components in the modelling process for establishing the reliability of models, because until the disagreements can be identified, nobody will know which model, if any, can be believed [43]. However, the disagreement among the predictions of different models and the degree of agreement with available observations is an important aspect of determining the uncertainty in the results.

Due to different assumptions, simplifications, approximations *etc.* made in developing different models, there can be -and usually there are- disagreements between model predictions. The process of analysing the disagreements among the results of different models is called *model intercomparison*. As Hall notes in Ref. [43], model intercomparison has an important role because it is essential to reconcile the disagreements between different models before we can have confidence in them, but it is UA that provides the basis for a meaningful model comparison.

The process of comparing model predictions with observational data to evaluate their accuracy is called *model validation*. Even though validation against observations provides one of the most convincing indications of the reliability of a model, it is important to emphasize that: 1. there is always some degree of uncertainty in the observed data; 2. a model can perform well with the past data but this does not necessarily mean that its future projections would be reliable as well.

The three GCC models simulated atmospheric CO<sub>2</sub> over the historical period 1750-2000, using the input factor values listed in Tables 3.2 - 3.5 in Chapter 3 and Table 4.1 in Chapter 4. These yearly atmospheric CO<sub>2</sub> predictions are shown in Figure 5.2 along with the historical measurements of atmospheric CO<sub>2</sub> concentrations from the Mauna Loa Observatory for the period 1958-2000 and from an ice core taken at Siple Station in Antarctica for the period 1744-1953.



**Figure 5.2.** Comparison of historical simulations of atmospheric CO<sub>2</sub> from three GCC models under the base-case scenario and the measured atmospheric CO<sub>2</sub> concentrations from Mauna Loa Observatory: 1958-2000 (Keeling & Whorf, 2001 [71]); and Siple Station: 1744-1953 (Friedli *et al.*, 1986 [37]).

The predictions of the 8-compartment models (Model I and Model II) bear strong similarities to each other. This is due largely to the similarity in the mathematical structure of these two models. The estimates from these two models compared to the 25-compartment model agrees rather well with the observed data in the first 200-220 years. Both of these models overestimate the Siple ice core data; Model I until around early 1900s and Model II until around 1870. Then, both fall below the historical data. Among the three models, Model II produces predictions closest to the observed data for the first 150 years, between years 1900 and 1970 Model I performs best, and for the most recent three decades the 25-compartment model predictions appear to be closest to the Mauna Loa data. The model predictions from the 25-compartment model fall below the historical

data and the other model estimates until around 1960s, and in the early 1970s this model's prediction rises above the other two models and agrees quite well with the Mauna Loa record.

For the 25-compartment model, the greatest deviation from the observed data occurs in 1843 for which the historical prediction underestimates the Siple ice core record by 25.5 ppm. As for the 8-compartment models, a maximum deviation of 35 ppm for Model I and 41.7 ppm for Model II occurs in year 2000.

From the comparison of model predictions with each other and with the observed data, it is clear that there is considerable uncertainty due to modelling and parametrization within model. Both of these sources of uncertainty are defined in Section 5.4.

Comparison of the model predictions with each other and with historical data is instructive and by doing so models can be tuned to simulate observed data well, but this does not necessarily mean that they will be capable of producing future projections well since the consequences of climate change are not known. The uncertainty in the future projections is discussed later in the chapter under scenario uncertainty.

Because our aim in this chapter is neither removing the disagreement between the models nor tuning the models so that they can simulate available data well, we only used model intercomparison and validation approach to show that there are uncertainties attached to the model results.

### 5.3 Uncertainties in GCC Models

The inherent uncertainty associated with most environmental and climatic systems is often acknowledged [109]. There is considerable uncertainty about the future role of the terrestrial and oceanic systems in the global carbon cycle. Concerning the terrestrial system, there is still considerable disagreement among scientists as to whether terrestrial vegetation is a source or a sink for CO<sub>2</sub>. Even

the amount of carbon stored in terrestrial system is not certain. Carbon estimated in plants range widely between 420 and 830 Gigatons ( $1Gt = 10^{12}kg$ ), depending on the methods used.

The oceans have a very high  $CO_2$  storage capacity, but lack of knowledge of the overall effect on the ocean environment and uncertainty about the injected  $CO_2$  into the oceans still exist.

UA of GCC models can identify carbon cycle components and processes with the greatest sensitivities and uncertainties. This information can then be used to determine which uncertainties have the greatest influence on future atmospheric  $CO_2$  concentrations and where further research and data collection could be most effectively applied to reduce uncertainties. Information gathered from UA can also be used in future model development.

## 5.4 Main Sources of Uncertainty

In a general sense, Draper identifies three main sources of uncertainty in any problem to be: (i) predictive uncertainty which is conditional on the scenario and model; (ii) scenario uncertainty about the inputs to the models; and (iii) model uncertainty (conditional on the scenario) about how to translate the inputs into forecasts (see [26]).

In another article Draper *et al.* [27] note that sources of uncertainty in complex prediction problems involve six ingredients: past data, future observables, scenarios, model (or structural), parametric, and predictive uncertainty.

Uncertainties in computer models can arise from a variety of different sources and attention has been devoted to examining the magnitude of uncertainty associated with model behavior. According to Chatfield [17], as modern computing power allows us to consider and compare increasingly large number of models, the problem of dealing with uncertainties in models is becoming increasingly serious. This issue is also emphasized by Kennedy & O'Hagan in Ref. [72]. They note that the widespread application of computer models brings together a widespread

concern about quantifying uncertainties attached to the model results. They define various sources of uncertainty in computer models which include parameter uncertainty, model inadequacy, residual variability, parametric variability, code uncertainty and observation error.

In the context of climate models, three main sources of uncertainty stated by King & Sale [73] are:

1. uncertainties in future energy and land-use emissions (i.e. *scenario uncertainty*);
2. uncertainties about the GCC reflected in the structural and conceptual differences between models (i.e. *model uncertainty*); and
3. measurement error and uncertainty in the parameters and variables within a particular model (i.e. *input factor uncertainty*).

In this chapter, our aim is to evaluate uncertainty in model predictions arising from these three main sources of uncertainty. First, we shall briefly describe these sources of uncertainty in the context of GCC modelling:

**Input Factor Uncertainty** can be defined as measurement error and uncertainty in the parameters and variables within a particular model.

**Scenario Uncertainty** can be defined as the uncertainties in future energy and land-use emissions.

**Model Uncertainty** can be defined as the uncertainties about the GCC reflected in the structural and conceptual differences between models.

## 5.5 Input Factor Uncertainty

In any model, the output from the model is the item of interest. The knowledge we have about the model input factors driving the model equations is not perfect. As a result, these input factors are described as being uncertain. Because of

uncertainties inherent in the input there are uncertainties attached to the output. As stated by Helton & Davis [49], uncertainty analysis can be defined as attempts to answer the question of “What is the uncertainty in the model response given the uncertainty in the input factors?”.

The values of model input factors can only be determined from the real world system to a certain extent, that is, there is always some degree of uncertainty about those values.

This type of uncertainty is quantified in a distribution of input factor values. Reducing uncertainty in influential factor(s) of a model, which can be identified through SA, can lead to significant reduction of uncertainty in model predictions. In Figure 5.3 we present an example showing how reduction in an influential input factor uncertainty yields a reduction in prediction uncertainty. This example is based on the atmosphere compartment of the 25-compartment model used in Chapter 4. In that chapter, we found that the area of the surface ocean (i.e. the input factor AREA) is the most important input factor for the atmosphere compartment. Figure 5.3 shows the substantial effect of reducing the uncertainty

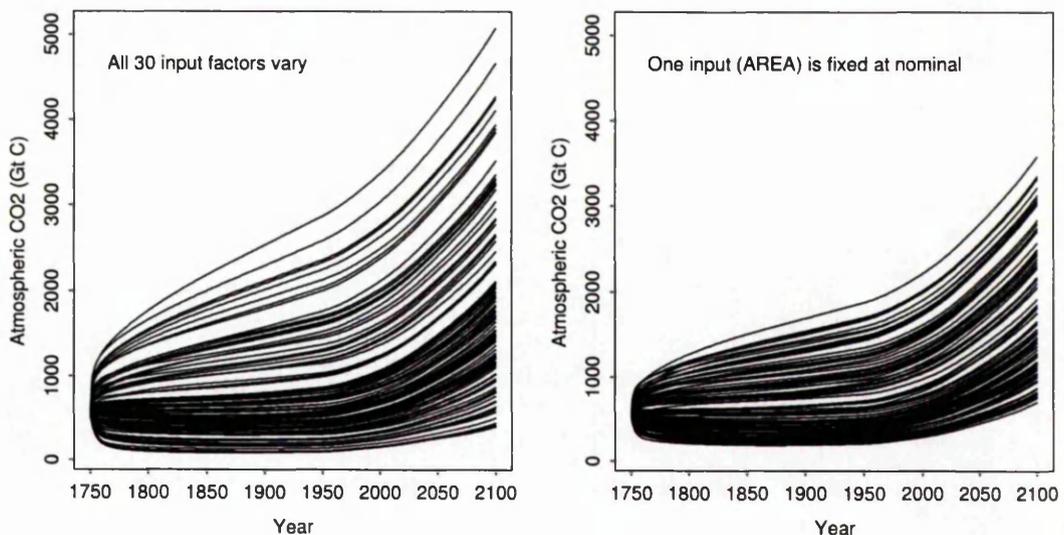


Figure 5.3. A simple example of prediction uncertainty bands.

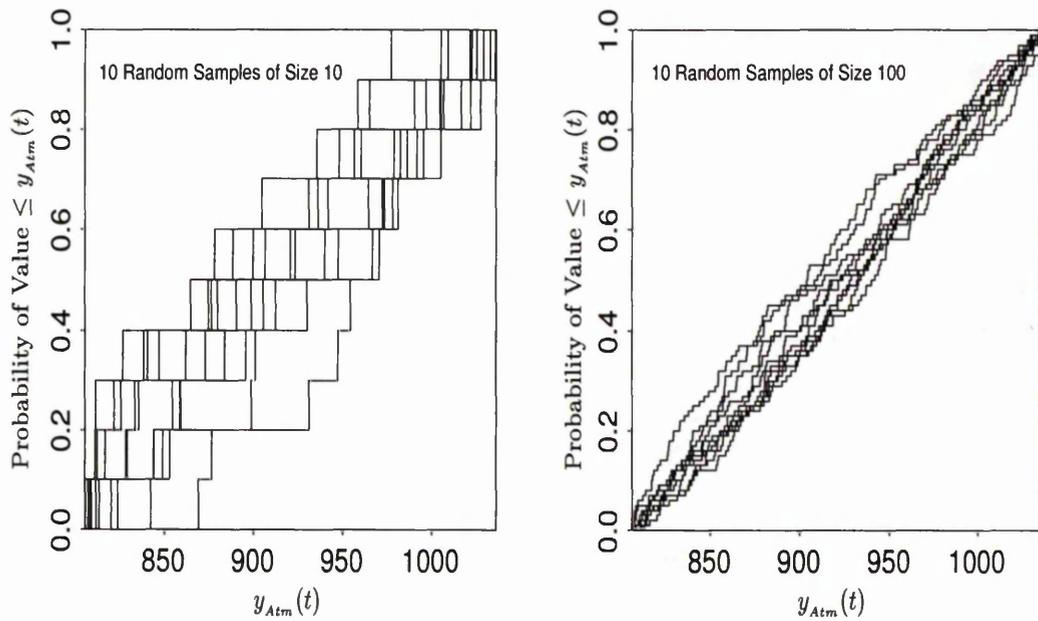
in AREA (i.e. setting it to its best estimate) on the uncertainty in the predicted CO<sub>2</sub> content of the atmosphere.

This type of uncertainty is of major interest in Chapters 3 and 4. In these chapters we examined sensitivity of model outputs to the uncertainties inherent in the model input factors. In this chapter, we investigate two different aspects of input factor uncertainty. First, the effect of sample size, and the sampling technique used to obtain input vectors, on model predictions. Then, the effect of different input factor distributions on model predictions.

### 5.5.1 Effect of Sample Size and Sampling Technique on Model Predictions

In uncertainty studies of computer models, it is often necessary to use a limited number of model runs, especially when dealing with complex models and/or models with large number of input factors and output variables. For this purpose, different sampling techniques have been developed and used.

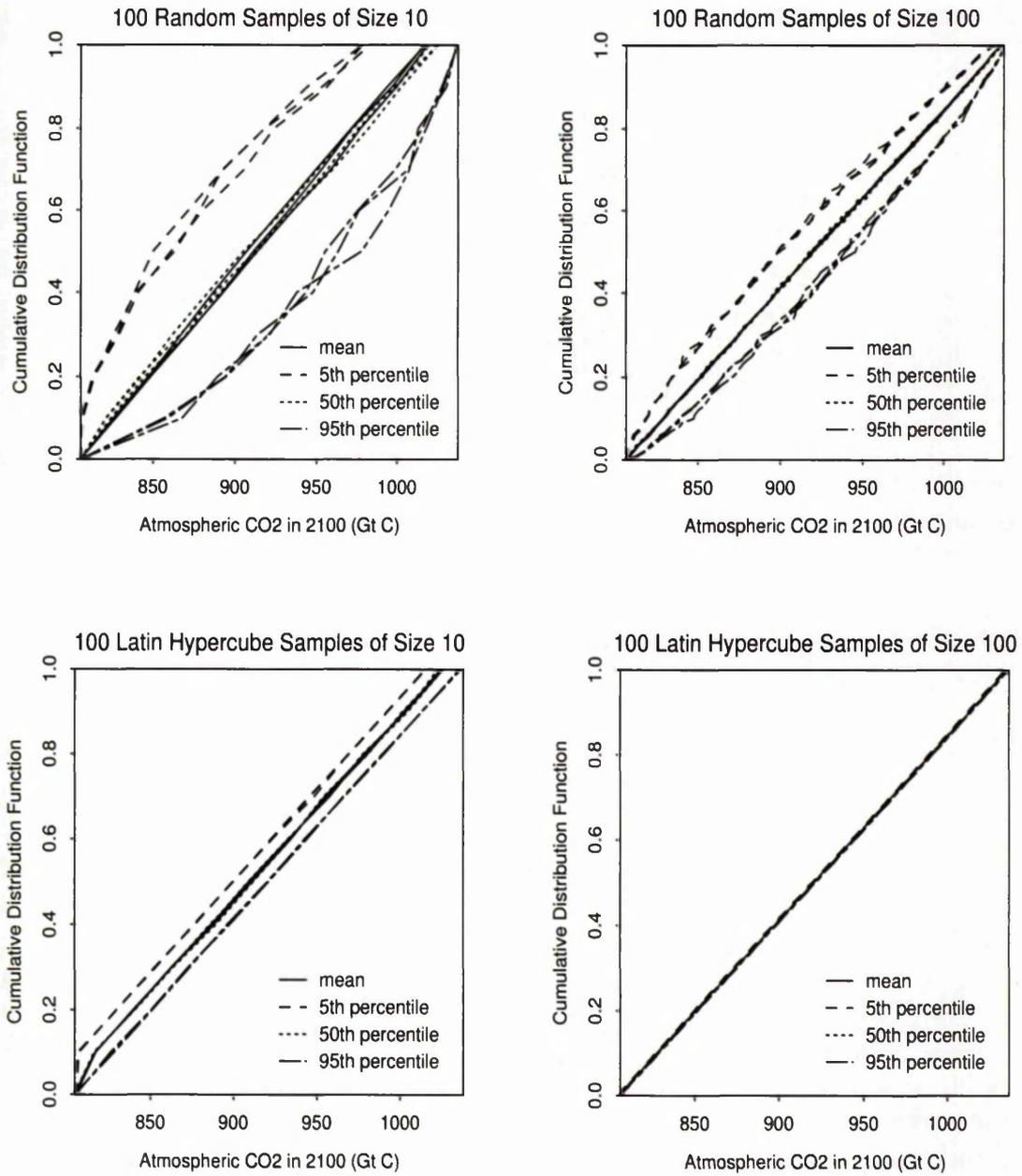
In this subsection we investigate the effect of sample size and sampling technique on the model predictions. The sampling techniques considered here are simple random sampling (SRS) and Latin hypercube sampling (LHS). First, to illustrate the effect of sample size on the variability in model predictions, considering SRS we generate 10 input samples of size  $N=10$  and  $N=100$  for the initial conditions of Model I under the assumption that they follow a uniform distribution on their assigned ranges given in Table 3.2 (see Chapter 3). For examining such effects we use cumulative distribution functions (CDFs). The CDFs for the atmospheric CO<sub>2</sub> predictions in 2100 (i.e. the output variable  $y_{Atm}(t = 2100)$ ) are shown in Figure 5.4. Examination of this figure shows that with increased sample size, it is reasonable to expect improvement in the results, that is, as the sample size increases the estimated CDF converges to the true CDF. Helton & Davis [48] use the term '*stability*' to refer to the amount of variation between results obtained with different samples generated by a particular sampling technique under



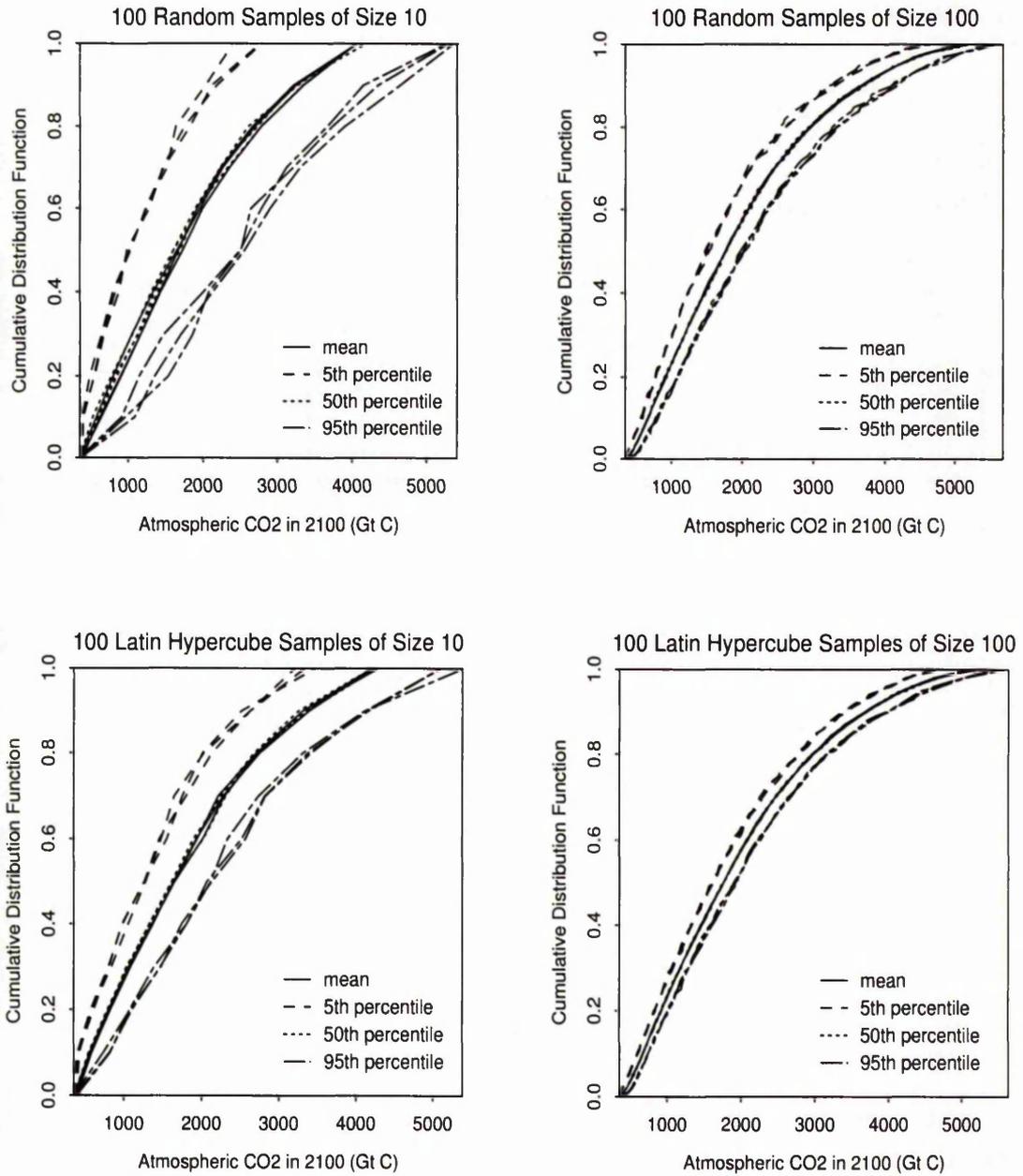
**Figure 5.4.** Example CDFs for  $y_{Atm}(t)$  from Model I estimated with random samples of size 10 and 100 under the assumption that the initial conditions of the compartments are uniformly distributed on their assigned ranges (see Table 3.2).

consideration. We shall adapt their terminology here.

For one of the 8-compartment models, Model I, and the 25-compartment model, using both SRS and LHS we present plots similar to those in Figure 5.4 for  $N=100$ . Because showing them all on the same figure is not very informative, we have summarised the distributions of the CDFs with mean and percentile curves. For a given probability value, the mean and the 5th, 50th and 95th percentiles of the corresponding  $y_{Atm}(t = 2100)$  were calculated. The results are given in Figures 5.5 and 5.6. The location of the percentile curves shows how stable the estimates of the CDFs are. This analysis is repeated three times to give three estimates of the mean and percentile curves. It is clear in Figures 5.5 and 5.6 that LHS is producing CDF estimates that are more stable than those produced with SRS.



**Figure 5.5.** Summary of distribution of CDFs for Atmospheric CO<sub>2</sub> predictions in 2100 from one of the 8-compartment GCC model (Model I) estimated with 3 replications of 100 simple random samples and 3 replications of 100 Latin hypercube samples of size 10 and 100 under the assumption that the initial conditions are uniformly distributed on their assigned uncertainty ranges (for the initial conditions and their assigned ranges see Table 3.2).



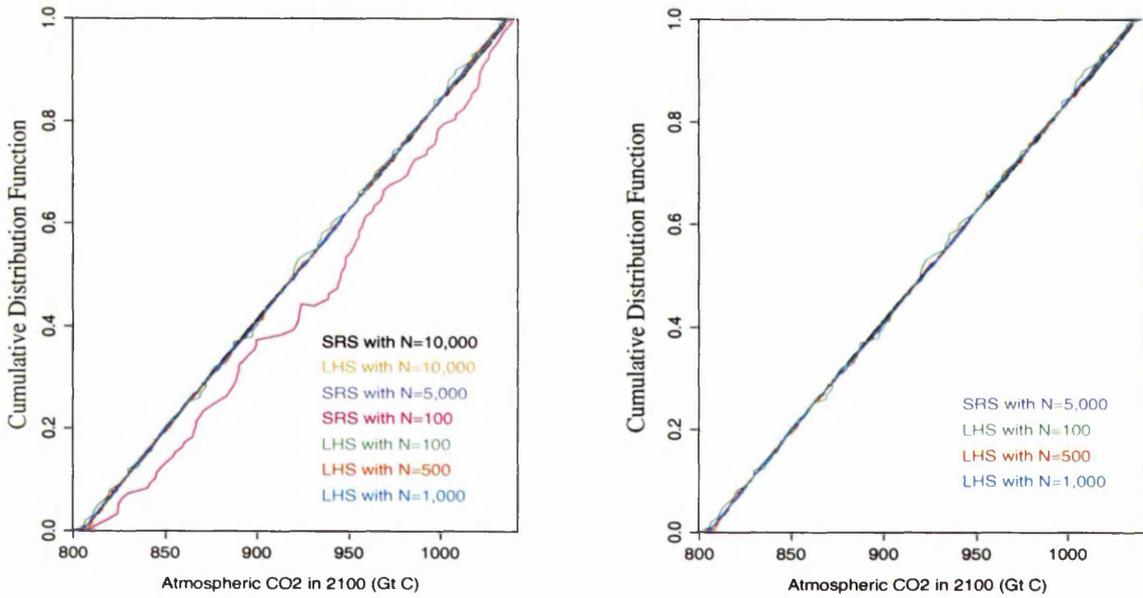
**Figure 5.6.** Summary of distribution of CDFs for  $y_{Atm}(t = 2100)$  from the 25-compartment GCC model estimated with 3 replications of 100 simple random samples and 3 replications of 100 Latin hypercube samples of size 10 and 100 under the assumption that the model input factors are uniformly distributed on their assigned uncertainty ranges (for the input factors and their assigned ranges see Table 4.1).

In SA chapters, because the three model codes are not computationally expensive to evaluate, and because SRS is easy to implement and easy to explain we have considered this sampling technique. Depending on the dimension of the input space (i.e. the number of input factors), to obtain a better coverage of the input space one can either use SRS with a sufficiently large sample size when large samples are computationally practicable or use LHS with a smaller sample size when large samples are not computationally practicable.

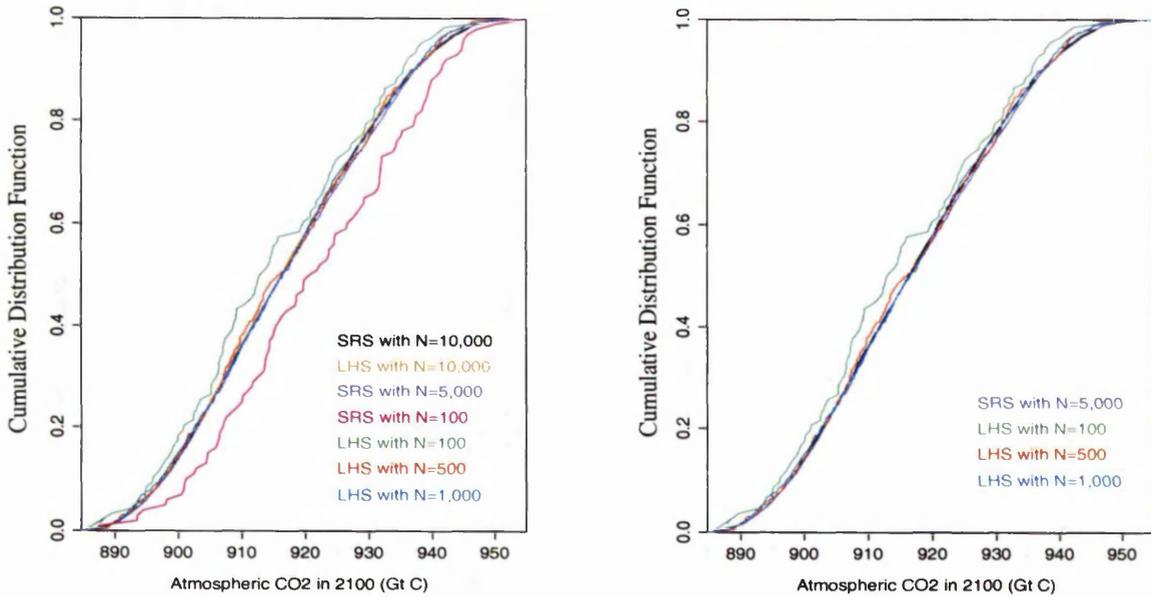
Having seen that LHS is producing CDF estimates that are more stable than those produced by SRS, it is of interest now to find out what sample size through the use of LHS would be needed to draw the same conclusion obtained through the use of SRS, and also to see if this changes from model to model. Considering SRS and LHS with various sample sizes we ran the model codes and calculated the output variable  $y_{Atm}(t = 2100)$ . The results from one of the 8-compartment models (Model I) are shown in Figure 5.7 (a) and (b) under the assumption that the initial conditions and the transfer coefficients follow uniform distribution in their specified ranges, respectively. As for the 25-compartment model the results are presented in Figure 5.8 under the assumption that all input factors follow uniform distribution in their assigned ranges. Appearing in the left panels of Figure 5.7 and Figure 5.8 are CDF estimates of  $y_{Atm}(t = 2100)$  obtained by running the model on: a SRS with  $N=10,000$ , with  $N=5000$  and with  $N=100$ ; LHS with  $N=10,000$ , with  $N=100$ , with  $N=500$  and with  $N=1000$ . The estimates based on  $N=10,000$  sample size result in a reasonably smooth curve - no matter which sampling scheme is used as their estimates are in great agreement - the estimated CDF of the output obtained through SRS with  $N=10,000$  is used as the "true" distribution function for purpose of examining the effect of using different sample sizes and sampling schemes.

The results of the previous two chapters were all based on SRS with  $N=5,000$ . As shown in left panels of Figures 5.7 and 5.8, the predictions based on SRS with  $N=5,000$  turned out to be in perfectly good agreement with the "true" CDF. Therefore, it is reasonable to treat the CDF obtained through SRS with  $N=5,000$

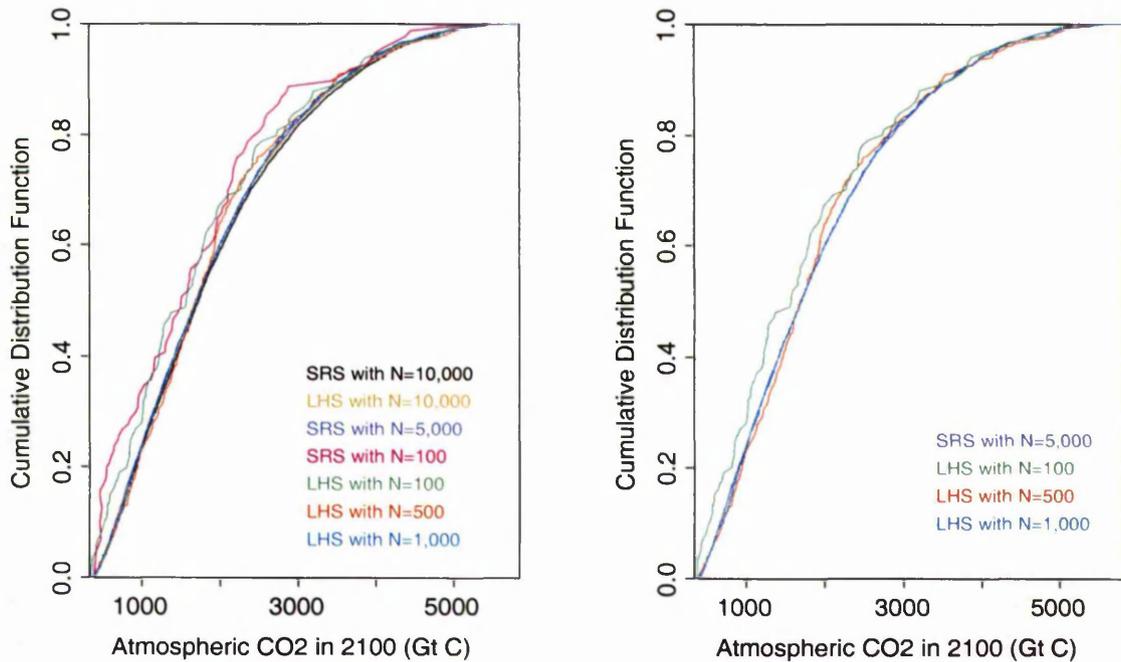
(a) Model I - Initial conditions



(b) Model I - Transfer coefficients



**Figure 5.7.** Comparison of estimated CDFs for output variable  $y_{Atm}(t = 2100)$  from Model I based on SRS and LHS with different sample sizes under the assumption that the model input factors: (a) Initial conditions, and (b) Transfer coefficients are uniformly distributed on their assigned uncertainty ranges (for the input factors and their assigned ranges see Tables 3.2 and 3.3).



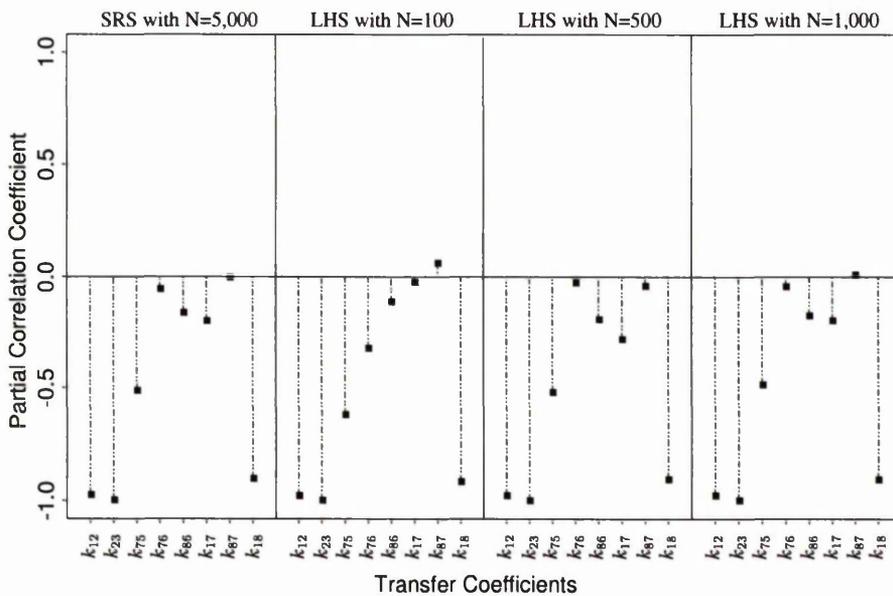
**Figure 5.8.** Comparison of estimated CDFs for output variable  $y_{Atm}(t = 2100)$  from the 25-compartment model based on SRS and LHS with different sample sizes under the assumption that the model input factors are uniformly distributed on their assigned uncertainty ranges (for the input factors and their assigned ranges see Table 4.1).

as “true” CDF.

As indicated in the left frames of Figure 5.7, the predictions based on SRS with  $N=100$  are consistently higher than the “true” estimates. With the 25-compartment model, the estimates based on both SRS and LHS with  $N=100$  are lower than the “true” estimates (see left frame in Figure 5.8). The same plots also show that compared to the estimates based on SRS with  $N=100$ , the estimates based on LHS with  $N=100$  are much closer to the “true” estimates. In the right panels of Figures 5.7 and 5.8, we compare the “true” CDF with CDFs based on LHS with sample sizes 100, 500 and 1,000 to investigate which sample size is required if LHS is considered in the analysis. It is clear from these figures that there is a slight but definite improvement in the quality of estimates as the sample size increases, but there appears to be no dramatic change in the CDFs based on

LHS with the three sample sizes.

As a further investigation, using partial correlation coefficient (PCC) as a diagnostic tool, we have shown how the importance order of the input factors change with the sample sizes in Figure 5.9. The output  $y_{Atm}(t = 2100)$  from Model I and the transfer coefficients as input factors are considered. It appears that LHS with  $N=500$  gives the importance order which agrees with the results obtained from SRS with  $N=5,000$ . The same is true when the initial conditions of the model are used in the analysis. As for the 25-compartment model the required sample size for LHS appears to be  $N=1000$ . Hereafter in the analyses we use LHS as the sampling scheme, and  $N=500$  (when 8-compartment models are concerned) and  $N=1000$  (when 25-compartment model is concerned) as the sample size.



**Figure 5.9.** Comparison of PCC values based on different sampling technique and sample sizes. The output variable considered here is  $y_{Atm}(t = 2100)$  calculated from Model I with the input factors (here the transfer coefficients) varied simultaneously under the condition that all factors follow uniform distribution on their assigned ranges.

We should emphasize that the results presented in this subsection apply to the GCC models considered in this thesis, and these results may not apply to just any model. As the results on the two different underlying models have shown, sample size requirement is a function of model complexity and the number of model input factors.

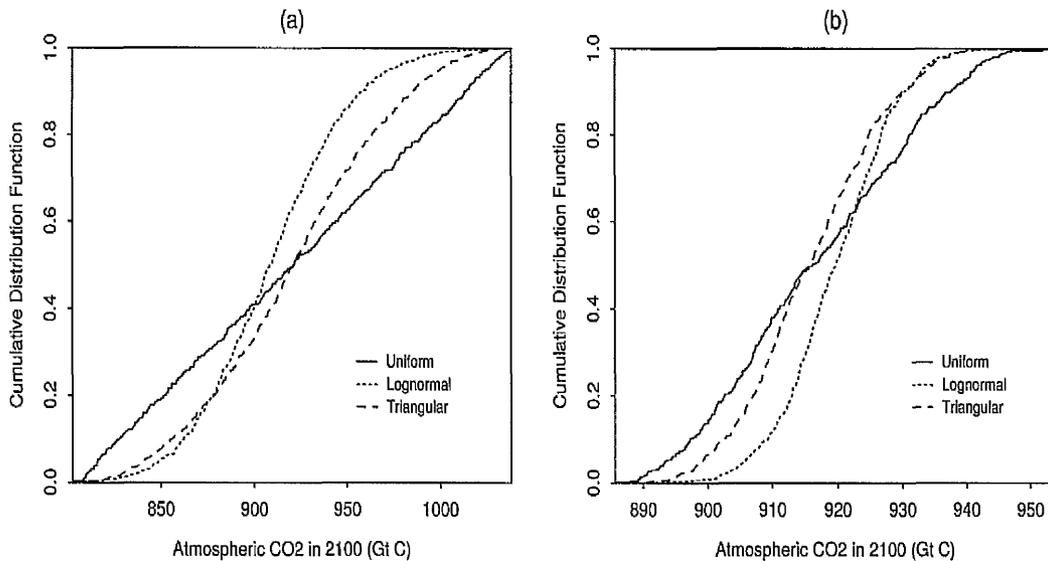
### 5.5.2 Effect of Input Factor Distribution on Model Predictions

In sampling based sensitivity/uncertainty studies, assumptions regarding the probability distributions of model input factors especially the ones that are very influential on the output variable(s) require careful consideration. There are various options available for probability distributions of the input factors; for example, uniform, lognormal, loguniform, triangular distributions. These probability distributions are often found to provide good representations for physical quantities. On the other hand, for many quantities such as concentrations, some probability distributions, like the normal distribution are theoretically inappropriate because negative values are allowed. For the models adapted in this thesis, there are no input distributions specified by the model developers. Since the uniform distribution's use is appropriate when we are able and willing to identify a range of possible values, but unable to decide which values within this range are more likely to occur than others, in the previous two chapters we have considered this distribution for all model input factors. A range for each factor is obtained by  $\pm 20\%$  of its nominal value gathered from the scientific literature.

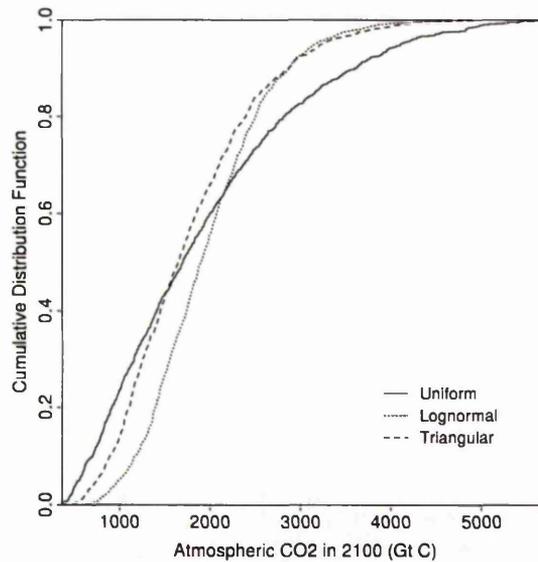
In this section, to assess the effect of input factor distribution on the model output uncertainty, we consider uniform, lognormal, and triangular distributions for the input factors of Model I and the 25-compartment model. While the assumptions regarding the probability distributions of the input factors are changed, the range of each input factor indicated in earlier chapters are kept the same.

For purposes of illustration, we consider  $y_{Atm}(t = 2100)$  as model output variable; IS92a as the emission scenario. Using LHS of size  $N=500$  for Model I and  $N=1000$  for the 25-compartment model input vectors are obtained and using the model codes  $y_{Atm}(t = 2100)$  from both models are calculated.

For Model I, for which we do not have a built-in calibration process in the model code and we have to maintain the steady-state condition, the effect of assumed probability distributions of initial conditions and of the transfer coefficients on the distributions of model outputs are assessed separately. Figures 5.10 (a) and (b) show the CDFs of  $y_{Atm}(t = 2100)$  from Model I, when the initial conditions and the transfer coefficients follow the three selected probability distributions. The results for the 25-compartment model are shown in Figure 5.11. In these figures, the solid line represents the CDF for  $y_{Atm}(t = 2100)$  predictions when input factors are assumed to follow uniform, the dotted line when factors are assumed to follow lognormal and the dashed lines when factors are assumed to follow triangular distributions on the same ranges assigned to each of



**Figure 5.10.** CDFs showing the effect of input factor - (a) Initial conditions, (b) Transfer coefficients - distributions on the estimated distribution of the output variable  $y_{Atm}(t = 2100)$  from Model I.



**Figure 5.11.** CDFs showing the effect of input factor distribution on the estimated distribution of the output variable  $y_{Atm}(t = 2100)$  from 25-compartment model.

the factors. Examination of these figures clearly indicates that the distribution associated with model input factors can have a significant effect on the distribution of model predictions. How much the output distribution function changes depends on the degree of change in the input factor distribution and the strength of the association between each output and each input factor. It can be seen that with both models under consideration here the change in the distribution is considerable, which indicates the importance of being as accurate as possible in specifying the input distributions.

For specified assumptions on input factor distributions, CDFs can be used to obtain cumulative probability estimates on model predictions. For example, with the assumption that all input factors follow the uniform distribution in their specified ranges, from Figure 5.11, the probability that atmospheric CO<sub>2</sub> in 2100 will be less than approximately 3500 Gt C is approximately 0.9. However such estimates are dependent on the ranges and distributions assumed for the input

factors.

In SA chapters, under the assumption that all input factors are uniformly distributed, we have identified the factors that are influential on the model output variables under consideration. As noted above, assumptions regarding the distributions of those factors require careful consideration. On the other hand, for any input factors that show little or no influence on the output variable, assumptions regarding their distributions are not as critical.

Next, we want to find out if input factor importance changes as a result of different distributional assumptions on the model input factors. Here we use an uncertainty importance measure introduced by Chun *et al.* [19], which is based on metric distance between two CDFs. This uncertainty measure is a useful tool to express the measure of uncertainty importance in terms of the relative impact of distributional changes of inputs on the output distribution. The authors evaluated this metric distance measure for both analytical and empirical distributions.

The form of the metric distance appropriate for our case where we have an empirical distribution generated by Monte Carlo simulation is as follows:

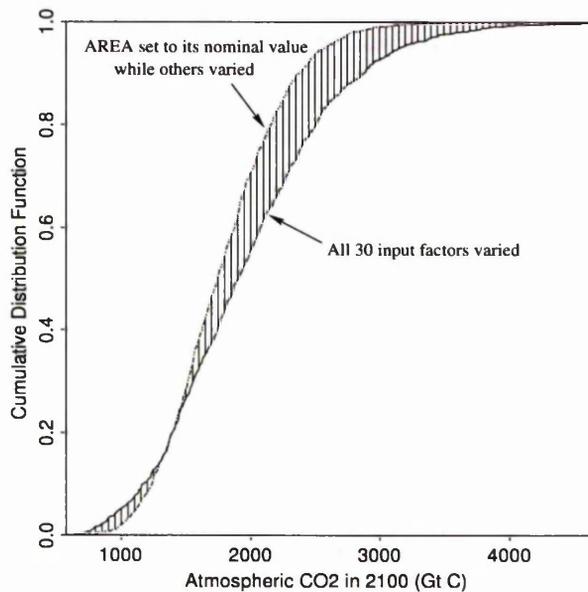
$$MD(i : o) = \frac{\sqrt{\frac{1}{N} \sum_{n=1}^N [y_{n/N}^i - y_{n/N}^o]^2}}{\frac{1}{N} \sum_{n=1}^N y_n^o}$$

where  $MD(i : o)$  is the normalized metric distance measure in terms of quantiles between the base case and its sensitivity case,  $y_{n/N}^o$  is the  $(n/N)$ th quantile of a CDF for the base case ( $0 < n < N$ ),  $y_{n/N}^i$  is the  $(n/N)$ th quantile of a CDF for its sensitivity case, and  $N$  is the number of simulations. Here, the base case refers to the case where an output distribution is obtained with all input distributions set to their assigned distribution, whereas the sensitivity case refers to the case where an output distribution is obtained with the  $i$ th input factor set to its nominal value while the other factors are varied. Normalization is carried out to make the

metric distance measure a dimensionless quantity.

Considering the output variable  $y_{Atm}(t = 2100)$  of the 25-compartment model we want to assess an importance order between the input factors based on the metric distance measure under three different distributional changes of the input factors. An example is given in Figure 5.12 showing the characteristics of this uncertainty measure graphically. The shaded area in this figure indicates the metric distance measure for the input factor AREA of the 25-compartment model and its relative impact on the distribution of the output variable  $y_{Atm}(t = 2100)$  when input distributions are taken to be lognormal.

The metric distance ( $MD$ ) measure for each 30 input factors have been calculated and their relative impacts on the distribution of the output variable have been ranked according to the magnitude of normalized  $MD$  measure obtained for each input factor. The ranked ten input factors are listed in Table 5.1 under



**Figure 5.12.** An example showing the metric distance measure between two CDFs. In this example the shaded area indicates the metric distance measure for the input factor AREA of the 25-compartment model and its relative impact on the distribution of the output variable  $y_{Atm}(t = 2100)$ . Simulations are based on LHS with 1000; emission scenario IS92a; and lognormally distributed input factors.

**Table 5.1.** Uncertainty importance rankings obtained by metric distance measure. The output variable  $y_{Aim}(t = 2100)$  from the 25-compartment model, three different input factor distributions; LHS with  $N=1000$ ; and emission scenario IS92a are considered in the calculations.

Rank	Uniform	Lognormal	Triangular
1	AREA	AREA	AREA
2	HM	HM	HM
3	TG	CSL0	TG
4	CSL0	TG	TSL
5	TSL	CW0	THG
6	CG0	TSL	CA0
7	CW0	TEMP0	CSL0
8	THG	THG	CW0
9	CL	CL	DELTP
10	TEMP0	CG0	CG0

the three input distributions. The importance order of input factors varies with different distributional assumptions. However, the  $MD$  values associated with AREA and HM factors are much higher than the  $MD$ s associated with the rest of the input factors, and no matter which input distribution is taken these two factors have been identified as the most important input factors for the output variable under consideration.

A desirable property of this measure is that it does not depend on assumptions about the form of the relationship (like, linearity) between model inputs and predictions. This measure can also be used as a screening tool. Even though it is easy to calculate, with a complex model involving a large number of input factors it can be quite laborious.

## 5.6 Scenario Uncertainty

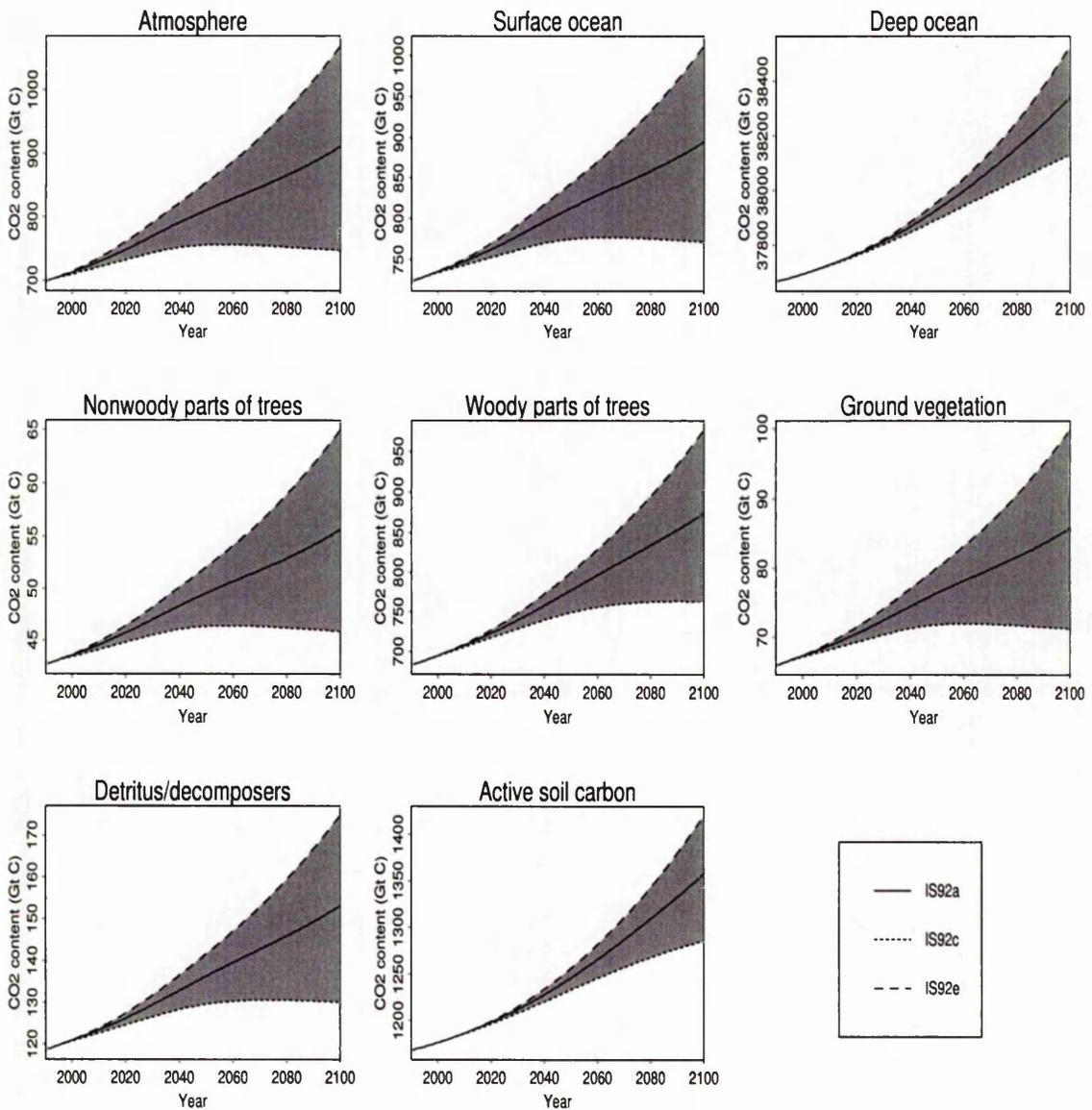
Another important source of uncertainty in GCC models is introduced by lack of knowledge or inability to predict the future conditions exactly. For example,

population growth, structural changes in economies, energy prices, fossil-fuel supplies, income are among the factors which could have a major influence on future levels of CO<sub>2</sub> emissions, and there is substantial uncertainty in all these factors. Scientists have been developing scenarios of future emissions reflecting different views of the future. These scenarios provide inputs to climate models.

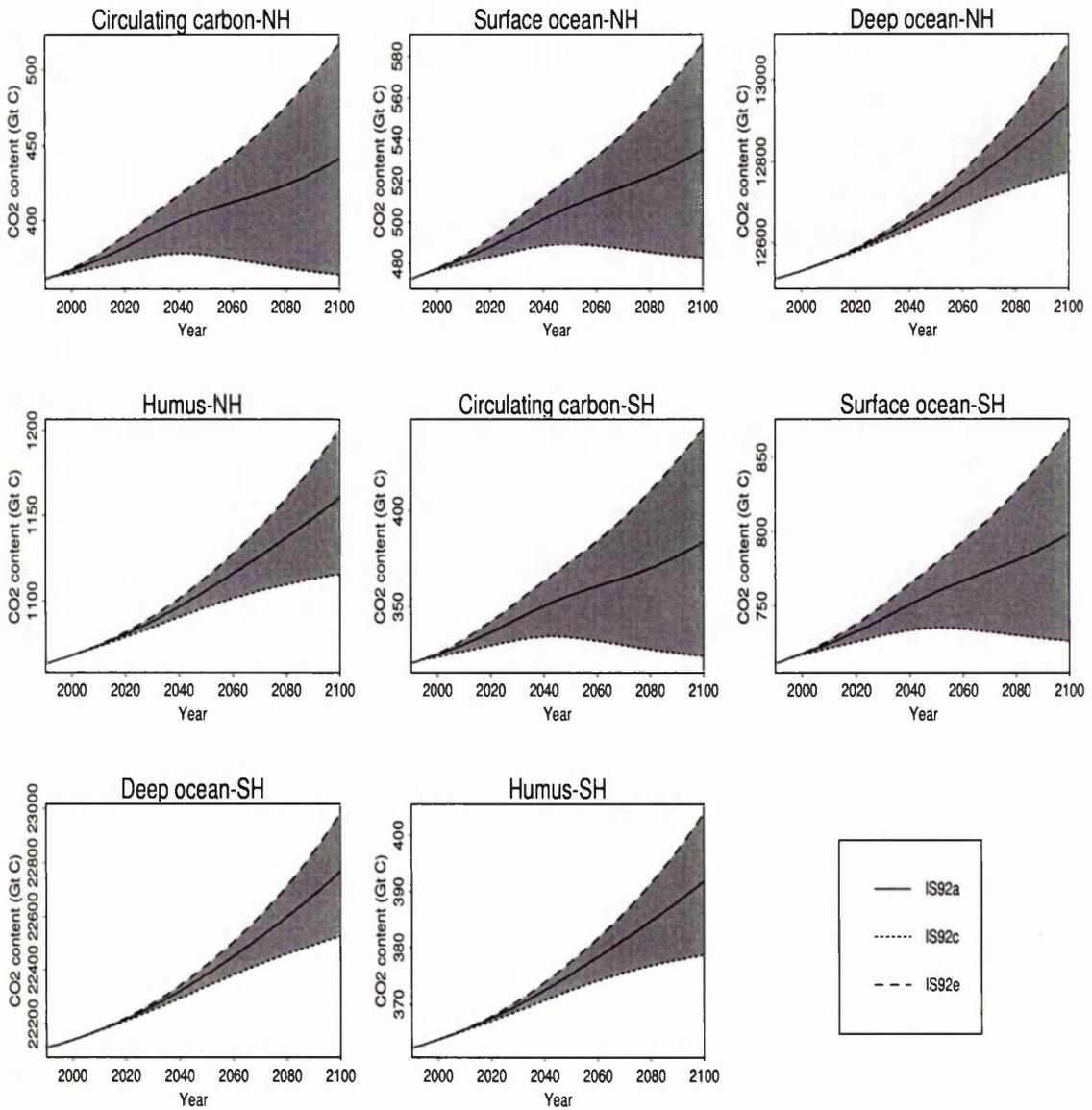
The three of the six emission scenarios developed by the Intergovernmental Panel on Climate Change (IPCC) in 1992 are considered in this thesis. These scenarios are IS92a (also known as the 'Business-as-usual' scenario), IS92c ('low' emission scenario) and IS92e ('high' emission scenario). A brief description of these three scenarios is given in Chapter 3 (see Section 3.2.2), and details of these scenarios can be found in Ref. [58].

