University of Glasgow

Ferguson, Ryan Funai (2026) *Interpolation of sparse data in hadron physics with Gaussian processes.* PhD thesis.

https://theses.gla.ac.uk/86320/

Copyright and moral rights for this work are retained by the author

A copy can be downloaded for personal non-commercial research or study, without prior permission or charge

This work cannot be reproduced or quoted extensively from without first obtaining permission from the author

The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given

Enlighten: Theses
https://theses.gla.ac.uk/
research-enlighten@glasgow.ac.uk

# Interpolation of Sparse Data in Hadron Physics with Gaussian Processes

Ryan Funai Ferguson

Submitted in fulfilment of the requirements for the
Degree of Doctor of Philosophy

Nuclear and Hadron Physics Group
School of Physics and Astronomy
College of Science and Engineering
University of Glasgow




2026

# Abstract

Hadron physics seeks to understand the structure and dynamics of strongly interacting matter governed by Quantum Chromodynamics; however, the interpretation of experimental data is complicated by their sparse, heterogeneous and multidimensional nature. This thesis develops a Gaussian Process (GP) framework for constructing continuous, uncertainty-aware representations of hadron physics observables across kinematic phase space.

A central contribution is a hyperparameter optimisation strategy designed for sparse datasets. Rather than relying on negative marginal log-likelihood minimisation, hyperparameters are determined using empirical coverage of the GP posterior, with the Wasserstein distance providing a quantitative optimisation criterion. Combined with Particle Swarm Optimisation and physics-informed parameter boundaries, this provides a systematic procedure for calibrating the predictive uncertainties. A discretised convex hull framework is also developed to distinguish interpolation from extrapolation regions in multidimensional kinematic space, incorporating experimental binning information and allowing the resolution to be controlled independently in each dimension.

The framework is validated using pseudodata and applied to polarisation observables in the $\gamma p \to K^0\Sigma^+$ reaction. The resulting GP surfaces provide continuous representations of the available measurements while retaining quantified predictive uncertainties across the energy–angle plane. A probabilistic multi-dataset combination method is also developed, in which independent GP regressions are combined into probability surfaces rather than reduced to a single averaged fit. Applied to $K^- p$ scattering data, this approach preserves non-trivial statistical structures, including separated regions of high probability, and provides a means of retaining differences between experimental constraints near the $\Lambda(1405)$ region.

Overall, the framework provides a statistical interface between sparse experimental measure-

ments and subsequent theoretical and phenomenological analyses. By retaining both the continuous representation of the observables and their associated uncertainties, it provides a basis for using sparse hadron physics data in analyses requiring information across kinematic phase space, while also identifying regions where the existing experimental constraints are comparatively weak.

# Contents

**Abstract** **i**

**Acknowledgements** **x**

**Declaration** **xii**

**1 Introduction** **1**

1.1 Machine Learning in Science . . . . . 1
1.2 Machine Learning in Physics . . . . . 3
1.3 Introduction to Gaussian Processes . . . . . 6
1.3.1 Development and Foundations of Gaussian Processes . . . . . 6
1.3.2 Gaussian Processes as Scientific Surrogate Models . . . . . 7
1.3.3 Gaussian Processes for Uncertainty Quantification . . . . . 8
1.3.4 Gaussian Processes for Experimental and Physical Data . . . . . 9
1.3.5 Summary of the Literature . . . . . 10
1.4 Gaussian Processes versus Neural Networks . . . . . 11
1.4.1 Neural Networks . . . . . 11
1.4.2 Comparison with Gaussian Processes . . . . . 12
1.5 Application to Hadron Physics . . . . . 14
1.5.1 The Issue of Sparse Experimental Data . . . . . 15
1.6 Aims and Thesis Structure . . . . . 17
1.6.1 Aims and Original Contributions . . . . . 17
1.6.2 Thesis Structure . . . . . 18

**2 Hadron Physics and Reaction Theory** **20**

2.1 Introduction to Hadron Physics . . . . . 20
2.2 The Quark Model and the Development of QCD . . . . . 21
2.3 Hadron Spectroscopy . . . . . 27
2.4 Reaction and Scattering Theory . . . . . 30
2.5 Reaction Kinematics and Coordinate Systems . . . . . 31
2.6 Polarisation Observables . . . . . 34

2.7 Structure and Limitations of Experimental Data . . . 36
2.7.1 Combining Experimental Datasets . . . 39
2.8 From Experimental Measurements to Physics Interpretation . . . 41
2.8.1 Experimental measurement . . . 42
2.8.2 Reaction reconstruction . . . 44
2.8.3 Determination of physical observables . . . 44
2.8.4 Reaction amplitude analysis . . . 44
2.8.5 Resonance and spectroscopy analysis . . . 45
2.8.6 Physics interpretation . . . 45

**3 Statistical and Data-Science Foundations** **47**
3.1 Probability Theory . . . 47
3.2 Bayesian Inference . . . 49
3.2.1 Bayesian and Frequentist Inference . . . 50
3.3 Regression . . . 51
3.4 Uncertainty Quantification . . . 54
3.5 Gaussian Processes . . . 57
3.5.1 Cholesky Decomposition . . . 63
3.6 GP Kernels . . . 64
3.6.1 Radial Basis Function Kernel . . . 66
3.6.2 Matérn kernel . . . 67
3.6.3 Comparison of RBF and Matérn Kernels . . . 68
3.7 Hyperparameter Optimisation . . . 69
3.7.1 Issues with Sparse Datasets . . . 72
3.8 Finding a Multidimensional Optimum . . . 74
3.8.1 Markov Chain Monte Carlo . . . 75
3.8.2 Particle Swarm Optimisation . . . 80
3.9 Comparing CDFs . . . 82
3.9.1 Kolmogorov-Smirnov Distance . . . 83
3.9.2 Cramér-von Mises Distance . . . 83
3.9.3 Anderson-Darling Distance . . . 84
3.9.4 Wasserstein Distance . . . 85
3.9.5 Comparison of CDF Measures . . . 86
3.10 Interpolation and Extrapolation . . . 86
3.11 Convex Hulls . . . 88

**4 Gaussian Process Framework for Hadron Data** **92**
4.1 Overview of the GP Framework . . . 92
4.2 Experimental Data . . . 93

4.3 GP Model Specification . . . 94
4.4 Hyperparameter Optimisation for Sparse Datasets . . . 96
4.4.1 Empirical Coverage . . . 97
4.4.2 Wasserstein Distance . . . 100
4.4.3 Physics-Informed Boundaries . . . 101
4.4.4 Combined Loss Function . . . 102
4.5 Optimisation Strategy . . . 106
4.6 Convex Hull Construction . . . 109
4.6.1 Additional Binning Information . . . 113
4.6.2 Computational Considerations . . . 117
4.7 Role of the Gaussian Process Framework . . . 119
4.8 Summary of the GP Framework . . . 121

**5 Validation of GP Framework** **123**
5.1 Construction of Pseudodata . . . 123
5.2 Uncertainty Calibration Test . . . 127
5.3 Recovery of Functional Form . . . 131
5.4 Summary . . . 134

**6 Results and Discussion** **136**
6.1 Application to Polarisation Observables . . . 136
6.1.1 GP Fits to 2D Polarisation Observables . . . 136
6.1.2 Results and Interpretation . . . 143
6.2 Application to Kaon Scattering Data . . . 144
6.2.1 World Dataset . . . 145
6.2.2 Individual Experimental Datasets . . . 148
6.2.3 Combined Probability Surfaces . . . 150
6.2.4 Interpretation of the Results . . . 153

**7 Conclusions and Future Work** **156**
7.1 Summary of the Thesis . . . 156
7.2 Principal Findings and Original Contributions . . . 157
7.3 Conclusions . . . 158
7.4 Future Work . . . 159
7.4.1 Higher Dimensional Observables . . . 159
7.4.2 Physics Informed Kernels . . . 161
7.4.3 Emulator . . . 162