In this section, we investigate how the uncertainty in the emission scenarios under consideration influence the uncertainty in model predictions.

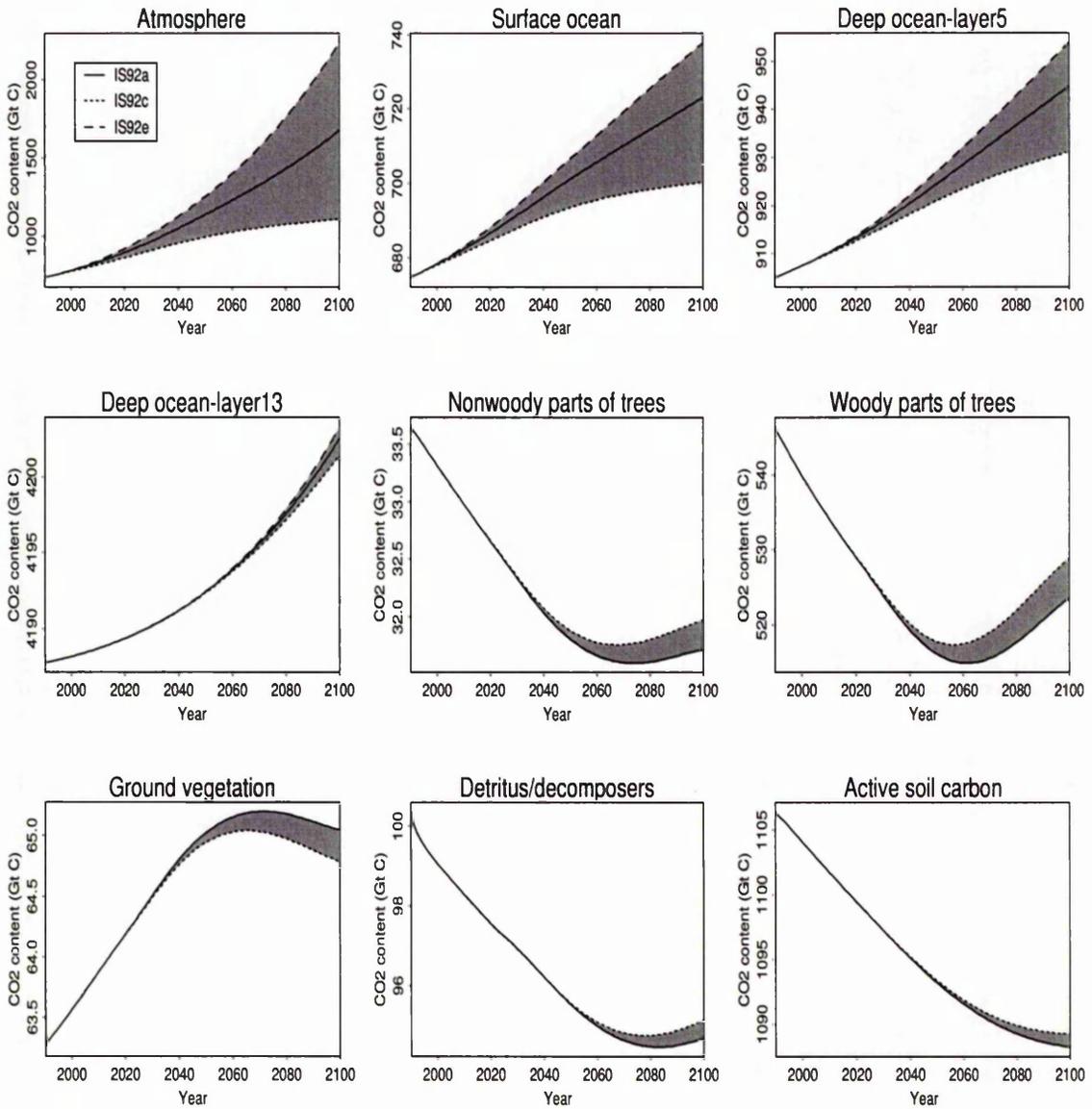
First, within each model, assuming zero uncertainty for the input factors, i.e. setting them to their nominal values, we have calculated the baseline CO<sub>2</sub> predictions for each compartment of the three models with the three IS92 scenarios. These baseline predictions are presented in Figures 5.13-5.15. Since the scenarios IS92c and IS92e are the two extreme IPCC-1992 scenarios the uncertainty ranges we see in these figures is for all IPCC-1992 scenarios. As seen in Figures 5.13 and 5.14 showing results for the 8-compartment models, Model I and Model II, respectively, the IS92e scenario yields a high estimate, the IS92a a median and the IS92c a low estimate of all compartmental CO<sub>2</sub> contents over the whole time period (from 1995 to 2100). However, for the 25-compartment model this is the case with the atmosphere and ocean compartments only. With the ground vegetation compartment the IS92a and IS92e give a high estimate and the IS92c a low estimate, and as for the other terrestrial compartments the IS92c yields a high estimate and the scenarios IS92a and IS92e a low estimate. The impact of different emission scenarios on the deeper layers of the ocean and the terrestrial ecosystem is not as rapid as it is on the atmosphere, surface layer and the upper layers of the ocean.



**Figure 5.13.** The range of CO<sub>2</sub> baseline predictions of each compartment of Model I, based on IPCC-IS92a,c,e emission scenarios. These calculations are based on a model simulation in which all model input factors are set to their nominal values.



**Figure 5.14.** The range of CO<sub>2</sub> baseline predictions of each compartment of Model II, based on IPCC-IS92a,c,e emission scenarios. These calculations are based on a model simulation in which all model input factors are set to their nominal values.



**Figure 5.15.** The range of CO<sub>2</sub> baseline predictions of the nine compartments of the 25-compartment model, based on IPCC-IS92a,c,e emission scenarios. These calculations are based on a model simulation in which all model input factors are set to their nominal values.

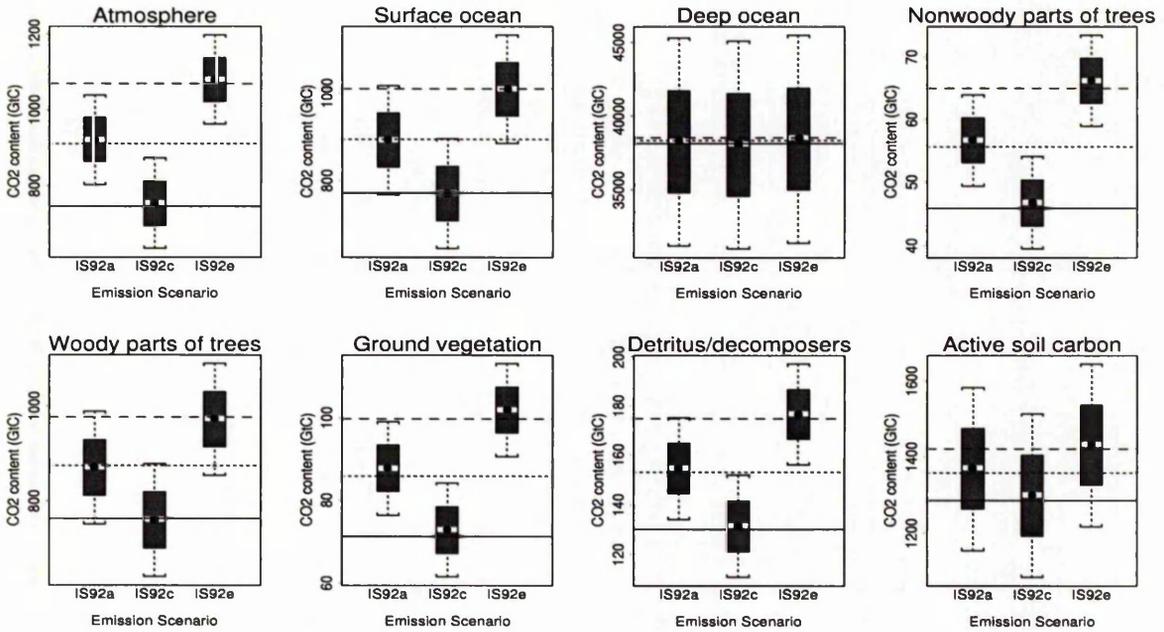
Next, taking the input factor uncertainty into account and varying them over their assigned ranges from uniform distribution, within each model, we want to investigate the results obtained considering the three emission scenarios. In the simulations the same set of input factor values is used with each scenario to enable a direct comparison among scenarios. The compartmental predictions are evaluated in the year 2100. A compact summary of the distributions of the output variables is provided by the boxplots in Figures 5.16 - 5.18 for the three models. The horizontal lines in these figures are the baseline values for the associated compartment in year 2100 under each scenario (dotted line: IS92a; solid line: IS92c; and dashed line: IS92e). The estimated mean values are also indicated in the boxplots by solid circles.

As examination of Figures 5.16 and 5.17 shows, there is some overlap between the boxplots (very much for the deep ocean and active soil carbon; and moderate for the other compartments) when the uncertainty in the initial conditions is considered. This indicates that on the basis of the overall output uncertainties no discrimination between the predictions for the three scenarios can be made. On the other hand, when the uncertainty in the transfer coefficients is considered there appears to be some overlap between the boxplots of the deep ocean and active soil carbon (humus compartments for Model II) compartments, but not for the others. The boxplots not overlapping indicate that on the basis of the overall output uncertainties discrimination between the predictions for the three scenarios can be made.

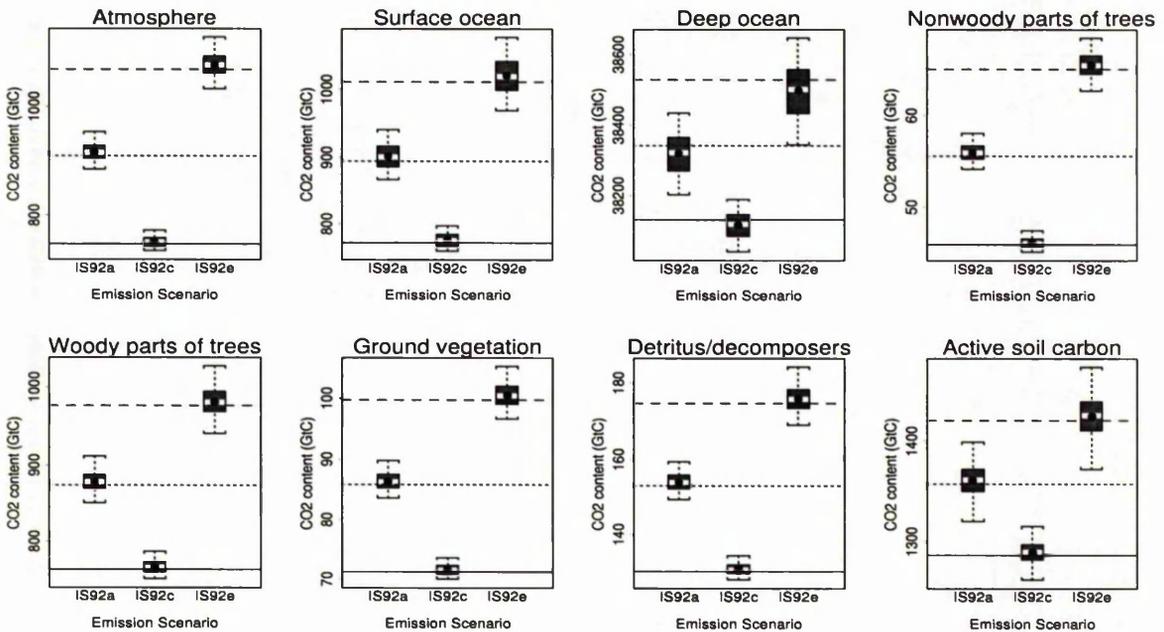
The boxplots for the 25-compartment model given in Figure 5.18 show that the boxplots of each compartment overlap quite a lot. For the deep ocean-layer13 compartment and all five terrestrial compartments the distributions of the outputs appear to behave almost identically no matter which emission scenario is used.

The contribution of the input factor uncertainties to the prediction uncertainties appears to be very high within each model. In Table 5.2, along with the baseline values of each compartment's year 2100 predictions under the three

(a) Initial conditions

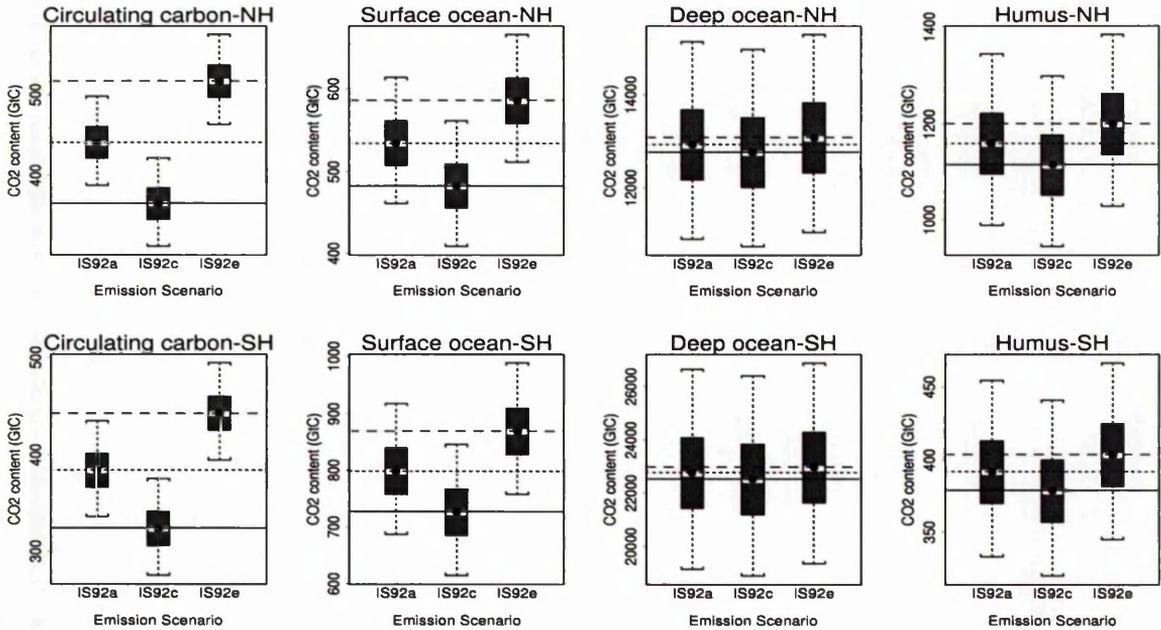


(b) Transfer coefficients

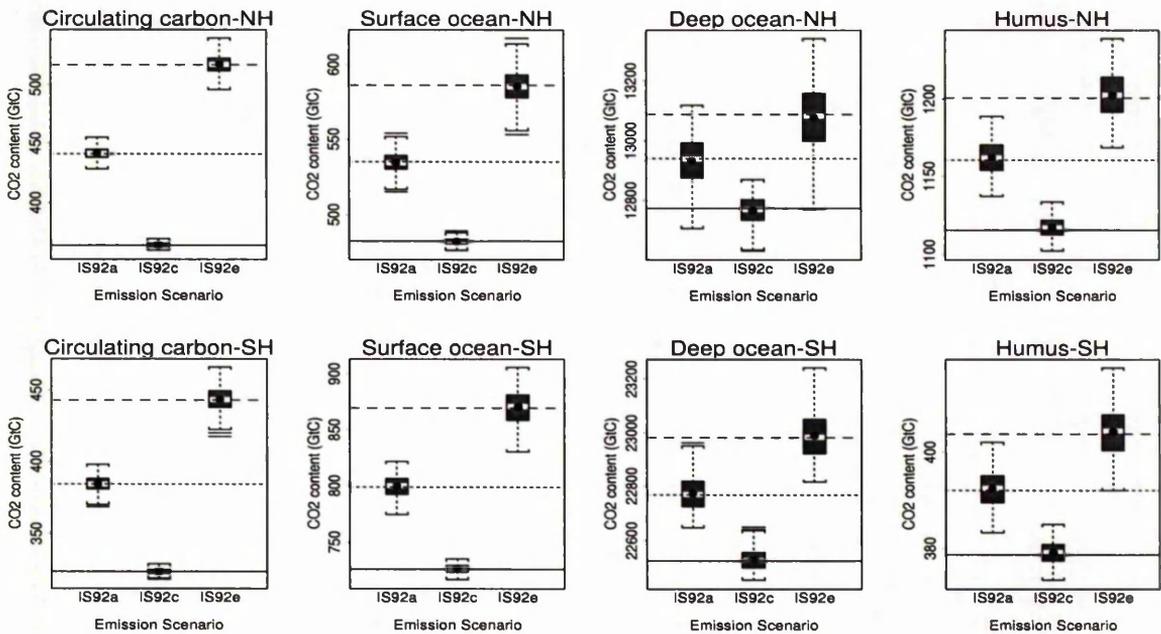


**Figure 5.16.** Boxplots showing the distribution of CO<sub>2</sub> predictions in year 2100 from Model I considering considering the three of the IPCC emission scenarios. 500 simulations of data generated using LHS and all input factors ((a) Initial conditions, (b) Transfer coefficients) considered to follow uniform distribution over their assigned ranges (see Tables 3.2 and 3.3). The horizontal lines show the baseline CO<sub>2</sub> content of each compartment in 2100 under each scenario: ···IS92a; —IS92c; ---IS92e. The means are indicated by solid circles.

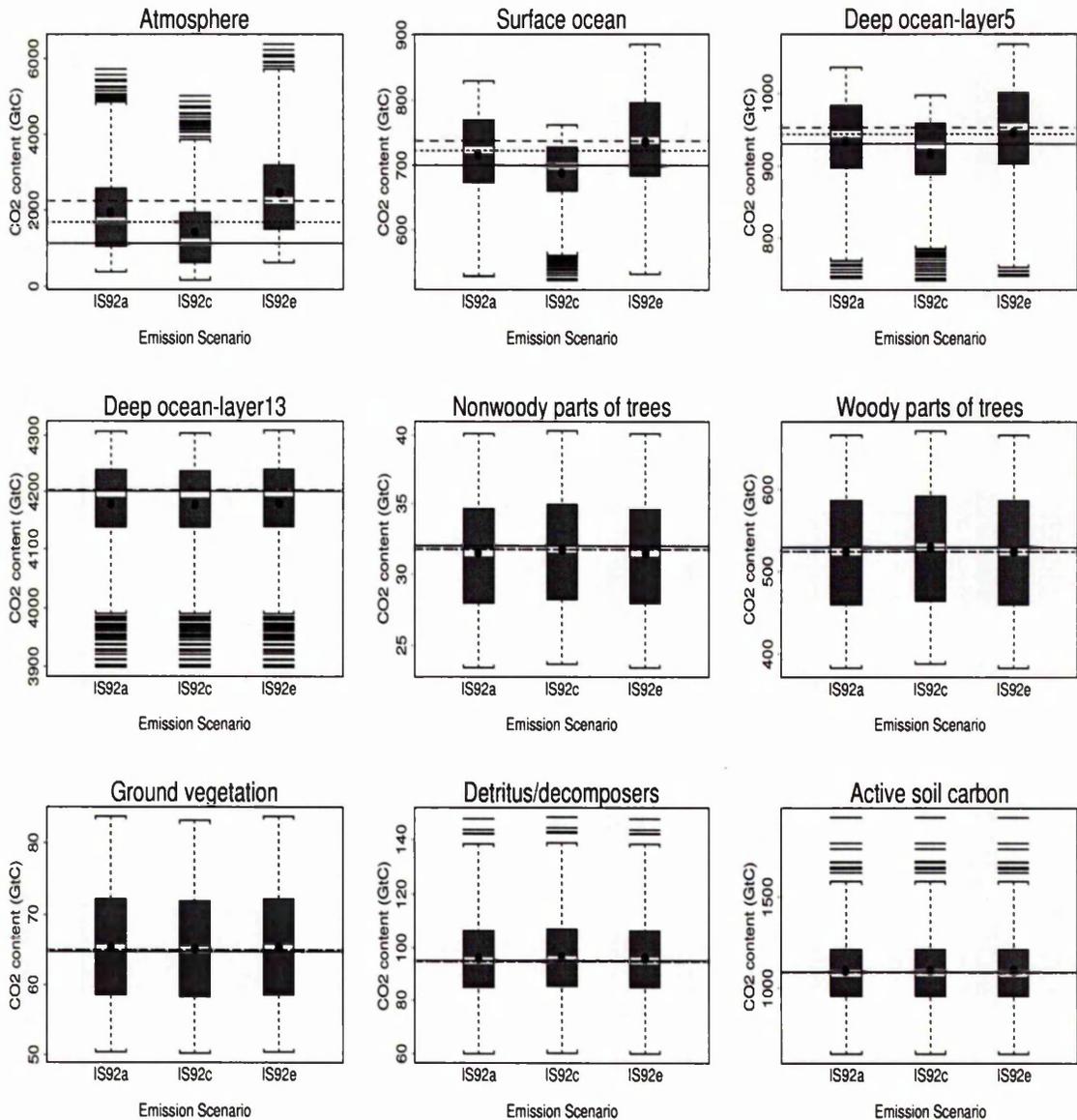
(a) Initial conditions



(b) Transfer coefficients



**Figure 5.17.** Boxplots showing the distribution of CO<sub>2</sub> predictions in year 2100 from Model II considering considering the three of the IPCC emission scenarios. 500 simulations of data generated using LHS and all input factors ((a) Initial conditions, (b) Transfer coefficients) considered to follow uniform distribution over their assigned ranges (see Tables 3.4 and 3.5). The horizontal lines show the baseline CO<sub>2</sub> content of each compartment in 2100 under each scenario: ···IS92a; —IS92c; ---IS92e. The means are indicated by solid circles.



**Figure 5.18.** Boxplots showing the distribution of CO<sub>2</sub> predictions in year 2100 from the 25-compartment model considering the three of the IPCC emission scenarios. 1000 simulations of data generated using LHS and all input factors considered to follow uniform distribution over their assigned ranges (see Table 4.1). The horizontal lines show the baseline CO<sub>2</sub> content of each compartment in 2100 under each scenario: ··· IS92a; —IS92c; ---IS92e. The means are indicated by solid circles.

scenarios, we give the uncertainty ranges due to only scenario uncertainty, and due to both scenario and input factor uncertainties.

As seen in this table by introducing input factor uncertainties in addition to scenario uncertainty overall uncertainty ranges increase dramatically for the

**Table 5.2.** Scenario baseline values in 2100; uncertainty ranges of compartmental predictions from 2100 as a result of scenario uncertainty; and uncertainty ranges of compartmental predictions from 2100 as a result of both scenario and input factor uncertainties.

	Compartment	IS92a	IS92c	IS92e	Scen. <sup>a</sup> Un. Range	Scen. & In.Factor <sup>b</sup> Un. Range
Model I	Atmosphere	746.7	910.51	1067.97	321.27	560.31 (394.12) <sup>c</sup>
	Surface ocean	771.80	894.42	1011.44	239.64	488.65 (319.35)
	Deep ocean	38130.16	38342.57	38527.41	397.25	14445.26 (610.21)
	N.woody parts trees	45.87	55.63	64.99	19.13	34.05 (23.32)
	Woody parts trees	763.19	874.18	976.54	213.35	449.04 (276.76)
	Ground vegetation	71.26	85.85	99.80	28.54	51.86 (35.25)
	Detritus/decomposers	130.09	153.10	174.80	44.71	86.12 (56.38)
	Active soil carbon	1285.88	1357.38	1420.06	134.19	561.41 (211.11)
Model II	Circulating carbon-(NH)	364.43	441.26	516.85	152.42	263.44 (179.61)
	Surface Ocean-(NH)	482.58	535.00	586.07	103.49	256.56 (140.96)
	Deep ocean-(NH)	12773.49	12940.26	13087.97	314.49	4557.71 (708.90)
	Humus-(NH)	1115.79	1160.47	1200.78	84.99	440.27 (136.59)
	Circulating carbon-(SH)	323.85	384.00	443.10	119.25	218.80 (147.42)
	Surface Ocean-(SH)	726.43	798.86	869.32	142.90	374.23 (188.18)
	Deep ocean-(SH)	22526.45	22768.65	22981.94	455.49	7961.98 (785.91)
	Humus-(SH)	378.71	391.91	403.76	25.04	146.50 (43.86)
25-compart. model	Atmosphere	1109.99	1674.90	2238.98	1128.99	6212.12
	Surface ocean	700.44	723.01	737.92	37.48	361.94
	Deep ocean	931.13	944.70	953.98	22.85	324.50
	Deep ocean	4201.45	4202.53	4203.30	1.85	410.52
	N.woody parts trees	31.97	31.72	31.72	0.25	16.96
	Woody parts trees	528.82	523.55	523.55	5.28	291.39
	Ground vegetation	64.79	65.05	65.05	0.26	33.40
	Detritus/decomposers	95.13	94.68	94.68	0.46	88.48
	Active soil carbon	1089.23	1088.24	1088.24	0.99	1283.69

<sup>a</sup>Uncertainty ranges as a result of scenario uncertainty

<sup>b</sup>Uncertainty ranges as a result of scenario and input factor uncertainties

<sup>c</sup>Uncertainty ranges given in brackets obtained when uncertainties in transfer coefficients are considered.

25-compartment model with minimum increase being in the atmosphere compartment which is over 450%. For the 8-compartment models change in the uncertainty ranges is also high especially when the initial conditions are taken as the input factors. In the table the uncertainty ranges given in brackets are based on the predictions when the uncertainty in the transfer coefficients are taken into account.

It is also of interest to investigate if the importance order of uncertain input factors changes depending on the emission scenario used. A quick examination of scatterplots (not included) showed that the relationship between the considered output variables (the compartmental CO<sub>2</sub> contents in 2100) and the input factors is either linear or there is no relationship. A stepwise regression analysis is used for the investigation. The analysis results revealed that even though there is slight change in the order in which the input factors are entered the regression models, the same set of important input factors are obtained under the three emission scenarios. For illustration purposes, we show the results from the analysis on one output variable from each model in Table 5.3.

**Table 5.3.** Comparison of Stepwise Regression Analyses with the IS92a,c,e IPCC emission scenarios for various output variables calculated from the three GCC models in year 2100 under the assumption that all input factors follow uniform distribution on their assigned ranges. The order at which the input factors were added to the corresponding model and the  $R^2$ -values for the regression models at each step are given.

	Compartment	Step	IS92a		IS92c		IS92e	
			Input	$R^2$	Input	$R^2$	Input	$R^2$
Model I	Nonwoody parts of trees	1	$k_{23}$	0.7331	$k_{23}$	0.8961	$k_{23}$	0.6021
		2	$k_{75}$	0.8588	$k_{12}$	0.9768	$k_{75}$	0.8285
		3	$k_{12}$	0.9710	$k_{75}$	0.9857	$k_{12}$	0.9483
		4	$k_{86}$	0.9829	$k_{18}$	0.9907	$k_{86}$	0.9682
		5	$k_{87}$	0.9887	$k_{86}$	0.9920	$k_{18}$	0.9810
		6	$k_{18}$	0.9911	$k_{87}$	0.9926	$k_{87}$	0.9903
		7	$k_{17}$	0.9915			$k_{17}$	0.9909
Model II	Humus-NH	1	$k_{14}$	0.7925	$k_{14}$	0.7345	$k_{14}$	0.8154
		2	$k_{67}$	0.8517	$k_{67}$	0.8227	$k_{56}$	0.8690
		3	$k_{56}$	0.9104	$k_{56}$	0.8912	$k_{12}$	0.9210
		4	$k_{12}$	0.9647	$k_{12}$	0.9519	$k_{67}$	0.9696
		5	$k_{23}$	0.9882	$k_{23}$	0.9864	$k_{23}$	0.9888
		6	$k_{15}$	0.9933	$k_{15}$	0.9914	$k_{15}$	0.9940
		7	$k_{65}$	0.9941	$k_{65}$	0.9922	$k_{65}$	0.9948
25-compt. model	Active soil carbon	1	TSL	0.4294	TSL	0.4288	TSL	0.4294
		2	CG0	0.5724	CG0	0.5716	CG0	0.5724
		3	TG	0.6985	TG	0.6971	TG	0.6985
		4	THG	0.8000	THG	0.7989	THG	0.8000
		5	CW0	0.8599	CW0	0.8591	CW0	0.8599
		6	TW	0.9173	THD	0.9169	TW	0.9173
		7	THD	0.9534	TW	0.9531	THD	0.9534

## 5.7 Model Uncertainty

As Kennedy and O'Hagan [72] put it "No model is perfect". Even if there is no input factor uncertainty, i.e. we know the true values of all the factors required to make a particular prediction of a process being modelled, the predicted value will not be equal to the value of the process. The authors call this discrepancy 'model inadequacy'.

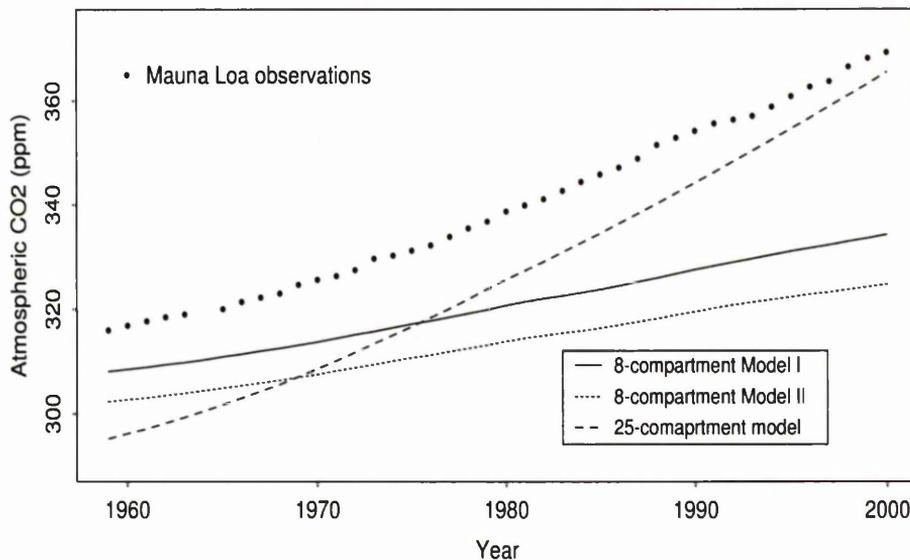
There is uncertainty due to model structure since there is no unique way

to abstract the real system into a few variables and equations. Many different sources of uncertainty in GCC models exist, for example, uncertainties can arise from making physical assumptions, or even entirely neglecting physical processes [43]. Despite the fact that these GCC models are attempting to represent the same system, there can be and there usually are disagreements between the results of the models because of uncertainties in model assumptions, initial conditions, model structure, *etc.*

With model uncertainty we simply want to examine and compare three GCC models. Simply the size of model uncertainties can be estimated by comparison with observations. However, as mentioned earlier in the chapter, we should keep in mind that observations can also introduce an important source of uncertainty. In Figure 5.19 the atmospheric CO<sub>2</sub> predictions of the three GCC models are plotted and compared with the Mouna Loa (MLO) observations over the period of the MLO record, from 1959 to 2000.

Figure 5.19 shows that all three GCC models underestimate the observed data. The MLO measurement show that the atmospheric CO<sub>2</sub> concentration has increased from 315.98 ppm in 1959 to 369.40 ppm in 2000 [71]. Model I predicts the atmospheric CO<sub>2</sub> to be 308.16 ppm in 1959 and 334.30 ppm in 2000. In these respective years, Model II predicts these concentrations as 302.27 ppm and 324.68 ppm. With both of these 8-compartment model the deviation from the observed data increases rapidly, even though they perform better than the 25-compartment model in the early years. The predictions with the 25-compartment model are 295.12 ppm in 1959 and 365.58 ppm in 2000.

The change in CO<sub>2</sub> concentration corresponds to an increase of approximately 84.43 Gt C in the mass of carbon in the atmosphere, approximately 59% of the release of carbon by fossil fuel combustion and land-use change between 1959 and 1991 [33]. The increase in CO<sub>2</sub> concentration is approximately 42.00 Gt C with Model I and 38.10 Gt C with Model II, and this corresponds to approximately 29% and 26%, respectively, of the carbon released. With the 25-compartment model the change in atmospheric CO<sub>2</sub> concentration is about 109.00 Gt C increase and



**Figure 5.19.** Comparison of predicted atmospheric CO<sub>2</sub> concentrations from the three GCC models under the base-case scenario and the measured atmospheric CO<sub>2</sub> concentrations from Mauna Loa Observatory: 1959-2000 (Keeling & Whorf, 2001 [71]).

this is approximately 76% of the carbon released into the atmosphere.

The observed increase in the carbon content of the atmosphere is less than the release by fossil fuel combustion and land-use change because atmosphere exchanges carbon with other reservoirs. With the 8-compartment models because of their highly linear mathematical structure, all model compartments take up carbon from the atmosphere quite rapidly. As a result we see a low rate of increase in the atmospheric CO<sub>2</sub> predicted by these models. The 8-compartment models utilized in this thesis are constructed using the linear, time-invariant compartmental modelling formulation while physical processes of GCC are entirely neglected, and these models found to be involving a large amount of uncertainty. The 25-compartment model, however, is a more realistic representation of carbon cycle processes. For example, it takes into account the turnover mechanism of carbon in the oceans, the storage of non-labile carbon in the terrestrial biota, the

depth distribution of  $^{14}\text{C}$ , *etc.*.

As noted by Draper [26] model uncertainty is conditional on scenarios and input factors. Now, assuming zero uncertainty for model input factors, i.e. setting them to their nominal values, and concentrating on the future projections we have calculated compartmental  $\text{CO}_2$  predictions for each model, under the three emission scenarios. The compartments of the three models are aggregated after the Monte Carlo simulations into: Atmosphere, Ocean, and Terrestrial Ecosystem. The time dependent behaviour of the predictions for these model components between the years 2000 and 2100 is presented in Figure 5.20.

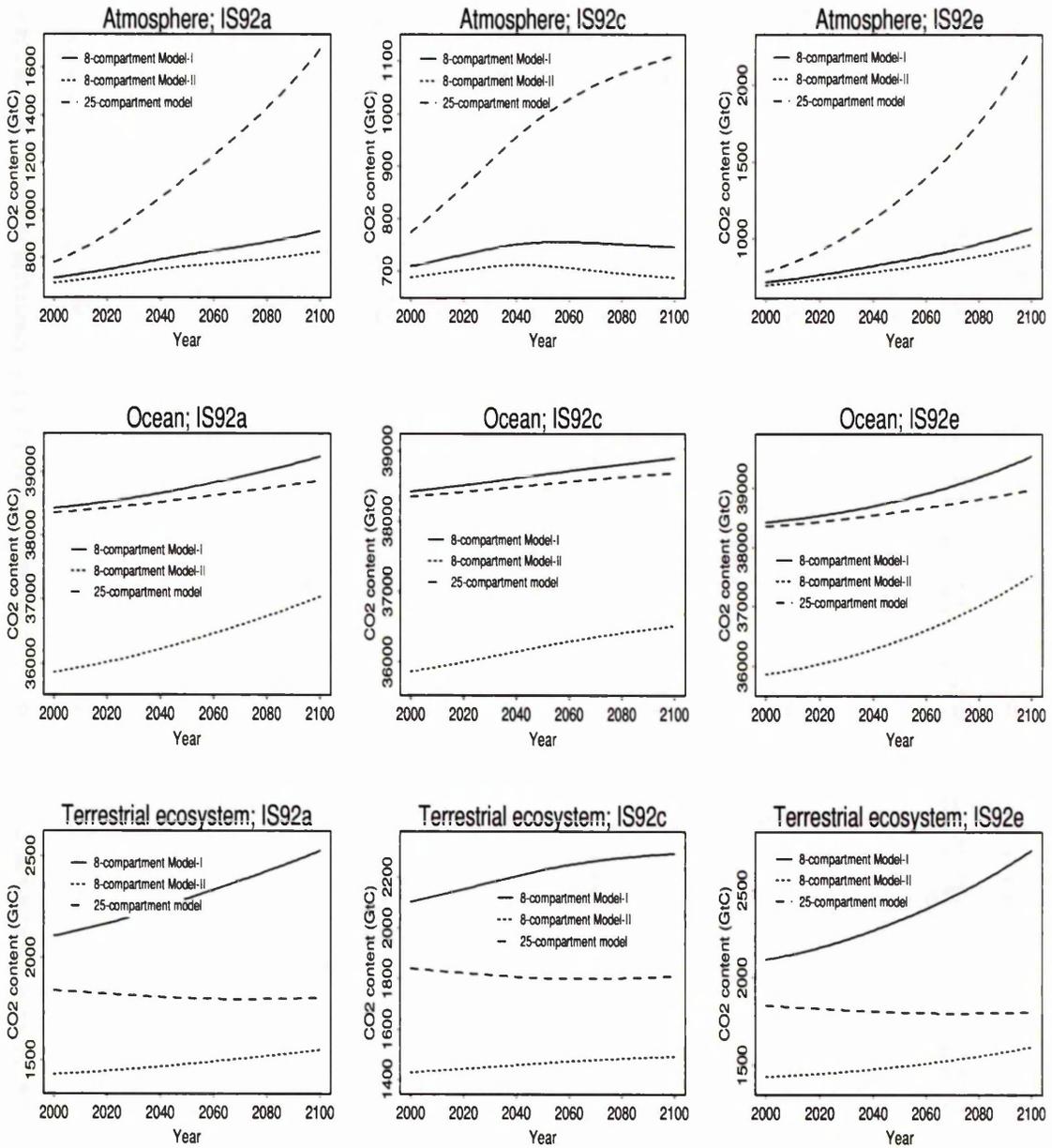
For the atmosphere component, considering the emission scenario IS92a, Model I and Model II do not yield substantially different results. With Model I the atmospheric  $\text{CO}_2$  content reaches to 910.50 Gt C by the year 2100 (about 28% increase from 2000 to 2100), and with Model II it reaches to 825.27 Gt C by 2100 (about 20% increase). However the 25-compartment model predicts the atmospheric  $\text{CO}_2$  content to be much higher, 1674.90 Gt C by 2100 (about 114% increase from 2000 to 2100).

The predictions of  $\text{CO}_2$  content of the ocean component obtained from Model I and the 25-compartment model appears to be closer together where as Model II predictions are much lower. All three models predict the  $\text{CO}_2$  content of ocean to increase with time, but the rate of increase varies with model, Model II predictions indicating a more rapid increase.

The baseline curve associated with the 25-compartment model show the amount of  $\text{CO}_2$  in the terrestrial ecosystem to decrease (from 1839.96 Gt C in 2000 to 1803.24 Gt C in 2100), but both 8-compartment models predict  $\text{CO}_2$  content of this reservoir to increase with time. This increase with Model I is more rapid than it is with Model II.

Although the predicted  $\text{CO}_2$  contents of the three reservoirs change with the emission scenarios, the tendency in model behaviour does not change.

Model uncertainty range appears to be increasing when the atmosphere and the terrestrial ecosystem components are concerned, and decreasing slightly when



**Figure 5.20.** Baseline predictions from the three GCC models with the IS92a,c,e emission scenarios for the time period 2000-2100. These calculations are based on a model simulation in which all model input factors are set to their nominal values.

the ocean component is concerned, and this is the case with all three emission scenarios.

Next, by allowing for input factor uncertainty and focusing on the year 2100 predictions from the three models under the three emission scenarios we calculate mean and coefficient of variation (CV) of the three model components. The results are presented in Table 5.4. When the two 8-compartment models are concerned the uncertainties in the initial conditions and in the transfer coefficients are taken into account separately in the analysis. In Table 5.4, the statistics calculated from the simulations in which transfer coefficient uncertainties taken into account are given in brackets. Because of differences in modelling assumptions, initial conditions and model structure the estimated average CO<sub>2</sub> content of the three reservoirs of the three models are different. Since there are some similarities between the 8-compartment models it may be reasonable to compare the estimated means and the CVs associated with these models. The average CO<sub>2</sub> content of all components in 2100 appears to be higher with Model I compared to Model II in both analysis based on initial condition and transfer coefficient uncertainties. The CV for each model component is also high with Model I compared to Model II.

Now, focusing on Model I and Model II when the uncertainty in the initial conditions is concerned we compare the CV for each component of the three models. This comparison shows that for the atmosphere component the 25-compartment model is the most variable (a CV of about 58% with IS92a, 71% with IS92c, and 49% with IS92e) where as Model II is the least variable (CVs of about 5-7% with all scenarios). When the ocean component is concerned the 25-compartment model is the least variable with a CV of around 3% with all three emission scenarios, where as Model I is the most variable with a CV of about 10% under all three scenarios. As for the terrestrial component, the variability is highest with the 25-compartment model (around 13% with all scenarios) and lowest with Model II (around 7% with all scenarios).

**Table 5.4.** Comparison of models in terms of estimated mean and coefficient of variation (CV) values based on year 2100 CO<sub>2</sub> contents of Atmosphere, Ocean and Terrestrial Ecosystem components calculated using the three GCC models with the three IPCC-1992 scenarios. The calculations are based on N=1000 model simulations with all input factors assumed to follow uniform distribution over their assigned ranges.

	Model Component	8-compartment Model I <sup>a</sup>		8-compartment Model II		25-comp. model	
		Mean	CV	Mean	CV	Mean	CV
IS92a	Atmosphere	920.89 (916.83)	7.29 (1.61)	825.25 (826.01)	5.86 (1.18)	1926.86	57.69
	Ocean	39212.26 (39220.85)	10.38 (0.11)	37043.26 (37039.95)	7.53 (0.05)	38628.22	2.89
	Terr. Ecosys.	2540.40 (2535.96)	8.33 (1.12)	1551.92 (1554.45)	7.06 (0.76)	1813.54	12.64
IS92c	Atmosphere	755.42 (750.61)	8.88 (1.19)	688.26 (688.62)	7.03 (0.52)	1402.77	70.94
	Ocean	38878.15 (38890.56)	10.46 (0.07)	36509.43 (36506.64)	7.64 (0.03)	38418.11	2.61
	Terr. Ecosys.	2311.30 (2303.77)	9.16 (0.85)	1494.04 (1496.46)	7.33 (0.47)	1820.23	12.60
IS92e	Atmosphere	1079.93 (1076.47)	6.21 (1.88)	959.93 (961.08)	5.04 (1.63)	2455.73	48.78
	Ocean	39513.22 (39518.55)	10.30 (0.13)	37525.79 (37522.00)	7.43 (0.07)	38778.30	3.10
	Terr. Ecosys.	2749.78 (2747.99)	7.70 (1.32)	1604.07 (1606.71)	6.83 (1.00)	1813.54	12.64

<sup>a</sup>With the two 8-compartment models, the statistics in brackets are from the simulations where only the uncertainty in the transfer coefficients are concerned.

## 5.8 Partitioning Uncertainty

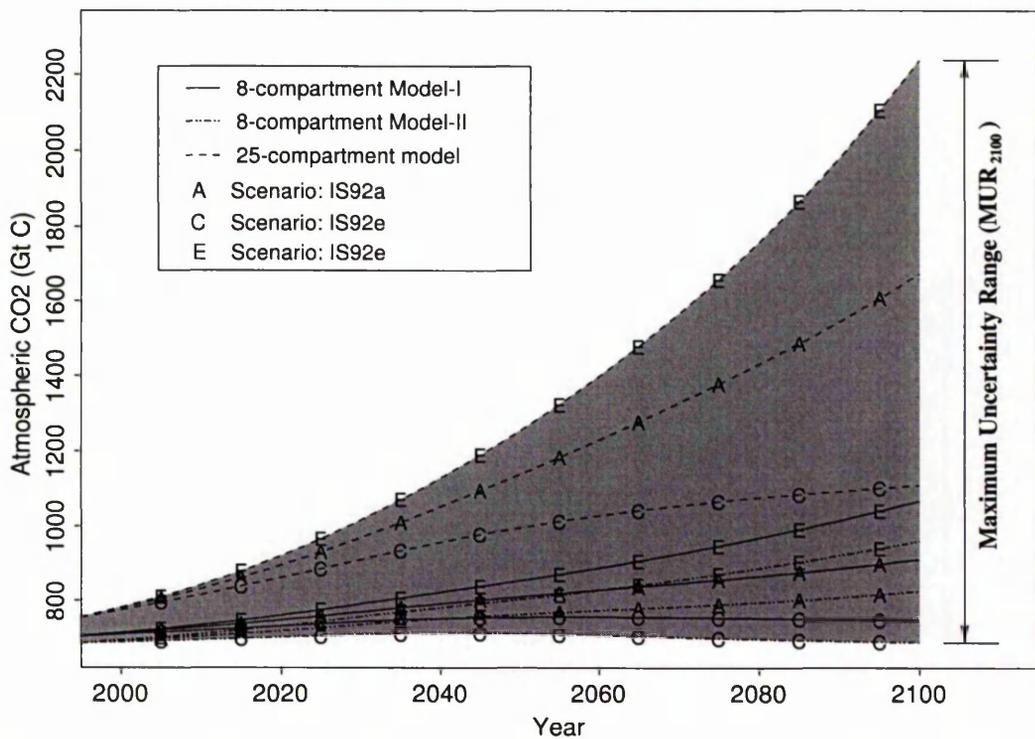
The discrepancies between the observed data and simulated atmospheric CO<sub>2</sub> content indicates that some of the sources of uncertainty must be large enough to be causing the discrepancies, but which sources of uncertainty are to blame for this?

In order to reduce the uncertainty in the model predictions an analysis of the contribution of the individual sources of uncertainty is of great interest. Once the main sources of uncertainty are identified, further research can be done in a way that the output uncertainties can be reduced.

First, assuming that there is no input factor uncertainty and focusing on the atmosphere compartment we calculate predictions using the three GCC models with the three emission scenarios to obtain an overall uncertainty range over time. In Figure 5.21 the shaded area indicates the maximum uncertainty range

( $MUR_t$ ) of atmospheric  $CO_2$  content over the period 1995-2100.

The  $MUR_{2100}$  which involves both model and scenario uncertainties in 2100 (pointed out in Figure 5.21) is  $2238.99 - 688.28 = 1550.71$  obtained by subtracting the maximum and minimum curves in year 2100. The uncertainty range ( $UR_t$ ) can be calculated when not all models and/or scenarios is considered. The reduced  $UR_{2100}$  when the 25-compartment model is not in operation is calculated and it appears that this model contributes 76% to the  $MUR_{2100}$ , without this model the  $MUR_{2100}$  goes down to 379.68. Since Model I is the 'median' model the  $MUR_{2100}$  do not change with this model not being in operation. The contribution of Model II to the  $MUR_{2100}$  is only 4%. Now, focusing on the scenarios, it



**Figure 5.21.** Maximum uncertainty range of the Atmospheric  $CO_2$  predictions based on the three GCC models with the IS92a,c,e emission scenarios for the period 1995-2100. These calculations are based on model simulations in which all model input factors are set to their nominal values.

is easily seen that IS92e is contributing the most to the  $MUR_{2100}$  (about 36%) and the contribution of IS92c is 9%. Without scenario IS92a which is the 'median' scenario the  $MUR_{2100}$  do not change.

In the previous sections we have discussed and investigated input factor uncertainty within model and scenario, scenario uncertainty within model, model uncertainty within scenario. Now, it is of interest to find out how much of the overall uncertainty about an output variable is attributable to the three sources of uncertainty.

First, focusing on scenario and input factor uncertainty, we wish to investigate whether differences in the predictions are attributable primarily to scenario or the input factors, for a given model. As the output variable the  $CO_2$  content of atmosphere in the year 2100 (i.e.,  $y_{Atm}(t = 2100)$ , for convenience we will use  $y$  to refer to this output variable in this section) is considered. Using an approach introduced by Draper (see [26]) and referred to as *model uncertainty audit*, we partition the overall predictive uncertainty about  $y$  into 'between scenario' and 'due to input factors within scenario' components, the second of which represents the component of uncertainty arising from lack of knowledge about the input factors.

With  $y$  as the output variable, and scenario  $i$  occurring with probability  $p_i$  and leading to estimated mean ( $\hat{\mu}_i$ ) and standard deviation ( $\hat{\sigma}_i$ ) of  $y$ , the overall mean and variance of the output variable are calculated using

$$\hat{\mu} = E_S[\hat{E}(y|S)] = \sum_{i=1}^s p_i \hat{\mu}_i, \quad (5.1)$$

and

$$\begin{aligned} \hat{\sigma}^2 &= V_S[\hat{E}(y|S)] + E_S[\hat{V}(y|S)] \\ &= \sum_{i=1}^s p_i (\hat{\mu}_i - \hat{\mu})^2 + \sum_{i=1}^s p_i \hat{\sigma}_i^2 = \hat{\sigma}_{BS}^2 + \hat{\sigma}_{WS}^2 \end{aligned} \quad (5.2)$$

respectively, where  $S$  stands for scenario and  $s = 3$  (three emission scenarios: IS92a, IS92c and IS92e). For each of the three models the scenario-specific means and standard deviation (SD) estimates are given in Table 5.5, together with two possible vectors of scenario probabilities. The first of these vectors (Case 1) gives a probability of 0.9 to the IS92a (Business-as-Usual scenario), and 0.05 to each of the IS92c and IS92e scenarios. We also consider another case (Case 2) where all three scenarios are equally probable. For the purpose of sensitivity analysis we have chosen these two cases but in reality for deciding on possible vectors of scenario probabilities expert opinion should be sought.

The calculation of the estimates given in Table 5.5 are based on  $N=1000$  model simulations in which all model input factors are assumed to follow uniform distribution over their assigned ranges. For the two 8-compartment models the uncertainties about the initial conditions are considered.

Then, considering each of the two cases, and applying Equations (5.1) and (5.2) with the estimates given in Table 5.5 we obtain the results summarised in Table 5.6. For the 25-compartment model, it can be seen that the percentage of

**Table 5.5.** Estimated scenario-specific means and standard deviations of Atmospheric  $\text{CO}_2$  content in 2100 from all three models, together with two sets of scenario probabilities.

Model	Scenario	Mean ( $\hat{\mu}_i$ )	SD ( $\hat{\sigma}_i$ )	Scenario prob. ( $p_i$ )	
				Case 1	Case 2
25-compartment model	IS92a	1926.9	1111.7	0.90	1/3
	IS92c	1402.9	995.3	0.05	1/3
	IS92e	2455.6	1198.0	0.05	1/3
8-compartment Model I	IS92a	920.9	67.0	0.90	1/3
	IS92c	755.4	67.0	0.05	1/3
	IS92e	1079.9	67.0	0.05	1/3
8-compartment Model II	IS92a	825.2	48.1	0.90	1/3
	IS92c	688.3	48.1	0.05	1/3
	IS92e	959.9	48.1	0.05	1/3

variance arising from scenario uncertainty is quite small in both cases of scenario probabilities, about 2% in Case 1 and 13% in Case 2. This confirms what we had already seen in Figure 5.18, where there was actually very little variation between the prediction uncertainty due to different scenarios, for this particular model. As for the two 8-compartment models, Table 5.6 shows that in Case 1 the contributions of scenario and input factor uncertainties to the overall uncertainty are about the same with input factor uncertainty contributing slightly more, but when the scenarios assumed to have equal probability of occurrence (Case 2) the situation is the other way around, that is the percentage of variance arising

**Table 5.6.** Results from partitioning the total uncertainty in predicted Atmospheric CO<sub>2</sub> content in 2100 into ‘between scenarios’ and ‘due to input factors within scenarios’ components as a function of scenario probabilities. The results are given for all three models.