**A PSO vs MCMC** **164**
A.1 Validation of MCMC Optimisation . . . 165

A.2 Comparison of PSO and MCMC . . . . . 166

**B Pull Distributions of Fitted Coefficients** **170**

**C List of Recurring Acronyms** **173**

**Bibliography** **174**

# List of Tables

5.1 Expected and measured coverage fractions for the pseudodata and corresponding GP points. . . . . 130
5.2 Means and variances of the Gaussian distributions fitted to the pull distributions of the fitted coefficients obtained using the original pseudodata and sampled GP datapoints. . . . . 133

A.1 Coverage results using MCMC optimisation. . . . . 165
A.2 Coefficient pull results using MCMC optimisation. . . . . 166
A.3 Comparison of PSO and MCMC loss scores. . . . . 167
A.4 Comparison of PSO and MCMC runtimes. . . . . 168

# List of Figures


1.1 Example neural network architecture. . . . . 12

2.1 Baryon Octet and Decuplet. . . . . 22
2.2 Pseudoscalar and Vector Meson Nonets. . . . . 23
2.3 QCD coupling constant $\alpha_s$. . . . . 25
2.4 Excited nucleon and $\Delta$ spectrum calculated using lattice QCD. . . . . 26
2.5 Pole positions of $\Lambda(1405)$ and related resonances in the complex energy plane. 29
2.6 Kinematics of the $\gamma p \to K^0\Sigma^+$ reaction. . . . . 33
2.7 Schematic construction of a bubble chamber. . . . . 38
2.8 Overview of the CLAS12 detector. . . . . 42

3.1 Example Empirical Coverage Curves. . . . . 56
3.2 Posterior GP samples. . . . . 62
3.3 Example GP fit. . . . . 63
3.4 Posterior samples for different kernels. . . . . 69
3.5 Example GP fits with inappropriate choices of the length scale. . . . . 70
3.6 Negative log-likelihood surface for $O_z$. . . . . 73
3.7 Example 2D corner plot. . . . . 78
3.8 Particle Position Evolution in Particle Swarm Optimisation (PSO). . . . . 82
3.9 Example GP Extrapolation Plot. . . . . 88
3.10 Example 2D Convex Hull. . . . . 91

4.1 Example Empirical Coverage Plot. . . . . 99
4.2 Example 1D Boundary Penalty Plot. . . . . 104
4.3 Loss Surface for $O_z$. . . . . 105
4.4 Example Convex Hull of a Pyramid. . . . . 112
4.5 2D Projection of Pyramid Facet. . . . . 113
4.6 Example 2D Convex Hull with Additional Binning. . . . . 117

5.1 Example of the uncertainty coverage test. . . . . 129
5.2 Comparison of functional form fits to pseudodata and GP datapoints. . . . . 132

6.1 $\Sigma$ polarisation observable. . . . . 138
6.2 $P$ polarisation observable. . . . . 139
6.3 $T$ polarisation observable. . . . . 140
6.4 $O_x$ polarisation observable. . . . . 141
6.5 $O_z$ polarisation observable. . . . . 142
6.6 GP fits to the world dataset for different $K^-p$ scattering final states. . . . . 147
6.7 Individual GP fits to experimental datasets for different $K^-p$ scattering final states. 149
6.8 Combined probability surfaces for $K^-p$ scattering final states. . . . . 152

7.1 Kinematics of DVCS. . . . . 160

A.1 Comparison of PSO and MCMC loss scores. . . . . 167
A.2 Comparison of PSO and MCMC runtimes. . . . . 168

B.1 Pull distributions of fitted coefficients using PSO. . . . . 171
B.2 Pull distributions of fitted coefficients using MCMC. . . . . 172

# Acknowledgements

This work has taken nearly 4 years to complete and there are so many people I wish I could thank for their help and support. Here are just a few that I want to give particular mention to.