Model	Summary of the Results	Results with Scenario probabilities	
		Case 1	Case 2
25-compartment model	Overall mean ( $\hat{\mu}$ )	1927.14	1928.47
	Overall variance ( $\hat{\sigma}^2$ )	1261285.00	1405265.00
	Between-scenario variance ( $\hat{\sigma}_{BS}^2$ )	27704.93	184697.40
	Within-scenario variance ( $\hat{\sigma}_{WS}^2$ )	1233581.00	1220568.00
	% of variance between scenarios	2.0	13.0
	% of variance due to input factors within scenarios	98.0	87.0
8-compartment Model I	Overall mean ( $\hat{\mu}$ )	920.58	918.73
	Overall variance ( $\hat{\sigma}^2$ )	7122.46	22041.39
	Between-scenario variance ( $\hat{\sigma}_{BS}^2$ )	2633.46	17552.39
	Within-scenario variance ( $\hat{\sigma}_{WS}^2$ )	4489.00	4489.00
	% of variance between scenarios	37.0	79.6
	% of variance due to input factors within scenarios	63.0	20.4
8-compartment Model II	Overall mean ( $\hat{\mu}$ )	825.09	824.47
	Overall variance ( $\hat{\sigma}^2$ )	4157.88	14608.31
	Between-scenario variance ( $\hat{\sigma}_{BS}^2$ )	1844.27	12294.7
	Within-scenario variance ( $\hat{\sigma}_{WS}^2$ )	2313.61	2313.61
	% of variance between scenarios	44.4	84.2
	% of variance due to input factors within scenarios	55.6	15.8

from scenario uncertainty is much higher, about 80% with Model I and 84% with Model II.

Above we have partitioned the prediction uncertainty between scenario and input factors within scenario, for each particular model, now we wish to go a step further and partition the prediction uncertainty between scenarios, between models within scenarios, and between input factors within models and scenarios, again following Draper's model uncertainty audit approach (see [26]). In this case, where we want to partition the overall uncertainty about  $y$  into three components, the situation is slightly more complicated.

There are  $s = 3$  scenarios,  $m = 3$  models and the models are given equal weights  $(w_1, w_2, w_3) = (1/3, 1/3, 1/3)$ . With  $i$  indexing scenarios and  $j$  models, the nine values of atmospheric CO<sub>2</sub> content in year 2100 (i.e.,  $\hat{y}_{ij}$ ) are calculated. Table 5.7 gives the scenario-specific means  $\hat{\mu}_i = \sum_{j=1}^m w_j \hat{y}_{ij}$  and standard deviations  $\hat{\sigma}_i = [\sum_{j=1}^m w_i (\hat{y}_{ij} - \hat{\mu}_i)^2]^{1/2}$  computed using these predictions, together with the probability assessments  $(p_1, p_2, p_3)$  for the three scenarios. As before the two cases with different sets of scenario probabilities is investigated here.

With  $y$  as the atmospheric CO<sub>2</sub> content in 2100,  $x$  as the means and standard deviations given in Table 5.7 and  $\hat{\sigma}_{ij}^2$  as the predictive variance conditional on the scenario and model, which are assumed to be independent, the overall mean and variance equations in this case are as follows:

$$\hat{\mu} = E_S[\hat{E}_M\{\hat{E}(y|x, M, S)\}] = \sum_{i=1}^s p_i \hat{\mu}_i, \quad (5.3)$$

and

$$\begin{aligned} \hat{\sigma}^2 &= V_S[\hat{E}_M\{\hat{E}(y|x, M, S)\}] + E_S[\hat{V}_M\{\hat{E}(y|x, M, S)\}] + E_S[\hat{E}_M\{\hat{V}(y|x, M, S)\}] \\ &= \sum_{i=1}^s p_i (\hat{\mu}_i - \hat{\mu})^2 + \sum_{i=1}^s p_i \hat{\sigma}_i^2 + \sum_{i=1}^s \sum_{j=1}^m p_i w_j \hat{\sigma}_{ij}^2 = \hat{\sigma}_{BS}^2 + \hat{\sigma}_{BMWS}^2 + \hat{\sigma}_{BPWMS}^2 \end{aligned} \quad (5.4)$$

respectively, where  $S$  standing for scenario and  $M$  standing for model.

**Table 5.7.** Estimated scenario-specific means and standard deviations of Atmospheric CO<sub>2</sub> content in 2100, together with two sets of scenario probabilities.

Scenario $i$	Mean ( $\hat{\mu}_i$ )	SD ( $\hat{\sigma}_i$ )	Scenario probability ( $p_i$ )	
			Case 1	Case 2
IS92a	1136.90	382.02	0.90	1/3
IS92c	848.32	186.56	0.05	1/3
IS92e	1422.30	579.17	0.05	1/3

As the results in Table 5.8 reveal in both cases, with different set of probabilities given to the three scenarios, what is determining the overall uncertainty appears to be the input factors within model. The input factor uncertainty attribute about 72% in Case 1 and 64% in Case 2 of the overall uncertainty. The least contribution to the overall uncertainty is due to scenarios, less than 2% in Case 1 and about 9% in Case 2. The uncertainty contribution of models to overall prediction uncertainty is moderate when we assume that they are equally likely.

**Table 5.8.** Results from partitioning the total uncertainty in predicted Atmospheric CO<sub>2</sub> content in 2100 into ‘between scenarios ( $BS$ )’, ‘between models within scenarios ( $BMWS$ )’ and ‘between predictions within models and scenarios ( $BPWMS$ )’ components as a function of scenario probabilities.

Summary of the Results	Results with Scenario probabilities	
	Case 1	Case 2
Overall mean ( $\hat{\mu}$ )	1136.74	1135.84
Overall variance ( $\hat{\sigma}^2$ )	571549.17	636087.42
Between scenario variance ( $\hat{\sigma}_{BS}^2$ )	8236.55	54909.41
Between models within-scenario variance ( $\hat{\sigma}_{BMWS}^2$ )	149854.78	172057.81
Between predictions within models & scenarios variance ( $\hat{\sigma}_{BPWMS}^2$ )	413457.84	409120.20
% of variance between scenarios	1.44	8.63
% of variance between models within scenarios	26.22	27.05
% of variance between predictions within models & scenarios	72.34	64.32

The results revealed that with equal weights given to the three models, no matter which of the two sets of scenario probabilities is taken, the uncertainty about the input factors within each model is the major source of uncertainty in the predictions. The contribution of model uncertainty to the overall uncertainty is reasonably high, while the contribution of scenario uncertainty on the overall prediction uncertainty is quite low.

## 5.9 Modeller Uncertainty

In this section, we identify another important source of uncertainty: modeller uncertainty. Firstly, we shall note that by the terminology '*modeller*' we do not mean model developer, we mean model user. Uncertainty due to modeller is not usually taken into account in uncertainty analysis, and the potential importance of modeller interpretation rarely seems to be recognised [2].

Uncertainty may arise because: different modellers' interpretation of both the model and the scenario can vary; modeller's limited experience of modelling concepts, and with incomplete knowledge of the situation to be modelled can contribute significantly to the model predictions.

We look at this aspect of uncertainty in the context of environmental radioactivity modelling, using real world data from Chernobyl accident. In this section, first a brief background of the accident and its consequence is included, then description of the study and data is given. Following that, the predictions made by different modellers are analysed using some statistical tools for addressing modeller uncertainty.

### 5.9.1 General Background

In 1986, a major accident occurred at the former Soviet Union's Chernobyl nuclear power station when two explosions destroyed the core of Unit 4 and the reactor building, and large amounts of radioactive substances were released to the atmosphere. The radioactive isotopes of iodine ( $^{131}\text{I}$ ) and cesium ( $^{137}\text{Cs}$ ) were

two of the radiologically most important isotopes widely dispersed and eventually deposited onto the surface of the earth over the ensuing ten days or so.

Radioactive material released from Chernobyl was distributed throughout the northern hemisphere, mainly across Europe and deposited by the actions of wind and rain. These substances were available for uptake by plants and animals, hence to foodstuffs and ultimately into man.

The radionuclide composition of the release and of the subsequent deposition on the ground also varied considerably during the accident due to variation in temperature and other parameters during the release.  $^{137}\text{Cs}$  was selected to characterise the magnitude of the ground deposition because: (1) it is easily measurable, and (2) it was the main contributor to the radiation doses received by the population once the short-lived  $^{131}\text{I}$  had decayed [85].

### 5.9.2 Description of the Study and Data

Researchers, regulatory agencies such as IAEA (International Atomic Energy Agency), NEA (Nuclear Energy Agency) and WHO (World Health Organization), environmental assessment groups within the programmes like VAMP (Validation of Environmental Model Predictions) and BIOMOVs (Biospheric Model Validation Study) have been using various kinds of computer codes to predict the movement and the level of radioactive material in various parts of the biota, over time.

These models are used mainly to evaluate situations in which only limited measurements are available. They can also be used to reconstruct past situations given the existence of partial or incomplete information. When decisions of major economic or social importance are based on model results, it is essential, for the sake of scientific and public credibility, to establish a degree of confidence in these results [104].

Three terrestrial food chain codes called CHERPAC, RUINS, and CLRP were obtained and run by modellers for several scenarios (such as BREMEN and FORT COLLINS) involving the transfer of radionuclides  $^{131}\text{I}$  and  $^{137}\text{Cs}$  from air to pasture and milk. The detailed descriptions of the scenarios and the manual for the

computer models can be found in Ref.[1]. In this exercise, we report on the results of the BREMEN scenario, and the CHERPAC computer code. The radionuclide considered here is  $^{137}\text{Cs}$ , a long-lived radionuclide deposited in the environment after the accident at Chernobyl.

A brief description of the BREMEN scenario and the CHERPAC computer code adopted for this study is as follows:

The BREMEN scenario is based on the data collected for the pasture-cow-milk pathway in Bremen/Germany after the Chernobyl accident.

CHERPAC (CHalk River Environmental Research Pathways Analysis Code) was developed by Peterson in 1994. It is a time-dependent stochastic code, and calculates output from a short-term release. Output can be calculated daily (for up to 60 days) concentrations of  $^{137}\text{Cs}$  in some foodstuffs (on fresh-weight pasture and in milk) or averaged monthly concentrations in others [93].

In September 1994, the CHERPAC code and description of the BREMEN scenario were sent out to the interested modellers. Ten participants submitted results from this code. The modellers involved in the experiment were provided with information about the accident, the particular region, Bremen, the effect of the accident on this particular region, and background information regarding the diet and pasture rotation for the one cow from which milk samples were taken. The participants were also given manuals explaining how to run the code and how to implement changes in input parameters for the model. These manuals did not have any information about the formulation of the code or how it was implemented numerically. Each modeller was provided with the same basic information.

The driving data for the CHERPAC code were concentrations of  $^{137}\text{Cs}$  aerosols on pasture and in milk at Bremen. The considered time periods for CHERPAC code are May 5 - June 27, 1986 for concentrations on pasture and May 14 - June 27, 1986 for concentrations in milk.

The participants were to use the same assumptions about the scenario in the code. Most of the modellers involved in this exercise were familiar with general modelling concepts, and models for the environmental transport of radioactivity. Modellers were encouraged to seek opinion and advice from experts who did not participate in the experiment, but were not allowed to discuss the exercise with participants. Questions asked by modellers regarding the code and the scenario were answered and the answers distributed to the participants unless answering those questions would have an effect on assumptions. Each modeller's results is labeled by a letter, A through J.

This test exercise was carried out as a so called "blind test", i. e. the modellers received a scenario description (input data), and after they had completed the calculations and submitted the results to the co-ordinator, they were provided with the observed data (test data) that matched the endpoints asked for in the scenario description.  $^{137}\text{Cs}$  concentrations in pasture ( $\text{Bq kg}^{-1}$ ) were measured in the time period May 5 - June 28, 1986 (at 20 time points).

It is well known that evaluation of the impact of radionuclide releases on humans and on the environment is important. Computer codes and the predictions made with them help us to assess the potential routes and the levels of those radionuclide releases. The fundamental purpose of this section is to investigate the uncertainty in the predictions due to modellers who used the same computer model, within BREMEN scenario to make predictions. For convenience, we shall use 'modeller prediction' terminology instead of 'model prediction made by modellers'. The steps followed here are: (i) intercomparison of modeller predictions; (ii) comparison of modeller predictions with the observed data; (iii) finding out if there is any apparent grouping amongst modeller predictions, and measured data.

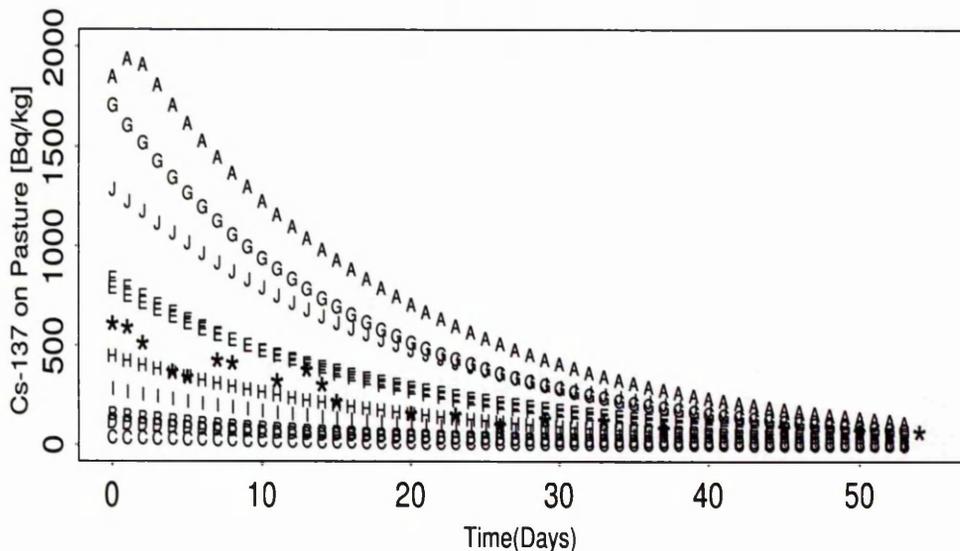
### 5.9.3 Testing Modeller Predictions

In order to compare the modeller predictions, first of all we assume that the model is being used by the different modellers with the same purpose. It is also

assumed that the modellers are independent of each other. Several graphical and statistical techniques have been used to make comparisons.

### 5.9.3.1 Exploratory Data Analysis

Predictions of daily concentrations on pasture are shown in Figure 5.22. The figure shows that there is a considerable decrease in concentrations for all modeller predictions and observed  $^{137}\text{Cs}$  concentrations. The potential influence of the modeller on model results can easily be seen. The uncertainty in the predictions due to modeller appears to be quite high. The modeller predictions especially E and F; B and D; A, G and J seem to give quite similar results. Modeller predictions H, I, E and F appear to be the ones closest to the observed concentrations. Some of the modellers, especially A, G and J, show a tendency for overestimating  $^{137}\text{Cs}$  activity in pasture compared to the observed data.



**Figure 5.22.** Measured concentrations, and model predictions calculated with *CHERPAC* on daily concentrations of  $^{137}\text{Cs}$  on Pasture from May 5 - June 27, 1986. The measured data are indicated by stars.

5.9.3.2 Modelling the Modeller Response

Now we perform a more quantitative analysis. The least squares method has been used to estimate the trend using the same model form for each modeller prediction. The main reason for modelling the changes in concentration is to provide summaries of the parameters, slopes and intercepts, to provide some idea about the variation in slopes or the intercepts across a group of modeller predictions. A simple linear relationship of log-transformed concentration against time is found to be sufficient to fit all ten modeller predictions, and the measured data. See Figure 5.23 for the plot of the log-transformed data. The model fitted to the predictions provided by each modeller, and the measurement data is:  $\log(\text{Concentration}_i) = \alpha + \beta \cdot \text{Day}_i + \epsilon_i$ , where  $\epsilon_i \sim N(0, \sigma^2)$ .

We can note that the logarithmic transformation on concentrations has led to a linear regression relation. The lowest  $R^2$  value of 99% shows that the regression equations explain almost all the variation in concentration for all the ten modeller

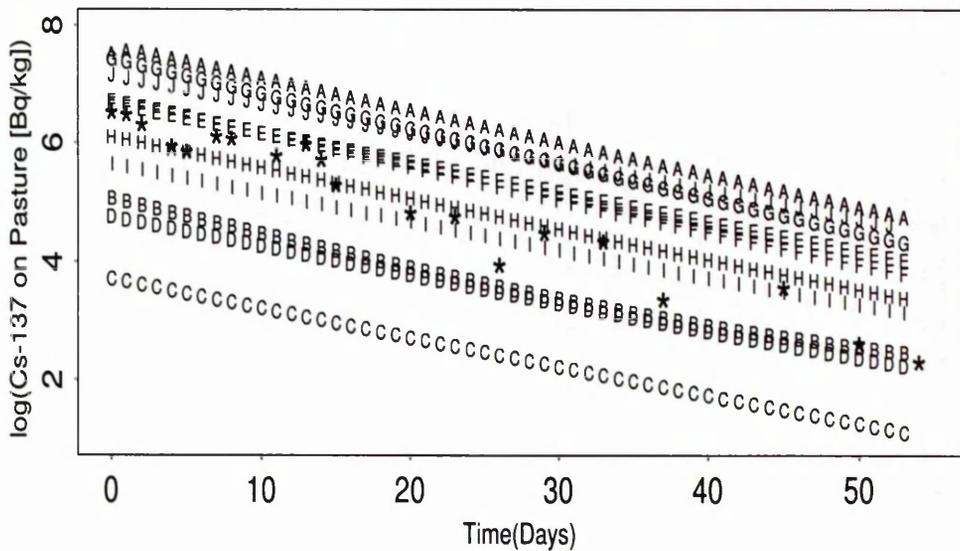
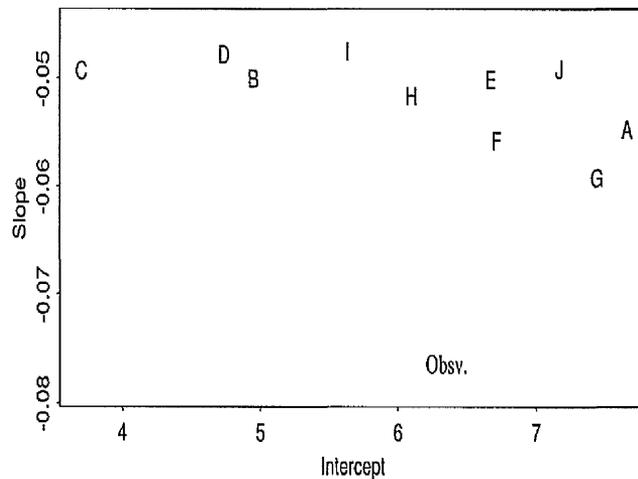


Figure 5.23. Log-transformed measured concentrations, and modeller predictions calculated with *CHERPAC* on daily concentrations of <sup>137</sup>Cs on Pasture. The measured data are indicated by stars.

predictions of PASTURE data. For the observed data  $R^2$  value is found to be 95% which is reasonably high.

The parameters of these linear relationships, the slopes and intercepts, are plotted in Figure 5.24. Each point in the figure is labeled by the corresponding modeller. In terms of the intercepts, we can see that there is quite a lot of variability. The intercepts are varying between 3.7 (associated with log-transformed modeller C predictions) and 7.6 (associated with log-transformed modeller A predictions). The mean and the standard deviation of the intercepts from the ten modellers predictions are 6.1 and 1.3, respectively, and the intercept of the log-transformed observations is 6.3. Comparison of the intercepts associated with the ten modellers' predictions and the observed data show that modellers H, E and F provided predictions which are closer to the observed data.



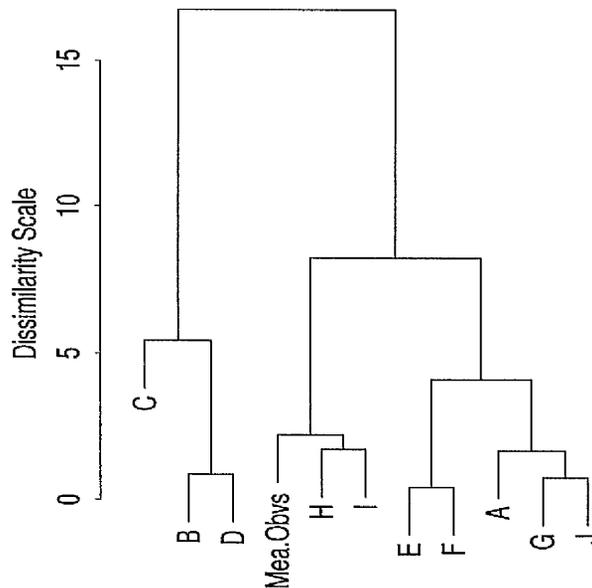
**Figure 5.24.** Scatterplots of estimated model parameters (Intercepts vs. Slopes) of log-transformed data.

In terms of the slopes, it is clear from Figure 5.24 that the slopes associated with the modellers appears to be close together and slightly higher than the slope of the observations. The mean and the standard deviation of the slopes from the modellers' predictions are -0.052 and 0.004, respectively. The slope of the observed data is -0.076.

### 5.9.3.3 Clustering

A cluster analysis is carried out to find the natural groupings, if any, of the ten modeller predictions. In order to carry out a cluster analysis first the similarity (or dissimilarity) of every pair of modeller predictions are measured. There are many ways of doing this. Standardized Euclidean distance is one of the most common measures of dissimilarity and it is used here. On the calculated distances, complete-linkage clustering has been used resulting in a plot of a clustering tree (often called a dendrogram), has been created for log-transformed data (See Figure 5.25).

In this dendrogram, we see a simple group structure, and a measure of “closeness” or “similarity” of the modeller predictions and the measurements.



**Figure 5.25.** Dendrogram of log-transformed measured concentrations and modeller predictions of daily  $^{137}\text{Cs}$  concentrations on Pasture.

## 5.10 Discussion

Uncertainty in predictions of models arises from, among other sources, the model structure representing the real world system, the scenario reflecting different views of the future, the uncertainty in the model input factors, the assumptions made by the model users. In this chapter we have looked at scenario uncertainty, model uncertainty and input factor uncertainty within the framework of the three GCC models. We also explored modeller uncertainty with a case study.

Within a model and a scenario, the prediction uncertainty can be partitioned between input factors, using methods like stepwise regression. But as McKay emphasizes [83] one should avoid using statements like “30% of the uncertainty in a dependent variable (say  $y$ ) is due to an input parameter (say  $x_1$ )” which presupposes a quantitative measure and can be quite misleading, depending on how well the probability distribution of  $y$  is summarised by the model. An example of a more precise statement would be “On average, the variance of  $y$  is 30% less when  $x_1$  is fixed than when it is varied; average is with respect to the distribution of  $x_1$ ”.

Prediction uncertainty due to input factors uncertainty was investigated in Chapters 3 and 4 in detail. In this chapter we have assessed the effect of sampling technique, sample size and distributional assumptions for the input factors on model predictions. First, with a graphical presentation we have demonstrated that holding an important input factor (or a set of important factors) at its best estimate value leads to a substantial reduction in prediction uncertainty. It is also shown that there is tendency for increasing uncertainty in the predictions with time.

To assure the full coverage of the input space it is important to use a sufficiently large number of runs in model simulations. It is shown in Section 5.5.1 that the Latin hypercube sampling produces more stable cumulative distribution function estimates of model output variables than those produced by the simple random sampling. It is also shown that increasing the sample size improves the quality of the estimate and hence the reliability of the results. The number of

model runs needed in the analysis depends on the complexity of the model and the number of model input factors.

Different distributional assumptions can have a dramatic effect on model output distribution, and also on input factor selection. Therefore, it is important to recognise this effect on analysis results and investigate its occurrence. An input factor may be identified as important when sampled on a specified range with one distribution and identified as relatively less important or even unimportant when sampled on the same range with a different distribution.

When investigating scenario uncertainty we only assessed the uncertainty caused by changing fossil fuel and land-use emissions and uncertainty in each model compartment as a result of those emissions. It would be of interest to assess the scenario uncertainty arising from scenario specific parameters as well.

When input factor uncertainty was taken into account, it was seen that for the 8-compartment models, on the basis of overall uncertainties discrimination between the predictions for the three scenarios can be made for most of the output variables considered in the analysis, but as for the 25-compartment model no discrimination between the predictions for the three scenarios can be made.

Uncertainty in the predictions due to model was found to be much higher with the 8-compartment models, due to the fact that they are quite simple, highly linear and do not capture physical processes we see in real world.

In order to reduce the uncertainty in model predictions, we need to know which source of uncertainty is determining the overall uncertainty. Such information can help to set priorities for future research, model and scenario improvements and developments.

First, focusing on uncertainty in the predictions resulting from scenario and input factor uncertainties, we have partitioned the uncertainty in  $y_{Atm}(t = 2100)$  between these two sources of uncertainty, within a particular model. For this we have considered two possible vectors of scenario probabilities. The results of this partitioning confirmed what we have found out when investigating scenario uncertainty. That is, for the 25-compartment model, the contribution of input

factor uncertainty to the prediction uncertainty is significantly high no matter which vector of scenario probabilities is considered. As for the 8-compartment models, we have seen that the contributions of input factor and scenario uncertainties to the prediction uncertainty changes with different cases of scenario probabilities. When the scenarios are given equal probability the uncertainty about the scenarios influences the prediction uncertainty much more. Again this confirms what we have seen in the scenario uncertainty section.

A further attempt involved the partitioning of the overall uncertainty into three components: between scenario, between models within scenarios, and between predictions within models and scenarios. Results of this investigation revealed that, for the particular models and scenarios under consideration, when equal weights given to the three models and with the chosen two sets of scenario probabilities the model input factors are dominating the prediction uncertainties, whereas the contribution of model uncertainty on the overall uncertainty is reasonable but the contribution of scenario uncertainty is very small.

With a case study we assessed the ten modellers' influence on model predictions by comparing differences in the predictions obtained by different modellers even though they used the same model and the same scenario description. We have seen that the potential influence of modeller's interpretation of model and scenario on the model results can be quite significant and should be taken into account.

# Chapter 6

## Conclusions

As Chapter 1 indicates our knowledge about the real world is always incomplete, and to understand the real world phenomena we need models. But the models, at best, are only approximations of the system being modeled, and thus are inherently uncertain. While quantifying the uncertainty associated with the process of building a model is not possible, there are ways of reducing, and perhaps more importantly, highlighting areas of uncertainty. Among many other sources of uncertainty, the key sources of uncertainties which we have explored in this thesis are: (i) structural uncertainty resulting from incomplete/improper mathematical formulation of conceptual models; (ii) scenario uncertainty resulting from lack of knowledge or inability to predict the future conditions; and (iii) input factor uncertainty resulting from vaguely defined, estimated input factors.

An important step towards reducing the uncertainty in model predictions and hence increasing the reliability of a model is to obtain more accurate information on the main sources of uncertainty. For this purpose it is necessary to carry out sensitivity and uncertainty analyses. These important modelling tools can lead to improved understanding of the fundamental processes being modeled.

Sensitivity analysis (SA) methods are broadly classified into three categories: screening methods, local SA methods and global SA methods. When dealing with a very complex, computationally expensive model with hundreds of input

parameters, screening methods can be used to identify the most important input factors which control most of the output uncertainty. Local SA methods provide information on the local impact of the input factors on the output variables. They are based on taking partial derivatives of the output functions with respect to the input factors at selected points in the parameter space. Even though local methods have the disadvantage of providing only local measures, they also have an important advantage of not being dependent on chosen distributions for input factors. Global SA methods which are widely used in the analysis of large complex computational models require information about the distributions, minimum and maximum values for the input factors as well as mean values and standard deviations.

In this thesis we have used sensitivity and uncertainty analysis tools with a view to understanding the uncertainty issues related to compartmental models, in particular some compartmental models which have been used for global carbon cycle (GCC) modelling.

We have considered three GCC models in this thesis, two of which are quite simple providing no detailed biological and chemical information on complex carbon cycle process. The third model consists of 25 compartments is a more complex model and believed to be a more realistic representation of carbon cycle processes. The application of sensitivity and uncertainty analyses techniques to these models revealed the following benefits and limitations:

- Even though a one-factor-at-a time (OAT) approach produces some benefits, it is a very inefficient way of performing a SA and is limited in application to models that are not computationally expensive and have a small number of input factors. Compared to computer models available in scientific literature with hundreds of input factors, the models we utilize in this thesis are relatively small. Even so, the application of OAT design was quite impractical.
- A very important disadvantage of OAT design is that it varies only one input factor at a time while fixing the others at their nominal values - frequently

not valid for the models in which the input factors are related and can result in a confused picture of how input factors affect model behaviour. This is the case with the two GCC models for which we have to take into account the fact that the system has to be in steady-state before introducing any perturbations to the system. So to maintain this steady-state condition, when one input factor is varied over its entire range at least one other input factor has to be changed to initialize the GCC model. As a result, the carbon content of each compartment changes during a simulation as a function of not only the factor varied but also the one(s) that are calculated, of course in addition to the time-dependent releases of carbon from fossil fuel emissions and land-use change.

- The use of an efficient screening design can help to avoid high computational and financial costs, as well as physical and human resources. The Morris design, we have applied to the three global carbon cycle models, is an OAT design but because it covers the entire space over which the input factors vary it is considered as a global screening design. It is also an economical screening design and should be preferred to the other screening designs discussed in this thesis.
- Even though the global SA methods require a higher number of model evaluations, they provide quantitative sensitivity measures, and should be preferred over local SA methods.
- When global SA methods were applied to the two 8-compartment models, in order to maintain a steady-state condition in the system, a subset of transfer coefficients were treated as uncertain input factors while others are used in the model calibration process. For this last setting, we have approached the necessary constraint from the linear algebra point of view using Gauss-Jordan elimination.
- Windowing analysis was performed on the 8-compartment models since

large uncertainty in model initial conditions was causing the model output to deviate from an acceptable trajectory. Although this problem was resolved by discarding ‘unsatisfactory’ simulations, a large number of model runs were needed to obtain an adequate sample of simulations. Windowing analysis is also being used in order to adjust the values of model input parameters to achieve an acceptable match between observed and predicted model conditions (for an example see [40]). According to King & Sale [73], windowing can also provide a correlation structure between the input factors as well as estimates of input factor ranges.

- A major objective in the SA of a model is to obtain a ranking of the input factors. SA techniques including sensitivity indices, standardised ranges, Morris method, derivative based method, regression based methods, correlation coefficients, non-parametric tests on partitioned data sets. It appears that the results we have from these techniques all very broadly agree. The rankings obtained from these techniques do not always agree but the subset of potentially important input factors that account for most of the variation in model output variables are identified by all the techniques considered. Thus, it is reasonable to conclude that most of the techniques would be appropriate for SA for the type of models considered in this thesis.
- Using stepwise regression technique or some other regression technique, relative input factor importance can readily be determined. In the application of such techniques, a criterion such as PRESS should be used to protect against overfitting the data. The standardized regression coefficients (SRC) and partial correlation coefficients (PCC) are used as a measure of input factor importance. The estimated coefficients proved to be a useful way of presenting sensitivity results for output variables which were functions of time. The sign of these coefficients indicates whether the output variable increases or decreases as the associated input variable increases. In this study, regression based global SA methods proved to perform well due to

the fact that the output variables under consideration were linear function of the input factors. However, when there is no linear relationship between the variables, the nonparametric equivalent of these methods should be used.

- In a time-dependent multi-output model a single ranking of model input factors is not possible. Uncertainty contributions of model input factors differ for various outputs and at different time points. In SA individual input factors were examined to determine whether their importance changes through time. We have found that different sets of input factors are important for different model outputs and also for different model output times. For instance, the uncertainty contribution of TSL (the terrestrial turnover time in the active soil compartment of the 25-compartment model) was found to be much higher when the output of the active soil compartment was considered (see Chapter 4).
- In this thesis, we also extended the use of some of the modern types of graphical techniques, such as star plots and dotplots, to present analysis results. These graphical tools prove to be good ways of picking out the most important input factors identified by the sensitivity measure at a glance.
- In Chapters 3 and 4 we have assumed that all model inputs follow a uniform distributions on uncertainty ranges obtained by applying  $\pm 20\%$  of their reference values. Such an assumption is usually made when knowledge about the input factor is poor. However, it should be recognised that changes in distributions and in uncertainty ranges may cause substantial changes in sensitivity analysis results. When needed and possible, expert elicitation should be used to construct the probability density functions for the uncertain model parameters.

As emphasized by Iman *et al.* [64] selection of uncertainty ranges and probability distributions for model input factors is an important issue and it is an area of sensitivity analysis where interaction among experts in the

process being modeled, model developers and those conducting SA is very important.

- The type of probability density function of the output variable depends on the mathematical structure of the model. In Chapter 5 we have shown that because of the highly linear and simple structure of the 8-compartment models the probability distribution of the output variables is not different from that of the input factors when all input factors are assigned the same probability distribution. On the other hand, with the 25-compartment model which has a more complex structure, we have seen that although the same probability distribution is assigned to all 30 input factors of the model, an output variable can follow a different probability distribution.
- The effect of sampling technique and sample size on the model predictions was investigated and it was shown that LHS gives more stable cumulative distribution functions of the output variables and requires fewer model evaluations. The number of model runs needed in the analysis is shown to be a function of model complexity and the number of model input factors.

There are many different sources of uncertainty and in the computer experiments that we have done we have been able to look at scenario uncertainty, model uncertainty and input factor uncertainty within the framework of the three GCC models. We also explored modeller uncertainty with a case study based on radioactivity modelling.

- We know that no model is right. When model uncertainty was investigated using the three GCC models it was found that there was considerable amounts of model uncertainty in the 8-compartment models and this is believed to be mostly related to the fact that the real world processes are not being reflected in the mathematical structure of the models.
- In the scenario uncertainty we explored the amount of uncertainty in the

predictions due to three emission scenarios both with and without input factor uncertainty included. The results showed that for the 25-compartment model the uncertainty in the predictions were not determined by the scenario uncertainty, but for the 8-compartment models, the uncertainty in the predictions was highly influenced by scenario uncertainty.

- The total prediction uncertainty of a model output variable (a compartmental CO<sub>2</sub> content at a certain time, in our case) is a combination of uncertainties from different sources like scenario, model structure, input factors within a particular model. It was the main interest of Chapter 5 to find out the relative contribution of these three main sources of uncertainties to the overall uncertainty. First, focusing on scenario and input factor within each model, we have partitioned the prediction uncertainty between scenario and input factors. The results of this analysis confirmed what we have found out from the analysis of scenario uncertainty.
- The overall uncertainty due to model structure, scenario and input factor uncertainty was partitioned between these three sources of uncertainty. The analysis results showed that, for the particular models and scenarios under consideration in this investigation, when equal weights are given to the three models and with the chosen two sets of scenario probabilities the model input factors dominate the overall prediction uncertainties, whereas the contribution of model uncertainty on the overall uncertainty was found to be reasonably high but the contribution of scenario uncertainty to be very small.
- Another source of uncertainty which is often ignored is modeller uncertainty. A comparison between the measured data and the ten sets of predictions made by ten modellers who used the same model and the same scenario description to make their predictions showed that there can be large discrepancies in the model results due to the modeller's interpretation of the model and scenario. This case study revealed how significant this source of

uncertainty can be.

We believe that sensitivity and uncertainty analysis is an essential tool box for every modeller. With the use of sensitivity and uncertainty analysis we can obtain valuable information which can help us in deciding which processes included in our model will benefit most from improvement. Thus, spending a large amount of effort in improving/characterizing aspects of the model which have very little influence on the predictions can be avoided.

Furthermore, there is no method of sensitivity and uncertainty analysis that we can rate as superior to another method, but different methods have both strengths and weaknesses. How to decide which technique to use depends on the model under consideration and the type of information is needed. However it is advisable to use more than one method and compare the results.

# Bibliography

- [1] Qualitative and quantitative guidelines for the comparison of environmental model predictions. Technical Report No 7, BIOMOVS II, Swedish Radiation Protection Ins., Sweden, 1995.
- [2] An overview of the BIOMOVS II study and its findings. Technical Report No 17, BIOMOVS II, Swedish Radiation Protection Ins., Sweden, 1996.
- [3] C. Baird. *Environmental Chemistry*. W. H. Freeman and Company, New York, 1995.
- [4] J. M. Barnola, D. Raynaud, Y. S. Korotkevich, and C. Lorius. Vostok ice core provides 160,000-year record of atmospheric CO<sub>2</sub>. *Nature*, 329(1):408–414, 1987.
- [5] W. E. Boyce and R. C. DiPrima. *Elementary Differential Equations*. John Wiley & Sons, New York, USA, 6th edition, 1997.
- [6] R. F. Brown. Compartmental system analysis: State of the art. *IEEE Transactions on Biomedical Engineering*, BME-27(1):1–11, 1980.
- [7] N. J. Bunce. *Introduction to Environmental Chemistry*. Weurz Pub., Winnipeg, Canada, 1993.
- [8] R. P. Bush, G. M. Smith, and I. F. White. Carbon-14 waste management. Technical Report EUR 8749EN, Commission of the European Communities, Brussels, Luxembourg, 1984.

- [9] F. Campolongo and R. Braddock. The use of graph theory in the sensitivity analysis of the model output: A second order screening method. *Reliability Engineering and System Safety*, 64:1–12, 1999.
- [10] F. Campolongo and A. Gabric. The parametric sensitivity of dimethylsulfide flux in the southern ocean. *J. Stat. Comput. Simul.*, 57:337–352, 1997.
- [11] F. Campolongo and J. P. C. Kleijnen. Screening methods. In A. Saltelli, K. Chan, and E. M. Scott, editors, *Sensitivity Analysis*, Probability and Statistics series. John Wiley & Sons, 2000.
- [12] F. Campolongo and A. Saltelli. Sensitivity analysis of an environmental model: An application of different analysis methods. *Reliability Engineering and System Safety*, 57:49–69, 1997.
- [13] F. Campolongo and A. Saltelli. Comparing different SA methods on a chemical reactions model. In A. Saltelli, K. Chan, and E. M. Scott, editors, *Sensitivity Analysis*, Probability and Statistics series. John Wiley & Sons, 2000.
- [14] F. Campolongo, A. Saltelli, T. Sørensen, and S. Tarantola. Hitchhiker's guide to sensitivity analysis. In A. Saltelli, K. Chan, and E. M. Scott, editors, *Sensitivity Analysis*, Probability and Statistics series. John Wiley & Sons, 2000.
- [15] F. Campolongo, S. Tarantola, and A. Saltelli. Tackling quantitatively large dimensionality problems. *Comp. Phy. Communications*, 117:75–85, 1999.
- [16] J. M. Chambers and T. J. Hastie, editors. *Statistical Models in S*. Chapman & Hall, London, 1993. AT&T Bell Laboratories.
- [17] C. Chatfield. Model uncertainty, data mining and statistical inference. *J. R. Statist. Soc. A*, 158(3):419–466, 1995.

- [18] W. Cheney and D. Kincaid. *Numerical Mathematics and Computing*. Brooks / Cole Publishing Company, California, 2nd edition, 1985.
- [19] M.H. Chun, S.J. Han, and N.I. Tak. An uncertainty importance measure using a distance metric for the change in a cumulative distribution function. *Reliability Engineering and System Safety*, 70:313–321, 2000.
- [20] W. S. Cleveland. *Visualizing data*. Hobart Press, Murray Hill, N.J., 1993.
- [21] C. Cobelli and G. Romanin-Jacur. Controllability, observability and structural identifiability of multi input and multi output biological compartmental systems. *IEEE Transactions on Biomedical Engineering*, BME-23(2):93–100, 1976.
- [22] W. J. Conover. *Practical Nonparametric Statistics*. John Wiley & Sons, New York, 1980.
- [23] R. M. Cooke and J. M. van Noortwijk. Graphical methods. In A. Saltelli, K. Chan, and E. M. Scott, editors, *Sensitivity Analysis, Probability and Statistics series*. John Wiley & Sons, 2000.
- [24] M. Crosetto, S. Tarantola, and A. Saltelli. Sensitivity and uncertainty in spatial modelling based on GIS. *Agriculture, Ecosystems and Environment*, 81:71–79, 2000.
- [25] E. P. Dougherty, J. T. Hwang, and Rabitz H. Further developments and applications of the Green's function method of sensitivity analysis in chemical kinetics. *Journal of Chemical Physics*, 71:1794–1808, 1979.
- [26] D. Draper. Assessment and propagation of model uncertainty. *Journal of The Royal Statistical Society*, 57-B(1):45–97, 1995.
- [27] D. Draper, A. Pereira, P. Prado, A. Saltelli, R. Cheal, S. Eguilior, B. Mendes, and S. Tarantola. Scenario and parametric uncertainty in

- GESAMAC: A methodological study in nuclear waste disposal risk assessment. *Computer Physics Communications*, 117:142–155, 1999.
- [28] S. H. C. du Toit, A. G. W. Steyn, and R. H. Stumpf. *Graphical Exploratory Data Analysis*. Springer-Verlag, New York, 1986.
- [29] A. M. Dunker. Efficient calculation of sensitivity coefficients for complex atmospheric models. *Atmospheric Environment*, 15(7):1155–1161, 1981.
- [30] W. R. Emanuel, G. G. Killough, W. M. Post, and H. H. Shugart. Modeling terrestrial ecosystems in the global carbon cycle with shifts in carbon storage capacity by land-use change. *Ecology*, 65(3):970–983, 1984.
- [31] W. R. Emanuel, G. G. Killough, W. M. Post, H. H. Shugart, and M. P. Stevenson. Computer implementation of a globally averaged model of the world carbon cycle. Technical Report TR010, DOE/NBB-0062, U. S. Department of Energy, Oak Ridge National Laboratory, Oak Ridge, TN, 1984.
- [32] I. G. Enting, T. M. L. Wigley, and M. Heimann. Future Emissions and Concentrations of Carbon Dioxide: Key Ocean/Atmosphere/Land Analyses. Technical Report No. 31, CSIRO Division of Atmospheric Research Technical Paper, 1994.
- [33] I. G. Enting, T. M. L. Wigley, and M. Heimann. Intergovernmental Panel on Climate Change (IPCC), Working Group 1, 1994: Modelling Results Relating Future Atmospheric CO<sub>2</sub> Concentrations to Industrial Emissions. Technical Report DB1009, CDIAC, <http://cdiac.esd.ornl.gov/ndps/db1009.html>, 1995.
- [34] M. Eslami. *Theory of Sensitivity in Dynamic System: An Introduction*. Springer-Verlag, Berlin Heidelberg, 1994.
- [35] B. S. Everitt and G. Dunn. *Applied Multivariate Data Analysis*. Edward Arnold, London, 1991.

- [36] P. M. Frank. *Introduction to System Theory*. Academic Press, New York, London, 1978.
- [37] H. Friedli, H. Löttscher, H. Oeschger, U. Siegenthaler, and B. Stauffer. Ice core records of the  $^{13}\text{C}/^{12}\text{C}$  ratio of atmospheric  $\text{CO}_2$  in the past two centuries. *Nature*, 324(20):237–238, 1986.
- [38] K. Godfrey. *Compartmental Models and Their Application*. Academic Press, London, 1983.
- [39] N. Goto, A. Sakoda, and M. Suzuki. Modeling of soil carbon dynamics as a part of carbon-cycle in terrestrial ecosystems. *Ecological Modelling*, 74(3-4):183–204, 1994.
- [40] T. M. Grieb, R. J. M. Hudson, N. Shang, R. C. Spear, S. A. Gherini, and R. A. Goldstein. Examination of model uncertainty and parameter interaction in a global carbon cycling model (GLOCO). *Environment International*, 25(6/7):787–803, 1999.
- [41] S. C. Gupta and V. K. Kapoor. *Fundamentals of Mathematical statistics*. Sultan Chand & Sons, New Delhi, 9th edition, 1994.
- [42] Y. Y. Haimes and J. H. Lambert. When and how can you specify a probability distribution when you don't know much? II. *Risk Analysis*, 19(1):43–46, 1999.
- [43] M. C. G. Hall. Estimating the reliability of climate model projections - steps toward a solution. In M. C. MacCracken and F. M. Luther, editors, *Projecting the Climatic Effects of Increasing Carbon Dioxide*, number DOE/ER-0237, pages 337–364. U. S. Department of Energy, Washington, D.C., 1985.
- [44] L. Hallstadius. Compartment modelling in nuclear medicine: A new program for the determination of transfer coefficients. *Nuclear Medicine Communications*, 7(6):405–414, 1986.

- [45] D. M. Hamby. A review of techniques for parameter sensitivity analysis of environmental models. *Environmental Monitoring and Assessment*, 32:135–154, 1994.
- [46] D. M. Hamby. A comparison of sensitivity analysis techniques. *Health Phys.*, 68(2):195–204, 1995.
- [47] J. C. Helton. Uncertainty and sensitivity analysis techniques for use in performance assessment for radioactive waste disposal. *Reliability Engineering and System Safety*, 42:327–367, 1993.
- [48] J. C. Helton and F. J. Davis. Sampling-based methods. In A. Saltelli, K. Chan, and E. M. Scott, editors, *Sensitivity Analysis*, Probability and Statistics series. John Wiley & Sons, 2000.
- [49] J. C. Helton and F. J. Davis. Sampling-based methods for uncertainty and sensitivity analysis. Technical Report SAND99-2240, Sandia National Laboratories, New Mexico, USA, 2000.
- [50] J. C. Helton, R. L. Iman, and J. B. Brown. Sensitivity analysis of the asymptotic behavior of a model for the environmental movement of radionuclides. *Ecological Modelling*, 28:243–278, 1985.
- [51] J. C. Helton, R. L. Iman, J. D. Johnson, and C. D. Leigh. Uncertainty and sensitivity analysis of a dry containment test problem for the MAEROS aerosol model. *Nuclear Science and Engineering*, 102(1):22–42, 1989.
- [52] J. C. Helton, J. D. Johnson, M. D. McKay, A. W. Shiver, and J. L. Sprung. Robustness of an uncertainty and sensitivity analysis of early exposure results with the maccs reactor accident consequence model. *Reliability Engineering and System Safety*, 48(2):129–148, 1995a.
- [53] J. C. Helton, J. D. Johnson, A. W. Shiver, and J. L. Sprung. Uncertainty and sensitivity analysis of early exposure results with the MACCS reactor

accident consequence model. *Reliability Engineering and System Safety*, 48(2):91-127, 1995b.

- [54] F. O. Hoffman and R. H. Gardner. Evaluation of uncertainties in radiological assessment models. In J. E. Till and H. R. Meyer, editors, *Radiological Assessment: A Textbook on Environmental Dose Assessment*, number NUREG/CR-3332, pages 11.1-11.55. U. S. Nuclear Regulatory Commission, 1983.
- [55] F. O. Hoffman and S. Kaplan. Beyond the domain of direct observation: How to specify a probability distribution that represents the "State of Knowledge" about uncertain inputs. *Risk Analysis*, 19(1):131-134, 1999.
- [56] T. Homma and A. Saltelli. *LISA Package User Guide, Part 1 : PREP (Statistical PREProcessor). Preparation of input sample for Monte Carlo simulations. Program description and User guide*, 1991.
- [57] S. C. Hora. Sensitivity, uncertainty, and decision analyses in the prioritization of research. *Journal of Statistical Computation and Simulation*, 57(1-4):175-196, 1996.
- [58] J. T. Houghton, B. A. Callander, and S. K. Varney, editors. *Climate Change 1992. The Supplementary Report to the IPCC Scientific Assessment*, Bracknell, UK, 1992. International Panel on Climate Change, WMO/UNEP.
- [59] R. L. Iman and W. J. Conover. Sensitivity analysis techniques: Self-teaching curriculum. Technical Report NUREG/CR-2350, SAND81-1978, SANDIA, USA, 1978.
- [60] R. L. Iman and W. J. Conover. Small sample sensitivity analysis techniques for computer models, with an application to risk assessment. *Communications in Statistics, Part A-Theory and Methods*, 9(17):1749-1842, 1980.

- [61] R. L. Iman and W. J. Conover. A distribution-free approach to inducing rank correlation among input variables. *Commun. Statist.-Simula. Computa.*, 11(3):311-334, 1982.
- [62] R. L. Iman and J. C. Helton. A comparison of uncertainty and sensitivity analysis techniques for computer models. Technical Report NUREG/CR-3904, SAND84-1461, U. S. Department of Energy, Washington, DC, 1985.
- [63] R. L. Iman and J. C. Helton. An investigation of uncertainty and sensitivity analysis techniques for computer models. *Risk Analysis*, 8(1):71-90, 1988.
- [64] R. L. Iman, J. C. Helton, and J. E. Campbell. An approach to sensitivity analysis of computer models: Part I - introduction, input variable selection and preliminary variable assessment. *Journal of Quality Technology*, 13(3):174-183, 1981.
- [65] R. L. Iman and M. J. Shortencarier. A FORTRAN 77 Program and User's Guide for The Generation of Latin Hypercube and Random Samples for Use With Computer Models. Technical Report NUREG/CR-3624, SAND83-2365, U. S. Department of Energy, Washington, DC, 1984.
- [66] J. A. Jacquez. *Compartmental Analysis in Biology and Medicine*. Michigan Press, Ann Arbor, Michigan, 2nd edition, 1985.
- [67] R. A. Johnson and D. W. Wichern. *Applied Multivariate Statistical Analysis*. Prentice Hall, Inc., Englewood Cliffs, New Jersey, 3rd edition, 1992.
- [68] C. D. Keeling. Global historical CO<sub>2</sub> emissions. In T. A. Boden, D. P. Kaiser, R. J. Sepanski, and F. W. Stoss, editors, *Trends '93: A Compendium of Data on Global Change*, number ORNL/CDIAC-65, pages 501-504. Carbon Dioxide Information Center, Oak Ridge National Laboratory, Oak Ridge, USA, 1994.
- [69] C. D. Keeling and T. P. Whorf. Atmospheric CO<sub>2</sub> records from sites in the SIO Air Sampling Network. In T. A. Boden, D. P. Kaiser, R. J. Sepanski,

and F. W. Stoss, editors, *Trends '93: A Compendium of Data on Global Change*, number ORNL/CDIAC-65, pages 18–19. Carbon Dioxide Information Center, Oak Ridge National Laboratory, Oak Ridge, USA, 1994.

- [70] C. D. Keeling and T. P. Whorf. Atmospheric CO<sub>2</sub> Concentrations – Mauna Loa Observatory, Hawaii, 1958-1997 (revised August 1998). Technical Report NDP-001, Carbon Dioxide Information Center, Oak Ridge National Laboratory, Oak Ridge, USA, 1997.
- [71] C. D. Keeling and T. P. Whorf. Atmospheric CO<sub>2</sub> concentrations (ppmv) air samples collected at Mauna Loa Observatory, Hawaii, 1958-2000. <http://cdiac.esd.ornl.gov/ftp/ndp001/maunaloa.co2>, 2001.
- [72] M. C. Kennedy and A. O'Hagan. Bayesian calibration of computer models. *J. R. Statist. Soc.-B*, 63:1–25, 2001.
- [73] A. W. King and M. J. Sale. A plan for intermodel comparison of atmospheric CO<sub>2</sub> projections with uncertainty analysis. Technical Report ORNL/CDIAC-32, Carbon Dioxide Information Center, Oak Ridge National Laboratory, Oak Ridge, TN, 1990.
- [74] J. P. C. Kleijnen and J. C. Helton. Statistical analyses of scatterplots to identify important factors in large-scale simulations, 1: Review and comparison of techniques. *Reliability Engineering and System Safety*, 65:147–185, 1999.
- [75] J. P. C. Kleijnen and J. C. Helton. Statistical analyses of scatterplots to identify important factors in large-scale simulations, 2: Robustness of techniques. *Reliability Engineering and System Safety*, 65:187–197, 1999.
- [76] C. J. Krebs. *Ecology: The Experimental Analysis of Distribution and Abundance*. Harper Collins College Publishers, New York, USA, 4th edition, 1994.

- [77] L. Ljung. *System Identification - Theory for the User*. Prentice Hall Ptr., Upper Saddle River, NJ, 2nd edition, 1999.
- [78] McCartney M. *Global and Local Effects of  $^{14}\text{C}$  Discharges from the Nuclear Fuel Cycle*. PhD thesis, Department of Chemistry, University of Glasgow, 1987.
- [79] G. Marland, R. J. Andres, and T. A. Boden. Global, regional, and national  $\text{CO}_2$  emissions. In T. A. Boden, D. P. Kaiser, R. J. Sepanski, and F. W. Stoss, editors, *Trends '93: A Compendium of Data on Global Change*, number ORNL/CDIAC-65, pages 505–584. Carbon Dioxide Information Center, Oak Ridge National Laboratory, Oak Ridge, USA, 1994.
- [80] G. Marland, R. J. Andres, T. A. Boden, C. A. Johnston, and A. L. Brenkert. Global, regional, and national  $\text{CO}_2$  emission estimates from fossil fuel burning, cement production, and gas flaring: 1751-1996 (revised March 1999). Technical Report NDP-030, CDIAC, <http://cdiac.esd.ornl.gov/ndps/ndp030.html>, 1999.
- [81] G. Marland and R. M. Rotty. Carbon dioxide emissions from fossil fuels: a procedure for estimation of results for 1950-82. *Tellus*, 36(b):232–261, 1984.
- [82] J. H. Matis and T. E. Wehrly. Stochastic models of compartmental systems. *Biometrics*, 35:199–220, 1979.
- [83] M. D. McKay. Evaluating prediction uncertainty. Technical Report NUREG/CR-6311, U.S. Nuclear Regulatory Commission and Los Alamos National Laboratory, Washington, DC, 1995.
- [84] M. D. McKay, R. J. Beckman, and W. J. Conover. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, 21(2):239–245, 1979. (published again in 2000, *Technometrics* 42(1):55-61).