My first and primary thanks has to go to Dave Ireland whose own interest first created this project and whose advice and expertise helped immeasurably during many Thursday morning meetings with many scrawlings on a whiteboard. I also want to thank Bryan McKinnon who was always on hand to keep me on track and provide a useful sounding board. The support and encouragement from both of you has resulted in my optimising the scope of this project. It was a pleasure to work with you.

A special thanks to Derek Glazier whose insights and suggestions and always answering any questions, no matter how trivial, have been of immense help. Thanks also to Maxim Mai and Feri Pittler for their collaboration on kaon scattering which sparked the idea that later became a key part of this thesis. Thanks also to my high school teachers Mr Hannah and Mr Whyte for starting me on this journey.

To my fellow survivors of the Dungeon that is 414 and also to the wider 4th floor: Alex, Andrew, Dave, Dillon, Gerry, Kathleen, Kayleigh, Nia, Richard, Will and Will; thanks for putting up with my stupid jokes, the countless card games, making FBF the lunch highlight of the week, Friday Films with Friends, the many trips to Curlers (and many more pubs in the city centre) and always being able to put a smile on my face. I could not have asked for a better group of people to have shared the PhD experience with.

To my friends and colleagues at the University of York: Alex, Anthony, Asli, Charlie, Stephen and, of course, The Dr Stuart Fegan (who is related to multiple Doctor Whos (not by blood));

thank you for making the annual IOP conferences a highlight of the year, helping organise the inaugural Glasgow Conference on Hadron Physics, making the STFC summer school an enjoyable experience and managing to make any trips to JLab somewhat bearable (margaronas and trivia did help).

To my friends outside of academia; thank you for helping me over the years with pub quizzes, lodge holidays, whisky tastings, late night gaming sessions, nights at the pub and general good times.

Finally, a heartfelt thanks to my mum and dad. Your love and support over all my years of studies have made this thesis possible and I cannot thank you enough.

*"A wise man once said 'Je ne sais pas'."*
– Remy du Raton

# Declaration

I hereby declare that the composition of this thesis, along with the work contained and described within it are my own unless explicitly stated. Where input/results from others have been given or collaboration with others has been made this is stated or referenced appropriately throughout the thesis. This work has not previously been submitted for any other degree to the University of Glasgow or any other institution.

Ryan Funai Ferguson

# Chapter 1

# Introduction

## 1.1 Machine Learning in Science

Machine Learning (ML) has become an increasingly important tool for scientific research, providing methods for extracting information from data, constructing predictive models and identifying relationships that may be difficult to describe using conventional analytical approaches. Rather than requiring an explicit functional form to be specified in advance, many ML methods learn relationships directly from available data. This has made ML particularly useful in scientific problems where the underlying system is complex, high-dimensional or only partially understood.

The application of ML to science is not fundamentally different from its application in other fields, but scientific problems introduce several requirements that influence the choice of method. Scientific measurements are often associated with uncertainties, correlations and experimental constraints, while the available data may be expensive to obtain and consequently limited in size. Furthermore, scientific modelling generally requires more than accurate predictions. In many applications, it is important to quantify the uncertainty associated with a prediction, particularly when the result is used for subsequent physical interpretation or as an input to another analysis. These requirements have motivated the development and application of a broad range of statistical and ML techniques within the physical sciences [1].

ML methods can broadly be divided according to the type of information available during training. In supervised learning, a model is trained using input–output pairs, with the objective of learning a mapping between measured or calculated quantities. Regression is a supervised learning problem in which the output is a continuous quantity and is therefore particularly relevant when constructing representations of physical observables. Classification instead considers discrete output categories. In unsupervised learning, the desired output is not specified, and the objective is instead to identify structure or patterns within the input data. Other approaches, including reinforcement learning, formulate learning as a sequential decision making problem in which an agent learns through interactions with an environment.