- [85] H. Métivier, P. Jacop, G. Souchkevitch, et al., editors. *CHERNOBYL Ten Years On Radiological and Health Impact*, Paris, France, November 1995. Nuclear Energy Agency, Organisation for Economic Co-Operation and Development (OECD).
- [86] J. Ll. Morris. *Computational methods in elementary numerical analysis*. John Wiley & Sons, Chichester, 1983.
- [87] M. D. Morris. Factorial sampling plans for preliminary computational experiments. *Technometrics*, 33(2):161–174, 1991.
- [88] R. J. Mulholland and M. S. Keener. Analysis of linear compartmental models for ecosystems. *Journal of Theor. Biol.*, 44:105–116, 1974.
- [89] A. Neftel, E. Moor, H. Oeschger, and B. Stauffer. Evidence from polar ice cores for the increase in atmospheric CO<sub>2</sub> in the past two centuries. *Nature*, 315:45–47, 1985.
- [90] M. J. Norušis. *SPSS<sup>x</sup> Advanced Statistics Guide*. McGraw-Hill Book Company, Chicago, Illinois, 1985.
- [91] R. V. O'Neill. A review of linear compartmental analysis in ecosystem science. In J. H. Matis, B. C. Patten, and G. C. White, editors, *Compartmental Analysis of Ecosystem Models*, volume 10 (whole part) of *Statistical Ecology*, pages 3–28, 1979.
- [92] V. Perincherry, S. Kikuchi, and Y. Hamamatsu. Uncertainties in the analysis of large-scale systems. In B. M. Ayyub and M. M. Gupta, editors, *Uncertainty Modelling and Analysis: Theory and Applications*, volume 17 of *Machine Intelligence and Pattern Recognition*, pages 73–96. Elsevier, 1994.
- [93] S. R. Peterson and S. Chouhan. *CHERPAC: User's Mini-Manual*, June 1994.

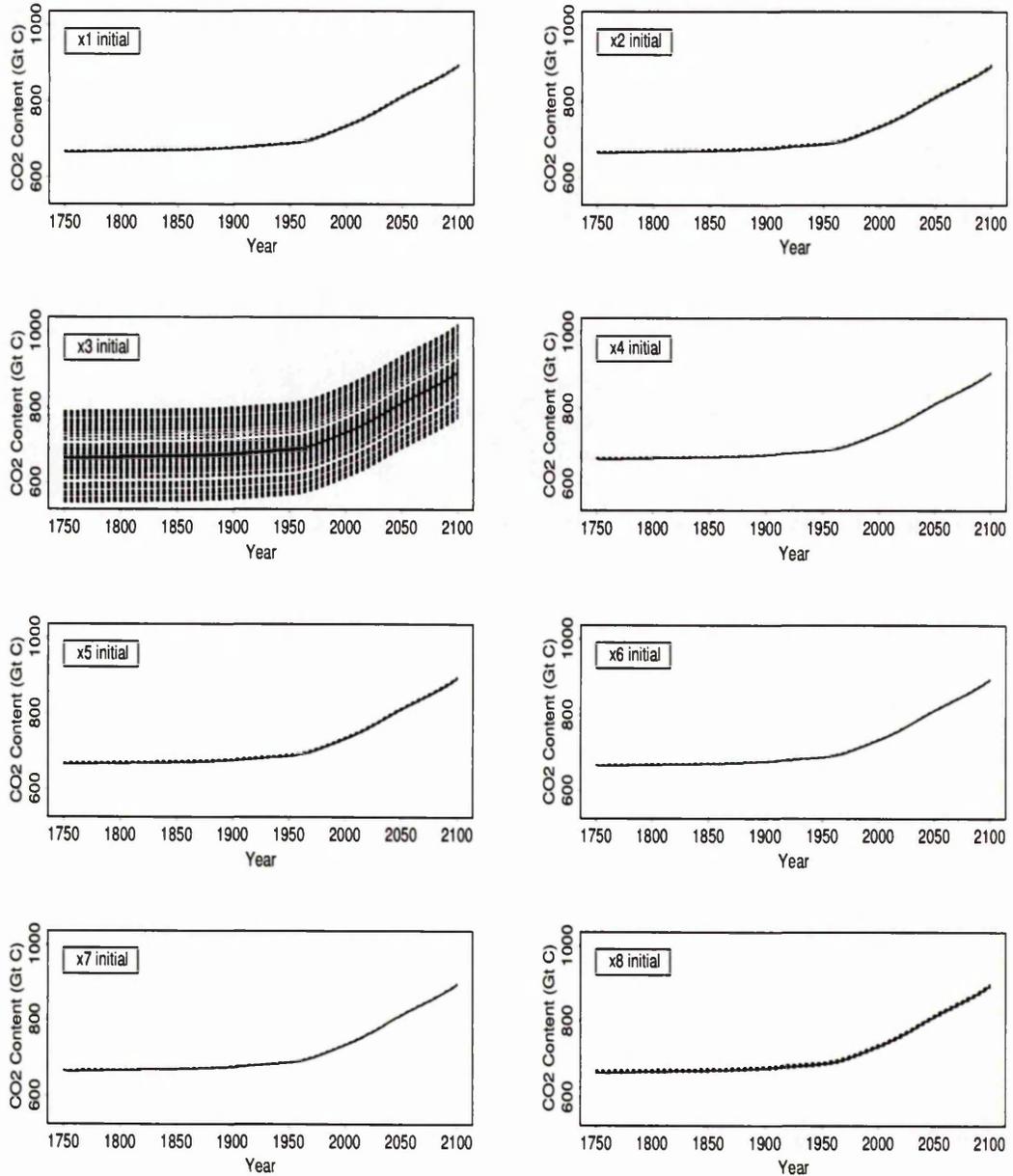
- [94] W. H. Press, B. P. Flannery, S. A. Teukolsky, and Vetterling W. T. *Numerical recipes: the art of scientific computing (FORTRAN version)*. Cambridge University Press, Cambridge, 1989.
- [95] K. L. Reed, K. A. Rose, and R. C. Whitmore. Latin hypercube analysis of parameter sensitivity in large model of outdoor recreation demand. *Ecological Modelling*, 24:159–169, 1984.
- [96] K. A. Rose. A simulation comparison and evaluation of parameter sensitivity methods applicable to large models. In *The Third International Conference on State-of-the-Art in Ecological Modeling*, pages 129–140. Colorado State University, 1982.
- [97] A. Saltelli. What is sensitivity analysis. In A. Saltelli, K. Chan, and E. M. Scott, editors, *Sensitivity Analysis*, Probability and Statistics series. John Wiley & Sons, 2000.
- [98] A. Saltelli, K. Chan, and E. M. Scott, editors. *Sensitivity Analysis*. Probability and Statistics series. John Wiley & Sons, West Sussex, England, 2000.
- [99] A. Saltelli and T. Homma. Sensitivity analysis for model output - performance of black box techniques on three international benchmark exercises. *Computational Statistics & Data Analysis*, 13:73–94, 1992.
- [100] A. Saltelli and J. Marivoet. Nonparametric statistics in sensitivity analysis for model output: A comparison of selected techniques. *Reliability Engineering and System Safety*, 28(2):229–253, 1990.
- [101] L. F. Shampine and M. K. Gordon. *Computer solution of ordinary differential equations: the initial value problem*. W. H. Freeman and Company, San Francisco, 1975.
- [102] W. Stein. Large sample properties of simulations using Latin hypercube sampling. *Technometrics*, 29(2):143–151, 1987.

- [103] J. A. Taylor. Fossil fuel emissions required to achieve atmospheric CO<sub>2</sub> stabilisation using anu-bace: A box-diffusion carbon cycle model. *Ecological Modelling*, 86:195–199, 1996.
- [104] K. M. Thiessen, F. O. Hoffman, A. Rantavaara, and S. Hossain. Environmental models undergo international test the science and art of exposure assessment modeling were tested using real-world data from the Chernobyl accident. *Env. Sci. & Technology / News*, 31(8):358–363, 1997.
- [105] R. Tomović and M. Vukobratović. *General Sensitivity Theory*. American Elsevier Publishing Company, Inc., New York, 1972.
- [106] J. R. Trabalka, editor. Atmospheric carbon dioxide and the global carbon cycle. Technical Report DOE/ER-0239, U. S. Department of Energy, Washington, D.C., 1985.
- [107] T. Turányi and H. Rabitz. Local methods. In A. Saltelli, K. Chan, and E. M. Scott, editors, *Sensitivity Analysis, Probability and Statistics series*. John Wiley & Sons, 2000.
- [108] R. K. White. Evaluating the reliability of predictions made using environmental transfer models. Technical Report Safety Series No. 10, IAEA, Vienna, 1991.
- [109] P. Young. Data-based mechanistic modelling, generalised sensitivity and dominant mode analysis. *Comp. Phy. Communications*, 117:113–129, 1999.
- [110] K. Zierler. A critique of compartmental analysis. *Ann. Rev. Biophys. Bioeng.*, 10:563–592, 1981.
- [111] D. Zwillinger. *Handbook of Differential Equations*. Academic Press, London, 2nd edition, 1992.

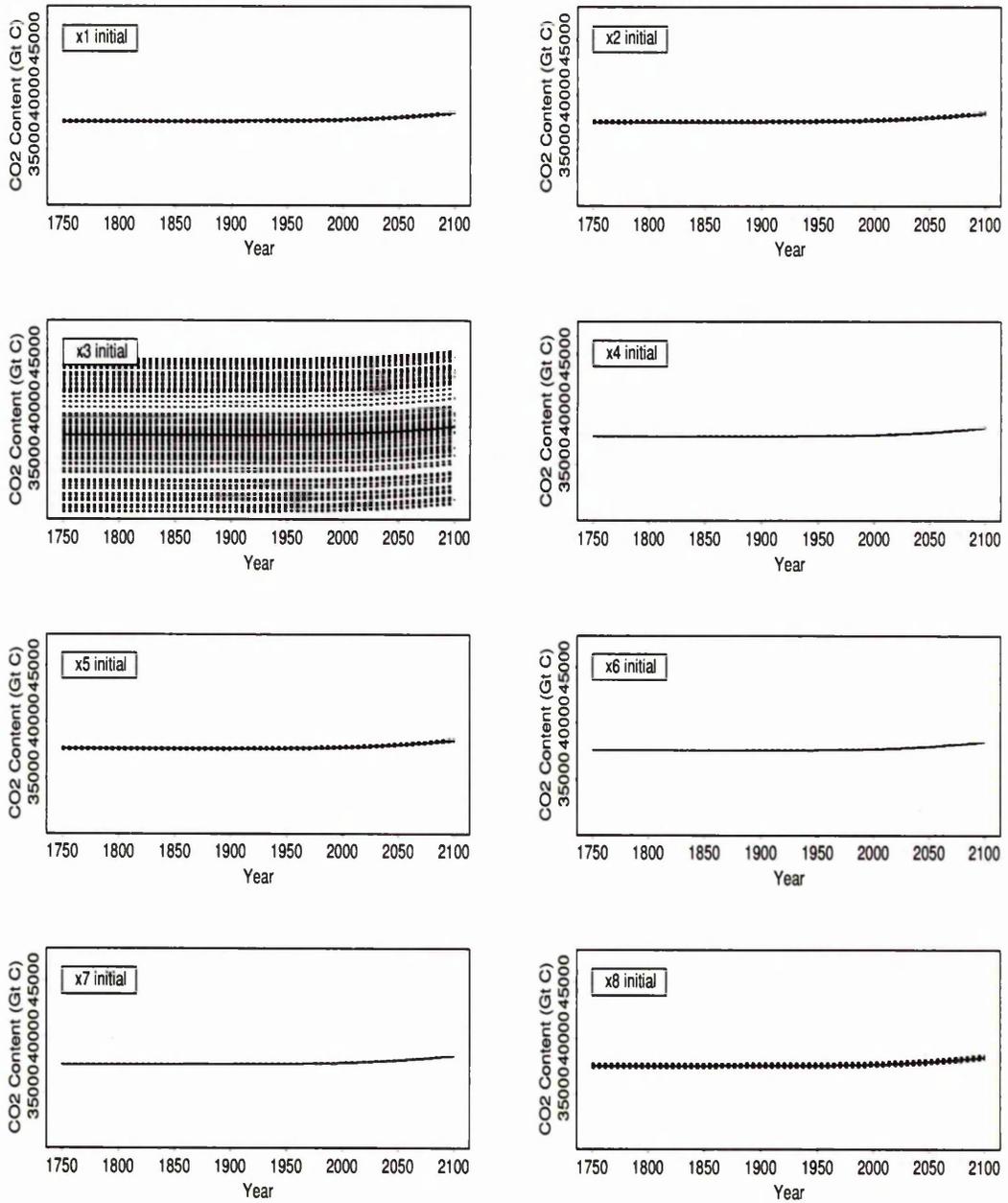
# Appendix A

## Simulation Results of Model I Summarised in Chapter 3

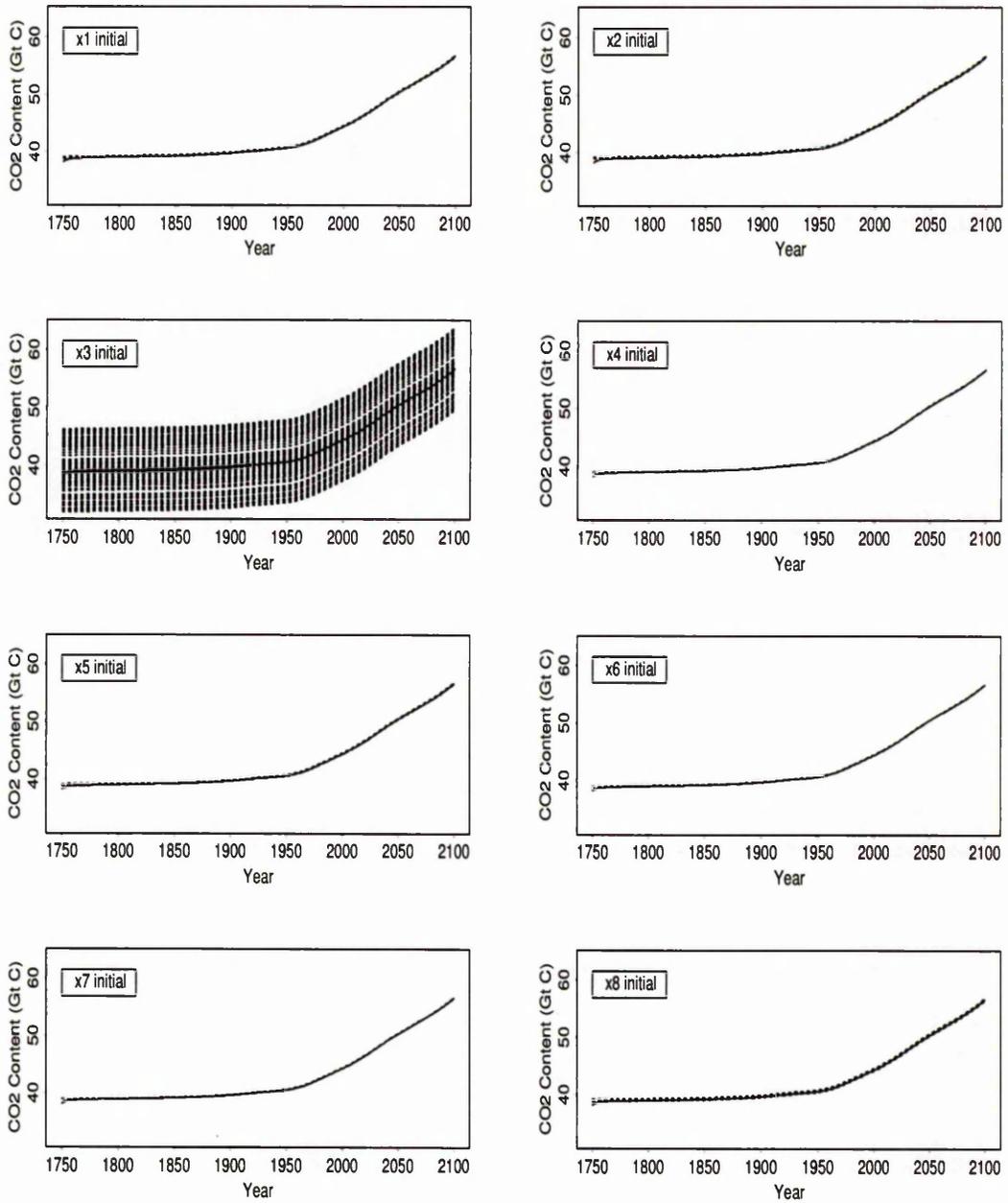
## A.1 Model I Initial Conditions



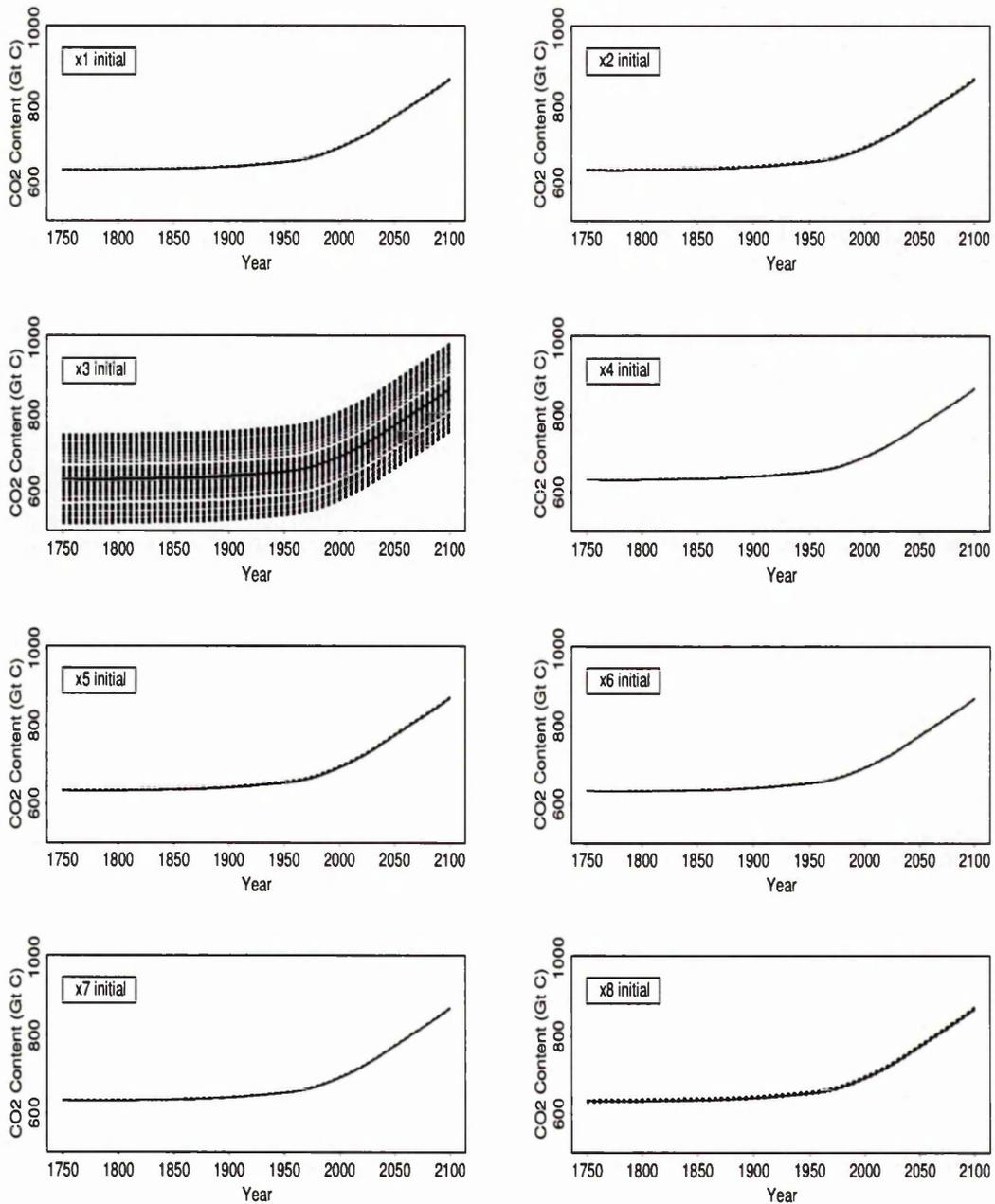
**Figure A.1.** Surface ocean CO<sub>2</sub> predictions resulting from varying initial compartmental content  $x_i^0$  of compartment  $i$  ( $i = 1, 2, \dots, 8$ ) OAT (given in top-left corner of each graph - see Table 3.2 for description of these input factors).  $N=100$  model runs, IS92a emission scenario is considered.



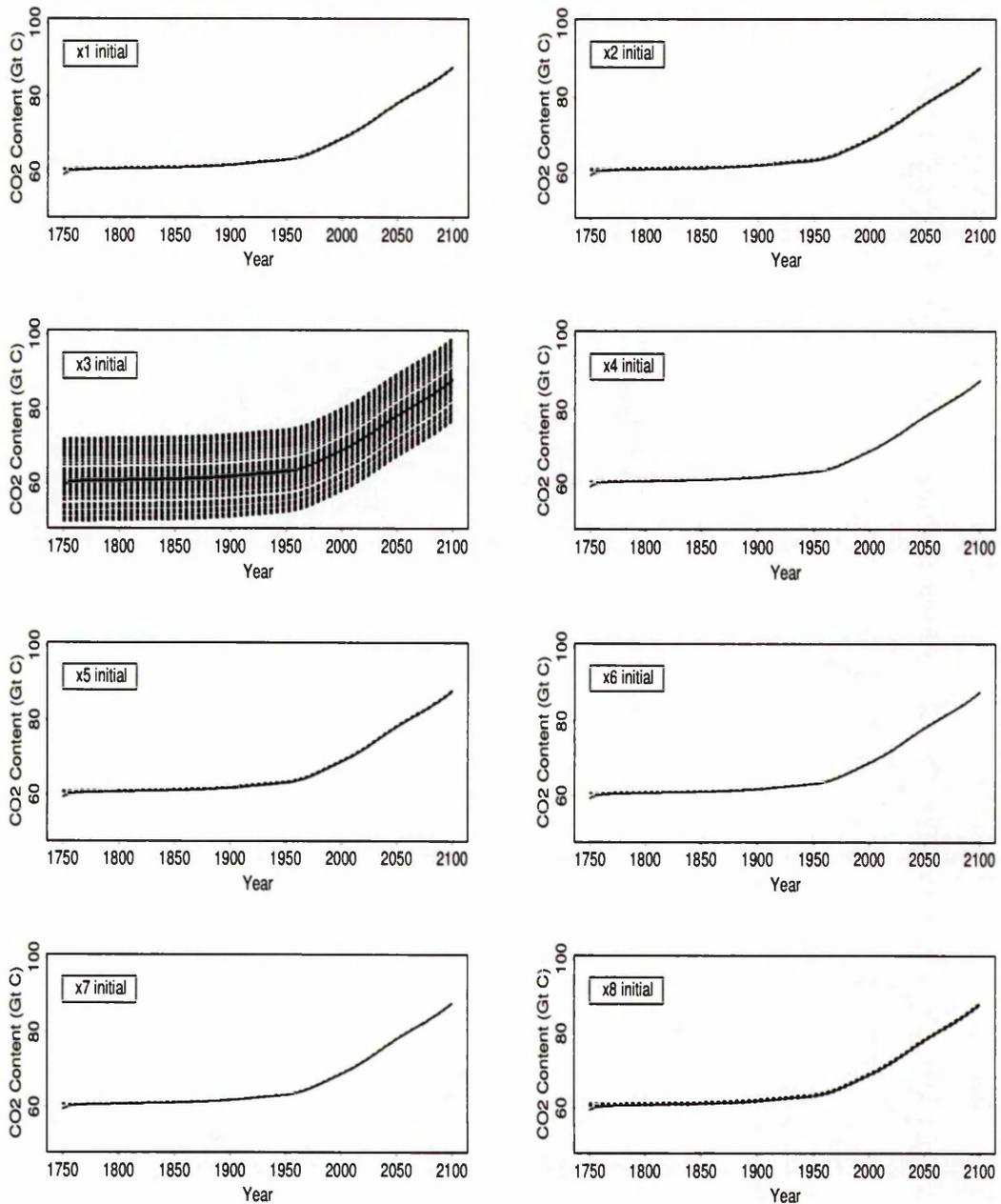
**Figure A.2.** Deep ocean CO<sub>2</sub> predictions resulting from varying initial compartmental content  $x_i^0$  of compartment  $i$  ( $i = 1, 2, \dots, 8$ ) OAT (given in top-left corner of each graph - see Table 3.2 for description of these input factors). N=100 model runs, IS92a emission scenario is considered.



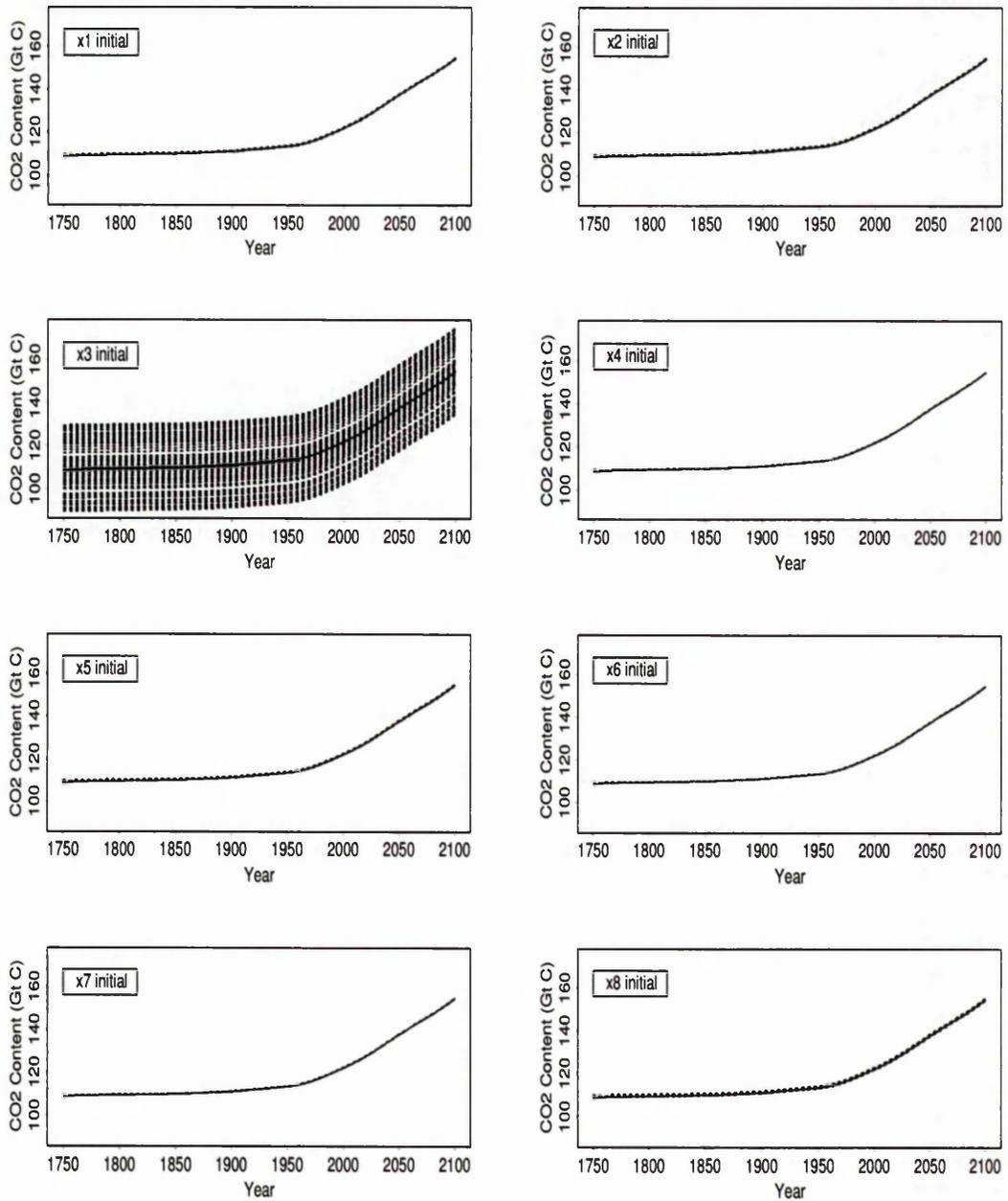
**Figure A.3.** Nonwoody parts of trees CO<sub>2</sub> predictions resulting from varying initial compartmental content  $x_i^0$  of compartment  $i$  ( $i = 1, 2, \dots, 8$ ) OAT (given in top-left corner of each graph - see Table 3.2 for description of these input factors). N=100 model runs, IS92a emission scenario is considered.



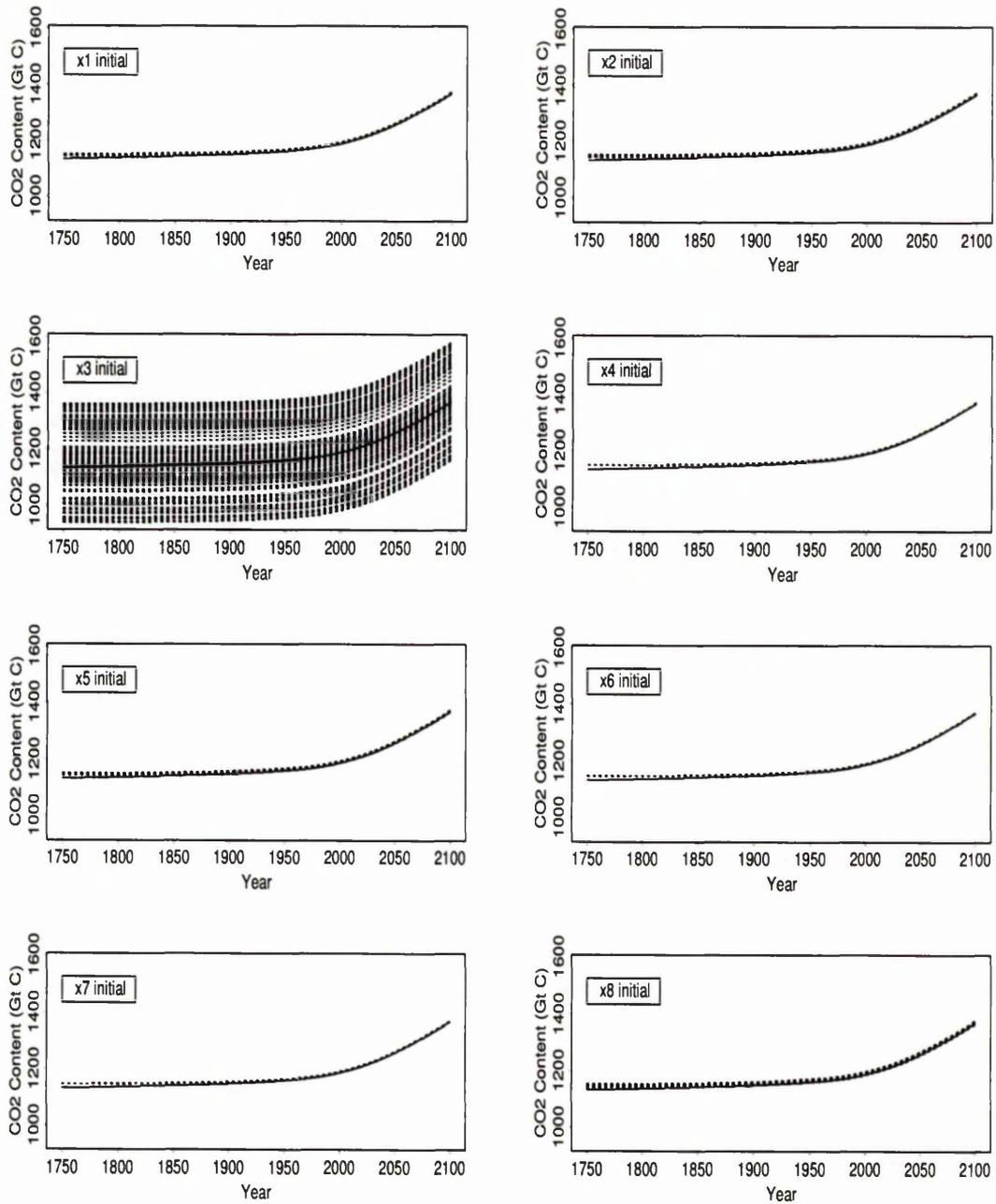
**Figure A.4.** Woody parts of trees CO<sub>2</sub> predictions resulting from varying initial compartmental content  $x_i^0$  of compartment  $i$  ( $i = 1, 2, \dots, 8$ ) OAT (given in top-left corner of each graph - see Table 3.2 for description of these input factors). N=100 model runs, IS92a emission scenario is considered.



**Figure A.5.** Ground vegetation CO<sub>2</sub> predictions resulting from varying initial compartmental content  $x_i^0$  of compartment  $i$  ( $i = 1, 2, \dots, 8$ ) OAT (given in top-left corner of each graph - see Table 3.2 for description of these input factors). N=100 model runs, IS92a emission scenario is considered.



**Figure A.6.** Detritus/decomposers CO<sub>2</sub> predictions resulting from varying initial compartmental content  $x_i^0$  of compartment  $i$  ( $i = 1, 2, \dots, 8$ ) OAT (given in top-left corner of each graph - see Table 3.2 for description of these input factors). N=100 model runs, IS92a emission scenario is considered.



**Figure A.7.** Active soil carbon CO<sub>2</sub> predictions resulting from varying initial compartmental content  $x_i^0$  of compartment  $i$  ( $i = 1, 2, \dots, 8$ ) OAT (given in top-left corner of each graph - see Table 3.2 for description of these input factors). N=100 model runs, IS92a emission scenario is considered.

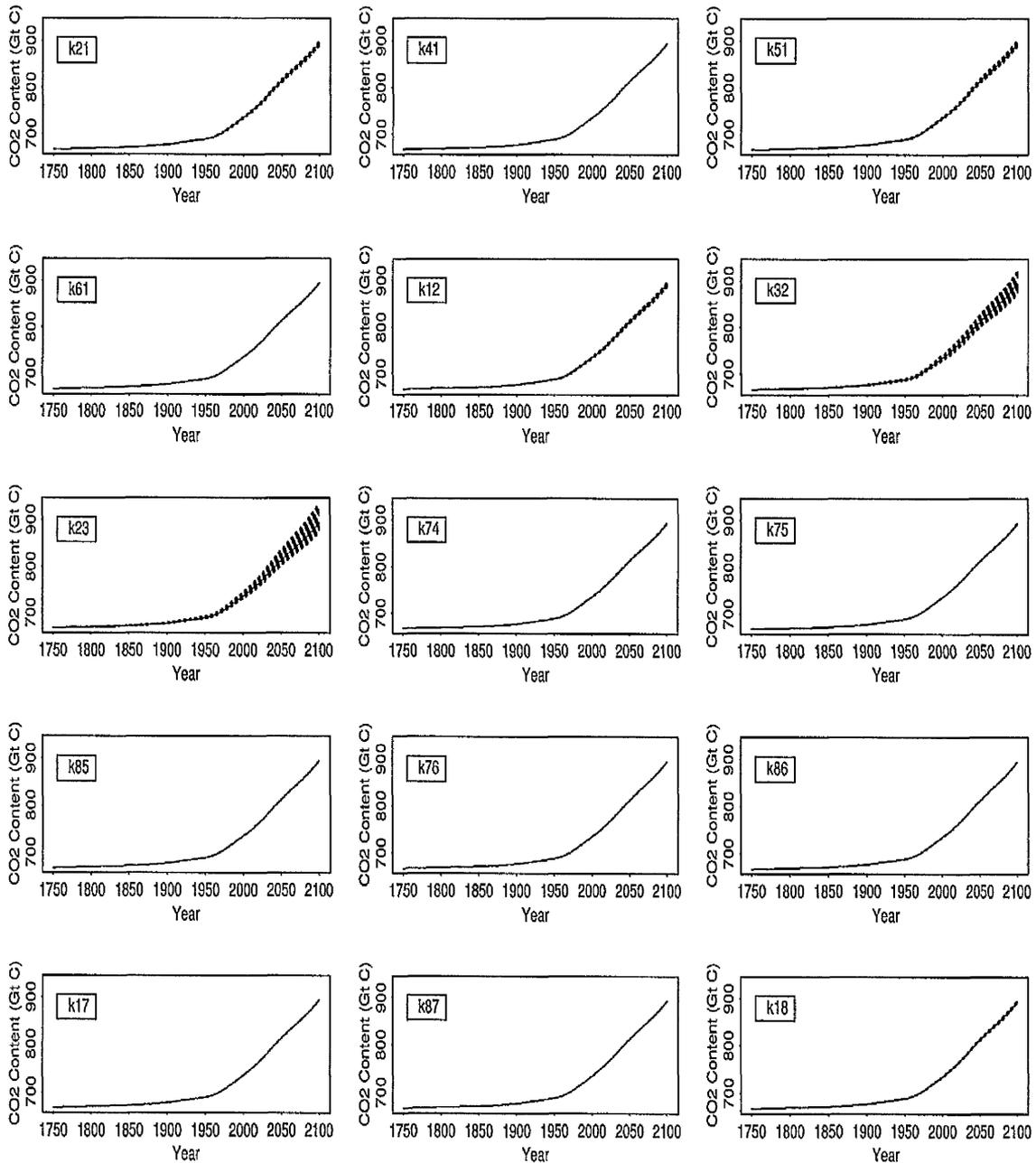
## A.2 Results of Morris Design on Model I Initial Conditions

### Conditions

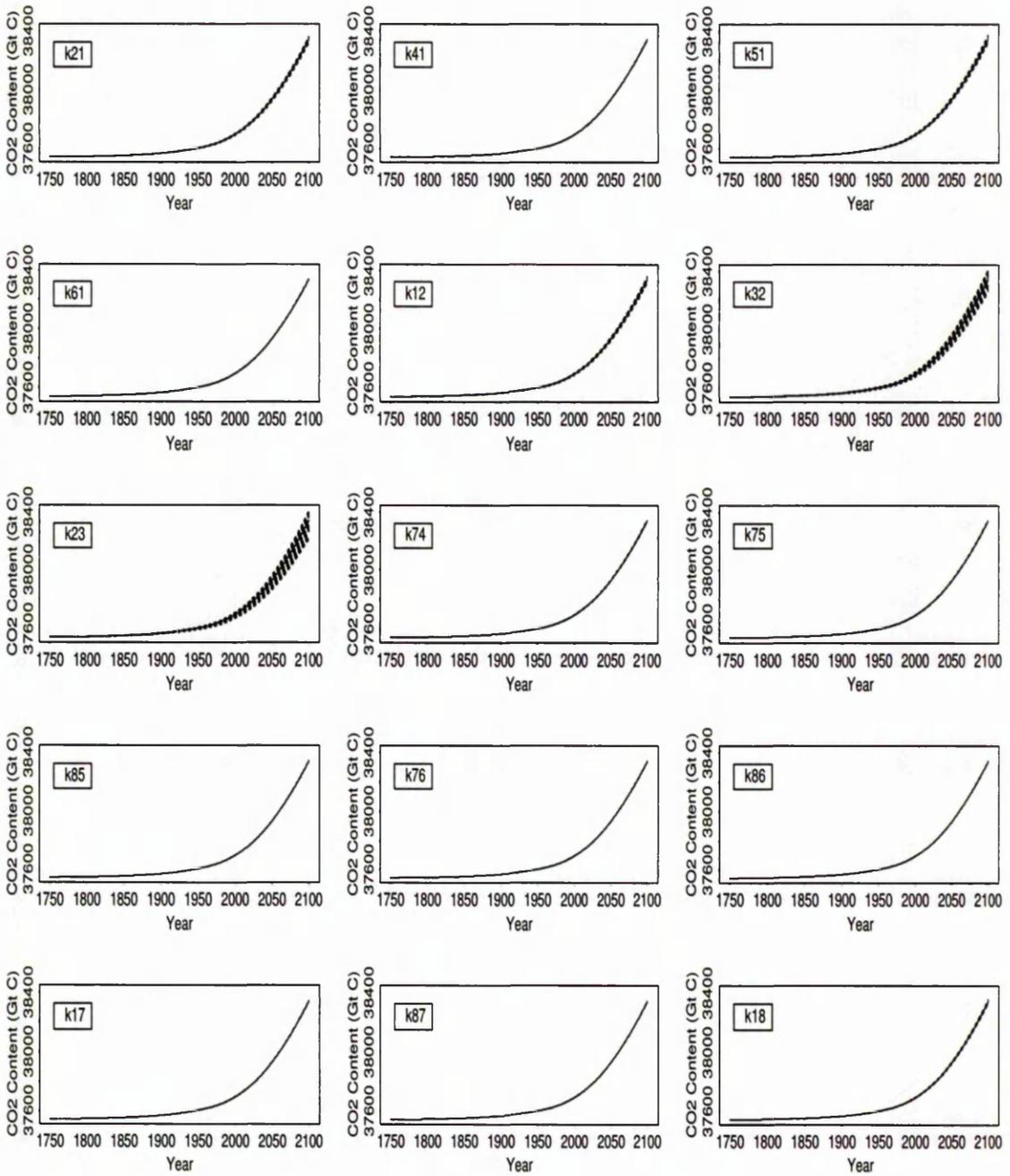
Table A.1. Morris estimated means and standard deviations associated with the initial conditions of Model I. (IS92a emission scenario considered).

Compartmental Output	Input Factor	Year 1900		Year 2000		Year 2100	
		mean	st.dev.	mean	st.dev.	mean	st.dev.
Atmosphere	E E E E E E E	1.70763E+00	2.10835E-06	1.70761E+00	2.81091E-05	1.70760E+00	2.10850E-06
		1.83101E+00	3.44247E-06	1.83099E+00	2.10818E-05	1.83099E+00	2.81058E-06
		1.02999E+02	1.04237E-05	1.02999E+02	3.44216E-05	1.02999E+02	1.09088E-05
		1.04835E-01	2.81091E-06	1.04840E-01	3.44265E-05	1.04833E-01	2.10818E-06
		1.74077E+00	3.22017E-06	1.74075E+00	2.81092E-05	1.74073E+00	3.22027E-06
		1.62755E-01	2.81091E-06	1.62767E-01	3.51364E-05	1.62753E-01	2.10819E-06
		2.96918E-01	3.22031E-06	2.96913E-01	3.22031E-05	2.96913E-01	2.10819E-06
		3.10421E+00	2.81046E-06	3.10414E+00	2.10819E-05	3.10410E+00	2.10865E-06
Surface Ocean	E E E E E E E	1.80805E+00	3.22031E-05	1.80804E+00	3.22037E-06	1.80804E+00	3.22047E-06
		1.93872E+00	2.81091E-05	1.93869E+00	3.44256E-06	1.93868E+00	3.51338E-06
		1.09057E+02	3.22487E-05	1.09057E+02	1.05471E-05	1.09057E+02	1.04237E-05
		1.11000E-01	1.57072E-09	1.10999E-01	2.81091E-06	1.11000E-01	0.00000E+00
		1.84314E+00	2.10819E-05	1.84313E+00	3.55413E-06	1.84311E+00	2.10835E-06
		1.72313E-01	3.22031E-05	1.72323E-01	3.51364E-06	1.72323E-01	3.51364E-06
		3.14380E-01	3.22031E-05	3.14377E-01	3.51364E-06	3.14375E-01	3.22030E-06
		3.28675E+00	3.22030E-05	3.28670E+00	7.10827E-08	3.28667E+00	0.00000E+00
Deep Ocean	E E E E E E E	1.01744E+02	3.21885E-05	1.01744E+02	2.10328E-05	1.01744E+02	3.59653E-06
		1.09096E+02	1.60842E-06	1.09096E+02	3.21281E-05	1.09097E+02	3.59653E-06
		6.13708E+03	5.91338E-04	6.13708E+03	5.95801E-04	6.13708E+03	6.08994E-04
		6.24622E+00	3.22031E-05	6.24624E+00	3.44266E-05	6.24625E+00	7.10827E-08
		1.03718E+02	3.43464E-05	1.03718E+02	3.51652E-05	1.03718E+02	3.21684E-06
		9.69715E+00	2.81085E-05	9.69717E+00	3.51358E-05	9.69718E+00	1.42165E-07
		1.76909E+01	2.81114E-05	1.76910E+01	3.22054E-05	1.76910E+01	2.06996E-06
		1.84950E+02	3.23288E-05	1.84950E+02	4.72776E-05	1.84950E+02	3.27265E-05
Nonwoody parts of Trees	E E E E E E E	1.05638E-01	3.22031E-06	1.05640E-01	3.44265E-05	1.05633E-01	3.51364E-05
		1.13271E-01	3.22031E-06	1.13273E-01	2.10819E-05	1.13273E-01	2.10819E-05
		6.37174E+00	2.10745E-06	6.37175E+00	3.22033E-05	6.37175E+00	3.22031E-05
		6.48533E-03	2.81091E-06	6.48667E-03	3.22031E-05	6.48000E-03	2.81091E-05
		1.07687E-01	2.10818E-06	1.07673E-01	2.10819E-05	1.07673E-01	2.10819E-05
		1.00673E-02	2.10819E-06	1.00667E-02	0.00000E+00	1.00667E-02	1.38833E-10
		1.83667E-02	2.77667E-10	1.83733E-02	3.44265E-05	1.83667E-02	3.51364E-05
		1.92033E-01	0.00000E+00	1.92033E-01	3.51364E-05	1.92033E-01	3.51364E-05
Woody parts of Trees	E E E E E E E	1.71489E+00	3.51347E-06	1.71487E+00	3.22056E-06	1.71486E+00	3.44265E-06
		1.83880E+00	3.44274E-06	1.83878E+00	2.81091E-06	1.83877E+00	3.22017E-06
		1.03436E+02	1.22493E-05	1.03437E+02	1.02989E-05	1.03437E+02	9.78362E-06
		1.05281E-01	2.10819E-06	1.05279E-01	2.10819E-06	1.05279E-01	2.81091E-06
		1.74818E+00	2.81091E-06	1.74815E+00	3.44274E-06	1.74813E+00	3.55413E-06
		1.63447E-01	2.22133E-09	1.63443E-01	3.51364E-06	1.63442E-01	3.22030E-06
		2.98180E-01	4.44267E-09	2.98176E-01	3.44265E-06	2.98174E-01	2.10819E-06
		3.11743E+00	2.81114E-06	3.11735E+00	3.22056E-06	3.11730E+00	2.81114E-06
Ground Vegetation	E E E E E E E	1.64616E-01	3.44265E-06	1.64614E-01	2.10818E-06	1.64600E-01	2.22133E-09
		1.76509E-01	3.22031E-06	1.76509E-01	3.44265E-06	1.76520E-01	2.81091E-05
		9.92907E+00	2.10865E-06	9.92909E+00	2.10865E-06	9.92909E+00	3.44272E-05
		1.01067E-02	1.38833E-10	1.01067E-02	0.00000E+00	1.01200E-02	2.81091E-05
		1.67809E-01	3.44265E-06	1.67807E-01	2.10819E-06	1.67807E-01	2.10819E-05
		1.56893E-02	3.44265E-06	1.56880E-02	2.81091E-06	1.57000E-02	3.51364E-05
		2.86240E-02	3.44265E-06	2.86227E-02	3.44265E-06	2.86267E-02	3.44265E-05
		2.99247E-01	0.00000E+00	2.99239E-01	2.10819E-06	2.99233E-01	3.51364E-05
Detritus/ Decomposers	E E E E E E E	2.96639E-01	2.81091E-06	2.96635E-01	3.22030E-06	2.96633E-01	3.51364E-05
		3.18073E-01	2.10820E-06	3.18070E-01	3.51364E-06	3.18073E-01	2.10819E-05
		17.89230E+00	3.41789E-06	17.89233E+00	2.10865E-06	17.89236E+00	3.44269E-05
		0.18213E-01	2.10819E-06	0.18211E-01	3.44265E-06	0.18220E-01	3.22031E-05
		3.02399E-01	2.10819E-06	3.02393E-01	2.10819E-06	3.02387E-01	2.81091E-05
		0.28273E-01	3.92680E-10	0.28271E-01	3.44265E-06	0.28273E-01	2.10819E-05
		0.51579E-01	2.81091E-06	0.51578E-01	3.22031E-06	0.51567E-01	3.51364E-05
		5.39247E-01	2.10819E-06	5.39233E-01	6.28288E-09	5.39220E-01	3.22031E-05
Active Soil Carbon	E E E E E E E	3.10655E+00	3.44201E-06	3.10647E+00	3.55413E-06	3.10653E+00	2.81091E-04
		3.33099E+00	2.10745E-06	3.33092E+00	3.21998E-06	3.33100E+00	3.51364E-04
		187.37016E+00	2.04718E-05	187.37078E+00	1.93010E-05	187.37127E+00	2.10795E-04
		0.190719E+00	2.81091E-06	0.19071E+00	3.51364E-06	0.19080E+00	2.81091E-04
		3.16690E+00	0.00000E+00	3.16678E+00	3.44274E-06	3.16680E+00	2.81091E-04
		0.29609E+00	3.22031E-06	0.29608E+00	3.44265E-06	0.29607E+00	2.10819E-04
		0.54016E+00	8.88534E-09	0.54014E+00	3.22029E-06	0.54033E+00	3.51364E-04
		5.64751E+00	2.80754E-06	5.64719E+00	2.81024E-06	5.64713E+00	3.22031E-04

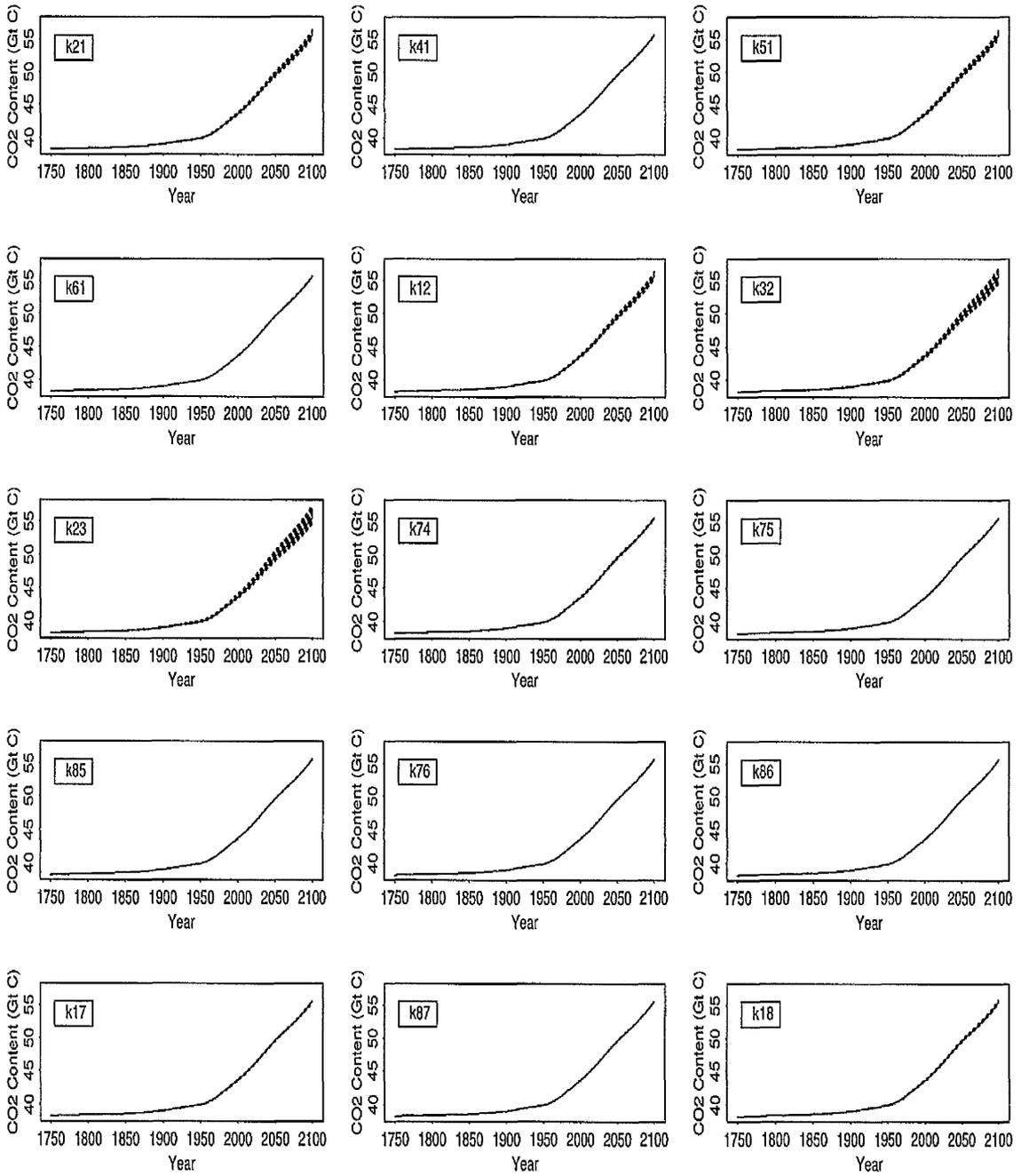
## A.3 Model I Transfer Coefficients



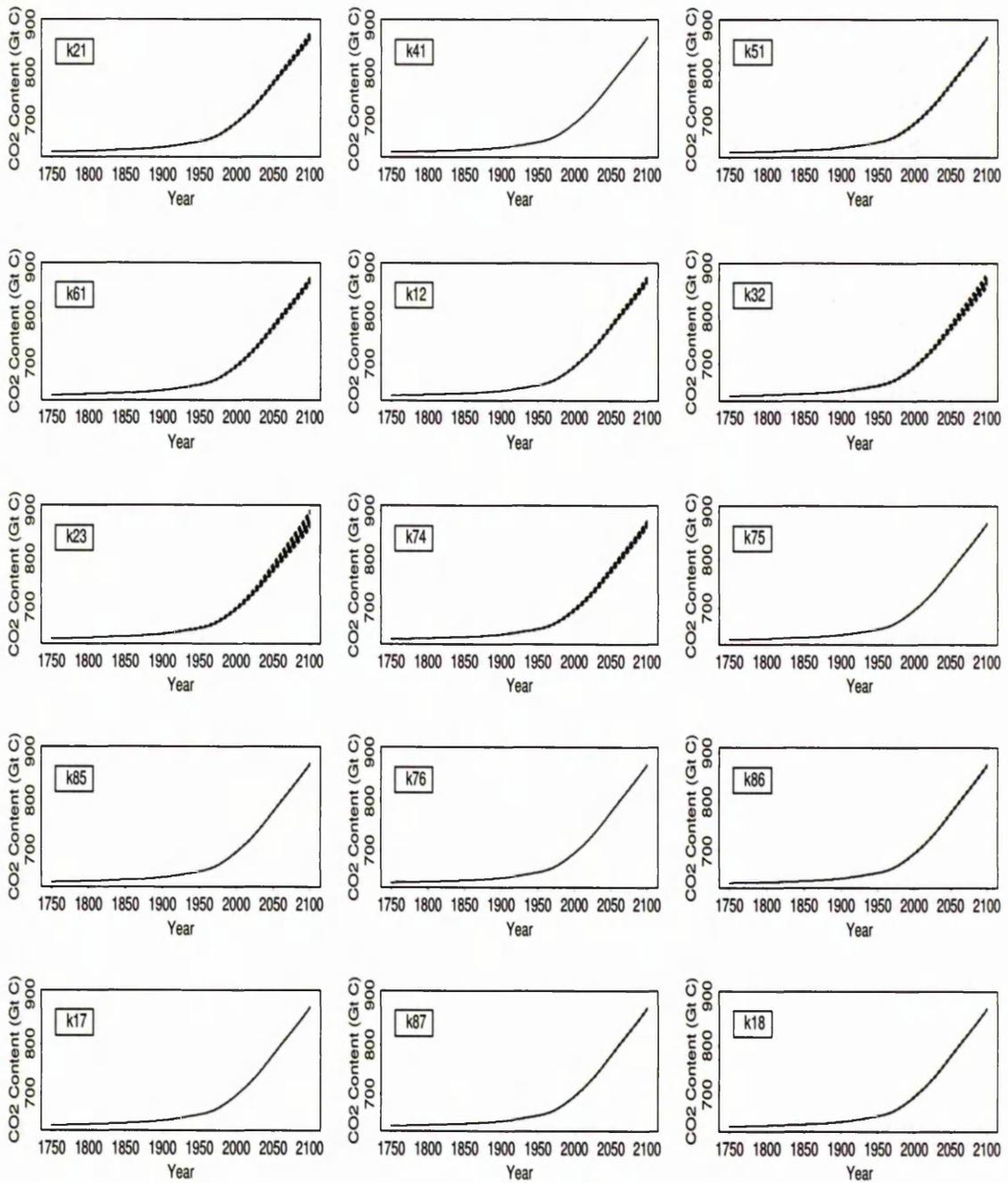
**Figure A.8.** Surface ocean CO<sub>2</sub> predictions resulting from varying transfer coefficients  $k_{ij}$  OAT (given in top-left corner of each graph - see Table 3.3 for description of these input factors). N=100 model runs, IS92a emission scenario is considered.