A wide range of ML methods have been applied to scientific data. Neural networks, for example, use layers of interconnected parameters to construct flexible non-linear mappings between input and output variables. Tree-based methods can provide effective models for classification and regression, while clustering methods can be used to identify groups or structures within datasets. Dimensionality reduction methods provide another approach for identifying lower-dimensional representations of complex datasets. The suitability of each method depends on the characteristics of the problem, including the dimensionality and size of the available dataset, the desired form of the output and the extent to which uncertainty must be represented.

An important development in scientific ML has been the use of data-driven methods alongside established physical knowledge. In many applications, ML is not intended to replace physical theory but rather to complement it. For example, ML models can be used as computationally efficient surrogates for expensive simulations, allowing regions of parameter space to be explored more efficiently. They can also be used to identify features in experimental measurements, reconstruct quantities that are difficult to measure directly, or accelerate components of numerical calculations. Physics-informed approaches provide another means of incorporating known physical relationships or constraints into the learning process, reducing the extent to which physically meaningful behaviour must be inferred from data alone [2].

The range of scientific applications of ML has expanded alongside the increasing scale and complexity of experimental and simulated datasets. Modern experiments and simulations can produce datasets containing millions of measurements and many correlated variables, creating challenges that are well suited to automated data-driven analysis. However, the availability of large datasets is not universal across scientific disciplines or applications. Some measurements are inherently expensive, difficult to reproduce or restricted to particular regions of parameter space. Consequently, scientific ML encompasses problems ranging from the analysis of extremely large datasets to situations in which only a relatively small number of measurements are

available.

These differences in data availability and scientific objectives are important when selecting an appropriate ML method. A method that performs well when large training datasets are available may not be equally suitable when measurements are sparse or irregularly distributed. Furthermore, different methods provide different approaches to representing uncertainty. For scientific applications, uncertainty is not simply a measure of predictive performance but can form an important part of the result itself. A model that provides a prediction without a corresponding estimate of its uncertainty may therefore be insufficient when the reliability of the prediction must be assessed or when the result is subsequently used in a physical analysis.

Probabilistic ML methods provide a framework in which uncertainty can be incorporated directly into the prediction. Rather than producing only a single estimated value, such methods can describe a distribution of possible predictions, allowing both the expected behaviour and the associated uncertainty to be considered. This is particularly useful when constructing continuous representations of experimentally measured quantities, since the available measurements generally constrain the model to different degrees across the relevant parameter space. Regions close to well-constrained measurements may therefore warrant smaller predictive uncertainties, while regions with limited experimental information should reflect the corresponding lack of knowledge.

The use of ML in science therefore encompasses a broad range of applications and methodologies, with the appropriate choice depending on both the characteristics of the available data and the scientific objective. For the problem considered in this thesis, particular importance is placed on regression of experimental data and on obtaining quantified uncertainties alongside the predicted observable. The use of ML in physics, and the range of problems to which these methods have been applied, is considered in the following section.

## 1.2 Machine Learning in Physics

The physical sciences have become an important area of application for ML, with methods being used across experimental data analysis, numerical simulation and theoretical modelling. A broad review by Carleo *et al.* describes applications of ML across a range of physical disciplines, including particle physics, nuclear physics, condensed matter and quantum systems [1]. The

motivations for these applications are varied, reflecting the different types of data and scientific problems encountered across physics.

In particle physics, ML has become particularly established in the analysis of experimental data. Modern experiments can produce extremely large numbers of events, with each event containing many measured or reconstructed quantities. ML methods can therefore be used to identify patterns within high-dimensional datasets and to perform tasks such as event classification, signal-background discrimination, particle identification and event reconstruction. Guest *et al.* provide a review of deep learning applications in high energy physics, covering applications throughout the experimental analysis chain [3]. Radovic *et al.* similarly review the use of ML across the energy and intensity frontiers of particle physics, demonstrating its application to both experimental analysis and detector related problems [4].