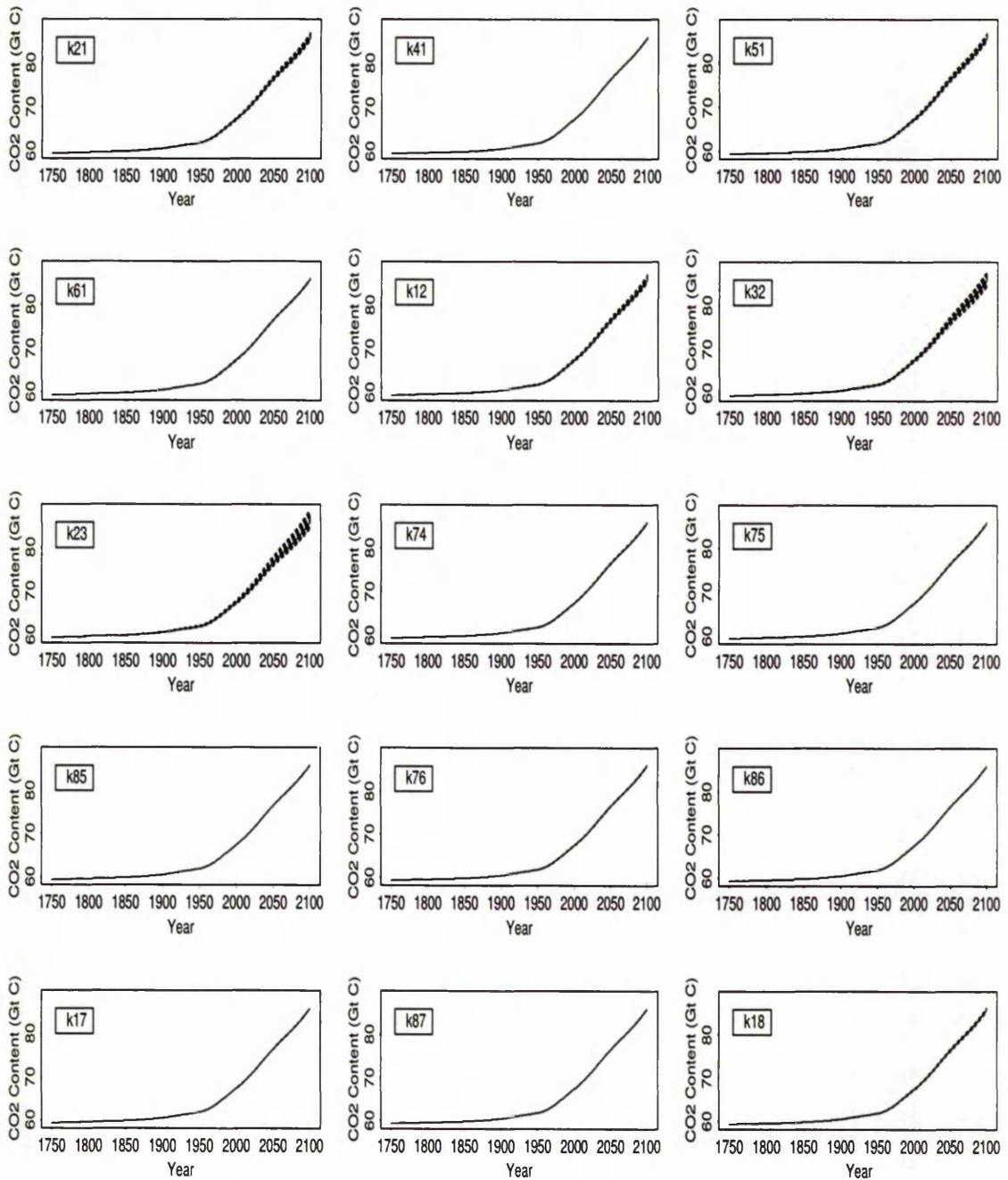
**Figure A.9.** Deep ocean CO<sub>2</sub> predictions resulting from varying transfer coefficients  $k_{ij}$  OAT (given in top-left corner of each graph - see Table 3.3 for description of these input factors). N=100 model runs, IS92a emission scenario is considered.



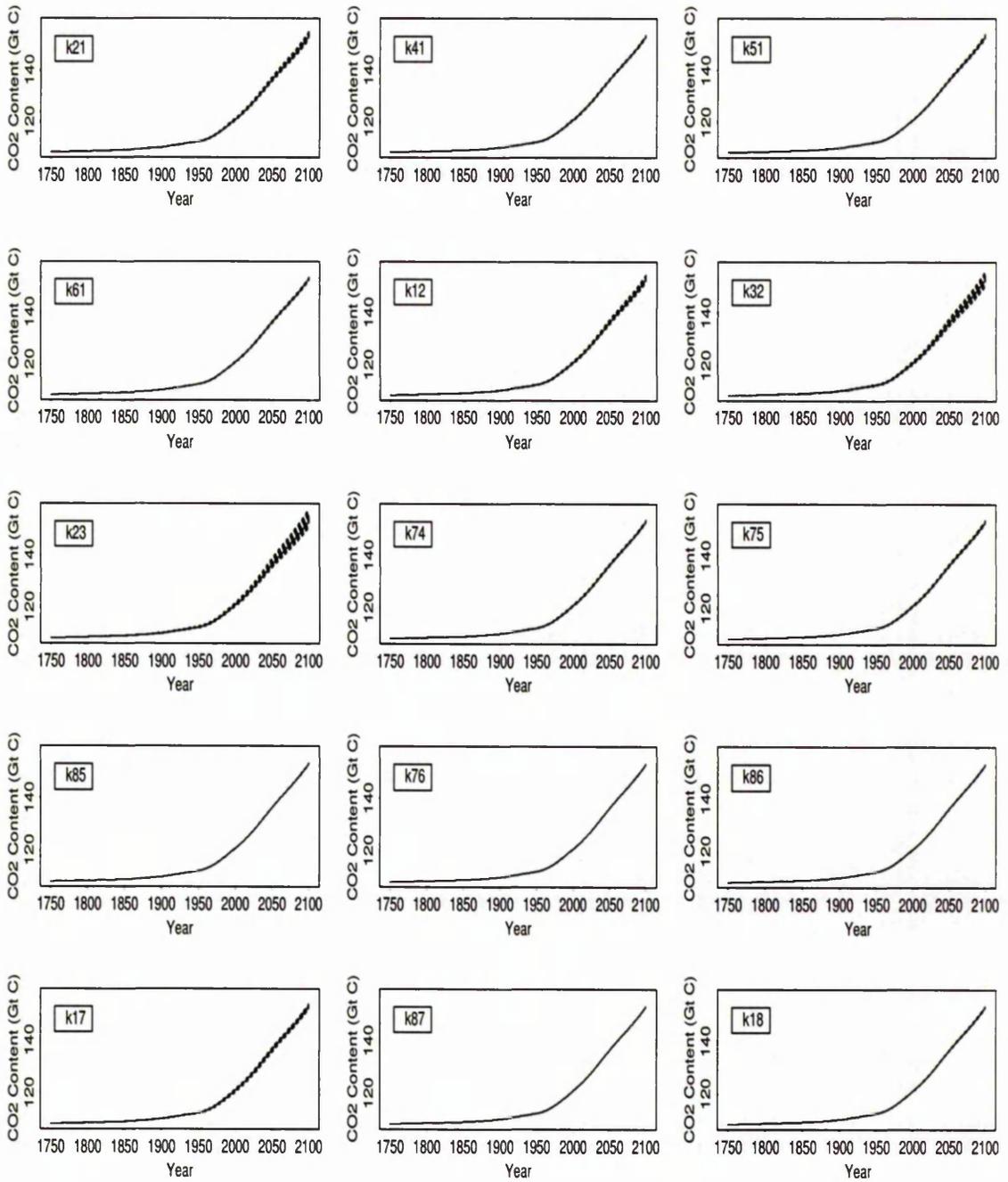
**Figure A.10.** Non-woody parts of trees CO<sub>2</sub> predictions resulting from varying transfer coefficients  $k_{ij}$  OAT (given in top-left corner of each graph - see Table 3.3 for description of these input factors). N=100 model runs, IS92a emission scenario is considered.



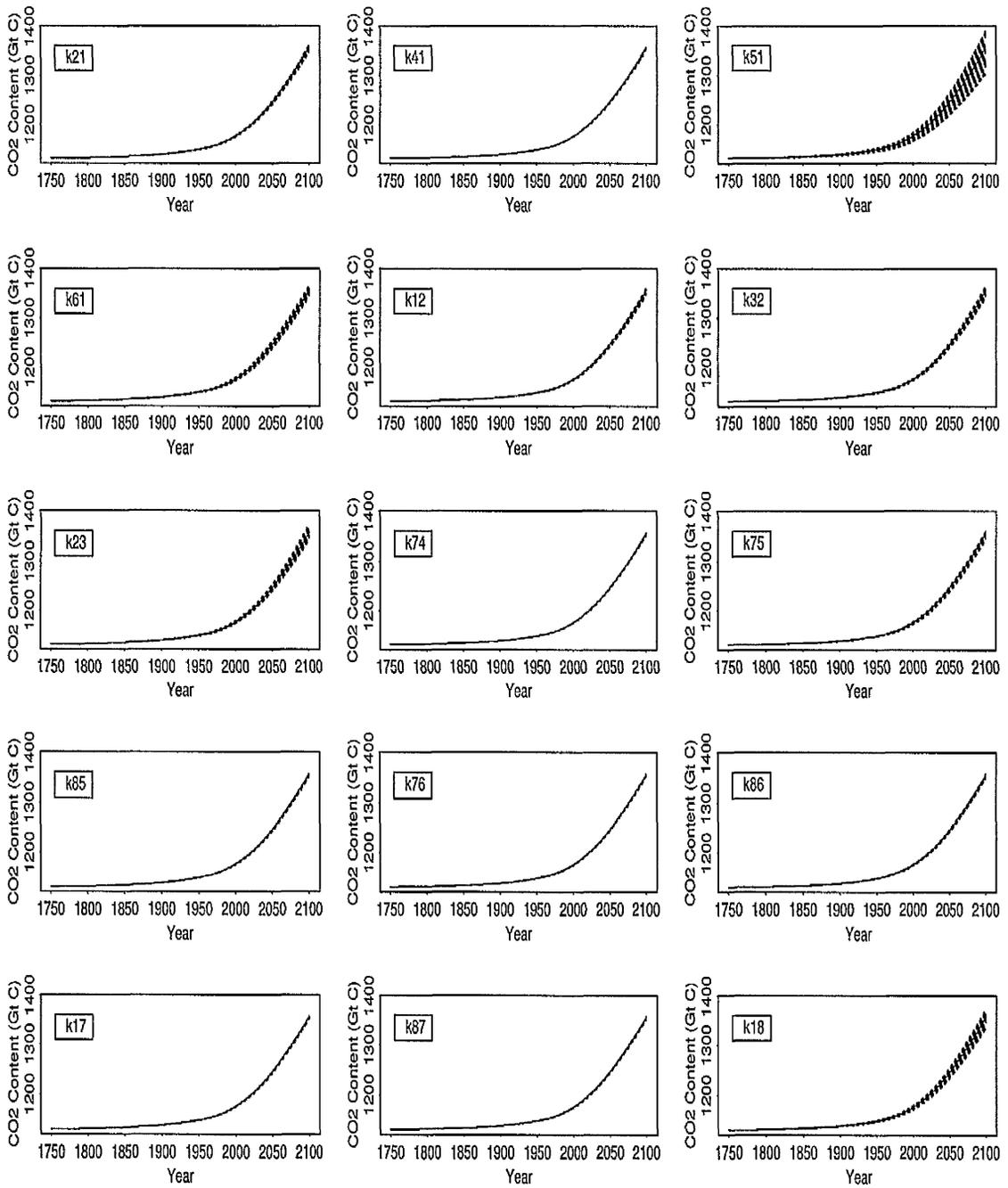
**Figure A.11.** Woody parts of trees CO<sub>2</sub> predictions resulting from varying transfer coefficients  $k_{ij}$  OAT (given in top-left corner of each graph - see Table 3.3 for description of these input factors). N=100 model runs, IS92a emission scenario is considered.



**Figure A.12.** Ground vegetation CO<sub>2</sub> predictions resulting from varying transfer coefficients  $k_{ij}$  OAT (given in top-left corner of each graph - see Table 3.3 for description of these input factors). N=100 model runs, IS92a emission scenario is considered.



**Figure A.13.** Detritus/decomposers CO<sub>2</sub> predictions resulting from varying transfer coefficients  $k_{ij}$  OAT (given in top-left corner of each graph - see Table 3.3 for description of these input factors). N=100 model runs, IS92a emission scenario is considered.



**Figure A.14.** Active soil carbon CO<sub>2</sub> predictions resulting from varying transfer coefficients  $k_{ij}$  OAT (given in top-left corner of each graph - see Table 3.3 for description of these input factors). N=100 model runs, IS92a emission scenario is considered.

## A.4 Results of Morris Design on Model I

### Transfer Coefficients

Table A.2. Morris estimated means and standard deviations associated with the selected transfer coefficients of Model I. (IS92a emission scenario considered).

Compartmental Output	Input Factor	Year 1900		Year 2000		Year 2100	
		mean	st.dev.	mean	st.dev.	mean	st.dev.
Atmosphere	k <sub>12</sub>	0.3501	0.0726	2.3683	0.4884	7.7045	1.6296
	k <sub>23</sub>	0.8670	0.0736	5.1667	0.4397	23.8652	2.0573
	k <sub>75</sub>	0.0925	0.0188	0.8271	0.1518	1.1521	0.3330
	k <sub>76</sub>	0.0060	0.0014	0.0501	0.0113	0.0759	0.0203
	k <sub>80</sub>	0.0118	0.0057	0.0701	0.0323	0.2969	0.1447
	k <sub>17</sub>	0.0290	0.0086	0.2289	0.0678	0.3698	0.1267
	k <sub>87</sub>	0.0040	0.0023	0.0443	0.0185	0.0460	0.0421
	k <sub>18</sub>	0.1935	0.0263	1.3359	0.1543	3.9234	0.6628
	Surface Ocean	k <sub>12</sub>	0.2531	0.0299	1.8600	0.2228	4.5337
k <sub>23</sub>		1.1865	0.0867	7.3525	0.5086	30.8632	2.4607
k <sub>75</sub>		0.0719	0.0162	0.6373	0.1261	0.9788	0.3126
k <sub>76</sub>		0.0048	0.0012	0.0407	0.0104	0.0630	0.0185
k <sub>80</sub>		0.0095	0.0048	0.0517	0.0258	0.2537	0.1272
k <sub>17</sub>		0.0234	0.0070	0.1938	0.0559	0.3064	0.1068
k <sub>87</sub>		0.0033	0.0018	0.0359	0.0142	0.0390	0.0361
k <sub>18</sub>		0.1525	0.0260	1.0215	0.1495	3.2554	0.6609
Deep Ocean		k <sub>12</sub>	0.6463	0.1589	3.9553	0.9661	17.5040
	k <sub>23</sub>	3.4076	0.2412	20.1411	1.4138	96.8655	7.2276
	k <sub>75</sub>	0.1816	0.0223	1.0583	0.1245	4.7081	0.6041
	k <sub>76</sub>	0.0111	0.0023	0.0695	0.0151	0.2789	0.0440
	k <sub>80</sub>	0.0213	0.0079	0.1023	0.0417	0.6660	0.2489
	k <sub>17</sub>	0.0549	0.0134	0.3489	0.0847	1.3017	0.3360
	k <sub>87</sub>	0.0079	0.0045	0.0534	0.0284	0.1855	0.0998
	k <sub>18</sub>	0.3495	0.0258	1.9219	0.1410	10.2683	0.6352
	Nonwoody parts of Trees	k <sub>12</sub>	0.0209	0.0043	0.1417	0.0285	0.4650
k <sub>23</sub>		0.0518	0.0049	0.3059	0.0303	1.4401	0.1335
k <sub>75</sub>		0.0114	0.0040	0.0935	0.0317	0.1470	0.0556
k <sub>76</sub>		0.0036	0.0043	0.0274	0.0347	0.0463	0.0568
k <sub>80</sub>		0.0007	0.0003	0.0042	0.0018	0.0183	0.0088
k <sub>17</sub>		0.0095	0.0086	0.0697	0.0678	0.1243	0.1139
k <sub>87</sub>		0.0012	0.0013	0.0099	0.0100	0.0126	0.0164
k <sub>18</sub>		0.0115	0.0016	0.0787	0.0094	0.2379	0.0400
Woody parts of Trees		k <sub>12</sub>	0.2467	0.0465	1.5732	0.2793	6.1783
	k <sub>23</sub>	0.6102	0.0412	3.3763	0.2360	18.4444	1.2536
	k <sub>75</sub>	0.3917	0.0886	2.9931	0.6014	7.0420	1.9492
	k <sub>76</sub>	0.0042	0.0009	0.0335	0.0071	0.0702	0.0175
	k <sub>80</sub>	0.1177	0.0308	0.8687	0.1950	2.2451	0.7238
	k <sub>17</sub>	0.0213	0.0065	0.1715	0.0507	0.3322	0.1147
	k <sub>87</sub>	0.0715	0.0230	0.5351	0.1502	1.3358	0.5277
	k <sub>18</sub>	0.1017	0.0636	1.0086	0.4332	1.1895	1.2306
	Ground Vegetation	k <sub>12</sub>	0.0314	0.0065	0.2122	0.0431	0.7052
k <sub>23</sub>		0.0774	0.0064	0.4518	0.0372	2.1818	0.1841
k <sub>75</sub>		0.0082	0.0017	0.0736	0.0132	0.1083	0.0318
k <sub>76</sub>		0.0081	0.0012	0.0627	0.0100	0.1045	0.0169
k <sub>80</sub>		0.0038	0.0011	0.0322	0.0090	0.0355	0.0173
k <sub>17</sub>		0.0026	0.0008	0.0218	0.0064	0.0341	0.0119
k <sub>87</sub>		0.0004	0.0002	0.0041	0.0016	0.0042	0.0039
k <sub>18</sub>		0.0174	0.0024	0.1177	0.0137	0.3653	0.0613
Detritus/ Decomposers		k <sub>12</sub>	0.0511	0.0103	0.3397	0.0665	1.1936
	k <sub>23</sub>	0.1263	0.0125	0.7242	0.0764	3.6497	0.3452
	k <sub>75</sub>	0.0328	0.0055	0.2880	0.0482	0.3623	0.0808
	k <sub>76</sub>	0.0025	0.0004	0.0194	0.0028	0.0342	0.0062
	k <sub>80</sub>	0.0074	0.0026	0.0512	0.0167	0.1609	0.0611
	k <sub>17</sub>	0.0592	0.0100	0.4737	0.0816	0.8521	0.1536
	k <sub>87</sub>	0.0010	0.0008	0.0081	0.0068	0.0274	0.0248
	k <sub>18</sub>	0.0117	0.0042	0.0562	0.0324	0.3218	0.0871
	Active Soil Carbon	k <sub>12</sub>	0.1993	0.0374	1.1800	0.2129	5.7912
k <sub>23</sub>		0.4884	0.0544	2.7636	0.3001	16.4211	1.7429
k <sub>75</sub>		0.0372	0.0142	0.2190	0.1080	0.9437	0.4202
k <sub>76</sub>		0.0242	0.0027	0.1783	0.0172	0.4642	0.0672
k <sub>80</sub>		0.1645	0.0473	1.1157	0.2768	3.6054	1.2133
k <sub>17</sub>		0.0623	0.0193	0.4214	0.1257	1.3677	0.4513
k <sub>87</sub>		0.0557	0.0320	0.3820	0.1998	1.2044	0.7741
k <sub>18</sub>		0.6342	0.0584	3.5234	0.3370	17.3505	1.6097

Table A.3. Pearson correlation coefficients (CC) for the outputs of Model I. The outputs from years 1900, 2000 and 2100 based on N=100 and N=5000 model runs are considered.

Compartmental Output	Input Factor	Correlation Coefficients (CC)					
		Year 1900		Year 2000		Year 2100	
		N=100	N=5000	N=100	N=5000	N=100	N=5000
Atmosphere	k <sub>12</sub>	-0.4264	-0.4096	-0.4599	-0.4457	-0.3663	-0.3463
	k <sub>23</sub>	-0.8898	-0.8775	-0.8605	-0.8448	-0.9282	-0.9199
	k <sub>75</sub>	-0.2384	-0.1028	-0.2842	-0.1501	-0.1795	-0.0464
	k <sub>76</sub>	-0.0201	0.0136	-0.0177	0.0115	-0.0219	0.0163
	k <sub>86</sub>	-0.0336	0.0109	-0.0274	0.0101	-0.0431	0.0125
	k <sub>17</sub>	0.0352	-0.0474	0.0199	-0.0562	0.0618	-0.0313
	k <sub>87</sub>	0.1347	0.0199	0.1291	0.0226	0.1439	0.0167
	k <sub>18</sub>	-0.2718	-0.2249	-0.2937	-0.2536	-0.2299	-0.1716
Surface Ocean	k <sub>12</sub>	0.2086	0.2324	0.2462	0.2721	0.1868	0.1626
	k <sub>23</sub>	-0.9557	-0.9586	-0.9416	-0.9444	-0.9747	-0.9771
	k <sub>75</sub>	-0.1442	-0.0673	-0.1667	-0.0947	-0.1199	-0.0356
	k <sub>76</sub>	0.0585	0.0142	0.0683	0.0127	0.0428	0.0161
	k <sub>86</sub>	-0.0527	0.0170	-0.0499	0.0174	-0.0565	0.0172
	k <sub>17</sub>	0.0618	-0.0456	0.0515	-0.0530	0.0773	-0.0344
	k <sub>87</sub>	0.1890	0.0091	0.1905	0.0099	0.1407	0.0091
	k <sub>18</sub>	-0.2516	-0.1311	-0.2654	-0.1413	-0.2237	-0.1085
Deep of Trees	k <sub>12</sub>	0.2574	0.2149	0.2643	0.2223	0.2471	0.2056
	k <sub>23</sub>	0.9635	0.9654	0.9630	0.9648	0.9654	0.9669
	k <sub>75</sub>	0.0442	-0.0657	0.0458	-0.0652	0.0481	-0.0604
	k <sub>76</sub>	0.0313	-0.0170	0.0312	-0.0174	0.0307	-0.0164
	k <sub>86</sub>	0.0732	-0.0319	0.0736	-0.0304	0.0722	-0.0327
	k <sub>17</sub>	-0.1249	0.0003	-0.1256	-0.0015	-0.1227	0.0040
	k <sub>87</sub>	-0.1685	-0.0180	-0.1672	-0.0174	-0.1700	-0.0187
	k <sub>18</sub>	-0.0318	-0.1146	-0.0243	-0.1060	-0.0333	-0.1182
Nonwoody parts of Trees	k <sub>12</sub>	-0.4140	-0.4010	-0.4435	-0.4333	-0.3634	-0.3457
	k <sub>23</sub>	-0.8709	-0.8458	-0.8331	-0.7974	-0.9230	-0.9113
	k <sub>75</sub>	-0.3042	-0.1905	-0.3587	-0.2467	-0.2160	-0.0916
	k <sub>76</sub>	-0.0420	-0.0276	-0.0420	-0.0377	-0.0324	-0.0041
	k <sub>86</sub>	-0.0396	0.0107	-0.0334	0.0101	-0.0459	0.0123
	k <sub>17</sub>	0.2151	0.1523	0.2286	0.1788	0.1526	0.0677
	k <sub>87</sub>	0.1316	0.0352	0.1243	0.0409	0.1430	0.0245
	k <sub>18</sub>	-0.2596	-0.2220	-0.2778	-0.2523	-0.2279	-0.1736
Woody parts of Trees	k <sub>12</sub>	-0.3730	-0.3511	-0.3607	-0.3468	-0.3641	-0.3369
	k <sub>23</sub>	-0.6995	-0.7397	-0.5731	-0.6318	-0.8456	-0.857
	k <sub>75</sub>	0.4567	0.5260	0.5686	0.6254	0.2593	0.3600
	k <sub>76</sub>	-0.1313	0.0068	-0.1540	0.0023	-0.0937	0.0121
	k <sub>86</sub>	-0.2306	-0.1295	-0.2569	-0.1561	-0.1802	-0.084
	k <sub>17</sub>	0.0945	-0.0283	0.0862	-0.0316	0.1019	-0.0227
	k <sub>87</sub>	0.0387	-0.0652	0.0060	-0.0801	0.0839	-0.0399
	k <sub>18</sub>	0.1622	0.1431	0.2616	0.2205	0.0378	0.0536
Ground Vegetation	k <sub>12</sub>	-0.4145	-0.4078	-0.4502	-0.4484	-0.3611	-0.3460
	k <sub>23</sub>	-0.8856	-0.8718	-0.8491	-0.8322	-0.9278	-0.9181
	k <sub>75</sub>	-0.2496	-0.1020	-0.3013	-0.1506	-0.1871	-0.0478
	k <sub>76</sub>	0.0740	0.1173	0.0995	0.1439	0.0234	0.0664
	k <sub>86</sub>	0.0177	0.0627	0.0385	0.0770	-0.0185	0.0371
	k <sub>17</sub>	0.0221	-0.0480	0.0009	-0.0592	0.0558	-0.0315
	k <sub>87</sub>	0.1570	0.0213	0.1557	0.0251	0.1547	0.0174
	k <sub>18</sub>	-0.2851	-0.2258	-0.3083	-0.2531	-0.2388	-0.1750
Detritus/ Decomposers	k <sub>12</sub>	-0.3962	-0.3769	-0.4056	-0.3845	-0.3682	-0.3471
	k <sub>23</sub>	-0.8217	-0.7824	-0.7470	-0.6846	-0.9092	-0.8948
	k <sub>75</sub>	-0.3079	-0.2071	-0.3695	-0.2841	-0.2028	-0.0836
	k <sub>76</sub>	-0.0451	0.0027	-0.0458	-0.0015	-0.0405	0.0108
	k <sub>86</sub>	-0.1074	-0.0355	-0.1144	-0.0440	-0.0955	-0.0219
	k <sub>17</sub>	0.4507	0.4127	0.5320	0.5166	0.2963	0.2199
	k <sub>87</sub>	0.1075	0.0185	0.0930	0.0217	0.1269	0.0118
	k <sub>18</sub>	-0.1105	-0.0974	-0.0821	-0.080	-0.1342	-0.0981
Active Soil Carbon	k <sub>12</sub>	-0.3351	-0.2453	-0.3440	-0.2566	-0.3350	-0.2461
	k <sub>23</sub>	-0.4902	-0.5218	-0.4952	-0.5213	-0.5767	-0.6034
	k <sub>75</sub>	0.0010	0.0323	-0.0171	0.0194	0.0015	0.0412
	k <sub>76</sub>	-0.0611	0.0149	-0.0529	0.0224	-0.0663	0.0092
	k <sub>86</sub>	0.0553	0.2044	0.1008	0.2444	0.0042	0.1574
	k <sub>17</sub>	0.1805	0.0607	0.1914	0.0759	0.1692	0.0402
	k <sub>87</sub>	0.1653	0.0869	0.1752	0.1010	0.1652	0.0705
	k <sub>18</sub>	0.7361	0.7786	0.7127	0.7606	0.6829	0.7318

**Table A.4.** Order of importance between the independent transfer coefficients resulted from stepwise regression on Model I.  $R^2$ -values obtained with the entry of listed transfer coefficients into the regression model are also given. The outputs from years 1900, 2000 and 2100 based on N=5000 model runs are considered.

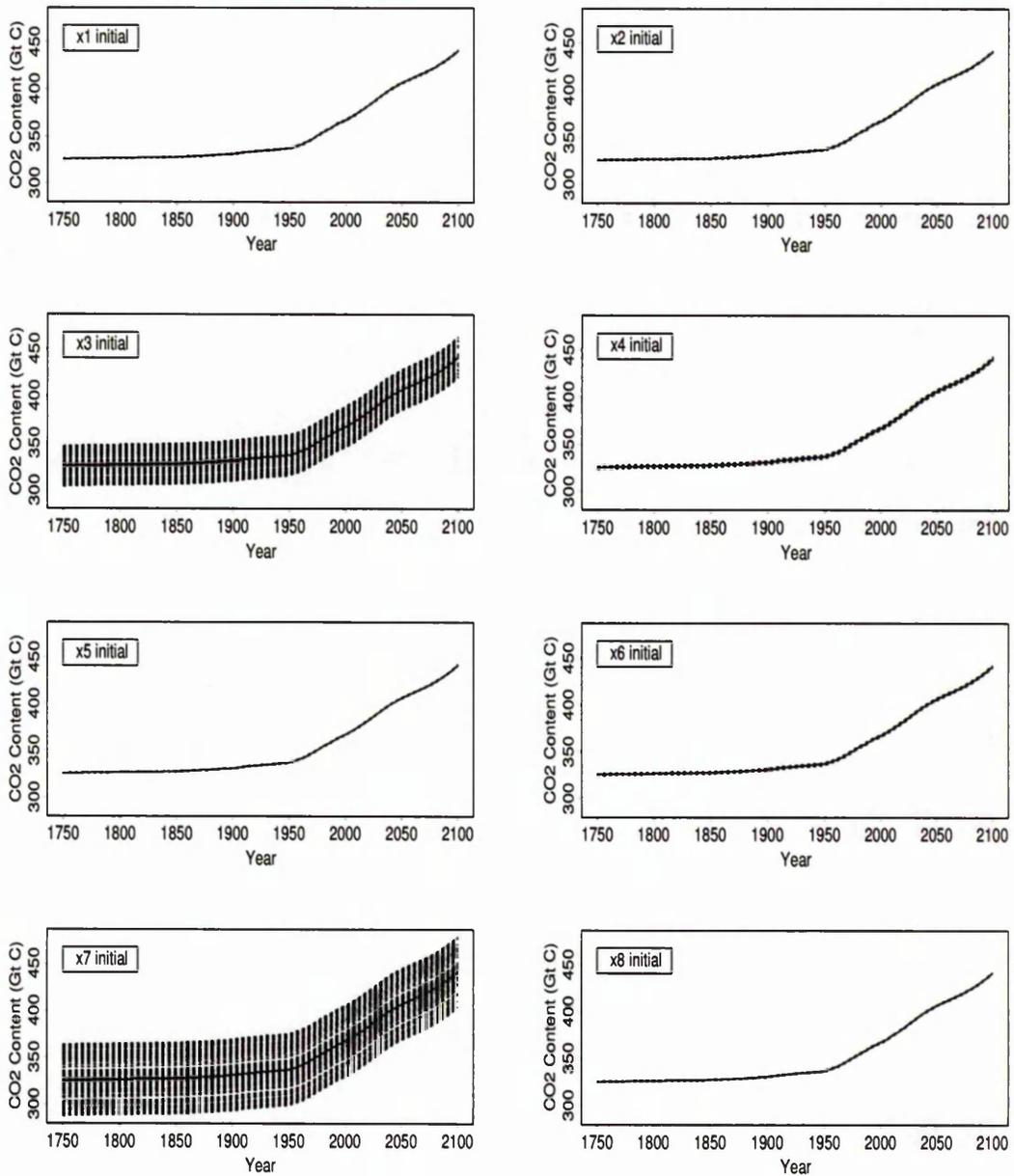
Compartmental Output	Step	Yr 1900	Yr 2000	Yr 2100
Atmosphere	1	$k_{23}$	$k_{23}$	$k_{23}$
	2	$k_{12}$	$k_{12}$	$k_{12}$
	3	$k_{18}$	$k_{18}$	$k_{18}$
	4	$k_{75}$	$k_{75}$	$k_{75}$
	5	$k_{17}$	$k_{17}$	$k_{17}$
	6	$k_{86}$	$k_{86}$	$k_{86}$
	7	$k_{76}$	$k_{76}$	$k_{76}$
		$R^2=0.9944$	$R^2=0.9944$	$R^2=0.9940$
Surface Ocean	1	$k_{23}$	$k_{23}$	$k_{23}$
	2	$k_{12}$	$k_{12}$	$k_{12}$
	3	$k_{18}$	$k_{18}$	$k_{18}$
	4	$k_{75}$	$k_{75}$	$k_{75}$
	5	$k_{17}$	$k_{17}$	$k_{17}$
	6	$k_{86}$	$k_{87}$	$k_{86}$
	7	$k_{87}$	$k_{86}$	
		$R^2=0.9956$	$R^2=0.9960$	$R^2=0.9947$
Deep Ocean	1	$k_{23}$	$k_{23}$	$k_{23}$
	2	$k_{12}$	$k_{12}$	$k_{12}$
	3	$k_{18}$	$k_{18}$	$k_{18}$
	4	$k_{75}$	$k_{75}$	$k_{75}$
	5	$k_{17}$	$k_{17}$	$k_{17}$
	6	$k_{86}$	$k_{86}$	$k_{86}$
	7	$k_{76}$	$k_{76}$	$k_{76}$
		$R^2=0.9956$	$R^2=0.9957$	$R^2=0.9949$
Nonwoody parts of Trees	1	$k_{23}$	$k_{23}$	$k_{23}$
	2	$k_{12}$	$k_{12}$	$k_{12}$
	3	$k_{18}$	$k_{75}$	$k_{18}$
	4	$k_{75}$	$k_{18}$	$k_{75}$
	5	$k_{17}$	$k_{17}$	$k_{17}$
	6	$k_{76}$	$k_{76}$	$k_{76}$
	7	$k_{87}$	$k_{87}$	$k_{86}$
		$R^2=0.9878$	$R^2=0.9844$	$R^2=0.9922$
Woody parts of Trees	1	$k_{23}$	$k_{23}$	$k_{23}$
	2	$k_{75}$	$k_{75}$	$k_{75}$
	3	$k_{12}$	$k_{12}$	$k_{12}$
	4	$k_{86}$	$k_{18}$	$k_{86}$
	5	$k_{18}$	$k_{86}$	$k_{87}$
	6	$k_{87}$	$k_{87}$	$k_{18}$
	7	$k_{17}$	$k_{17}$	$k_{17}$
		$R^2=0.9907$	$R^2=0.9907$	$R^2=0.9912$
Ground Vegetation	1	$k_{23}$	$k_{23}$	$k_{23}$
	2	$k_{12}$	$k_{12}$	$k_{12}$
	3	$k_{18}$	$k_{18}$	$k_{18}$
	4	$k_{75}$	$k_{75}$	$k_{75}$
	5	$k_{76}$	$k_{76}$	$k_{76}$
	6	$k_{86}$	$k_{86}$	$k_{86}$
	7	$k_{17}$	$k_{17}$	$k_{17}$
		$R^2=0.9944$	$R^2=0.9943$	$R^2=0.9940$
Detritus/Decomposers	1	$k_{23}$	$k_{23}$	$k_{23}$
	2	$k_{17}$	$k_{17}$	$k_{12}$
	3	$k_{12}$	$k_{12}$	$k_{17}$
	4	$k_{75}$	$k_{75}$	$k_{75}$
	5	$k_{18}$	$k_{18}$	$k_{18}$
	6	$k_{86}$	$k_{86}$	$k_{86}$
	7	$k_{76}$	$k_{76}$	$k_{76}$
		$R^2=0.9930$	$R^2=0.9920$	$R^2=0.9935$
Active Soil Carbon	1	$k_{18}$	$k_{18}$	$k_{18}$
	2	$k_{23}$	$k_{23}$	$k_{23}$
	3	$k_{12}$	$k_{12}$	$k_{12}$
	4	$k_{86}$	$k_{86}$	$k_{86}$
	5	$k_{17}$	$k_{17}$	$k_{17}$
	6	$k_{87}$	$k_{87}$	$k_{87}$
	7	$k_{76}$	$k_{76}$	$k_{76}$
		$R^2=0.9942$	$R^2=0.9936$	$R^2=0.9940$

## Appendix B

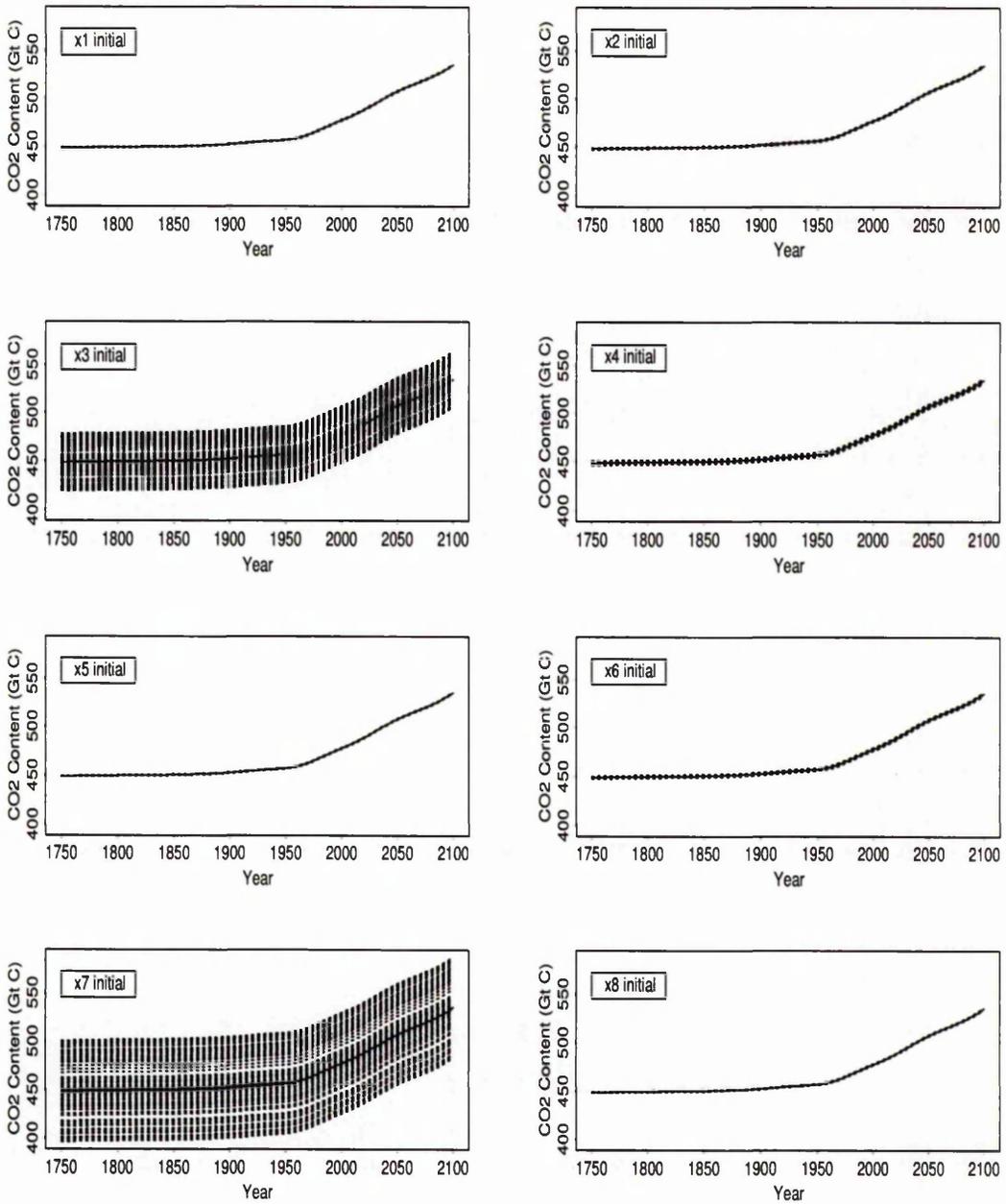
### Summary of Model II Simulation Results

## B.1 Screening Methods

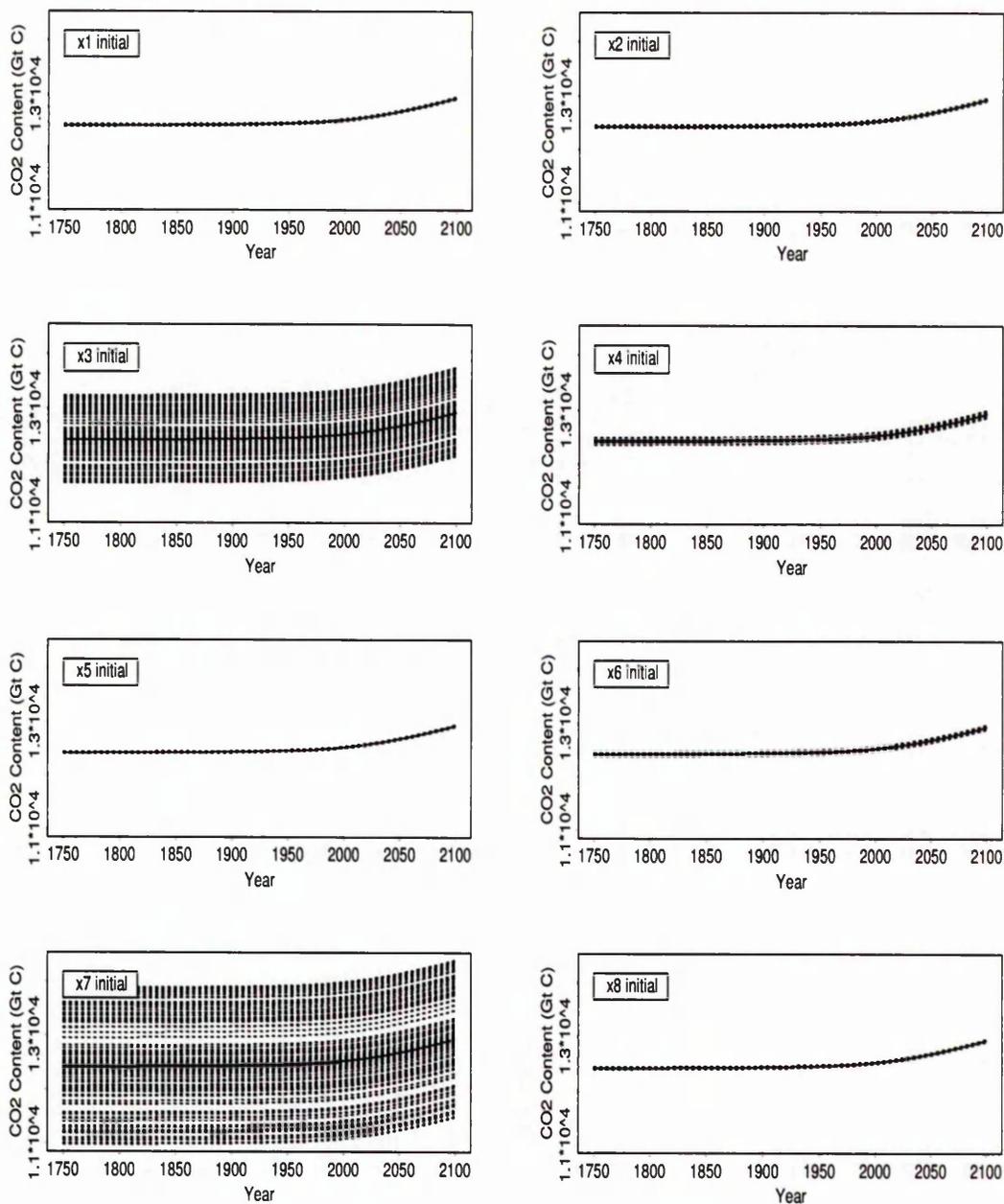
### B.1.1 Initial Conditions



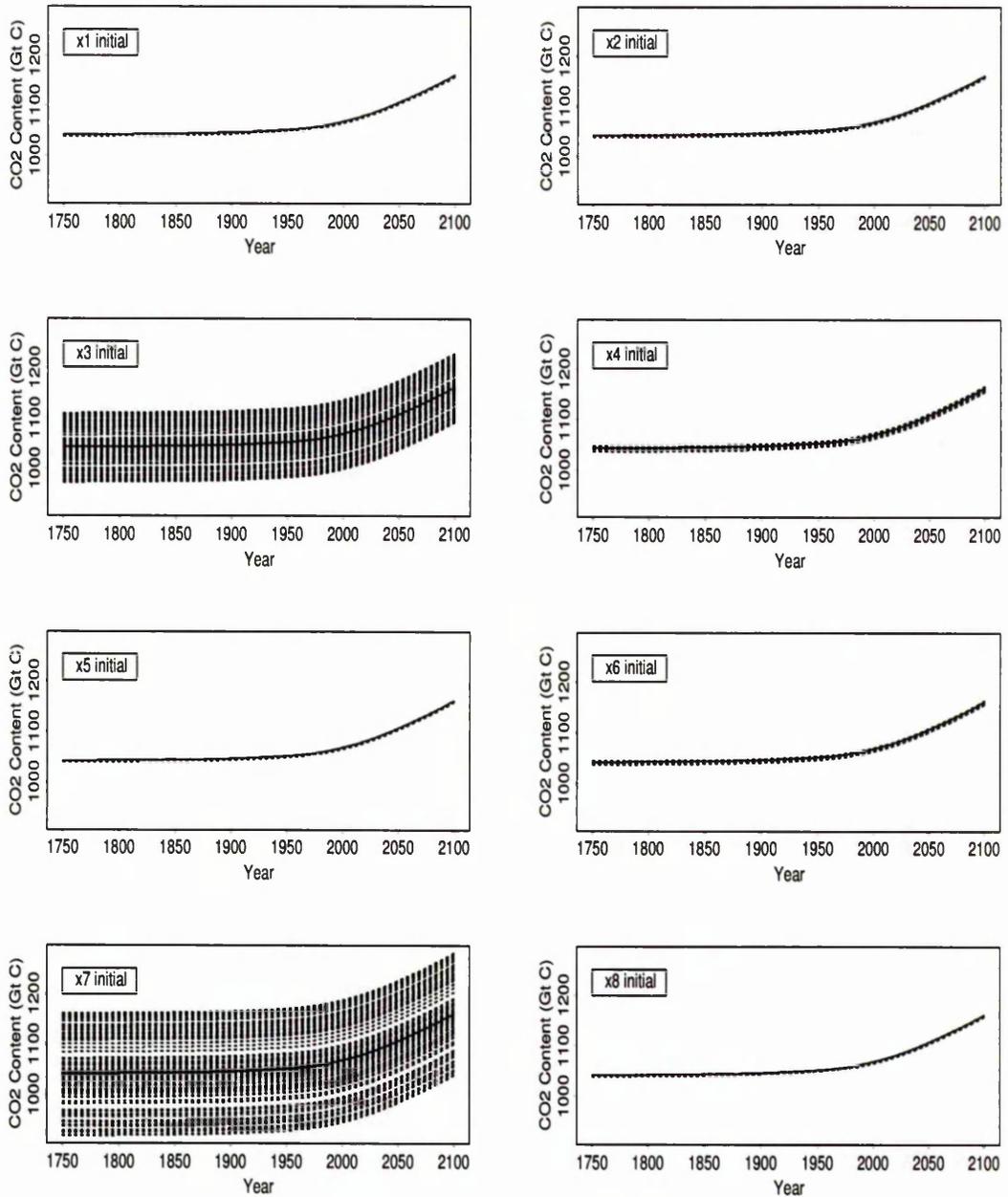
**Figure B.1.** Circulating carbon-(NH) CO<sub>2</sub> predictions resulting from varying initial condition  $x_i^0$  of compartment  $i$  ( $i = 1, 2, \dots, 8$ ) OAT (given in top-left corner of each graph - see Table 3.4 in Chapter 3 for description of these factors).  $N=100$  model runs, IS92a input scenario is considered. In each graph solid line represents the base-line case and dashed lines represent the predictions.



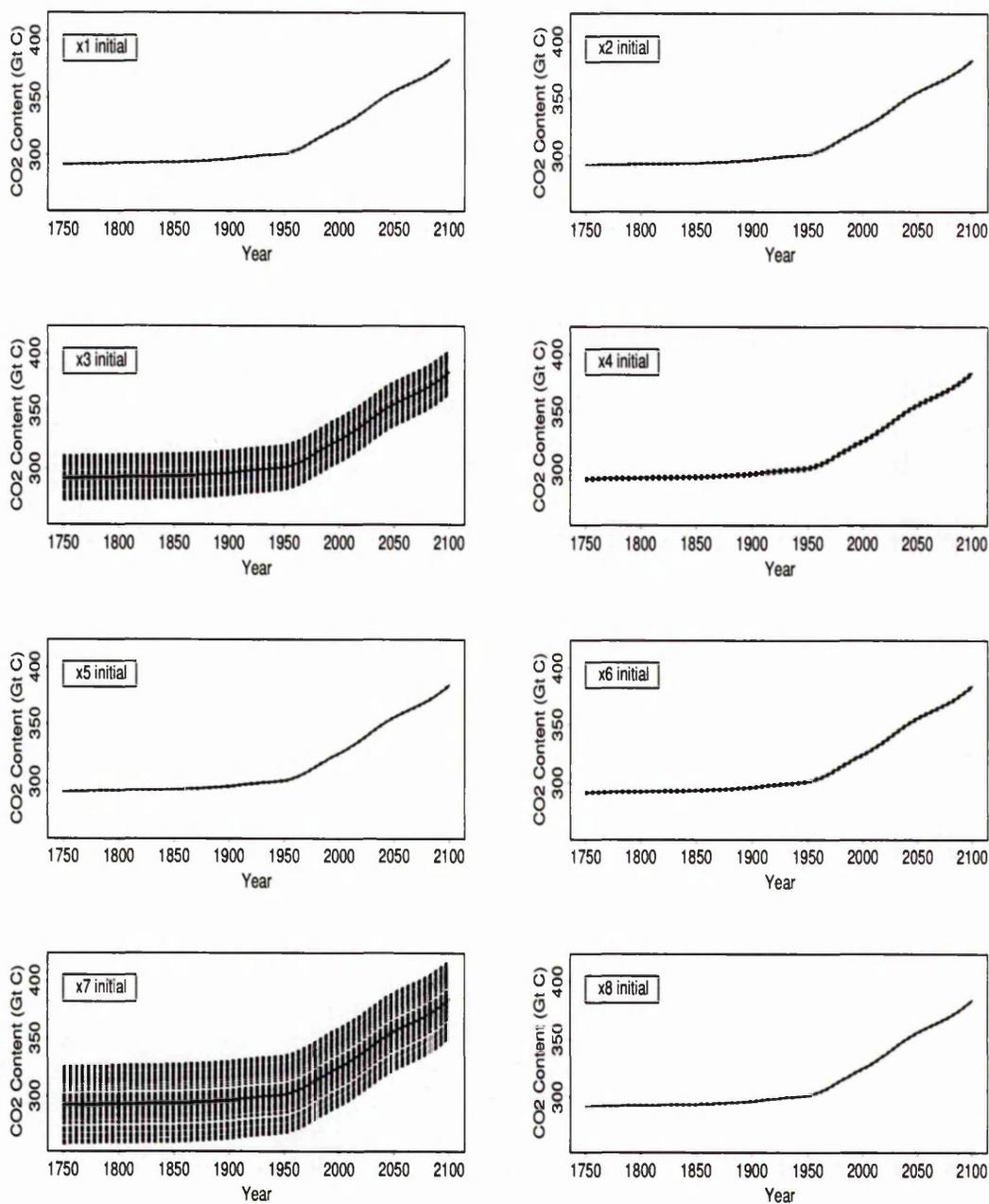
**Figure B.2.** Surface ocean-(NH) CO<sub>2</sub> predictions resulting from varying initial condition  $x_i^0$  of compartment  $i$  ( $i = 1, 2, \dots, 8$ ) OAT (given in top-left corner of each graph - see Table 3.4 for description of these factors). N=100 model runs, IS92a input scenario is considered. In each graph solid line represents the base-line case and dashed lines represent the predictions.



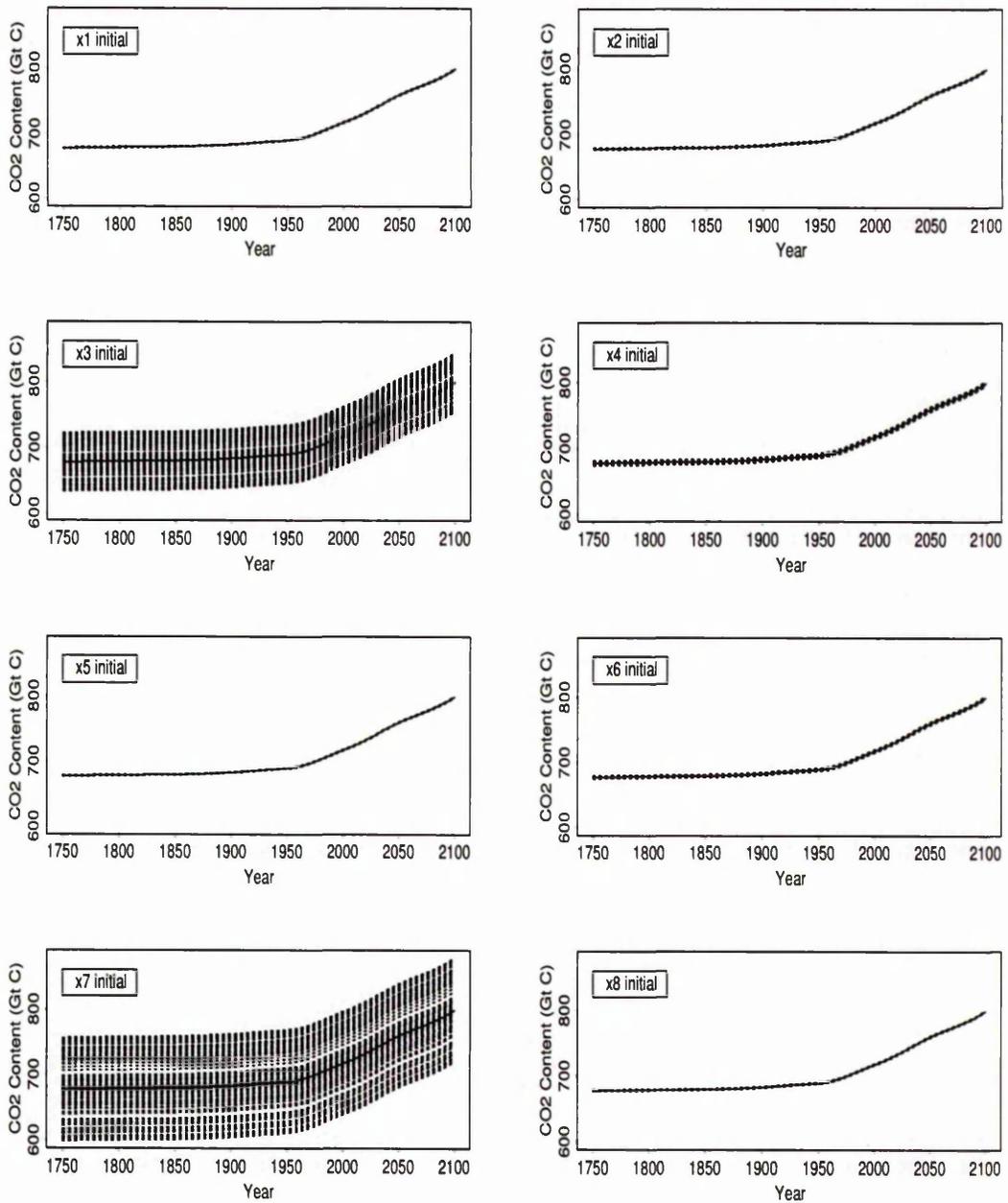
**Figure B.3.** Deep ocean-(NH) CO<sub>2</sub> predictions resulting from varying initial condition  $x_i^0$  of compartment  $i$  ( $i = 1, 2, \dots, 8$ ) OAT (given in top-left corner of each graph - see Table 3.4 for description of these factors). N=100 model runs, IS92a input scenario is considered. In each graph solid line represents the baseline case and dashed lines represent the predictions.



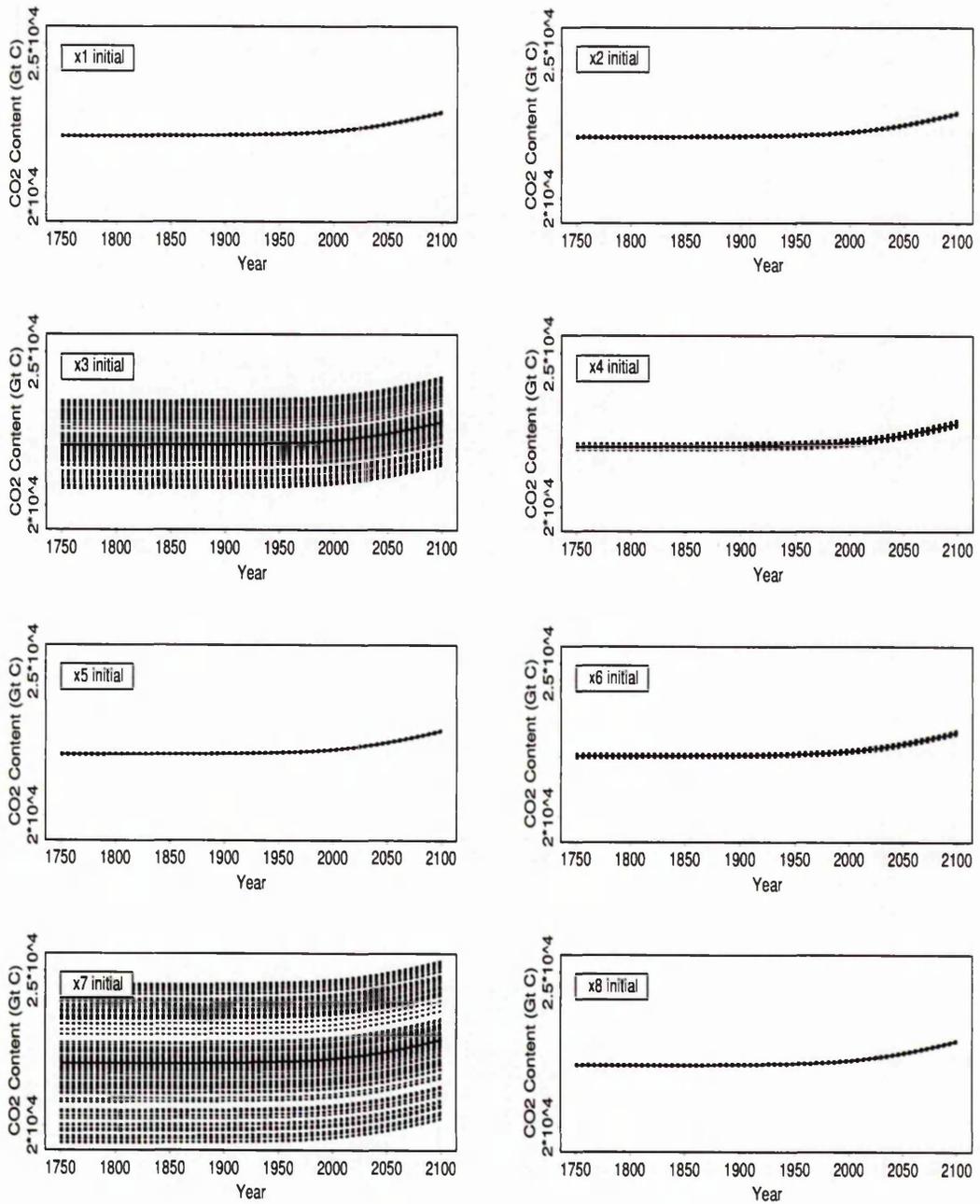
**Figure B.4.** Humus-(NH) CO<sub>2</sub> predictions resulting from varying initial condition  $x_i^0$  of compartment  $i$  ( $i = 1, 2, \dots, 8$ ) OAT (given in top-left corner of each graph - see Table 3.4 for description of these factors).  $N=100$  model runs, IS92a input scenario is considered. In each graph solid line represents the base-line case and dashed lines represent the predictions.



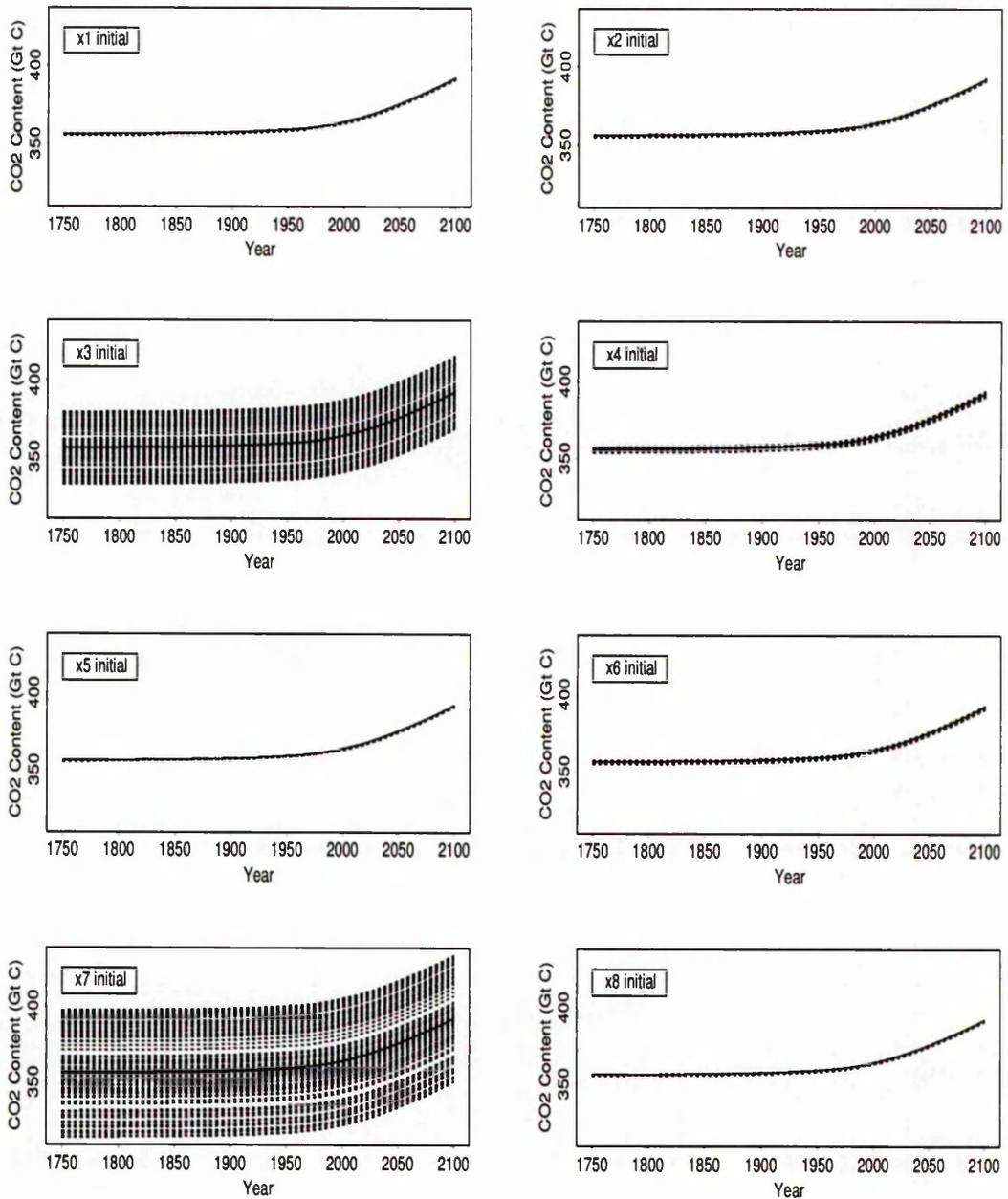
**Figure B.5.** Circulating carbon-(SH) CO<sub>2</sub> predictions resulting from varying initial condition  $x_i^0$  of compartment  $i$  ( $i = 1, 2, \dots, 8$ ) OAT (given in top-left corner of each graph - see Table 3.4 for description of these factors).  $N=100$  model runs, IS92a input scenario is considered. In each graph solid line represents the base-line case and dashed lines represent the predictions.



**Figure B.6.** Surface ocean-(SH) CO<sub>2</sub> predictions resulting from varying initial condition  $x_i^0$  of compartment  $i$  ( $i = 1, 2, \dots, 8$ ) OAT (given in top-left corner of each graph - see Table 3.4 for description of these factors). N=100 model runs, IS92a input scenario is considered. In each graph solid line represents the base-line case and dashed lines represent the predictions.



**Figure B.7.** Deep ocean-(SH)  $\text{CO}_2$  predictions resulting from varying initial condition  $x_i^0$  of compartment  $i$  ( $i = 1, 2, \dots, 8$ ) OAT (given in top-left corner of each graph - see Table 3.4 for description of these factors).  $N=100$  model runs, IS92a input scenario is considered. In each graph solid line represents the base-line case and dashed lines represent the predictions.



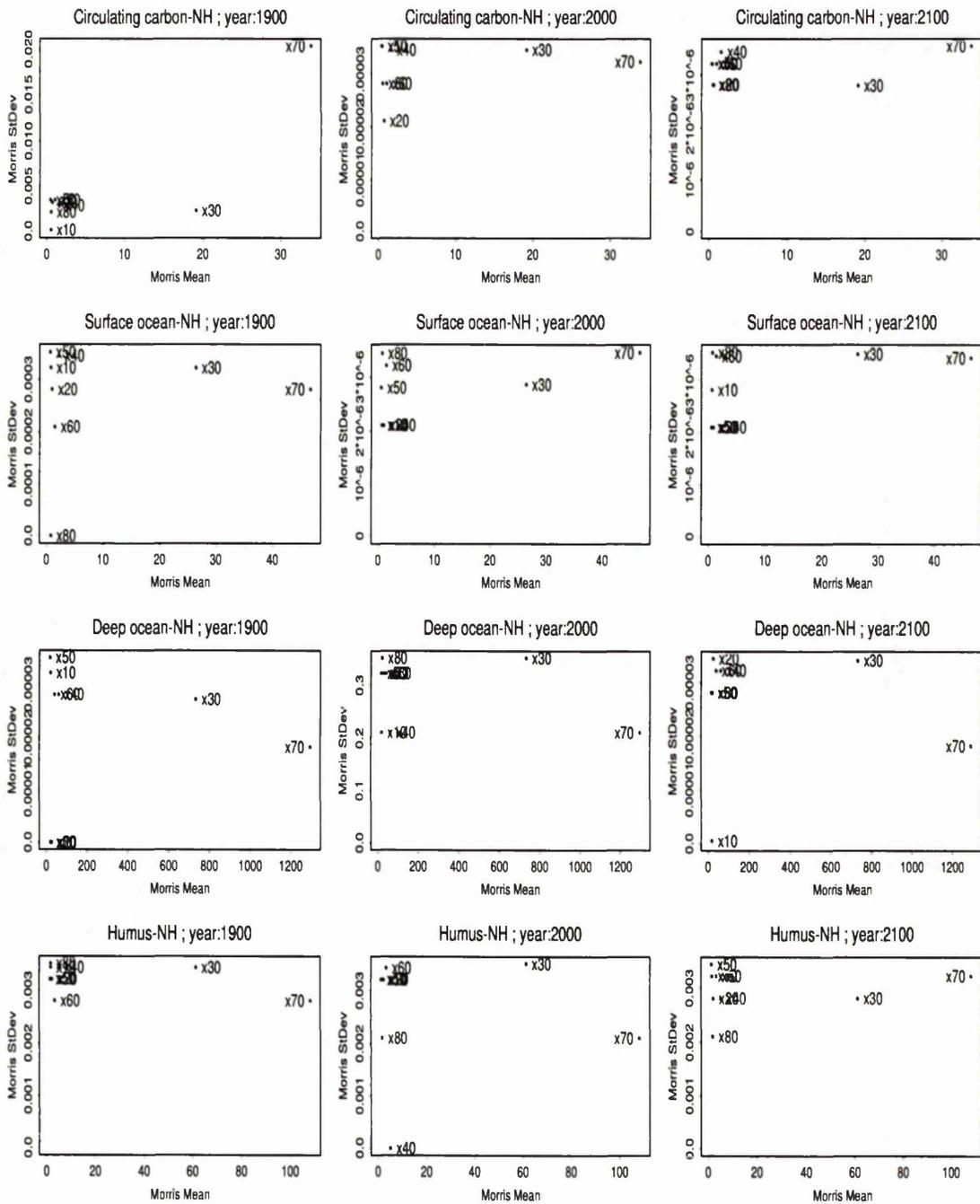
**Figure B.8.** Humus-(SH) CO<sub>2</sub> predictions resulting from varying initial condition  $x_i^0$  of compartment  $i$  ( $i = 1, 2, \dots, 8$ ) OAT (given in top-left corner of each graph - see Table 3.4 for description of these factors).  $N=100$  model runs, IS92a input scenario is considered. In each graph solid line represents the base-line case and dashed lines represent the predictions.

Table B.1. Sensitivity Rankings from SI - Model II initial conditions are varied OAT.

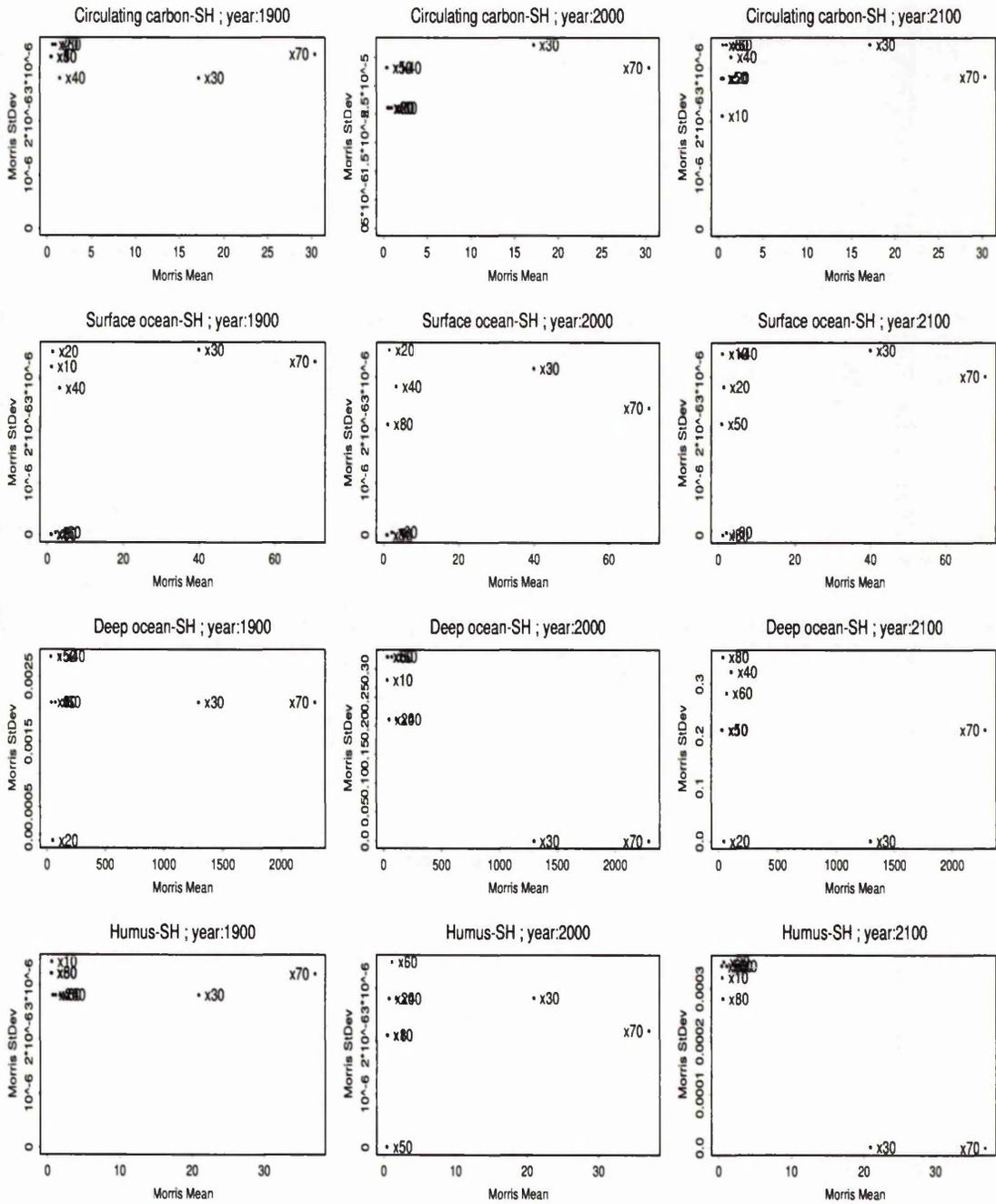
Compartmental Output	Input Factor	Sensitivity Index (SI)					
		Year 1900		Year 2000		Year 2100	
		N=100	N=5000	N=100	N=5000	N=100	N=5000
Circulating Carbon-(NH) <sup>a</sup>		7	7	6	7	7	7
		5	5	4	5	5	5
		2	2	1	2	2	2
		3	3	2	3	3	3
		8	8	7.5	8	8	8
		4	4	3	4	4	4
		1	1	7.5	1	1	1
	6	6	5	6	6	6	
Surface Ocean-(NH)		7	7	6	7	7	7
		5	5	4	5	5	5
		2	2	1	2	2	2
		3	3	2	3	3	3
		8	8	7.5	8	8	8
		4	4	3	4	4	4
		1	1	7.5	1	1	1
	6	6	5	6	6	6	
Deep Ocean-(NH)		7	7	6	7	7	7
		5	5	4	5	5	5
		2	2	1	2	2	2
		3	3	2	3	3	3
		8	8	7.5	8	8	8
		4	4	3	4	4	4
		1	1	7.5	1	1	1
	6	6	5	6	6	6	
Humus-(NH)		7	7	6	7	7	7
		5	5	4	5	5	5
		2	2	1	2	2	2
		3	3	2	3	3	3
		8	8	7.5	8	8	8
		4	4	3	4	4	4
		1	1	7.5	1	1	1
	6	6	5	6	6	6	
Circulating Carbon-(SH) <sup>b</sup>		7	7	6	7	7	7
		5	5	4	5	5	5
		2	2	1	2	2	2
		3	3	2	3	3	3
		8	8	7.5	8	8	8
		4	4	3	4	4	4
		1	1	7.5	1	1	1
	6	6	5	6	6	6	
Surface Ocean-(SH)		7	7	6	7	7	7
		5	5	4	5	5	5
		2	2	1	2	2	2
		3	3	2	3	3	3
		8	8	7.5	8	8	8
		4	4	3	4	4	4
		1	1	7.5	1	1	1
	6	6	5	6	6	6	
Deep Ocean-(SH)		7	7	6	7	7	7
		5	5	4	5	5	5
		2	2	1	2	2	2
		3	3	2	3	3	3
		8	8	7.5	8	8	8
		4	4	3	4	4	4
		1	1	7.5	1	1	1
	6	6	5	6	6	6	
Humus-(SH)		7	7	6	7	7	7
		5	5	4	5	5	5
		2	2	1	2	2	2
		3	3	2	3	3	3
		8	8	7.5	8	8	8
		4	4	3	4	4	4
		1	1	7.5	1	1	1
	6	6	5	6	6	6	

Table B.2. Sensitivity Rankings from SR - Model II initial conditions are varied OAT.

Compartmental Output	Input Factor	Sensitivity Index (SR)					
		Year 1900		Year 2000		Year 2100	
		N=100	N=5000	N=100	N=5000	N=100	N=5000
Circulating Carbon-(NH) <sup>c</sup>	B	7	7	6	7	7	7
	B	5	5	4	5	5	5
	B	2	2	1	2	2	2
	B	3	3	2	3	3	3
	B	8	8	7.5	8	8	8
	B	4	4	3	4	4	4
	B	1	1	7.5	1	1	1
Surface Ocean-(NH)	B	7	7	6	7	7	7
	B	5	5	4	5	5	5
	B	2	2	1	2	2	2
	B	3	3	2	3	3	3
	B	8	8	7.5	8	8	8
	B	4	4	3	4	4	4
	B	1	1	7.5	1	1	1
Deep Ocean-(NH)	B	7	7	6	7	7	7
	B	5	5	4	5	5	5
	B	2	2	1	2	2	2
	B	3	3	2	3	3	3
	B	8	8	7.5	8	8	8
	B	4	4	3	4	4	4
	B	1	1	7.5	1	1	1
Humus-(NH)	B	7	7	6	7	7	7
	B	5	5	4	5	5	5
	B	2	2	1	2	2	2
	B	3	3	2	3	3	3
	B	8	8	7.5	8	8	8
	B	4	4	3	4	4	4
	B	1	1	7.5	1	1	1
Circulating Carbon-(SH) <sup>d</sup>	B	7	7	6	7	7	7
	B	5	5	4	5	5	5
	B	2	2	1	2	2	2
	B	3	3	2	3	3	3
	B	8	8	7.5	8	8	8
	B	4	4	3	4	4	4
	B	1	1	7.5	1	1	1
Surface Ocean-(SH)	B	7	7	6	7	7	7
	B	5	5	4	5	5	5
	B	2	2	1	2	2	2
	B	3	3	2	3	3	3
	B	8	8	7.5	8	8	8
	B	4	4	3	4	4	4
	B	1	1	7.5	1	1	1
Deep Ocean-(SH)	B	7	7	6	7	7	7
	B	5	5	4	5	5	5
	B	2	2	1	2	2	2
	B	3	3	2	3	3	3
	B	8	8	7.5	8	8	8
	B	4	4	3	4	4	4
	B	1	1	7.5	1	1	1
Humus-(SH)	B	7	7	6	7	7	7
	B	5	5	4	5	5	5
	B	2	2	1	2	2	2
	B	3	3	2	3	3	3
	B	8	8	7.5	8	8	8
	B	4	4	3	4	4	4
	B	1	1	7.5	1	1	1
	B	7	7	6	7	7	7
	B	5	5	4	5	5	5
	B	2	2	1	2	2	2
	B	3	3	2	3	3	3
	B	8	8	7.5	8	8	8
	B	4	4	3	4	4	4
	B	1	1	7.5	1	1	1
	B	7	7	6	7	7	7
	B	5	5	4	5	5	5
	B	2	2	1	2	2	2
	B	3	3	2	3	3	3
	B	8	8	7.5	8	8	8
	B	4	4	3	4	4	4
	B	1	1	7.5	1	1	1
	B	7	7	6	7	7	7
	B	5	5	4	5	5	5
	B	2	2	1	2	2	2
	B	3	3	2	3	3	3
	B	8	8	7.5	8	8	8
	B	4	4	3	4	4	4
	B	1	1	7.5	1	1	1
	B	7	7	6	7	7	7
	B	5	5	4	5	5	5
	B	2	2	1	2	2	2
	B	3	3	2	3	3	3
	B	8	8	7.5	8	8	8
	B	4	4	3	4	4	4
	B	1	1	7.5	1	1	1
	B	7	7	6	7	7	7
	B	5	5	4	5	5	5
	B	2	2	1	2	2	2
	B	3	3	2	3	3	3
	B	8	8	7.5	8	8	8
	B	4	4	3	4	4	4
	B	1	1	7.5	1	1	1
	B	7	7	6	7	7	7
	B	5	5	4	5	5	5
	B	2	2	1	2	2	2
	B	3	3	2	3	3	3
	B	8	8	7.5	8	8	8
	B	4	4	3	4	4	4
	B	1	1	7.5	1	1	1
	B	7	7	6	7	7	7
	B	5	5	4	5	5	5
	B	2	2	1	2	2	2
	B	3	3	2	3	3	3
	B	8	8	7.5	8	8	8
	B	4	4	3	4	4	4
	B	1	1	7.5	1	1	1
	B	7	7	6	7	7	7
	B	5	5	4	5	5	5
	B	2	2	1	2	2	2
	B	3	3	2	3	3	3
	B	8	8	7.5	8	8	8
	B	4	4	3	4	4	4
	B	1	1	7.5	1	1	1
	B	7	7	6	7	7	7
	B	5	5	4	5	5	5
	B	2	2	1	2	2	2
	B	3	3	2	3	3	3
	B	8	8	7.5	8	8	8
	B	4	4	3	4	4	4
	B	1	1	7.5	1	1	1
	B	7	7	6	7	7	7
	B	5	5	4	5	5	5
	B	2	2	1	2	2	2
	B	3	3	2	3	3	3
	B	8	8	7.5	8	8	8
	B	4	4	3	4	4	4
	B	1	1	7.5	1	1	1
	B	7	7	6	7	7	7
	B	5	5	4	5	5	5
	B	2	2	1	2	2	2
	B	3	3	2	3	3	3
	B	8	8	7.5	8	8	8
	B	4	4	3	4	4	4
	B	1	1	7.5	1	1	1
	B	7	7	6	7	7	7
	B	5	5	4	5	5	5
	B	2	2	1	2	2	2
	B	3	3	2	3	3	3
	B	8	8	7.5	8	8	8
	B	4	4	3	4	4	4
	B	1	1	7.5	1	1	1
	B	7	7	6	7	7	7
	B	5	5	4	5	5	5
	B	2	2	1	2	2	2
	B	3	3	2	3	3	3
	B	8	8	7.5	8	8	8
	B	4	4	3	4	4	4
	B	1	1	7.5	1	1	1
	B	7	7	6	7	7	7
	B	5	5	4	5	5	5
	B	2	2	1	2	2	2
	B	3	3	2	3	3	3
	B	8	8	7.5	8	8	8
	B	4	4	3	4	4	4
	B	1	1	7.5	1	1	1
	B	7	7	6	7	7	7
	B	5	5	4	5	5	5
	B	2	2	1	2	2	2
	B	3	3	2	3	3	3
	B	8	8	7.5	8	8	8
	B	4	4	3	4	4	4
	B	1	1	7.5	1	1	1
	B	7	7	6	7	7	7
	B	5	5	4	5	5	5
	B	2	2	1	2	2	2
	B	3	3	2	3	3	3
	B	8	8	7.5	8	8	8
	B	4	4	3	4	4	4
	B	1	1	7.5	1	1	1
	B	7	7	6	7	7	7
	B	5	5	4	5	5	5
	B	2	2	1	2	2	2
	B	3	3	2	3	3	3
	B	8	8	7.5	8	8	8
	B	4	4	3	4	4	4
	B	1	1	7.5	1	1	1
	B	7	7	6	7	7	7
	B	5	5	4	5	5	5
	B	2	2	1	2	2	2
	B	3	3	2	3	3	3
	B	8	8	7.5	8	8	8
	B	4	4	3	4	4	4
	B	1	1	7.5	1	1	1
	B	7	7	6	7	7	7
	B	5	5	4	5	5	5
	B	2	2	1	2	2	2
	B	3	3	2	3	3	3
	B	8	8	7.5	8	8	8
	B	4	4	3	4	4	4
	B	1	1	7.5	1	1	1
	B	7	7	6	7	7	7
	B	5	5	4	5	5	5
	B	2	2	1	2	2	2
	B	3	3	2	3	3	3
	B	8	8	7.5	8	8	8
	B	4	4	3	4	4	4
	B	1	1	7.5	1	1	1
	B	7	7	6	7	7	7
	B	5	5	4	5	5	5
	B	2	2	1	2	2	2
	B	3	3	2	3	3	3
	B	8	8	7.5	8	8	8
	B	4	4	3	4	4	4
	B	1	1	7.5	1	1	1
	B	7	7	6	7	7	7
	B	5	5	4	5	5	5
	B	2	2	1	2	2	2
	B	3	3	2	3	3	3
	B	8	8	7.5	8	8	8
	B	4	4	3	4	4	4
	B	1	1	7.5	1	1	1
	B	7	7	6	7	7	7
	B						



**Figure B.9.** Morris screening results on Circulating carbon-(NH), Surface ocean-(NH), Deep ocean-(NH) and Humus-(NH) compartments of Model II in 1900, 2000 and 2100. Mean and standard deviations are associated with the initial conditions considered in the analysis.



**Figure B.10.** Morris screening results on Circulating carbon-(SH), Surface ocean-(SH), Deep ocean-(SH) and Humus-(SH) compartments of Model II in 1900, 2000 and 2100. Mean and standard deviations are associated with the initial conditions considered in the analysis.

**Table B.3.** Results of Morris experiment on Model II. Initial conditions are ranked in order of importance according to the SA measures of Morris mean  $\mu$ . Rankings given here are the same for the three chosen years - 1900, 2000 and 2100.

Compartmental Output	Input Factor	Morris Rank
Circulating Carbon-(NH)	$x_1^0$	7
	$x_2^0$	5
	$x_3^0$	2
	$x_4^0$	3
	$x_5^0$	8
	$x_6^0$	4
	$x_7^0$	1
	$x_8^0$	6
Surface Ocean-(NH)	$x_1^0$	7
	$x_2^0$	5
	$x_3^0$	2
	$x_4^0$	3
	$x_5^0$	8
	$x_6^0$	4
	$x_7^0$	1
	$x_8^0$	6
Deep Ocean-(NH)	$x_1^0$	7
	$x_2^0$	5
	$x_3^0$	2
	$x_4^0$	3
	$x_5^0$	8
	$x_6^0$	4
	$x_7^0$	1
	$x_8^0$	6
Humus-(NH)	$x_1^0$	7
	$x_2^0$	5
	$x_3^0$	2
	$x_4^0$	3
	$x_5^0$	8
	$x_6^0$	4
	$x_7^0$	1
	$x_8^0$	6

Compartmental Output	Input Factor	Morris Rank
Circulating Carbon-(SH)	$x_1^0$	7
	$x_2^0$	5
	$x_3^0$	2
	$x_4^0$	3
	$x_5^0$	8
	$x_6^0$	4
	$x_7^0$	1
	$x_8^0$	6
Surface Ocean-(SH)	$x_1^0$	7
	$x_2^0$	5
	$x_3^0$	2
	$x_4^0$	3
	$x_5^0$	8
	$x_6^0$	4
	$x_7^0$	1
	$x_8^0$	6
Deep Ocean-(SH)	$x_1^0$	7
	$x_2^0$	5
	$x_3^0$	2
	$x_4^0$	3
	$x_5^0$	8
	$x_6^0$	4
	$x_7^0$	1
	$x_8^0$	6
Humus-(SH)	$x_1^0$	7
	$x_2^0$	5
	$x_3^0$	2
	$x_4^0$	3
	$x_5^0$	8
	$x_6^0$	4
	$x_7^0$	1
	$x_8^0$	6

B.1.2 Transfer Coefficients

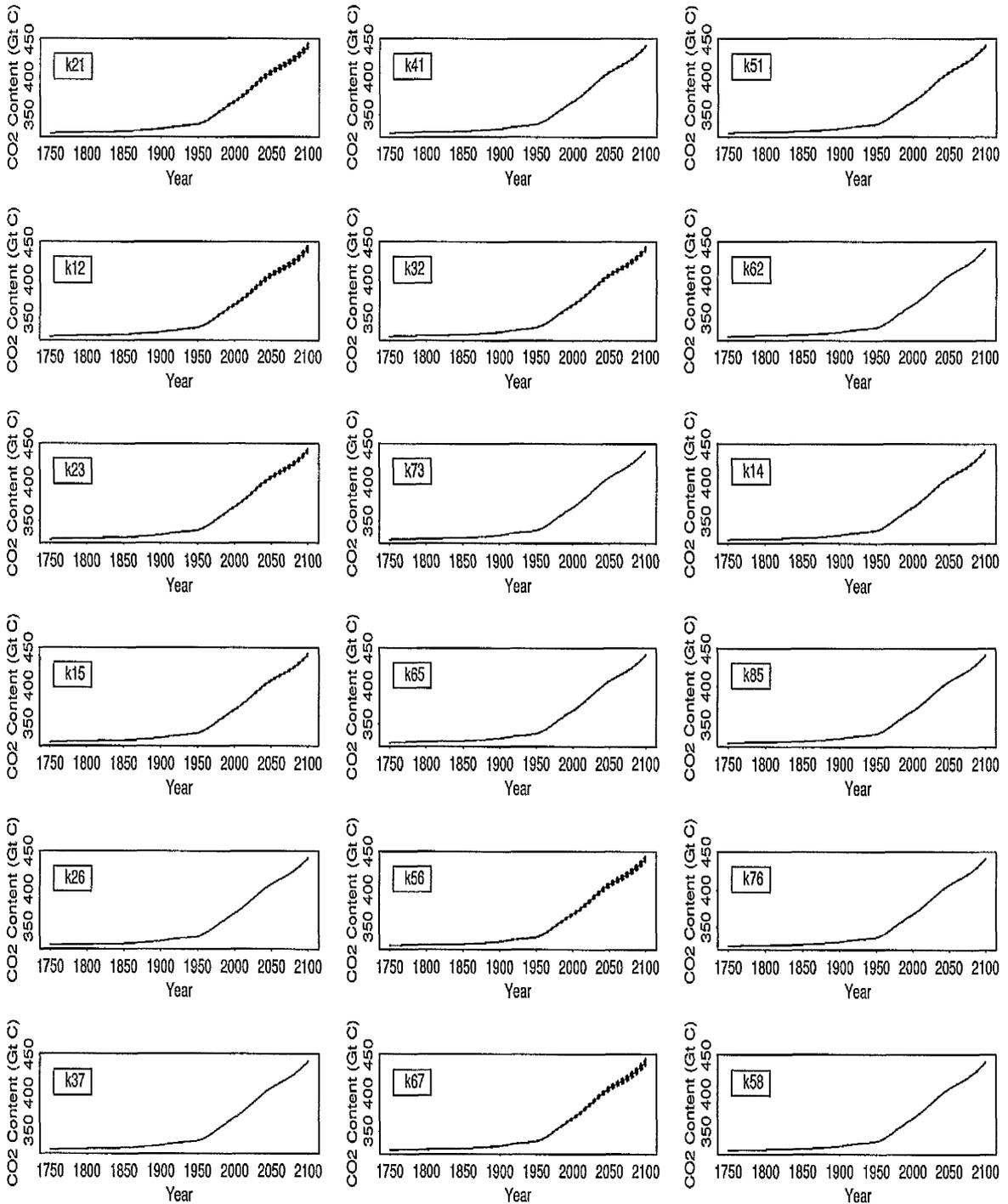
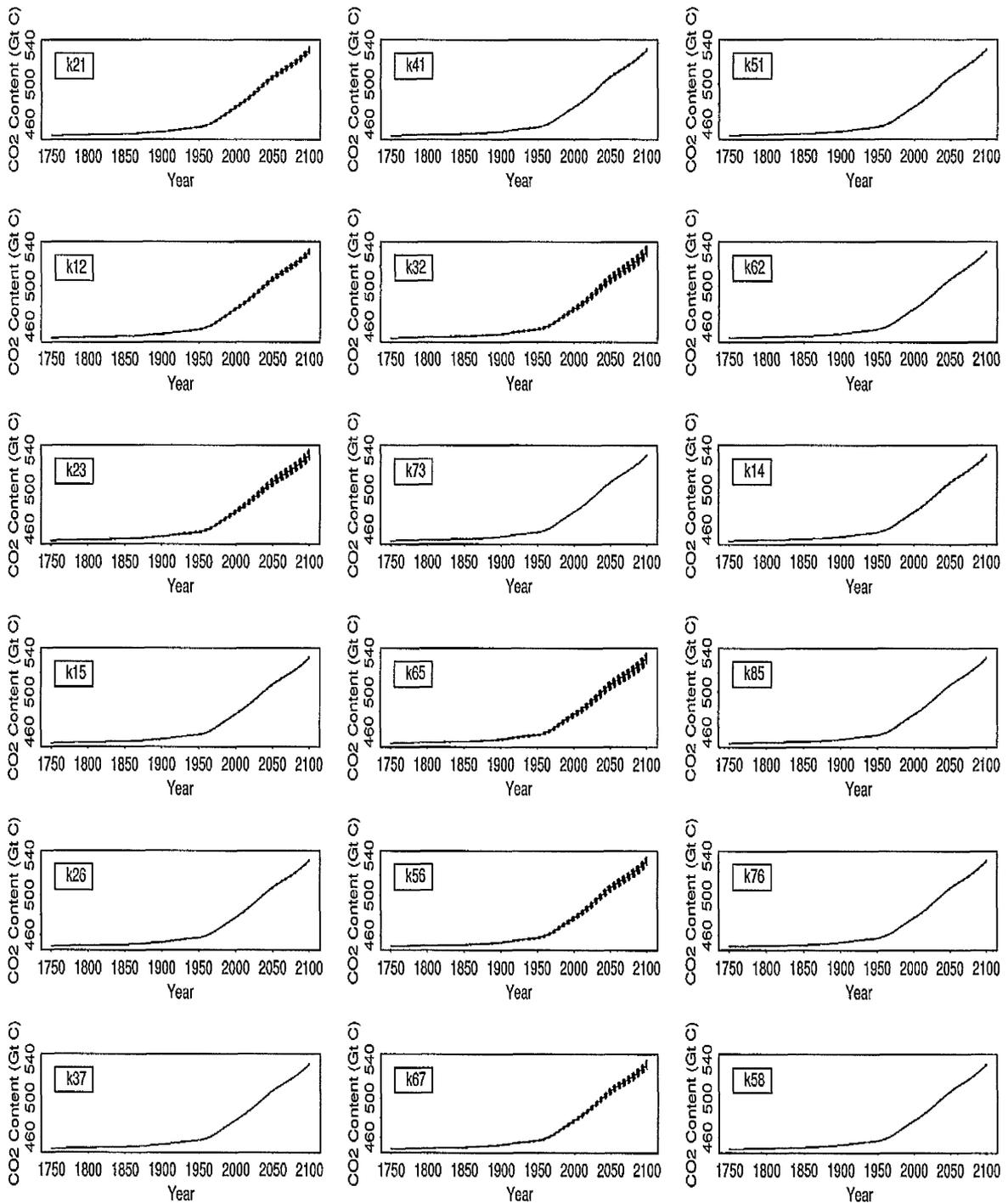
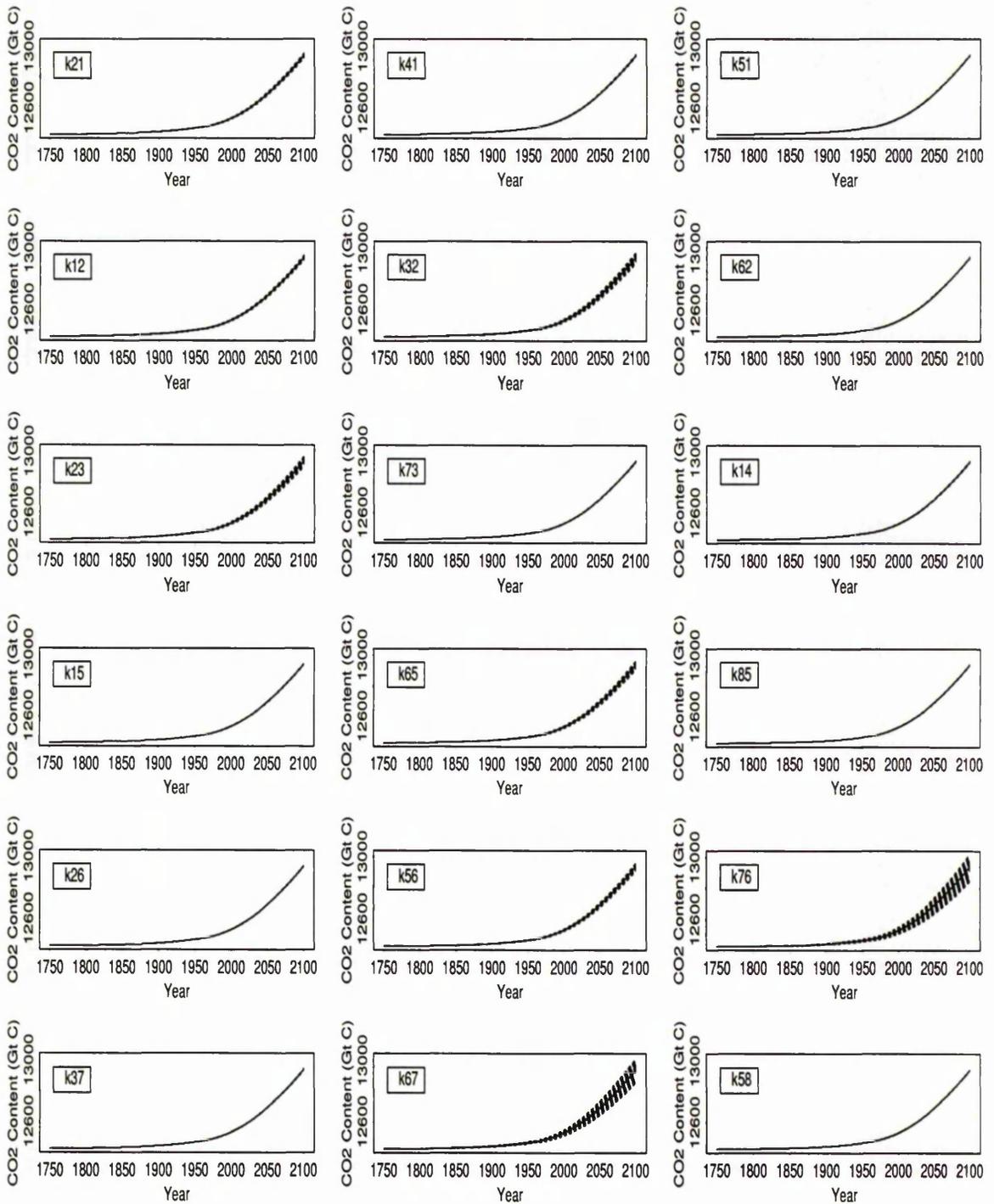


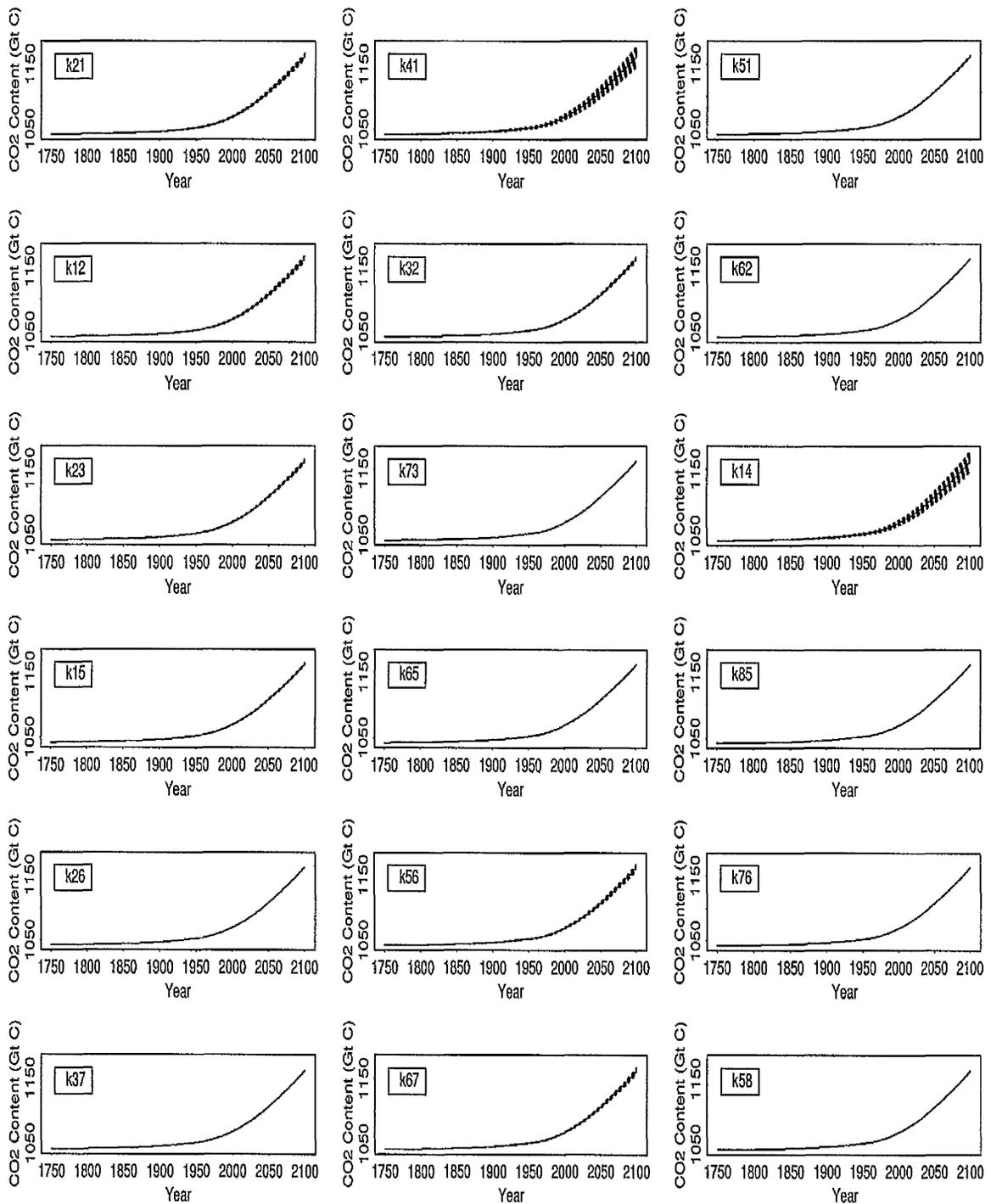
Figure B.11. Circulating carbon-(NH) CO<sub>2</sub> predictions resulting from varying transfer coefficients  $k_{ij}$  OAT (given in top-left corner of each graph - see Table 3.5 for description of these factors). N=100 model runs, IS92a input scenario is considered.