The use of ML in physics is not limited to the analysis of large experimental datasets. ML methods have also been applied to theoretical and computational problems, where they can be used to identify relationships between physical quantities, accelerate numerical calculations and assist with parameter inference. In such applications, ML can provide a data-driven representation of a physical system without requiring the full underlying calculation to be performed at every evaluation. This can be particularly useful when calculations are computationally expensive or when a large parameter space must be explored repeatedly [1].

Nuclear physics provides another substantial area of application, encompassing problems in nuclear structure, nuclear reactions, many body theory and the analysis of experimental data. Boehnlein *et al.* provide a comprehensive review of these applications, demonstrating how ML has been used for tasks including prediction of nuclear properties, classification, parameter estimation and the analysis of complex nuclear datasets [5]. These applications illustrate that the role of ML in physics extends beyond replacing computationally expensive calculations, with ML methods also being used to extract information from data and assist in problems where the underlying physical relationships are difficult to model directly.

A further development within scientific ML has been the incorporation of established physical knowledge into data-driven models. Rather than treating the available data as the only source of information, physical constraints can be incorporated into the learning process. Physics-informed approaches provide one example of this strategy, in which known equations or other physical relationships are incorporated into the model or its training procedure [2]. Such approaches illustrate the broader principle that ML can be combined with established physical

theory rather than being used independently of it.

These applications demonstrate that ML is used for substantially different purposes within physics. In experimental particle physics, the emphasis is often on extracting information from large and complex datasets. In theoretical and computational applications, ML may instead be used to construct efficient surrogate models or assist with parameter inference. In nuclear physics, applications span both experimental and theoretical problems, while physics-informed approaches seek to combine data-driven methods with established physical knowledge. Consequently, the suitability of a particular ML method depends strongly on the characteristics of the data and the scientific objective.

The problem considered in this thesis differs from many of these established applications. The objective is not primarily to classify events, reconstruct detector information or replace an expensive theoretical calculation. Instead, the aim is to construct a continuous representation of experimentally measured physical observables across a multidimensional kinematic space. The available measurements are sparse and irregularly distributed, and individual measurements have associated experimental uncertainties.

These characteristics place particular emphasis on regression and uncertainty quantification. The desired model must provide an estimate of the observable at points where no direct measurement is available, while also quantifying the uncertainty associated with that prediction. This requirement distinguishes the problem from many applications in which the primary objective is classification or point prediction and motivates consideration of probabilistic regression methods.

Gaussian Processes provide one such approach, combining flexible regression with a probabilistic description of the predicted observable. Their use in scientific and physical applications, and in particular their application to problems involving uncertainty quantification and limited experimental data, is considered in the following section.

## 1.3 Introduction to Gaussian Processes

The Gaussian Process (GP) has become an established approach to probabilistic regression and uncertainty quantification within scientific ML. In contrast to methods that produce only a single prediction for a given input, a GP defines a probability distribution over possible functions. Predictions therefore consist of both an estimated value and an associated uncertainty. This probabilistic formulation has made GPs useful in a range of scientific applications, including the construction of surrogate models, the interpolation of experimental measurements and the representation of uncertainties in theoretical calculations.

### 1.3.1 Development and Foundations of Gaussian Processes

The modern formulation of GPs for machine learning builds on earlier work in Gaussian Process regression and related kernel methods. Rasmussen and Williams provide a comprehensive treatment of GPs for ML, establishing the probabilistic framework in which an unknown function is described through a distribution over functions [6]. Seeger subsequently reviewed the application of GPs to ML problems including regression, classification and probabilistic modelling [7]. These works established GPs as a general framework for learning relationships between input and output variables while retaining a probabilistic description of the resulting predictions.

A central feature of this framework is the use of a covariance function, or kernel, to describe the relationship between function values at different input locations. The kernel determines how information from observations is correlated across the input space. Consequently, a GP can use measurements at known locations to constrain predictions at locations where no direct observation is available, while the resulting uncertainty reflects the extent to which those predictions are constrained by the available information. The choice of covariance function therefore provides a means of incorporating assumptions about the expected behaviour of the underlying function and is discussed further in Section 3.6.