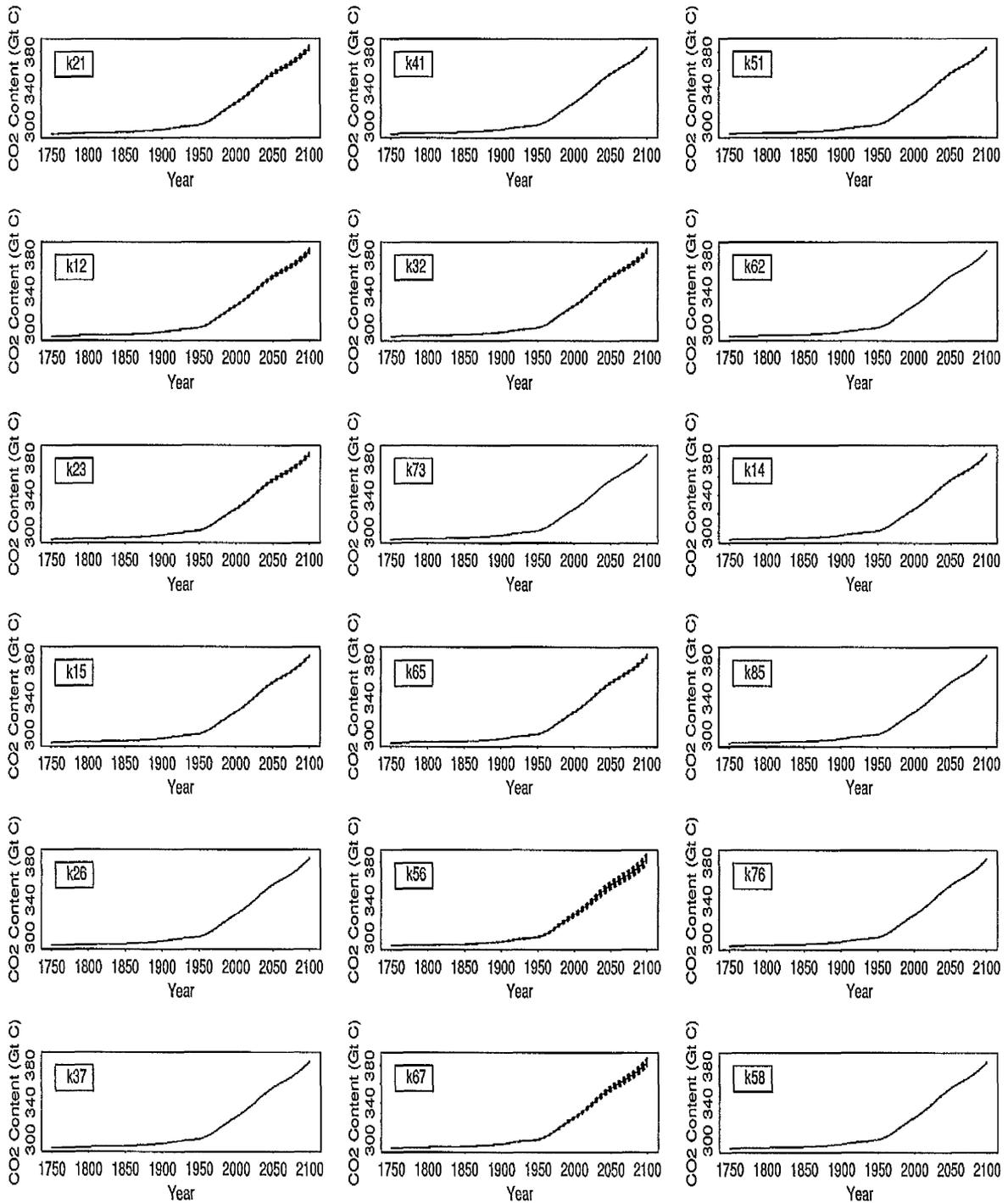
**Figure B.12.** Surface Ocean-(NH) CO<sub>2</sub> predictions resulting from varying transfer coefficients  $k_{ij}$  OAT (given in top-left corner of each graph - see Table 3.5 for description of these factors). N=100 model runs, IS92a input scenario is considered.



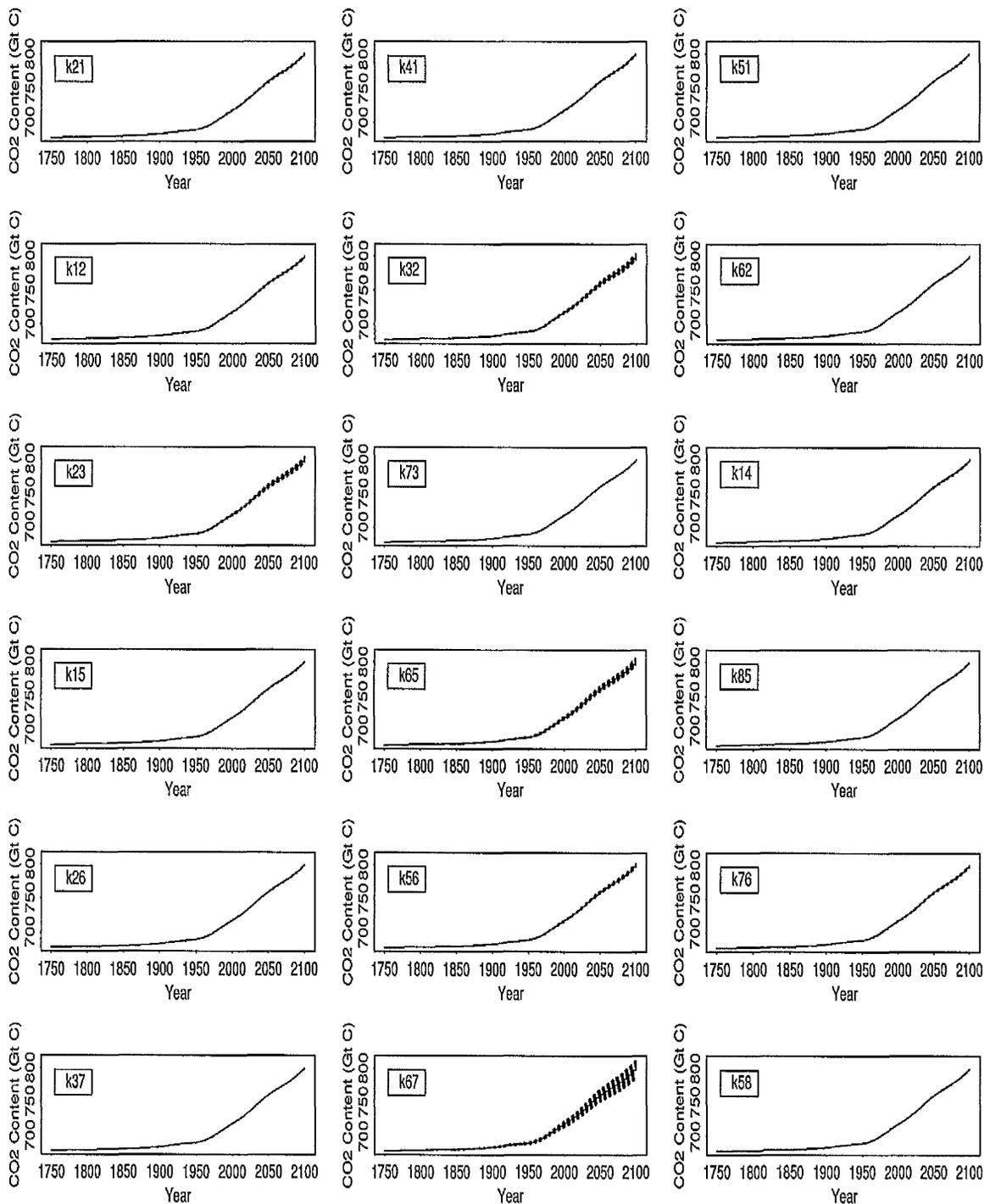
**Figure B.13.** Deep Ocean-(NH) CO<sub>2</sub> predictions resulting from varying transfer coefficients  $k_{ij}$  OAT (given in top-left corner of each graph - see Table 3.5 for description of these factors). N=100 model runs, IS92a input scenario is considered.



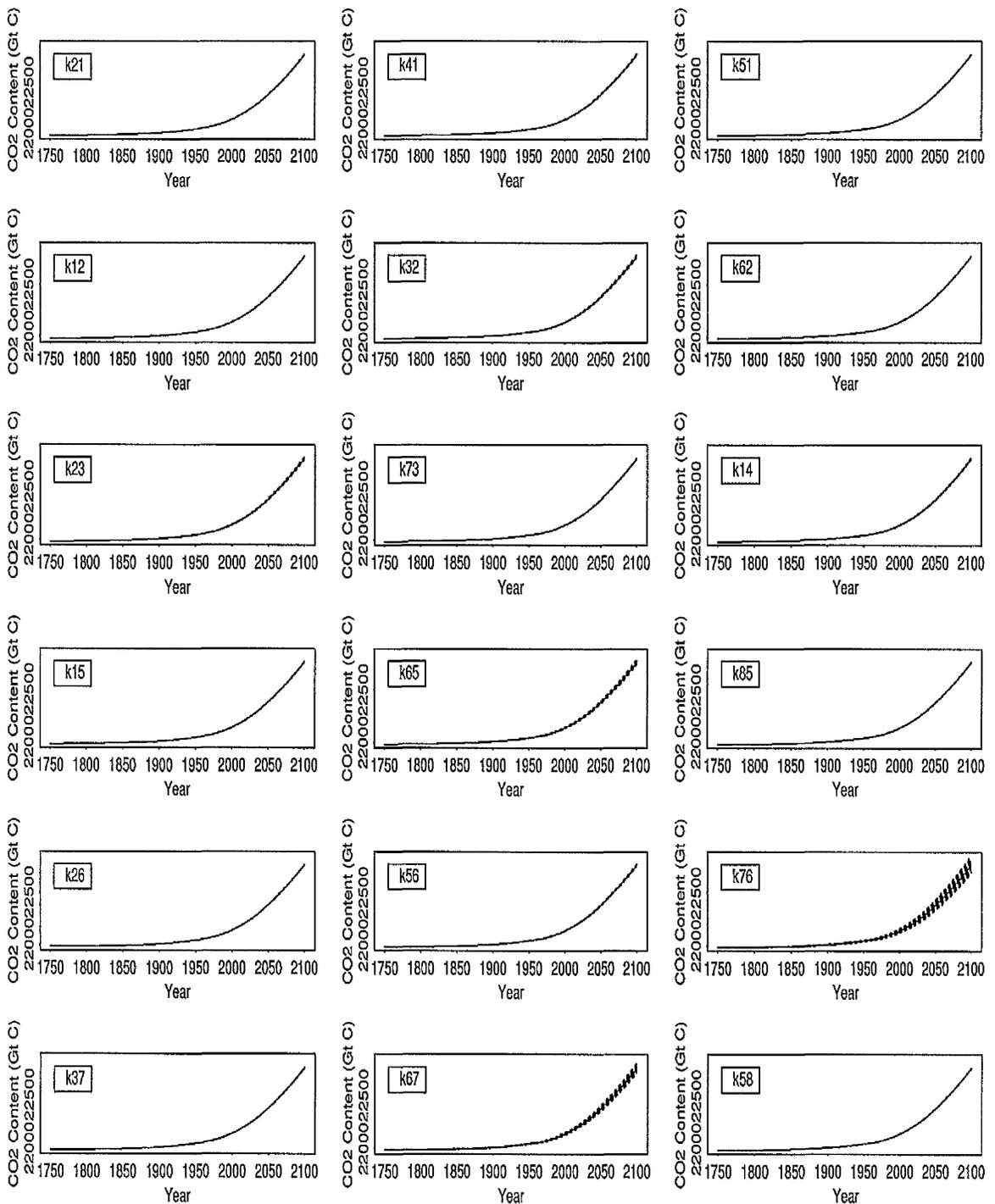
**Figure B.14.** Humus-(NH) CO<sub>2</sub> predictions resulting from varying transfer coefficients  $k_{ij}$  OAT (given in top-left corner of each graph - see Table 3.5 for description of these factors). N=100 model runs, IS92a input scenario is considered.



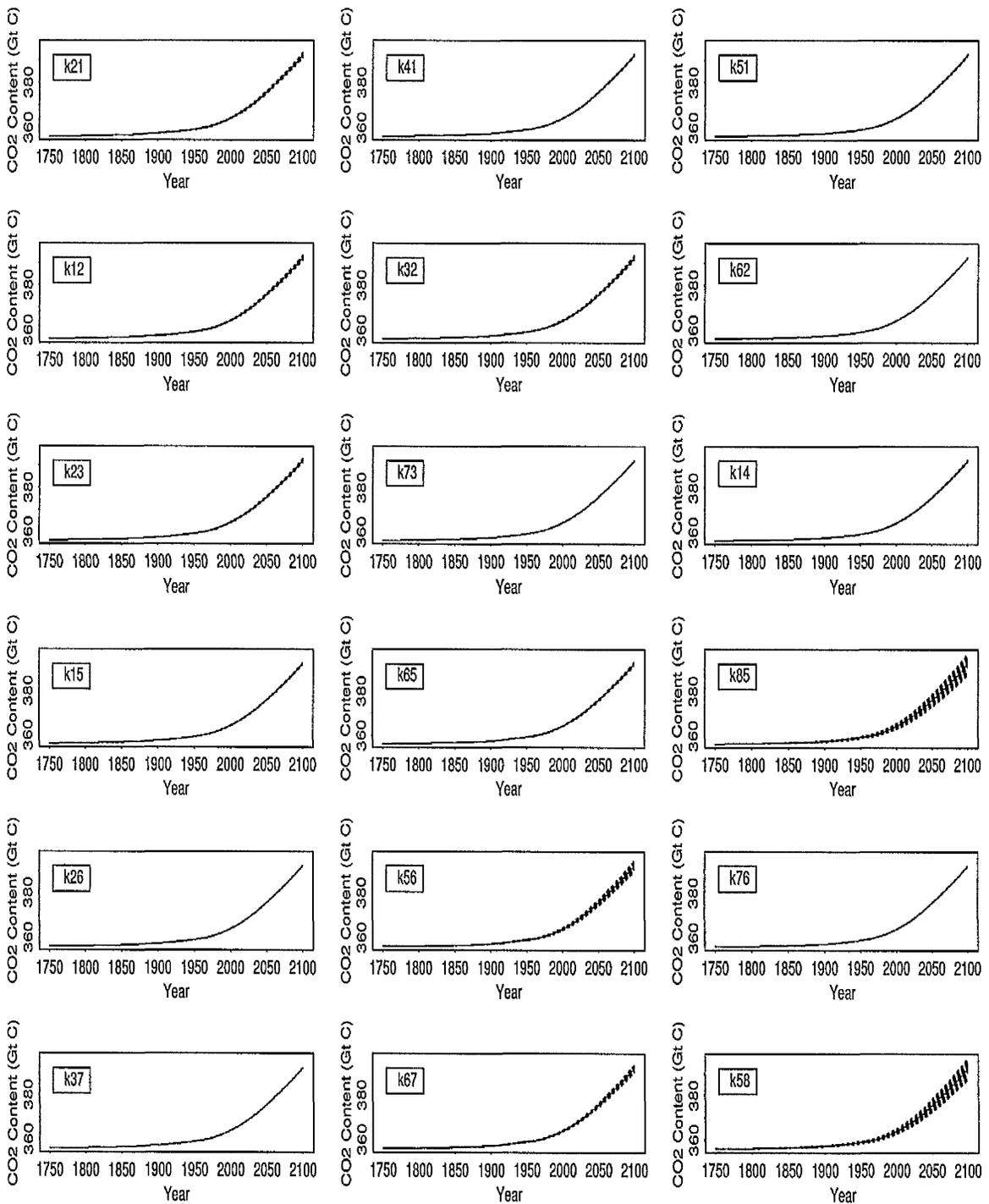
**Figure B.15.** Circulating Carbon-(SH) CO<sub>2</sub> predictions resulting from varying transfer coefficients  $k_{ij}$  OAT (given in top-left corner of each graph - see Table 3.5 for description of these factors). N=100 model runs, IS92a input scenario is considered.



**Figure B.16.** Surface Ocean-(SH) CO<sub>2</sub> predictions resulting from varying transfer coefficients  $k_{ij}$  OAT (given in top-left corner of each graph - see Table 3.5 for description of these factors). N=100 model runs, IS92a input scenario is considered.



**Figure B.17.** Deep Ocean-(SH) CO<sub>2</sub> predictions resulting from varying transfer coefficients  $k_{ij}$  OAT (given in top-left corner of each graph - see Table 3.5 for description of these factors). N=100 model runs, IS92a input scenario is considered.



**Figure B.18.** Humus-(SH) CO<sub>2</sub> predictions resulting from varying transfer coefficients  $k_{ij}$  OAT (given in top-left corner of each graph - see Table 3.5 for description of these factors). N=100 model runs, IS92a input scenario is considered.

Table B.4. Sensitivity Rankings from SI - Model II transfer coefficients  $k_{ij}$  are varied OAT.

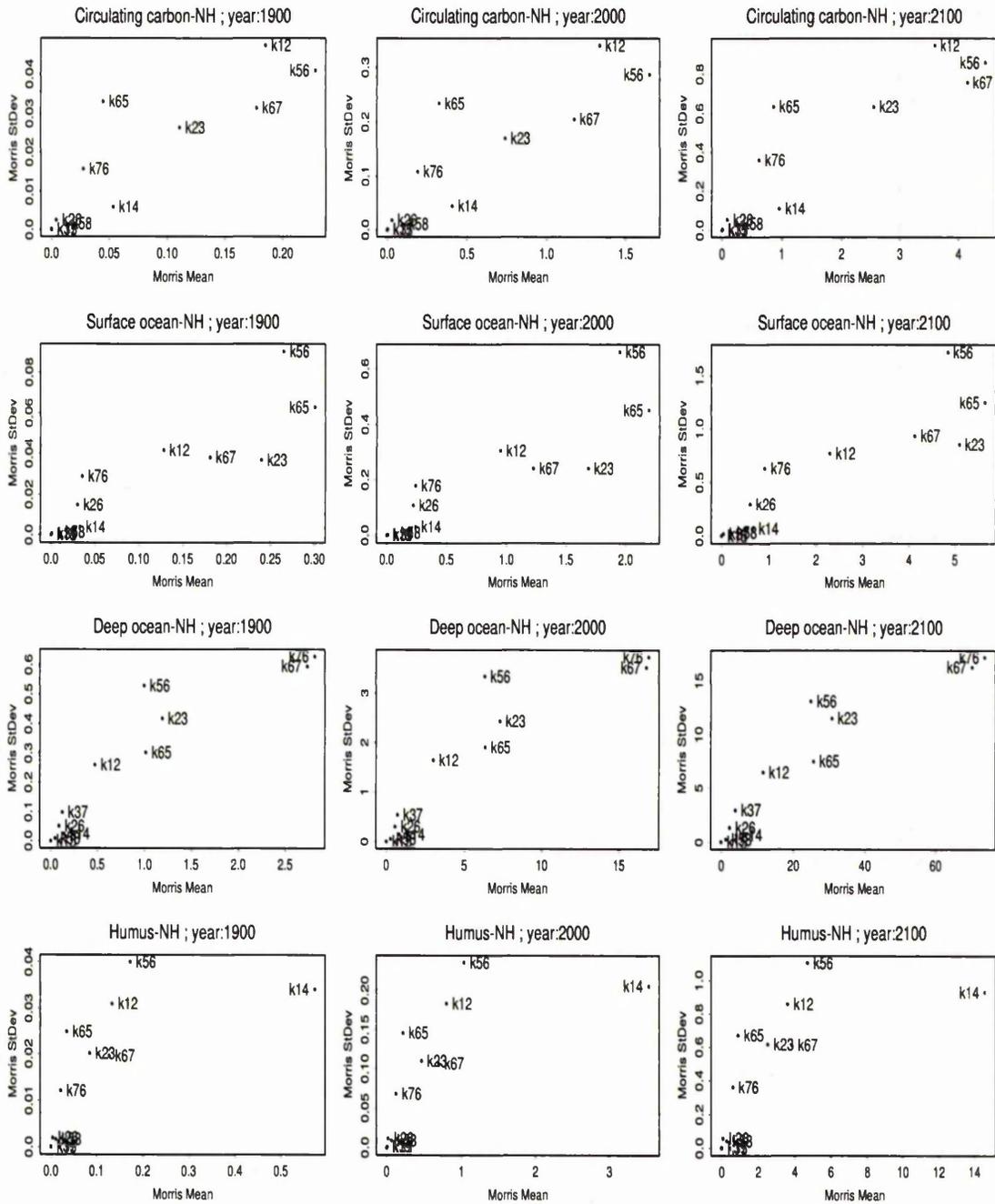
Compartment	Input Factor	Sensitivity Index (SI)					
		Yr: 1900		Yr: 2000		Yr: 2100	
		N=100	N=5000	N=100	N=5000	N=100	N=5000
Circulating Carbon-(NH)	k <sub>21</sub>	2	2	2	2	3	3
	k <sub>41</sub>	10	9.5	10	10	10	10
	k <sub>51</sub>	8	7.5	8	8	8	8
	k <sub>12</sub>	3	3	3.5	3	4	4
	k <sub>32</sub>	5	5	6	5	5	5
	k <sub>62</sub>	15.5	15.5	15	15.5	15	15.5
	k <sub>23</sub>	6	6	3.5	6	6	6
	k <sub>73</sub>	18	17.5	18	17.5	18	17.5
	k <sub>14</sub>	9	9.5	9	9	9	9
	k <sub>15</sub>	7	7.5	7	7	7	7
	k <sub>65</sub>	11	11	11	11	11	11
	k <sub>85</sub>	13	12.5	13	13	13	13
	k <sub>26</sub>	15.5	15.5	16	15.5	16	15.5
	k <sub>56</sub>	1	1	1	1	2	2
	k <sub>76</sub>	14	14	14	14	14	14
	k <sub>37</sub>	17	17.5	17	17.5	17	17.5
k <sub>67</sub>	4	4	5	4	1	1	
k <sub>58</sub>	12	12.5	12	12	12	12	
Surface Ocean-(NH)	k <sub>21</sub>	6	6	5	6	6	6
	k <sub>41</sub>	9	9	9	9	9	9
	k <sub>51</sub>	11	10.5	11	11	12	12
	k <sub>12</sub>	7	7	6.5	7	7	7
	k <sub>32</sub>	2	2	2	2	2	1
	k <sub>62</sub>	12	12.5	12	12	13	13
	k <sub>23</sub>	4	4	6.5	4	3	3
	k <sub>73</sub>	18	17.5	18	17.5	18	18
	k <sub>14</sub>	8	8	8	8	8	8
	k <sub>15</sub>	10	10.5	10	10	11	11
	k <sub>65</sub>	1	1	1	1	2	2
	k <sub>85</sub>	16	15.5	15	16	16	16
	k <sub>26</sub>	13	12.5	13	13	14	14
	k <sub>56</sub>	3	3	3	3	4	4
	k <sub>76</sub>	14	14	16	16	10	10
	k <sub>37</sub>	17	17.5	17	17.5	17	17
k <sub>67</sub>	5	5	4	5	5	5	
k <sub>58</sub>	15	15.5	14	14	15	15	
Deep Ocean-(NH)	k <sub>21</sub>	7	7	6	7	7	7
	k <sub>41</sub>	10	10	10	10	10	10
	k <sub>51</sub>	14	14	14	14	14	14
	k <sub>12</sub>	8	8	7.5	8	8	8
	k <sub>32</sub>	3	3	3	3	3	3
	k <sub>62</sub>	15	15	15	15	15	15
	k <sub>23</sub>	4	4	7.5	4	4	4
	k <sub>73</sub>	12	12	12	12	12	12
	k <sub>14</sub>	9	9	9	9	9	9
	k <sub>15</sub>	13	13	13	13	13	13
	k <sub>65</sub>	5	5	4	5	5	5
	k <sub>85</sub>	18	18	18	18	18	18
	k <sub>26</sub>	16	16	16	16	16	16
	k <sub>56</sub>	6	6	5	6	6	6
	k <sub>76</sub>	1	1	1	1	1	1
	k <sub>37</sub>	11	11	11	11	11	11
k <sub>67</sub>	2	2	2	2	2	2	
k <sub>58</sub>	17	17	17	17	17	17	
Humus-(NH)	k <sub>21</sub>	4	4	4	4	5	5
	k <sub>41</sub>	2	2	2	2	2	2
	k <sub>51</sub>	10	9.5	10	10	10	10
	k <sub>12</sub>	5	5	5.5	5	6	6
	k <sub>32</sub>	7	7	8	7	7	7
	k <sub>62</sub>	15	15.5	15	15.5	15	15.5
	k <sub>23</sub>	8	8	5.5	8	8	8
	k <sub>73</sub>	18	17.5	18	17.5	18	18
	k <sub>14</sub>	1	1	1	1	1	1
	k <sub>15</sub>	9	9.5	9	9	9	9
	k <sub>65</sub>	11	11	11	11	11	11
	k <sub>85</sub>	13	12.5	13	13	13	13
	k <sub>26</sub>	16	15.5	16	15.5	16	15.5
	k <sub>56</sub>	3	3	3	3	3	3
	k <sub>76</sub>	14	14	14	14	14	14
	k <sub>37</sub>	17	17.5	17	17.5	17	17
k <sub>67</sub>	6	6	7	6	4	4	
k <sub>58</sub>	12	12.5	12	12	12	12	

Compartment	Input Factor	Sensitivity Index (SI)					
		Yr: 1900		Yr: 2000		Yr: 2100	
		N=100	N=5000	N=100	N=5000	N=100	N=5000
Circulating Carbon-(SH)	k <sub>21</sub>	3	3	3	3	3	3
	k <sub>41</sub>	11	10.5	11	11	11	11
	k <sub>51</sub>	9	9	9	9	9	9
	k <sub>12</sub>	4	4	4.5	4	4	4
	k <sub>32</sub>	5	5	6	5	5	5
	k <sub>62</sub>	15	15.5	15	15.5	15	15.5
	k <sub>23</sub>	6	6	4.5	6	6	6
	k <sub>73</sub>	17.5	17.5	18	17.5	18	17.5
	k <sub>14</sub>	10	10.5	10	10	10	10
	k <sub>15</sub>	8	8	8	8	8	8
	k <sub>65</sub>	7	7	7	7	7	7
	k <sub>85</sub>	14	14	14	14	14	14
	k <sub>26</sub>	16	15.5	16	15.5	16	15.5
	k <sub>56</sub>	1	1	1	1	1	1
	k <sub>76</sub>	12	12	12	12	12	12
	k <sub>37</sub>	17.5	17.5	17	17.5	17	17.5
k <sub>67</sub>	2	2	2	2	2	2	
k <sub>58</sub>	13	13	13	13	13	13	
Surface Ocean-(SH)	k <sub>21</sub>	7	7	6	7	7	7
	k <sub>41</sub>	10	9.5	10	10	10	10
	k <sub>51</sub>	12	12	12	12	12	12
	k <sub>12</sub>	8	8	7.5	8	8	8
	k <sub>32</sub>	3	3	3	3	3	3
	k <sub>62</sub>	13	13.5	13	13	13	13
	k <sub>23</sub>	4	4	7.5	4	4	4
	k <sub>73</sub>	18	17.5	18	17.5	18	18
	k <sub>14</sub>	9	9.5	9	9	9	9
	k <sub>15</sub>	11	11	11	11	11	11
	k <sub>65</sub>	2	2	2	2	2	2
	k <sub>85</sub>	16	16	16	16	16	16
	k <sub>26</sub>	14	13.5	14	14	14	14
	k <sub>56</sub>	5	5	4	5	5	5
	k <sub>76</sub>	6	6	5	6	6	6
	k <sub>37</sub>	17	17.5	17	17.5	17	17
k <sub>67</sub>	1	1	1	1	1	1	
k <sub>58</sub>	15	15	15	15	15	15	
Deep Ocean-(SH)	k <sub>21</sub>	11	11	10	11	9	9
	k <sub>41</sub>	8	8	7	8	8	8
	k <sub>51</sub>	10	10	9	10	13	12
	k <sub>12</sub>	12	12	11.5	12	10	10
	k <sub>32</sub>	4	4	4	4	4	4
	k <sub>62</sub>	17	17.5	17	17	17	17
	k <sub>23</sub>	5	5	11.5	5	5	5
	k <sub>73</sub>	14	14	14	14	14	14
	k <sub>14</sub>	7	7	6	7	7	7
	k <sub>15</sub>	9	9	8	9	12	11
	k <sub>65</sub>	3	3	3	3	3	3
	k <sub>85</sub>	16	16	16	16	16	16
	k <sub>26</sub>	18	17.5	18	18	18	18
	k <sub>56</sub>	6	6	5	6	6	6
	k <sub>76</sub>	1	1	1	1	1	1
	k <sub>37</sub>	13	13	13	13	11	13
k <sub>67</sub>	2	2	2	2	2	2	
k <sub>58</sub>	15	15	15	15	15	15	
Humus-(SH)	k <sub>21</sub>	5	5	5	5	5	5
	k <sub>41</sub>	13	12.5	13	12.5	13	13
	k <sub>51</sub>	11	10.5	11	10.5	11	11
	k <sub>12</sub>	6	6	6.5	6	6	6
	k <sub>32</sub>	7	7	8	7	7	7
	k <sub>62</sub>	15.5	15.5	15	15.5	15	15.5
	k <sub>23</sub>	8	8	6.5	8	8	8
	k <sub>73</sub>	17.5	17.5	17.5	17.5	18	17.5
	k <sub>14</sub>	12	12.5	12	12.5	12	12
	k <sub>15</sub>	10	10.5	10	10.5	10	10
	k <sub>65</sub>	9	9	9	9	9	9
	k <sub>85</sub>	2	2	2	2	2	2
	k <sub>26</sub>	15.5	15.5	16	15.5	16	15.5
	k <sub>56</sub>	3	3	3	3	3	3
	k <sub>76</sub>	14	14	14	14	14	14
	k <sub>37</sub>	17.5	17.5	17.5	17.5	17	17.5
k <sub>67</sub>	4	4	4	4	4	4	
k <sub>58</sub>	1	1	1	1	1	1	

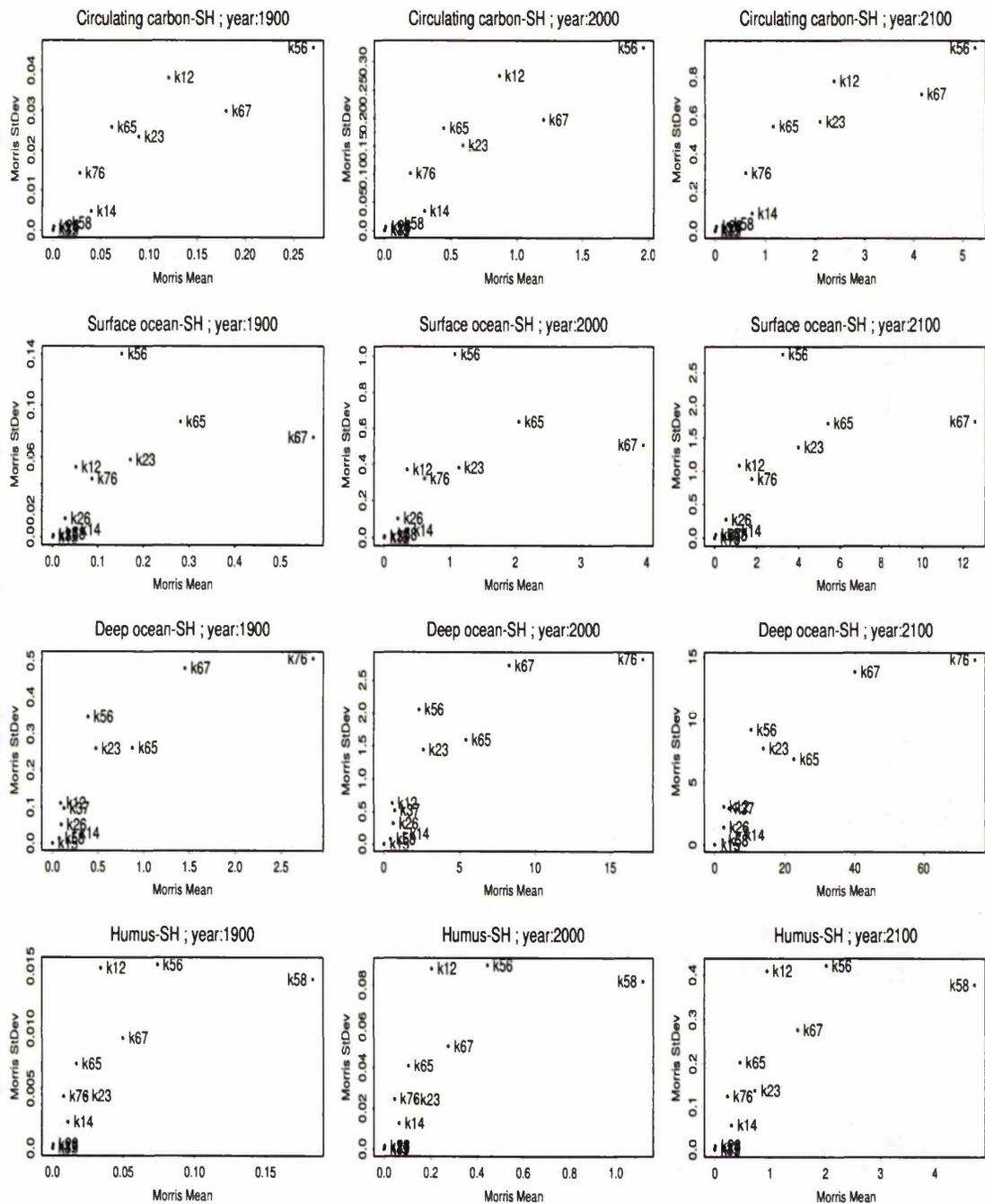
Table B.5. Sensitivity Rankings from SR - Model II transfer coefficients  $k_{ij}$  are varied OAT.

Compartment	Input Factor	Standardized Range (SR)					
		Yr: 1900		Yr: 2000		Yr: 2100	
		N=100	N=5000	N=100	N=5000	N=100	N=5000
Circulating Carbon-(NH)	k21	2	2	2	2	2	2
	k41	10	9.5	10	9.5	10	10
	k51	8	7.5	8	8	8	8
	k12	3	3	3.5	3	3	3
	k32	5	5	6	5	5	5
	k62	15.5	15.5	15	15.5	15	15.5
	k23	6	6	3.5	6	6	6
	k73	18	18	18	18	18	18
	k14	9	9.5	9	9.5	9	9
	k15	7	7.5	7	7	7	7
	k65	11	11	11	11	11	11
	k85	12	12	12	12	12	12
	k26	15.5	15.5	16	15.5	16	15.5
	k56	1	1	1	1	1	1
	k76	13	13	13	13	13	13
	k37	17	17	17	17	17	17
k87	4	4	5	4	4	4	
k58	14	14	14	14	14	14	
Surface Ocean-(NH)	k21	6	6	5	6	6	6
	k41	9	8	9	9	9	8.5
	k51	11	10.5	11	11	12	12
	k12	7	7	6.5	7	7	7
	k32	2	2	2	2	1	1
	k62	12	12.5	12	12	13	13
	k23	4	4	6.5	4	3	3
	k73	18	17.5	18	17.5	18	18
	k14	8	9	8	8	8	8.5
	k15	10	10.5	10	10	11	11
	k65	1	1	1	1	2	2
	k85	16	15.5	15	15	16	16
	k26	13	12.5	13	13	14	14
	k56	3	3	3	3	4	4
	k76	14	14	16	16	10	10
	k37	17	17.5	17	17.5	17	17
k87	5	5	4	5	5	5	
k58	15	15.5	14	14	15	15	
Deep Ocean-(NH)	k21	7	7	6	7	7	7
	k41	10	9.5	10	10	10	10
	k51	14	14	14	14	14	14
	k12	8	8	7.5	8	8	8
	k32	3	3	3	3	3	3
	k62	15	15	15	15	15	15
	k23	4	4	7.5	4	4	4
	k73	12	12	12	12	12	12
	k14	9	9.5	9	9	9	9
	k15	13	13	13	13	13	13
	k65	5	5	4	5	5	5
	k85	18	18	18	18	18	18
	k26	16	16	16	16	16	16
	k56	6	6	5	6	6	6
	k76	1	1	1	1	1	1
	k37	11	11	11	11	11	11
k87	2	2	2	2	2	2	
k58	17	17	17	17	17	17	
Humus-(NH)	k21	4	4	4	4	5	5
	k41	2	2	2	2	2	2
	k51	10	9.5	10	10	10	10
	k12	5	5	5.5	5	6	6
	k32	7	7	8	7	7	7
	k62	15	15.5	15	15.5	15	15.5
	k23	8	8	5.5	8	8	8
	k73	18	17.5	18	17.5	18	18
	k14	1	1	1	1	1	1
	k15	9	9.5	9	9	9	9
	k65	11	11	11	11	11	11
	k85	13	12.5	13	13	13	13
	k26	16	15.5	16	15.5	16	15.5
	k56	3	3	3	3	3	3
	k76	14	14	14	14	14	14
	k37	17	17.5	17	17.5	17	17
k87	6	6	7	6	4	4	
k58	12	12.5	12	12	12	12	

Compartment	Input Factor	Standardized Range (SR)					
		Yr: 1900		Yr: 2000		Yr: 2100	
		N=100	N=5000	N=100	N=5000	N=100	N=5000
Circulating Carbon-(SH)	k21	3	3	3	3	3	3
	k41	11	10.5	11	11	11	10.5
	k51	9	9	9	9	9	9
	k12	4	4	4.5	4	4	4
	k32	5	5	6	5	5	5
	k62	15	15.5	15	15.5	15	15.5
	k23	6	6	4.5	6	6	6
	k73	17.5	17.5	18	17.5	18	17.5
	k14	10	10.5	10	10	10	10.5
	k15	8	8	8	8	8	8
	k65	7	7	7	7	7	7
	k85	14	14	14	14	14	14
	k26	16	15.5	16	15.5	16	15.5
	k56	1	1	1	1	1	1
	k76	12	12	12	12	12	12
	k37	17.5	17.5	17	17.5	17	17.5
k87	2	2	2	2	2	2	
k58	13	13	13	13	13	13	
Surface Ocean-(SH)	k21	7	7	6	7	7	7
	k41	10	9.5	10	10	10	10
	k51	12	12	12	12	12	12
	k12	8	8	7.5	8	8	8
	k32	3	3	3	3	3	3
	k62	13	13.5	13	13	13	13
	k23	4	4	7.5	4	4	4
	k73	18	17.5	18	17.5	18	18
	k14	9	9.5	9	9	9	9
	k15	11	11	11	11	11	11
	k65	2	2	2	2	2	2
	k85	16	16	16	16	16	16
	k26	14	13.5	14	14	14	14
	k56	5	5	4	5	5	5
	k76	6	6	5	6	6	6
	k37	17	17.5	17	17.5	17	17
k87	1	1	1	1	1	1	
k58	15	15	15	15	15	15	
Deep Ocean-(SH)	k21	11	11	10	11	9	9
	k41	8	8	7	8	8	8
	k51	10	10	9	10	13	12
	k12	12	12	11.5	12	10	10
	k32	4	4	4	4	4	4
	k62	17	17.5	17	17	17	17
	k23	5	5	11.5	5	5	5
	k73	14	14	14	14	14	14
	k14	7	7	6	7	7	7
	k15	9	9	8	9	12	11
	k65	3	3	3	3	3	3
	k85	16	16	16	16	16	16
	k26	18	17.5	18	18	18	18
	k56	6	6	5	6	6	6
	k76	1	1	1	1	1	1
	k37	13	13	13	13	11	13
k87	2	2	2	2	2	2	
k58	15	15	15	15	15	15	
Humus-(SH)	k21	5	5	5	5	5	5
	k41	13	12.5	13	12.5	13	12
	k51	11	10.5	11	10.5	11	11
	k12	6	6	6.5	6	6	6
	k32	7	7	8	7	7	7
	k62	15.5	15.5	15	15.5	15	15.5
	k23	8	8	6.5	8	8	8
	k73	17.5	17.5	17.5	17.5	18	17.5
	k14	12	12.5	12	12.5	12	13
	k15	10	10.5	10	10.5	10	10
	k65	9	9	9	9	9	9
	k85	2	2	2	2	2	2
	k26	15.5	15.5	16	15.5	16	15.5
	k56	3	3	3	3	3	3
	k76	14	14	14	14	14	14
	k37	17.5	17.5	17.5	17.5	17	17.5
k87	4	4	4	4	4	4	
k58	1	1	1	1	1	1	



**Figure B.19.** Morris screening results on Circulating carbon-(NH), Surface ocean-(NH), Deep ocean-(NH) and Humus-(NH) compartments of Model II in 1900, 2000 and 2100. Mean and standard deviations are associated with the transfer coefficients considered in the analysis.



**Figure B.20.** Morris screening results on Circulating carbon-(SH), Surface ocean-(SH), Deep ocean-(SH) and Humus-(SH) compartments of Model II in 1900, 2000 and 2100. Mean and standard deviations are associated with the transfer coefficients considered in the analysis.

**Table B.6.** Results of Morris experiment on Model II. Initial conditions are ranked in order of importance according to the SA measures of Morris mean  $\mu$ . Rankings are the same for the three chosen years - 1900, 2000 and 2100.

Compartmental Output	Input Factor	Morris Rank
Circulating Carbon-(NH)	$k_{12}$	2
	$k_{23}$	4
	$k_{14}$	5
	$k_{15}$	11
	$k_{65}$	6
	$k_{26}$	9
	$k_{56}$	1
	$k_{76}$	7
	$k_{37}$	10
Surface Ocean-(NH)	$k_{67}$	3
	$k_{58}$	8
	$k_{12}$	5
	$k_{23}$	3
	$k_{14}$	7
	$k_{15}$	11
	$k_{65}$	1
	$k_{26}$	8
	$k_{56}$	2
Deep Ocean-(NH)	$k_{76}$	6
	$k_{37}$	10
	$k_{67}$	4
	$k_{58}$	9
	$k_{12}$	6
	$k_{23}$	3
	$k_{14}$	7
	$k_{15}$	11
	$k_{65}$	4
Humus-(NH)	$k_{26}$	9
	$k_{56}$	2
	$k_{76}$	7
	$k_{37}$	10
	$k_{67}$	4
	$k_{58}$	8
	$k_{12}$	3
	$k_{23}$	5
	$k_{14}$	1

Compartmental Output	Input Factor	Morris Rank
Circulating Carbon-(SH)	$k_{12}$	3
	$k_{23}$	4
	$k_{14}$	6
	$k_{15}$	11
	$k_{65}$	5
	$k_{26}$	9
	$k_{56}$	1
	$k_{76}$	7
	$k_{37}$	10
Surface Ocean-(SH)	$k_{67}$	2
	$k_{58}$	8
	$k_{12}$	6
	$k_{23}$	3
	$k_{14}$	7
	$k_{15}$	11
	$k_{65}$	2
	$k_{26}$	8
	$k_{56}$	4
Deep Ocean-(SH)	$k_{76}$	5
	$k_{37}$	10
	$k_{67}$	1
	$k_{58}$	9
	$k_{12}$	9
	$k_{23}$	4
	$k_{14}$	6
	$k_{15}$	11
	$k_{65}$	3
Humus-(SH)	$k_{26}$	8
	$k_{56}$	5
	$k_{76}$	1
	$k_{37}$	7
	$k_{67}$	2
	$k_{58}$	10
	$k_{12}$	4
	$k_{23}$	5
	$k_{14}$	7

## B.2 Local SA

### B.2.1 Initial Conditions

Table B.7. Rankings of the transfer coefficients of Model II based on standardised sensitivity coefficients evaluated in 1900, 2000 and 2100.

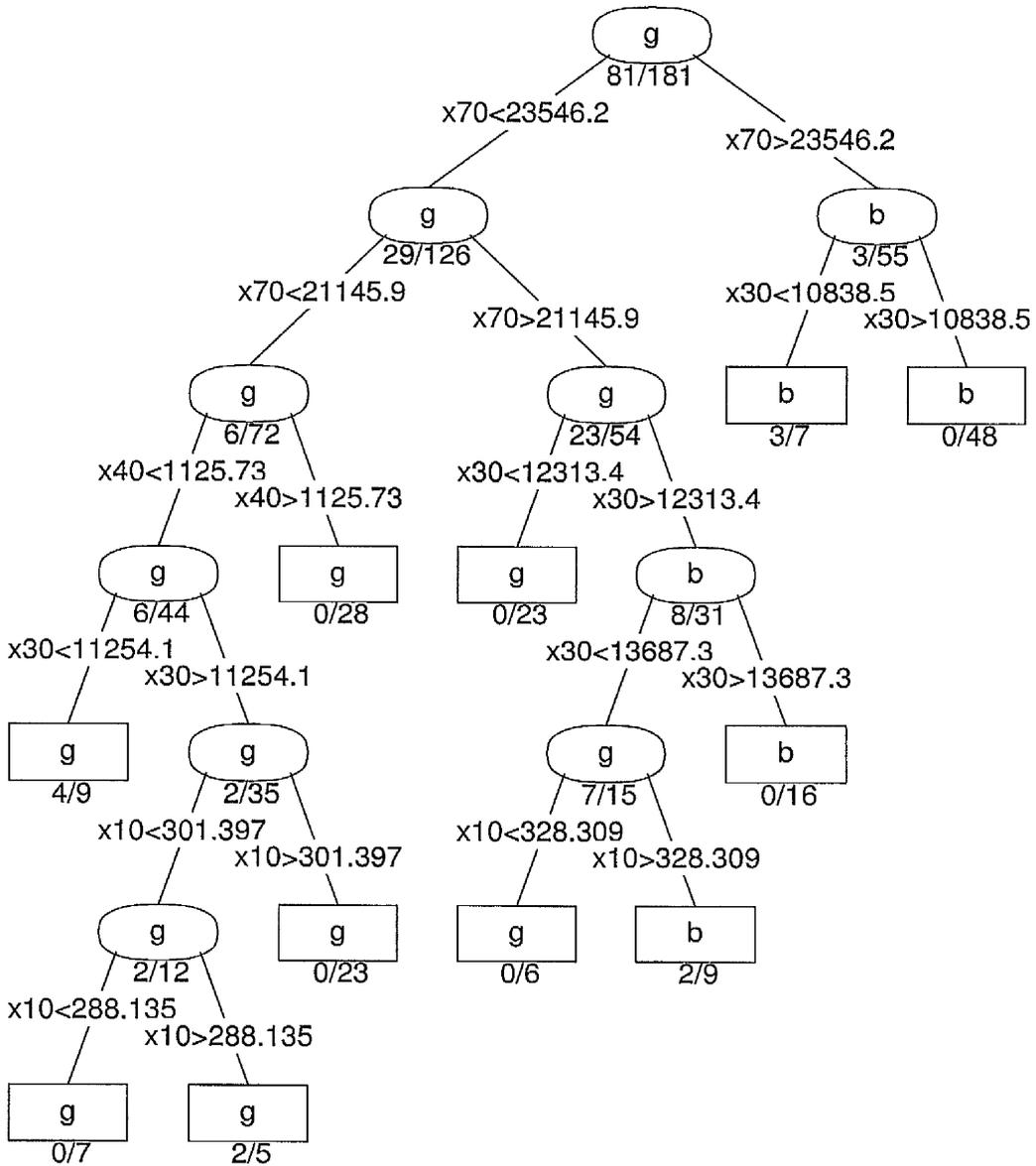
Compartment Output	Local Sensitivity	Yr 1900	Yr 2000	Yr 2100
Circulating Carbon-(NH) (y1)	$\partial y_1 / \partial \omega_1^0$	7	7	7
	$\partial y_1 / \partial \omega_2^0$	6	6	6
	$\partial y_1 / \partial \omega_3^0$	2	2	2
	$\partial y_1 / \partial \omega_4^0$	3	3	3
	$\partial y_5 / \partial \omega_1^0$	8	8	8
	$\partial y_1 / \partial \omega_6^0$	4.5	4	4
	$\partial y_1 / \partial \omega_7^0$	1	1	1
	$\partial y_1 / \partial \omega_8^0$	4.5	5	5
Surface Ocean-(NH) (y2)	$\partial y_2 / \partial \omega_1^0$	7	7	7
	$\partial y_2 / \partial \omega_2^0$	5	5	5
	$\partial y_2 / \partial \omega_3^0$	2	2	2
	$\partial y_2 / \partial \omega_4^0$	3	3	3
	$\partial y_2 / \partial \omega_5^0$	8	8	8
	$\partial y_2 / \partial \omega_6^0$	4	4	4
	$\partial y_2 / \partial \omega_7^0$	1	1	1
	$\partial y_2 / \partial \omega_8^0$	6	6	6
Deep Ocean-(NH) (y3)	$\partial y_3 / \partial \omega_1^0$	6	6	6
	$\partial y_3 / \partial \omega_2^0$	5	5	5
	$\partial y_3 / \partial \omega_3^0$	1	2	2
	$\partial y_3 / \partial \omega_4^0$	4	3	3
	$\partial y_3 / \partial \omega_5^0$	7	7	7
	$\partial y_3 / \partial \omega_6^0$	3	4	4
	$\partial y_3 / \partial \omega_7^0$	2	1	1
	$\partial y_3 / \partial \omega_8^0$	8	8	8
Humus-(NH) (y4)	$\partial y_4 / \partial \omega_1^0$	5	5	7
	$\partial y_4 / \partial \omega_2^0$	6	6	5
	$\partial y_4 / \partial \omega_3^0$	3	3	2
	$\partial y_4 / \partial \omega_4^0$	1	2	3
	$\partial y_4 / \partial \omega_5^0$	7	8	8
	$\partial y_4 / \partial \omega_6^0$	4	4	4
	$\partial y_4 / \partial \omega_7^0$	2	1	1
	$\partial y_4 / \partial \omega_8^0$	8	7	6
Circulating Carbon-(SH) (y5)	$\partial y_5 / \partial \omega_1^0$	7	7	7
	$\partial y_5 / \partial \omega_2^0$	6	6	6
	$\partial y_5 / \partial \omega_3^0$	2	2	2
	$\partial y_5 / \partial \omega_4^0$	3	3	3
	$\partial y_5 / \partial \omega_5^0$	8	8	8
	$\partial y_5 / \partial \omega_6^0$	5	4	4
	$\partial y_5 / \partial \omega_7^0$	1	1	1
	$\partial y_5 / \partial \omega_8^0$	4	5	5
Surface Ocean-(SH) (y6)	$\partial y_6 / \partial \omega_1^0$	7	7	7
	$\partial y_6 / \partial \omega_2^0$	6	5	5
	$\partial y_6 / \partial \omega_3^0$	2	2	2
	$\partial y_6 / \partial \omega_4^0$	3	3	3
	$\partial y_6 / \partial \omega_5^0$	8	8	8
	$\partial y_6 / \partial \omega_6^0$	4	4	4
	$\partial y_6 / \partial \omega_7^0$	1	1	1
	$\partial y_6 / \partial \omega_8^0$	5	6	6
Deep Ocean-(SH) (y7)	$\partial y_7 / \partial \omega_1^0$	6	6	6
	$\partial y_7 / \partial \omega_2^0$	5	5	5
	$\partial y_7 / \partial \omega_3^0$	2	2	2
	$\partial y_7 / \partial \omega_4^0$	4	4	3
	$\partial y_7 / \partial \omega_5^0$	7	7	7
	$\partial y_7 / \partial \omega_6^0$	3	3	4
	$\partial y_7 / \partial \omega_7^0$	1	1	1
	$\partial y_7 / \partial \omega_8^0$	8	8	8
Humus-(SH) (y8)	$\partial y_8 / \partial \omega_1^0$	6	7	7
	$\partial y_8 / \partial \omega_2^0$	8	6	6
	$\partial y_8 / \partial \omega_3^0$	3	3	2
	$\partial y_8 / \partial \omega_4^0$	4	4	4
	$\partial y_8 / \partial \omega_5^0$	7	8	8
	$\partial y_8 / \partial \omega_6^0$	5	5	5
	$\partial y_8 / \partial \omega_7^0$	2	1	1
	$\partial y_8 / \partial \omega_8^0$	1	2	3

### B.2.2 Transfer Coefficients



### B.3 Global SA

#### B.3.1 Initial Conditions



**Figure B.21.** Classification tree showing the performance of the windowing data. The class of the predicted response variable (g-for 'good', b-for 'bad') is centered in the node. The number underneath each terminal node is the misclassification error rate.