The probabilistic nature of GPs has led to their use in scientific problems where uncertainty in a prediction is important in addition to the prediction itself. This is particularly relevant when a GP is used as a representation of a physical quantity rather than simply as a tool for classification or point prediction. The following sections consider how this framework has been applied in practice across scientific and physical applications.

### 1.3.2 Gaussian Processes as Scientific Surrogate Models

One of the established applications of GPs in science is the construction of surrogate models for computationally expensive calculations. In many areas of physics, detailed numerical calculations may require substantial computational resources, making repeated evaluation impractical when exploring a large parameter space or performing statistical inference. A GP can instead be trained using a finite set of evaluations of the original calculation and subsequently used as a computationally cheaper approximation.

An early example of this approach in materials science was provided by Bartók *et al.*, who introduced Gaussian Approximation Potentials for modelling atomic interactions using GP methods [8]. Their approach used results from quantum mechanical calculations to construct a model of the potential-energy surface. The resulting GP representation could then be used in subsequent calculations without requiring the original electronic structure calculation at every evaluation. This demonstrated the ability of GPs to provide a flexible representation of a complex physical relationship while substantially reducing the computational cost of subsequent evaluations.

Similar approaches have been adopted in nuclear physics. Higdon *et al.* demonstrated the use of GP emulators to approximate computationally intensive nuclear physics calculations [9]. Their approach allowed a relatively small number of evaluations of the underlying calculation to be used to construct an emulator that could subsequently be evaluated efficiently during parameter estimation. Importantly, the emulator also provided an estimate of the uncertainty associated with its prediction, allowing the approximation uncertainty to be incorporated into the subsequent statistical analysis.

McDonnell *et al.* applied a related GP emulation approach to nuclear physics calculations, using the emulator to quantify uncertainties in theoretical predictions and to investigate the information provided by additional measurements [10]. In this case, the GP was therefore not simply being used as a computational replacement for the original calculation; its probabilistic formulation allowed uncertainty information to be retained while exploring the relationship between model parameters and predicted observables.

These studies illustrate an important role for GPs in scientific modelling. A surrogate model can provide a computationally efficient approximation to an expensive calculation, while the probabilistic formulation of the GP allows the uncertainty associated with that approximation to

be quantified. The GP therefore provides information about both the predicted behaviour and the reliability of the prediction.

### 1.3.3 Gaussian Processes for Uncertainty Quantification

The ability to represent uncertainty has also motivated the use of GPs independently of their role as computational surrogates. In many scientific calculations, predictions at different input values are not statistically independent. For example, theoretical predictions at neighbouring energies or parameter values may share common sources of uncertainty. Treating these uncertainties independently can therefore fail to capture important correlations within the predictions. GPs provide a natural framework for describing such correlated uncertainties. Rather than assigning an independent uncertainty to each prediction, the covariance kernel describes how the uncertainty at different input locations is related. This allows the uncertainty to be represented as a continuous and correlated function over the relevant input space.

Acharya and Bacca used a GP to model the uncertainty in theoretical predictions for a low energy nuclear reaction [11]. Their approach treated the theoretical error as a correlated function of the relevant input energy, allowing information about the uncertainty at one energy to be related to the uncertainty at another. This demonstrates an important application of the GP framework in which the quantity being modelled is not necessarily the physical observable itself, but the uncertainty associated with a theoretical prediction.

Similar approaches have been applied to nuclear matter calculations. Drischler *et al.* used GPs to quantify uncertainties and correlations in predictions of the nuclear matter equation of state [12]. By representing the predictions probabilistically, their approach allowed correlations across the relevant input variables to be retained when assessing the uncertainty of the theoretical result. Keller *et al.* subsequently used a GP emulator to efficiently represent nuclear matter calculations while retaining information about the uncertainty in the resulting predictions [13].