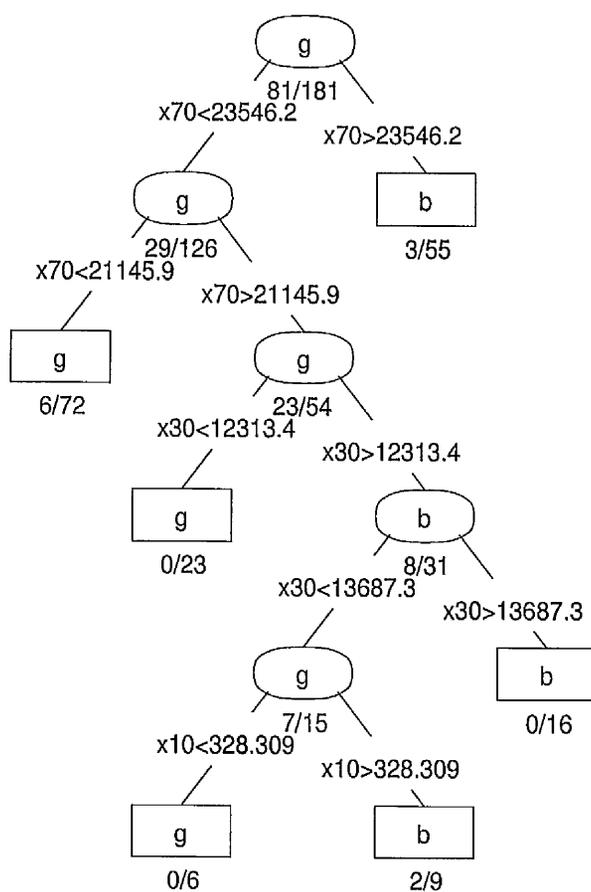
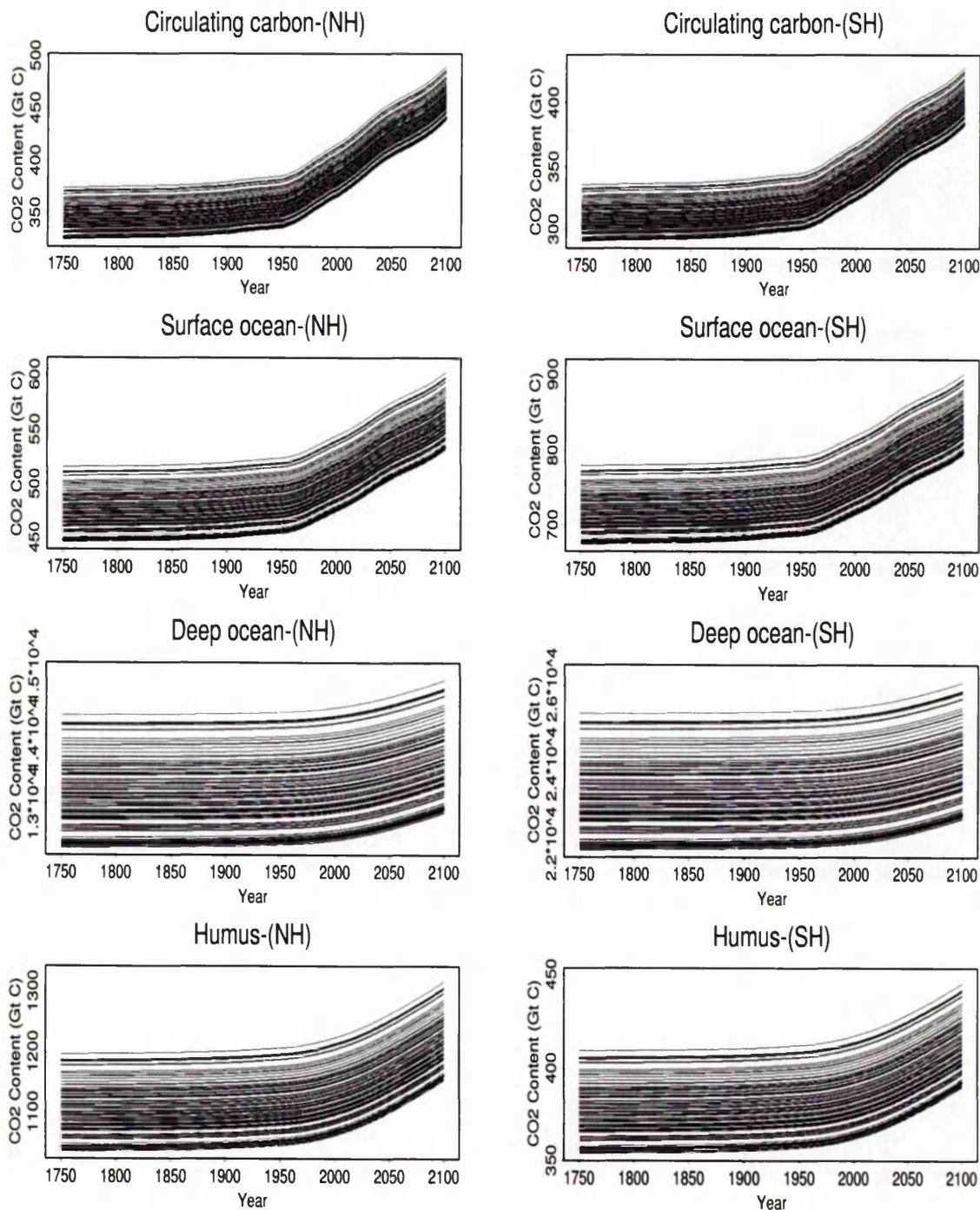
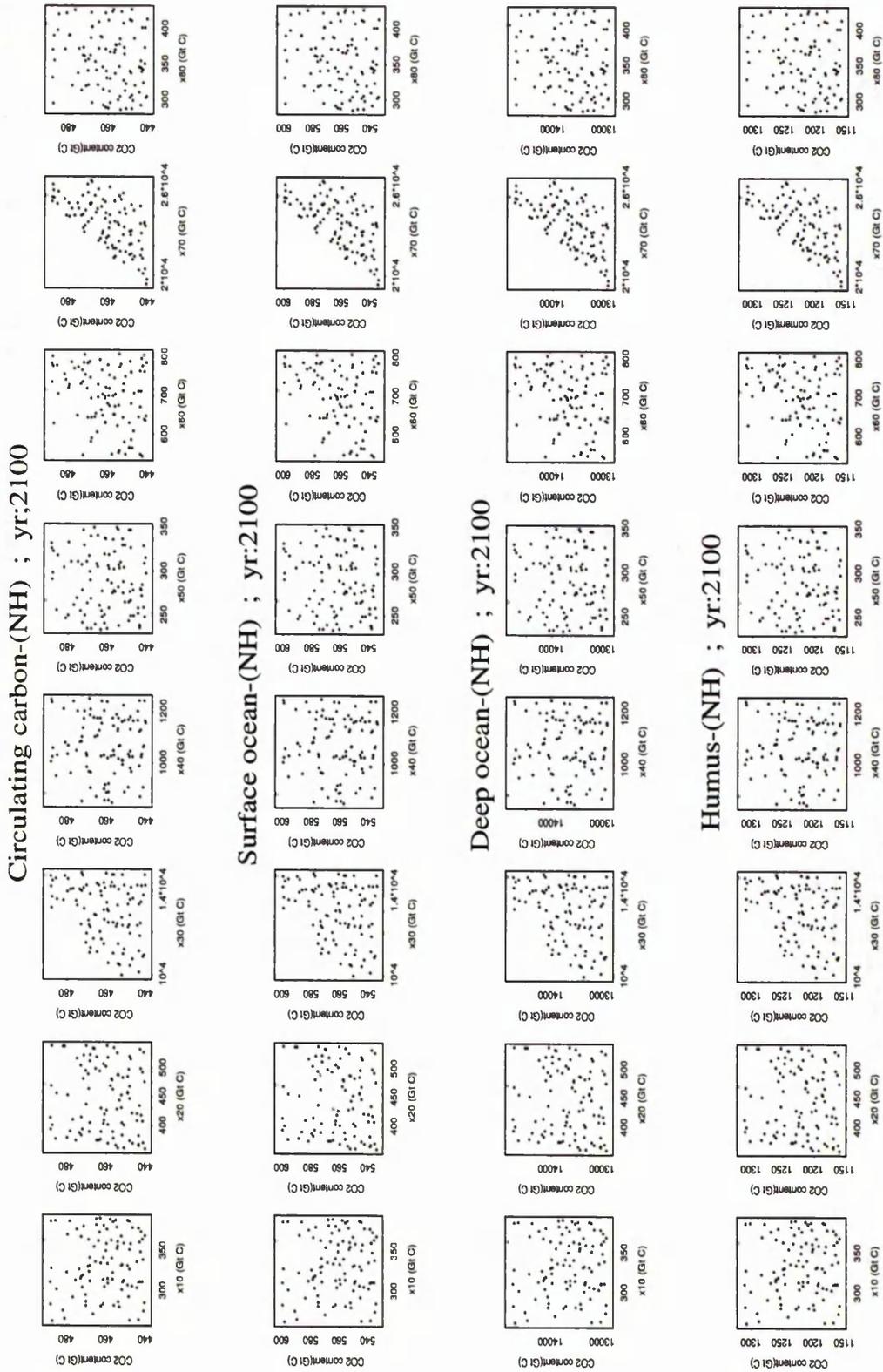


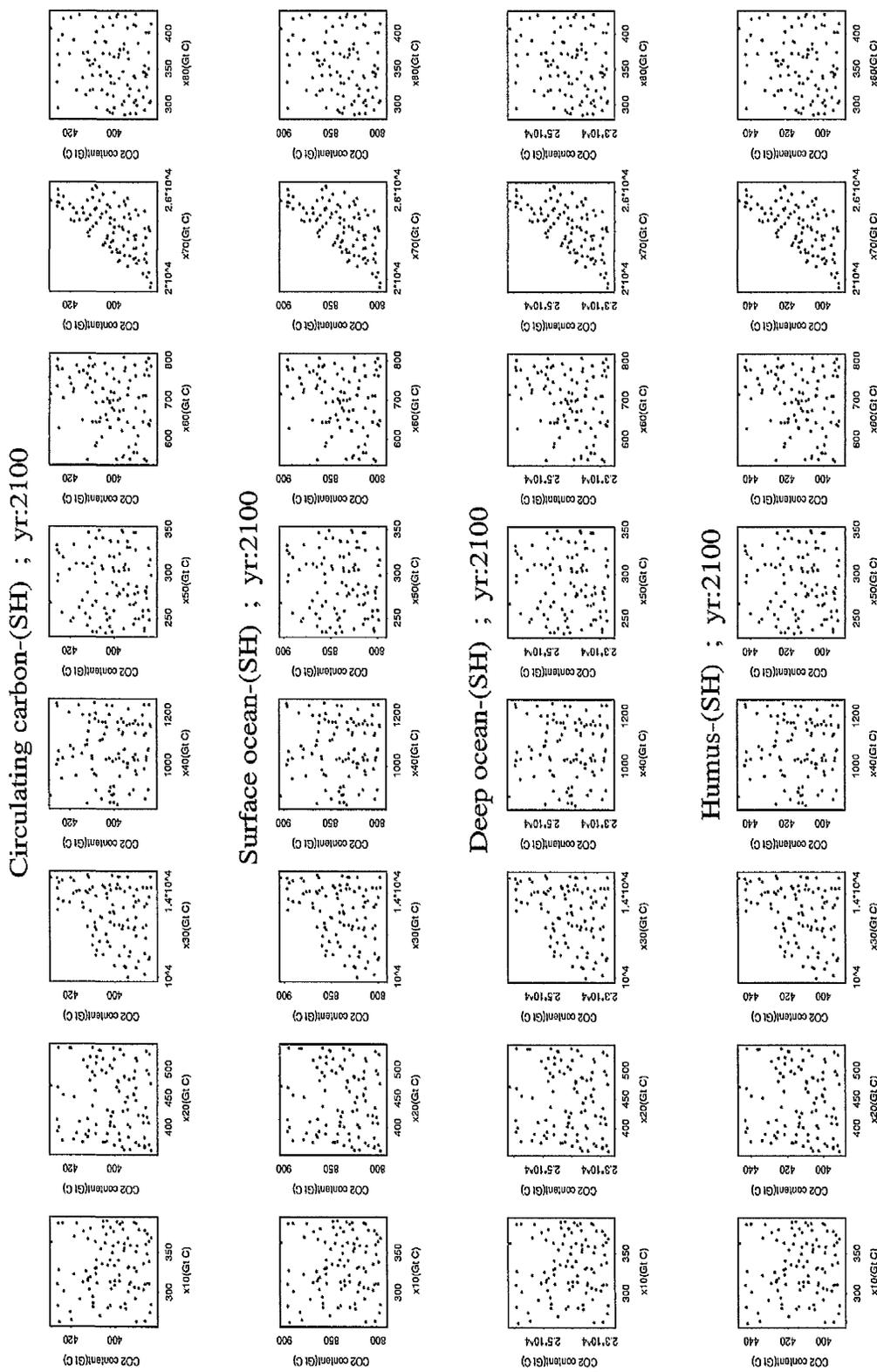
Figure B.22. A pruned version of the classification tree given in Figure B.21.



**Figure B.23.** Time dependent behaviour of all output variables predicted by Model II following windowing analysis as a result of varying all input factors ( $x_i^0, i = 1, \dots, 8$ ) simultaneously. Emission scenario IS92a is considered in the model calculations.



**Figure B.24.** Scatterplots of predicted CO<sub>2</sub> content of Circulating carbon-(NH), Surface ocean-(NH), Deep ocean-(NH) and Humus-(NH) compartments in 2100 versus each compartment's initial condition. N=100 model runs and IS92a emission scenario are considered.



**Table B.9.** Rankings of absolute Pearson correlation coefficients (CC) for the outputs of Model II. The outputs from years 1900, 2000 and 2100 based on N=5,000 model runs are considered. The eight initial conditions are ranked in order of importance, and the rankings associated with the initial conditions for which CC with each output variable have *p*-values less than 0.1 are highlighted.

Compart. Output	Input Factor	Yr 1900	Yr 2000	Yr 2100
Circulating Carbon-(NH)	$x_1^0$	6	6	6
	$x_2^0$	3	3	3
	$x_3^0$	2	2	2
	$x_4^0$	4	4	4
	$x_5^0$	7	7	7
	$x_6^0$	5	5	5
	$x_7^0$	1	1	1
	$x_8^0$	8	8	8
Surface ocean-(NH)	$x_1^0$	6	6	6
	$x_2^0$	3	3	3
	$x_3^0$	2	2	2
	$x_4^0$	4	4	4
	$x_5^0$	7	7	7
	$x_6^0$	5	5	5
	$x_7^0$	1	1	1
	$x_8^0$	8	8	8
Deep ocean-(NH)	$x_1^0$	6	6	6
	$x_2^0$	3	3	3
	$x_3^0$	2	2	2
	$x_4^0$	4	4	4
	$x_5^0$	7	7	7
	$x_6^0$	5	5	5
	$x_7^0$	1	1	1
	$x_8^0$	8	8	8
Humus-(NH)	$x_1^0$	6	6	6
	$x_2^0$	3	3	3
	$x_3^0$	2	2	2
	$x_4^0$	4	4	4
	$x_5^0$	7	7	7
	$x_6^0$	5	5	5
	$x_7^0$	1	1	1
	$x_8^0$	8	8	8

Compart. Output	Input Factor	Yr 1900	Yr 2000	Yr 2100
Circulating Carbon-(SH)	$x_1^0$	6	6	6
	$x_2^0$	3	3	3
	$x_3^0$	2	2	2
	$x_4^0$	4	4	4
	$x_5^0$	7	7	7
	$x_6^0$	5	5	5
	$x_7^0$	1	1	1
	$x_8^0$	8	8	8
Surface ocean-(SH)	$x_1^0$	6	6	6
	$x_2^0$	3	3	3
	$x_3^0$	2	2	2
	$x_4^0$	4	4	4
	$x_5^0$	7	7	7
	$x_6^0$	5	5	5
	$x_7^0$	1	1	1
	$x_8^0$	8	8	8
Deep ocean-(SH)	$x_1^0$	6	6	6
	$x_2^0$	3	3	3
	$x_3^0$	2	2	2
	$x_4^0$	4	4	4
	$x_5^0$	7	7	7
	$x_6^0$	5	5	5
	$x_7^0$	1	1	1
	$x_8^0$	8	8	8
Humus-(SH)	$x_1^0$	6	6	6
	$x_2^0$	3	3	3
	$x_3^0$	2	2	2
	$x_4^0$	4	4	4
	$x_5^0$	7	7	7
	$x_6^0$	5	5	5
	$x_7^0$	1	1	1
	$x_8^0$	8	8	8

The most important input factor for all compartments in all three years is  $x_7^0$  with a CC value of about 0.636 and the second most important factor is  $x_3^0$  with a CC value of about 0.357. The absolute CC values associated with the other input factors and the output variables vary between 0.007 and 0.044.

**Table B.10.** Rankings of absolute Standardized Regression Coefficients (SRC) for the outputs of Model II. The outputs from years 1900, 2000 and 2100 based on N=5,000 model runs are considered. The initial conditions are ranked in order of importance, and the most important factors with SRC values greater than 0.1 are highlighted.

Compart. Output	Input Factor	SRC Ranks		
		Yr 1900	Yr 2000	Yr 2100
Circulating Carbon-(NH)	$x_1^0$	7	7	7
	$x_2^0$	5	5	5
	$x_3^0$	2	2	2
	$x_4^0$	3	3	3
	$x_5^0$	8	8	8
	$x_6^0$	4	4	4
	$x_7^0$	1	1	1
	$x_8^0$	6	6	6
Surface ocean-(NH)	$x_1^0$	7	7	7
	$x_2^0$	5	5	5
	$x_3^0$	2	2	2
	$x_4^0$	3	3	3
	$x_5^0$	8	8	8
	$x_6^0$	4	4	4
	$x_7^0$	1	1	1
	$x_8^0$	6	6	6
Deep ocean-(NH)	$x_1^0$	7	7	7
	$x_2^0$	5	5	5
	$x_3^0$	2	2	2
	$x_4^0$	3	3	3
	$x_5^0$	8	8	8
	$x_6^0$	4	4	4
	$x_7^0$	1	1	1
	$x_8^0$	6	6	6
Humus-(NH)	$x_1^0$	7	7	7
	$x_2^0$	5	5	5
	$x_3^0$	2	2	2
	$x_4^0$	3	3	3
	$x_5^0$	8	8	8
	$x_6^0$	4	4	4
	$x_7^0$	1	1	1
	$x_8^0$	6	6	6

Compart. Output	Input Factor	SRC Ranks		
		Yr 1900	Yr 2000	Yr 2100
Circulating Carbon-(SH)	$x_1^0$	7	7	7
	$x_2^0$	5	5	5
	$x_3^0$	2	2	2
	$x_4^0$	3	3	3
	$x_5^0$	8	8	8
	$x_6^0$	4	4	4
	$x_7^0$	1	1	1
	$x_8^0$	6	6	6
Surface ocean-(SH)	$x_1^0$	7	7	7
	$x_2^0$	5	5	5
	$x_3^0$	2	2	2
	$x_4^0$	3	3	3
	$x_5^0$	8	8	8
	$x_6^0$	4	4	4
	$x_7^0$	1	1	1
	$x_8^0$	6	6	6
Deep ocean-(SH)	$x_1^0$	7	7	7
	$x_2^0$	5	5	5
	$x_3^0$	2	2	2
	$x_4^0$	3	3	3
	$x_5^0$	8	8	8
	$x_6^0$	4	4	4
	$x_7^0$	1	1	1
	$x_8^0$	6	6	6
Humus-(SH)	$x_1^0$	7	7	7
	$x_2^0$	5	5	5
	$x_3^0$	2	2	2
	$x_4^0$	3	3	3
	$x_5^0$	8	8	8
	$x_6^0$	4	4	4
	$x_7^0$	1	1	1
	$x_8^0$	6	6	6

The most important input factor for all compartments in all three years is  $x_7^0$  with an SRC of about 0.98 and the second most important factor is  $x_3^0$  with an SRC of about 0.88. The absolute SRC values associated with the other input factors and the output variables are quite small, between 0.021 and 0.078.

The absolute PCCs produce identical rankings of input factor importance, and as with the SRCs, except for  $x_7^0$  and  $x_3^0$  the other initial conditions under consideration have very small PCCs, below 0.025.

**Table B.11.** Stepwise regression analyses for output variable  $y_1(t)$  (i.e. Circulating carbon-(NH) compartment) of Model II in 1900, 2000 and 2100. Calculations are based on  $N=5,000$  model runs and IS92a emission scenario.

Circulating carbon-(NH)									
Step	$y_1(t = 1900)$			$y_1(t = 2000)$			$y_1(t = 2100)$		
	Variable	$R^2$	PRESS	Variable	$R^2$	PRESS	Variable	$R^2$	PRESS
1	$x_7^0$	0.4038	533678.00	$x_7^0$	0.4037	533679.00	$x_7^0$	0.4038	533680.00
2	$x_3^0$	0.9882	10571.60	$x_3^0$	0.9884	10571.00	$x_3^0$	0.9883	10570.60
3	$x_4^0$	0.9944	5004.83	$x_4^0$	0.9944	5004.76	$x_4^0$	0.9944	5004.71
4	$x_6^0$	0.9972	2603.49	$x_6^0$	0.9971	2603.43	$x_6^0$	0.9971	2603.39
5	$x_2^0$	0.9983	1555.74	$x_2^0$	0.9982	1555.69	$x_2^0$	0.9982	1555.65
6	$x_8^0$	0.9990	927.21	$x_8^0$	0.9990	927.20	$x_8^0$	0.9990	927.19
7	$x_1^0$	0.9995	417.02	$x_1^0$	0.9995	417.02	$x_1^0$	0.9995	417.02

The analysis results for the other seven compartments are very similar to the results presented in the above table for the circulating carbon-(NH) compartment, in the sense that the same variables were selected with  $R^2$ -values that are quite similar and the order of variable selection does not change.

B.3.2 Transfer Coefficients

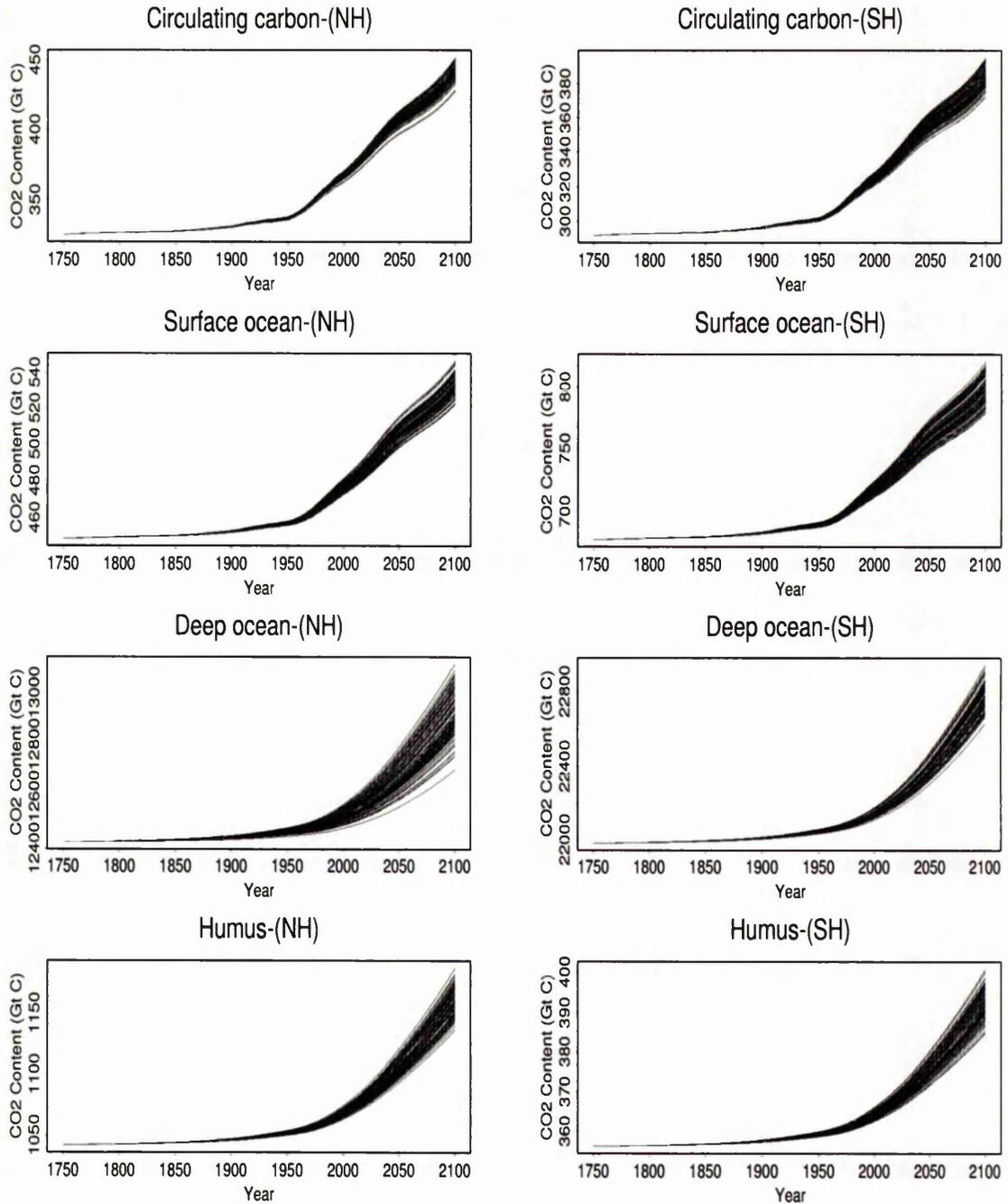
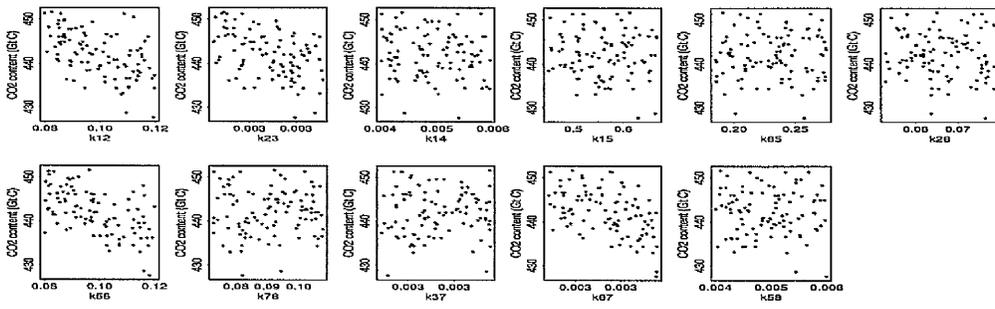
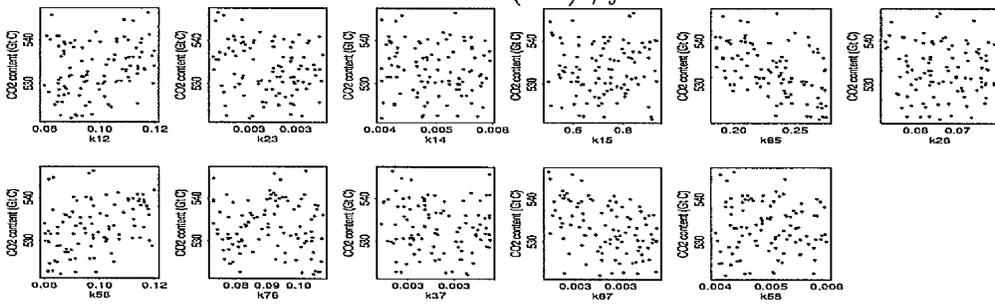


Figure B.26. Time dependent behaviour of all output variables predicted by Model II as a result of varying selected 11 free transfer coefficients simultaneously. Emission scenario IS92a is considered in the model calculations.

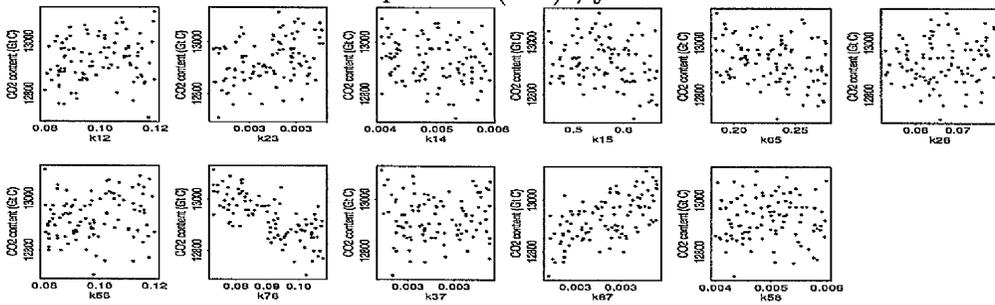
Circulating carbon-(NH) ; yr:2100



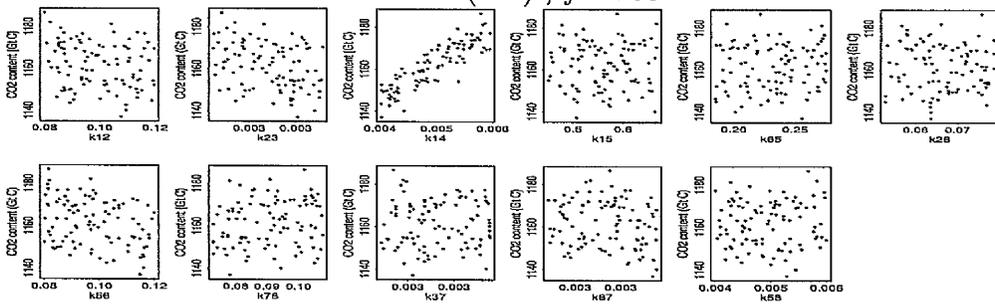
Surface ocean-(NH) ; yr:2100



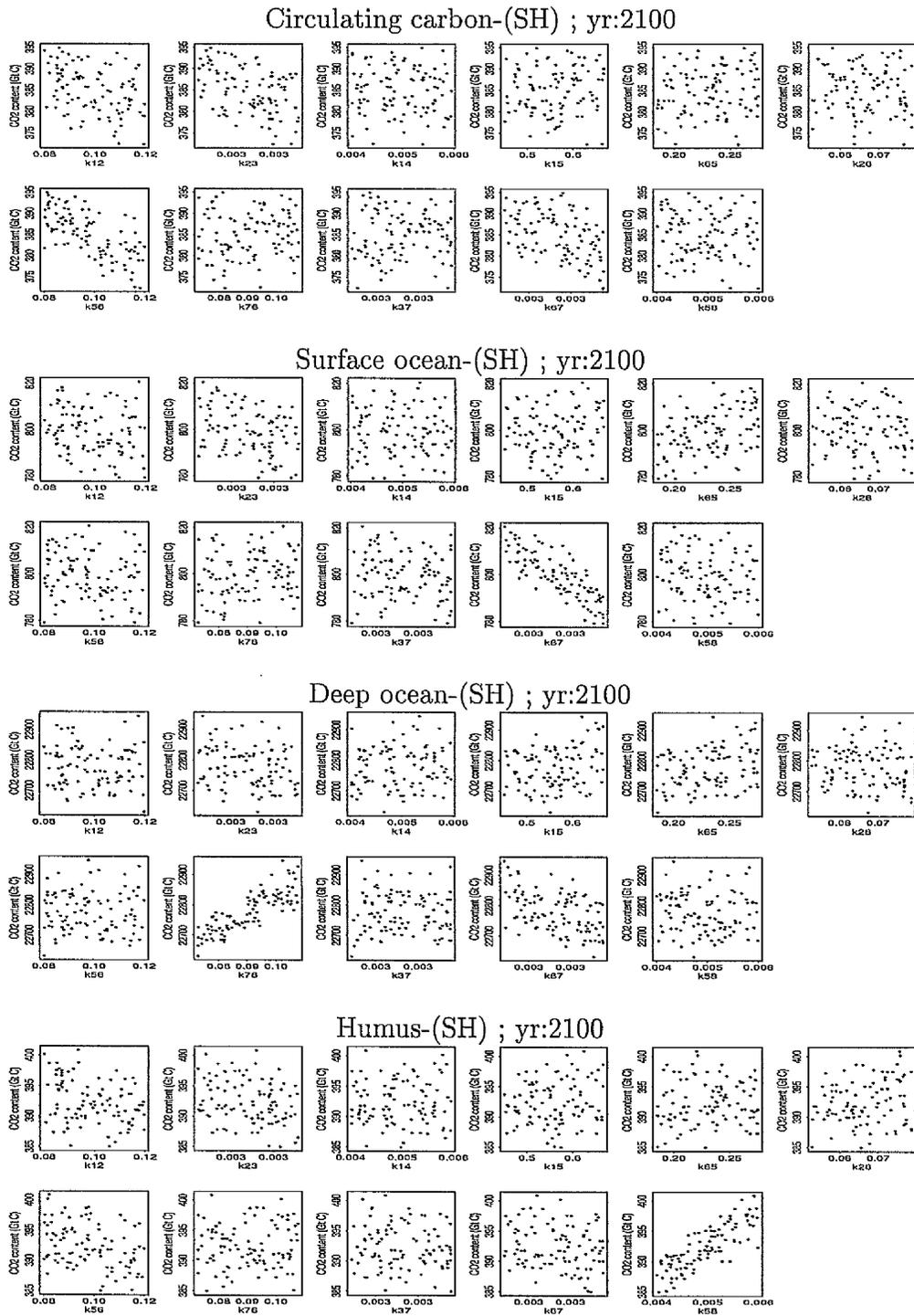
Deep ocean-(NH) ; yr:2100



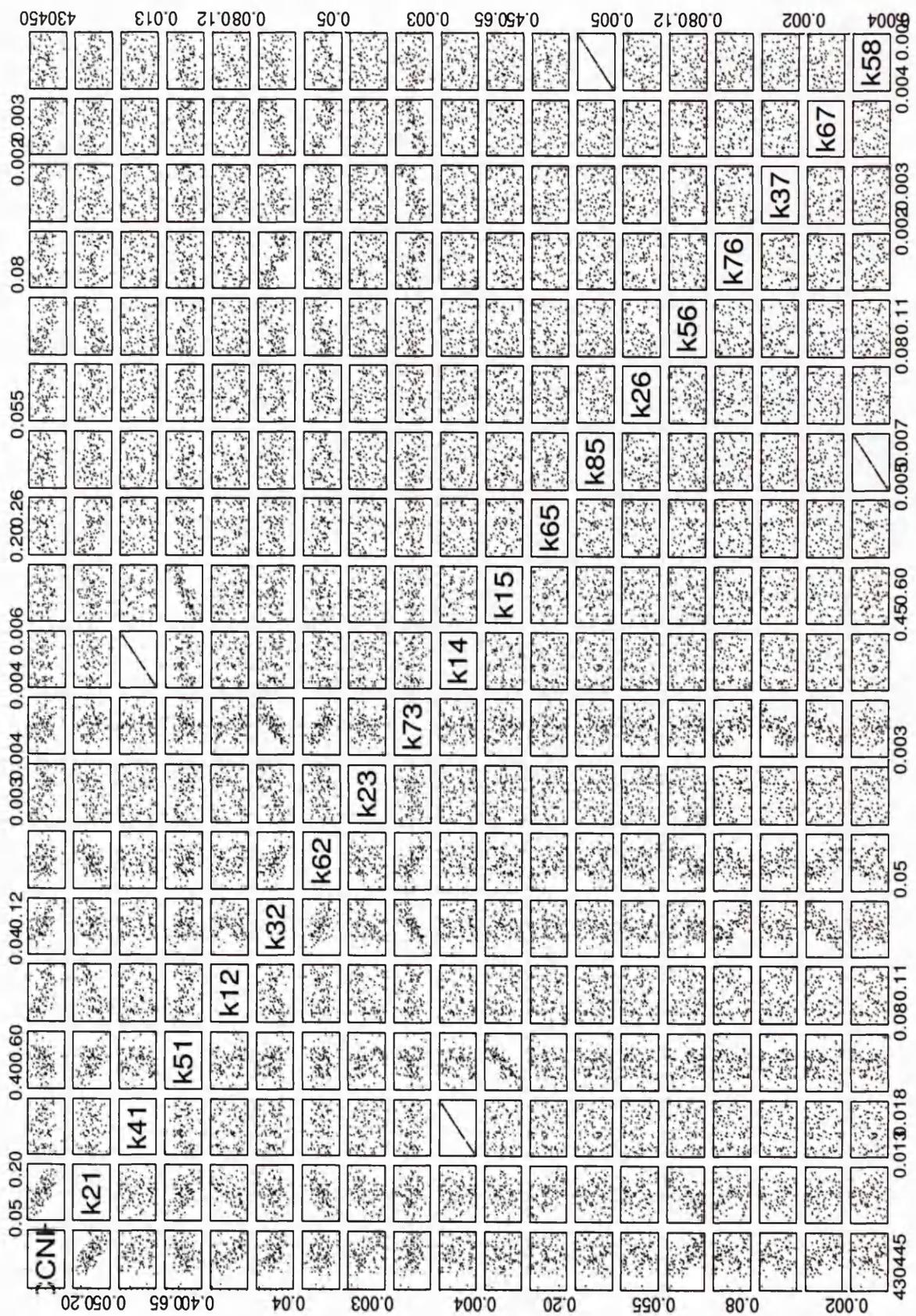
Humus-(NH) ; yr:2100



**Figure B.27.** Scatterplots of predicted CO<sub>2</sub> content of Circulating carbon-(NH), Surface ocean-(NH), Deep ocean-(NH) and Humus-(NH) compartments in 2100 versus each free transfer coefficient. N=100 model runs, and IS92a emission scenario is considered.



**Figure B.28.** Scatterplots of predicted CO<sub>2</sub> content of Circulating carbon-(SH), Surface ocean-(SH), Deep ocean-(SH) and Humus-(SH) compartments in 2100 versus each free transfer coefficient. N=100 model runs, and IS92a emission scenario is considered.



**Figure B.29.** Scatterplot matrix of predicted CO<sub>2</sub> content of Circulating carbon-(NH) in 2100 and each transfer coefficient (see Table 3.5 in Chapter 3). N=100 model runs, and IS92a emission scenario is used.

**Table B.12.** Rankings of absolute Pearson correlation coefficients (CC) for the outputs of Model II. The outputs from years 1900, 2000 and 2100 based on N=5,000 model runs are considered. The eleven free input factors are ranked in order of importance, and the rankings associated with the input factors for which CC with each output variable have *p*-values less than 0.1 are highlighted.

Compart. Output	Input Factor	CC Ranks		
		Yr 1900	Yr 2000	Yr 2100
Circulating carbon-(NH)	k <sub>12</sub>	2	2	3
	k <sub>23</sub>	4	4	4
	k <sub>14</sub>	6	6	6
	k <sub>15</sub>	5	5	5
	k <sub>65</sub>	7	7	7
	k <sub>26</sub>	10	10	10
	k <sub>56</sub>	1	1	2
	k <sub>76</sub>	8	8	9
	k <sub>37</sub>	11	11	11
	k <sub>67</sub>	3	3	1
k <sub>58</sub>	9	9	8	
Surface ocean-(NH)	k <sub>12</sub>	5	5	5
	k <sub>23</sub>	3	3	3
	k <sub>14</sub>	6	6	6
	k <sub>15</sub>	7	7	8
	k <sub>65</sub>	1	1	1
	k <sub>26</sub>	8	8	9
	k <sub>56</sub>	2	2	2
	k <sub>76</sub>	9	10	7
	k <sub>37</sub>	11	11	11
	k <sub>67</sub>	4	4	4
k <sub>58</sub>	10	9	10	
Deep ocean-(NH)	k <sub>12</sub>	6	6	6
	k <sub>23</sub>	3	3	3
	k <sub>14</sub>	7	7	7
	k <sub>15</sub>	11	11	11
	k <sub>65</sub>	5	5	5
	k <sub>26</sub>	8	8	8
	k <sub>56</sub>	4	4	4
	k <sub>76</sub>	1	1	1
	k <sub>37</sub>	9	9	10
	k <sub>67</sub>	2	2	2
k <sub>58</sub>	10	10	9	
Humus-(NH)	k <sub>12</sub>	3	3	4
	k <sub>23</sub>	5	5	5
	k <sub>14</sub>	1	1	1
	k <sub>15</sub>	6	6	6
	k <sub>65</sub>	7	7	7
	k <sub>26</sub>	9	9	9
	k <sub>56</sub>	2	2	2
	k <sub>76</sub>	11	11	11
	k <sub>37</sub>	8	8	8
	k <sub>67</sub>	4	4	3
k <sub>58</sub>	10	10	10	

Compart. Output	Input Factor	CC Ranks		
		Yr 1900	Yr 2000	Yr 2100
Circulating carbon-(SH)	k <sub>12</sub>	3	3	3
	k <sub>23</sub>	4	4	4
	k <sub>14</sub>	7	7	7
	k <sub>15</sub>	6	6	6
	k <sub>65</sub>	5	5	5
	k <sub>26</sub>	11	11	11
	k <sub>56</sub>	1	1	1
	k <sub>76</sub>	8	8	8
	k <sub>37</sub>	10	10	10
	k <sub>67</sub>	2	2	2
k <sub>58</sub>	9	9	9	
Surface ocean-(SH)	k <sub>12</sub>	5	5	5
	k <sub>23</sub>	3	3	3
	k <sub>14</sub>	7	7	7
	k <sub>15</sub>	8	8	9
	k <sub>65</sub>	2	2	2
	k <sub>26</sub>	11	11	11
	k <sub>56</sub>	4	4	4
	k <sub>76</sub>	6	6	6
	k <sub>37</sub>	10	10	10
	k <sub>67</sub>	1	1	1
k <sub>58</sub>	9	9	8	
Deep ocean-(SH)	k <sub>12</sub>	8	8	8
	k <sub>23</sub>	4	4	4
	k <sub>14</sub>	6	6	6
	k <sub>15</sub>	9	9	9
	k <sub>65</sub>	3	3	3
	k <sub>26</sub>	10	10	10
	k <sub>56</sub>	5	5	5
	k <sub>76</sub>	1	1	1
	k <sub>37</sub>	11	11	11
	k <sub>67</sub>	2	2	2
k <sub>58</sub>	7	7	7	
Humus-(SH)	k <sub>12</sub>	4	4	4
	k <sub>23</sub>	5	5	5
	k <sub>14</sub>	8	8	8
	k <sub>15</sub>	7	7	7
	k <sub>65</sub>	6	6	6
	k <sub>26</sub>	10	11	10
	k <sub>56</sub>	2	2	2
	k <sub>76</sub>	11	9	11
	k <sub>37</sub>	9	10	9
	k <sub>67</sub>	3	3	3
k <sub>58</sub>	1	1	1	

**Table B.13.** Rankings of absolute Standardized Regression Coefficients (SRC) for the outputs of Model II. The outputs from years 1900, 2000 and 2100 based on N=5,000 model runs are considered. The input factors (Transfer Coefficients) are ranked in order of importance, and the most important factors with SRC values greater than 0.1 are highlighted.

Compart. Output	Input Factor	SRC Ranks		
		Yr 1900	Yr 2000	Yr 2100
Circulating carbon-(NH)	<i>k</i> <sub>12</sub>	2	2	3
	<i>k</i> <sub>23</sub>	4	4	4
	<i>k</i> <sub>14</sub>	6	6	6
	<i>k</i> <sub>15</sub>	5	5	5
	<i>k</i> <sub>65</sub>	7	7	7
	<i>k</i> <sub>26</sub>	10	10	10
	<i>k</i> <sub>56</sub>	1	1	2
	<i>k</i> <sub>76</sub>	9	9	9
	<i>k</i> <sub>37</sub>	11	11	11
	<i>k</i> <sub>67</sub>	3	3	1
<i>k</i> <sub>58</sub>	8	8	8	
Surface ocean-(NH)	<i>k</i> <sub>12</sub>	5	5	5
	<i>k</i> <sub>23</sub>	3	3	3
	<i>k</i> <sub>14</sub>	6	6	6
	<i>k</i> <sub>15</sub>	7	7	8
	<i>k</i> <sub>65</sub>	1	1	1
	<i>k</i> <sub>26</sub>	9	8	9
	<i>k</i> <sub>56</sub>	2	2	2
	<i>k</i> <sub>76</sub>	8	9	7
	<i>k</i> <sub>37</sub>	11	11	11
	<i>k</i> <sub>67</sub>	4	4	4
<i>k</i> <sub>58</sub>	10	10	10	
Deep ocean-(NH)	<i>k</i> <sub>12</sub>	6	6	6
	<i>k</i> <sub>23</sub>	3	3	3
	<i>k</i> <sub>14</sub>	7	7	7
	<i>k</i> <sub>15</sub>	9	9	9
	<i>k</i> <sub>65</sub>	4	4	4
	<i>k</i> <sub>26</sub>	11	11	11
	<i>k</i> <sub>56</sub>	5	5	5
	<i>k</i> <sub>76</sub>	1	1	1
	<i>k</i> <sub>37</sub>	8	8	8
	<i>k</i> <sub>67</sub>	2	2	2
<i>k</i> <sub>58</sub>	10	10	10	
Humus-(NH)	<i>k</i> <sub>12</sub>	3	3	4
	<i>k</i> <sub>23</sub>	5	5	5
	<i>k</i> <sub>14</sub>	1	1	1
	<i>k</i> <sub>15</sub>	6	6	6
	<i>k</i> <sub>65</sub>	7	7	7
	<i>k</i> <sub>26</sub>	10	10	10
	<i>k</i> <sub>56</sub>	2	2	2
	<i>k</i> <sub>76</sub>	9	9	9
	<i>k</i> <sub>37</sub>	11	11	11
	<i>k</i> <sub>67</sub>	4	4	3
<i>k</i> <sub>58</sub>	8	8	8	

Compart. Output	Input Factor	SRC Ranks		
		Yr 1900	Yr 2000	Yr 2100
Circulating carbon-(SH)	<i>k</i> <sub>12</sub>	3	3	3
	<i>k</i> <sub>23</sub>	4	4	4
	<i>k</i> <sub>14</sub>	7	7	7
	<i>k</i> <sub>15</sub>	6	6	6
	<i>k</i> <sub>65</sub>	5	5	5
	<i>k</i> <sub>26</sub>	11	11	11
	<i>k</i> <sub>56</sub>	1	1	1
	<i>k</i> <sub>76</sub>	9	9	9
	<i>k</i> <sub>37</sub>	10	10	10
	<i>k</i> <sub>67</sub>	2	2	2
<i>k</i> <sub>58</sub>	8	8	8	
Surface ocean-(SH)	<i>k</i> <sub>12</sub>	5	5	5
	<i>k</i> <sub>23</sub>	3	3	3
	<i>k</i> <sub>14</sub>	7	7	7
	<i>k</i> <sub>15</sub>	8	8	8
	<i>k</i> <sub>65</sub>	2	2	2
	<i>k</i> <sub>26</sub>	10	10	10
	<i>k</i> <sub>56</sub>	4	4	4
	<i>k</i> <sub>76</sub>	6	6	6
	<i>k</i> <sub>37</sub>	11	11	11
	<i>k</i> <sub>67</sub>	1	1	1
<i>k</i> <sub>58</sub>	9	9	9	
Deep ocean-(SH)	<i>k</i> <sub>12</sub>	8	8	7
	<i>k</i> <sub>23</sub>	4	4	4
	<i>k</i> <sub>14</sub>	6	6	6
	<i>k</i> <sub>15</sub>	7	7	8
	<i>k</i> <sub>65</sub>	3	3	3
	<i>k</i> <sub>26</sub>	11	11	11
	<i>k</i> <sub>56</sub>	5	5	5
	<i>k</i> <sub>76</sub>	1	1	1
	<i>k</i> <sub>37</sub>	9	9	9
	<i>k</i> <sub>67</sub>	2	2	2
<i>k</i> <sub>58</sub>	10	10	10	
Humus-(SH)	<i>k</i> <sub>12</sub>	4	4	4
	<i>k</i> <sub>23</sub>	5	5	5
	<i>k</i> <sub>14</sub>	8	8	8
	<i>k</i> <sub>15</sub>	7	7	7
	<i>k</i> <sub>65</sub>	6	6	6
	<i>k</i> <sub>26</sub>	11	11	11
	<i>k</i> <sub>56</sub>	2	2	2
	<i>k</i> <sub>76</sub>	9	9	9
	<i>k</i> <sub>37</sub>	10	10	10
	<i>k</i> <sub>67</sub>	3	3	3
<i>k</i> <sub>58</sub>	1	1	1	

The PCCs have also been calculated. Even though the PCCs tend to be larger than the SRCs, the rankings of input factor importance provided by both of these statistics are identical.

Table B.14. Stepwise regression analyses for all compartments of Model II in 1900, 2000 and 2100. Calculations are based on N=5,000 model runs and IS92a emission scenario.

Compart. Output	Yr 1900			Yr 2000		Yr 2100	
	Step	Variable	R <sup>2</sup>	Variable	R <sup>2</sup>	Variable	R <sup>2</sup>
Circulating carbon-(NH)	1	k <sub>56</sub>	0.3157	k <sub>56</sub>	0.3308	k <sub>67</sub>	0.2964
	2	k <sub>12</sub>	0.6103	k <sub>12</sub>	0.6396	k <sub>56</sub>	0.5806
	3	k <sub>67</sub>	0.8391	k <sub>67</sub>	0.8430	k <sub>12</sub>	0.8312
	4	k <sub>23</sub>	0.9328	k <sub>23</sub>	0.9283	k <sub>23</sub>	0.9447
	5	k <sub>15</sub>	0.9597	k <sub>15</sub>	0.9564	k <sub>15</sub>	0.9648
	6	k <sub>14</sub>	0.9792	k <sub>14</sub>	0.9790	k <sub>14</sub>	0.9790
	7	k <sub>65</sub>	0.9827	k <sub>65</sub>	0.9829	k <sub>65</sub>	0.9818
Surface ocean-(NH)	1	k <sub>65</sub>	0.3194	k <sub>65</sub>	0.3269	k <sub>65</sub>	0.2925
	2	k <sub>56</sub>	0.5611	k <sub>56</sub>	0.5801	k <sub>23</sub>	0.5303
	3	k <sub>23</sub>	0.7699	k <sub>23</sub>	0.7793	k <sub>56</sub>	0.7424
	4	k <sub>67</sub>	0.8817	k <sub>67</sub>	0.8783	k <sub>67</sub>	0.8914
	5	k <sub>12</sub>	0.9648	k <sub>12</sub>	0.9656	k <sub>12</sub>	0.9612
	6	k <sub>14</sub>	0.9696	k <sub>14</sub>	0.9705	k <sub>14</sub>	0.9659
	7	k <sub>15</sub>	0.9712	k <sub>15</sub>	0.9722	k <sub>76</sub>	0.9683
Deep ocean-(NH)	1	k <sub>76</sub>	0.4487	k <sub>76</sub>	0.4375	k <sub>76</sub>	0.4607
	2	k <sub>67</sub>	0.8222	k <sub>67</sub>	0.8141	k <sub>67</sub>	0.8300
	3	k <sub>23</sub>	0.8901	k <sub>23</sub>	0.8842	k <sub>23</sub>	0.8958
	4	k <sub>65</sub>	0.9330	k <sub>65</sub>	0.9296	k <sub>65</sub>	0.9357
	5	k <sub>56</sub>	0.9710	k <sub>56</sub>	0.9709	k <sub>56</sub>	0.9702
	6	k <sub>12</sub>	0.9834	k <sub>12</sub>	0.9844	k <sub>12</sub>	0.9813
	7	k <sub>14</sub>	0.9846	k <sub>14</sub>	0.9854	k <sub>14</sub>	0.9826
Humus-(NH)	1	k <sub>14</sub>	0.8203	k <sub>14</sub>	0.8344	k <sub>14</sub>	0.7883
	2	k <sub>56</sub>	0.8776	k <sub>56</sub>	0.8893	k <sub>56</sub>	0.8522
	3	k <sub>12</sub>	0.9310	k <sub>12</sub>	0.9415	k <sub>12</sub>	0.9106
	4	k <sub>67</sub>	0.9722	k <sub>67</sub>	0.9752	k <sub>67</sub>	0.9660
	5	k <sub>23</sub>	0.9892	k <sub>23</sub>	0.9893	k <sub>23</sub>	0.9884
	6	k <sub>15</sub>	0.9945	k <sub>15</sub>	0.9949	k <sub>15</sub>	0.9935
	7	k <sub>65</sub>	0.9951	k <sub>65</sub>	0.9956	k <sub>65</sub>	0.9942
Circulating carbon-(SH)	1	k <sub>56</sub>	0.5394	k <sub>56</sub>	0.5612	k <sub>56</sub>	0.4837
	2	k <sub>67</sub>	0.7536	k <sub>67</sub>	0.7527	k <sub>67</sub>	0.7555
	3	k <sub>12</sub>	0.8672	k <sub>12</sub>	0.8699	k <sub>12</sub>	0.8627
	4	k <sub>23</sub>	0.9201	k <sub>23</sub>	0.9165	k <sub>23</sub>	0.9315
	5	k <sub>65</sub>	0.9611	k <sub>65</sub>	0.9597	k <sub>65</sub>	0.9655
	6	k <sub>15</sub>	0.9762	k <sub>15</sub>	0.9753	k <sub>15</sub>	0.9769
	7	k <sub>14</sub>	0.9852	k <sub>14</sub>	0.9753	k <sub>14</sub>	0.9769
Surface ocean-(SH)	1	k <sub>67</sub>	0.6931	k <sub>67</sub>	0.6811	k <sub>67</sub>	0.7192
	2	k <sub>65</sub>	0.8493	k <sub>65</sub>	0.8525	k <sub>65</sub>	0.8424
	3	k <sub>23</sub>	0.9140	k <sub>23</sub>	0.9120	k <sub>23</sub>	0.9174
	4	k <sub>56</sub>	0.9507	k <sub>56</sub>	0.9495	k <sub>56</sub>	0.9529
	5	k <sub>12</sub>	0.9588	k <sub>12</sub>	0.9575	k <sub>12</sub>	0.9615
	6	k <sub>76</sub>	0.9659	k <sub>76</sub>	0.9642	k <sub>76</sub>	0.9691
	7	k <sub>14</sub>	0.9703	k <sub>14</sub>	0.9689	k <sub>14</sub>	0.9729
Deep ocean-(SH)	1	k <sub>76</sub>	0.7069	k <sub>76</sub>	0.7204	k <sub>76</sub>	0.6924
	2	k <sub>67</sub>	0.8947	k <sub>67</sub>	0.8943	k <sub>67</sub>	0.8937
	3	k <sub>65</sub>	0.9507	k <sub>65</sub>	0.9553	k <sub>65</sub>	0.9444
	4	k <sub>23</sub>	0.9668	k <sub>23</sub>	0.9690	k <sub>23</sub>	0.9631
	5	k <sub>56</sub>	0.9744	k <sub>56</sub>	0.9763	k <sub>56</sub>	0.9708
	6	k <sub>14</sub>	0.9786	k <sub>14</sub>	0.9800	k <sub>14</sub>	0.9756
	7	k <sub>15</sub>	0.9797	k <sub>15</sub>	0.9812	k <sub>12</sub>	0.9766
Humus-(SH)	1	k <sub>58</sub>	0.7290	k <sub>58</sub>	0.7408	k <sub>58</sub>	0.6977
	2	k <sub>56</sub>	0.8735	k <sub>56</sub>	0.8855	k <sub>56</sub>	0.8493
	3	k <sub>67</sub>	0.9308	k <sub>67</sub>	0.9339	k <sub>67</sub>	0.9233
	4	k <sub>12</sub>	0.9618	k <sub>12</sub>	0.9641	k <sub>12</sub>	0.9570
	5	k <sub>23</sub>	0.9759	k <sub>23</sub>	0.9758	k <sub>23</sub>	0.9754
	6	k <sub>65</sub>	0.9869	k <sub>65</sub>	0.9872	k <sub>65</sub>	0.9864
	7	k <sub>15</sub>	0.9910	k <sub>15</sub>	0.9917	k <sub>15</sub>	0.9902