These applications demonstrate a broader role for GPs in scientific uncertainty quantification. The uncertainty is not treated simply as an independent error attached to each prediction, but as a quantity that can vary across the input space and possess correlations between different predictions. This ability to describe both the predicted function and its uncertainty is one of the principal characteristics distinguishing GPs from conventional deterministic regression methods.

### 1.3.4 Gaussian Processes for Experimental and Physical Data

GPs have also been applied directly to experimental measurements, where the objective is to construct a continuous representation of a physical quantity from observations made at discrete input locations. In such applications, the GP can provide predictions at locations where no direct measurement is available while retaining an estimate of the uncertainty associated with those predictions.

Noack *et al.* reviewed the use of GPs for autonomous data acquisition at large scale synchrotron and neutron facilities [14]. In this setting, GP models were used to describe experimentally measured quantities and to guide the selection of subsequent measurements. The predictive uncertainty of the GP therefore provided information not only about the reliability of the current model, but also about where additional experimental data could be most informative. This illustrates how uncertainty estimates can become an active part of the experimental process rather than simply an output reported after the analysis.

Iwamoto applied GP regression directly to experimental nuclear cross-section data, using measurements at discrete energies to generate predictions at energies where measurements were unavailable [15]. The resulting GP provided both an interpolated estimate of the cross-section and an associated uncertainty. This represents a particularly direct application of GP regression to experimental data, where the objective is to construct a continuous representation of a measured physical quantity rather than to emulate an expensive theoretical calculation.

GP methods have also been applied to physical reaction observables and hadron related quantities. Acharya and Bacca considered theoretical reaction predictions as discussed above, demonstrating the use of GPs for modelling correlated uncertainties in a physical observable [11]. Gal *et al.* used ML methods including GPs and neural networks to investigate relationships between hadron masses and related measured quantities [16]. Their work demonstrates that GP methods can be incorporated into ML approaches to problems involving experimentally measured hadron quantities.

More recently, Wang *et al.* applied two-dimensional GP regression to experimental data from exclusive charmonium photoproduction [17]. The GP used two kinematic variables as inputs to construct a continuous representation of an experimentally measured cross-section and was subsequently used to extrapolate the observable into a kinematic region where measurements

were not directly available. The associated GP uncertainty provided an estimate of the reliability of the extrapolated prediction. This study demonstrates the application of GP regression to multidimensional experimental data in a hadron reaction and illustrates the usefulness of the GP framework when predictions are required beyond the locations of the available measurements.

### 1.3.5 Summary of the Literature

The literature demonstrates that Gaussian Processes have been applied to a broad range of scientific problems, but several common features emerge across these applications. Bartók *et al.*, Higdon *et al.* and McDonnell *et al.* demonstrate the use of GPs as computationally efficient surrogate models for expensive scientific calculations [8–10]. Drischler *et al.*, Keller *et al.* and Acharya and Bacca demonstrate the use of GPs to represent and propagate correlated uncertainties in theoretical predictions [11–13]. Noack *et al.* and Iwamoto demonstrate applications to experimental measurements, including continuous prediction and the use of uncertainty to guide data acquisition [14, 15]. Finally, Gal *et al.* and Wang *et al.* demonstrate applications of ML and GP regression to hadron related physical quantities and experimental reaction data [16, 17].

Although the specific applications differ substantially, a recurring feature of these studies is that the GP provides more than a deterministic estimate of the quantity of interest. Its probabilistic formulation allows information about the uncertainty and correlations of the prediction to be retained. This is particularly valuable when the model is used to interpolate or extrapolate a physical quantity, when the underlying calculation is expensive, or when the reliability of the prediction is itself important to the scientific analysis.

The literature therefore establishes Gaussian Processes as a mature method for scientific regression and uncertainty quantification, with applications ranging from computational modelling to experimental data analysis. The following section compares this probabilistic approach with neural networks, another widely used ML method for flexible function approximation, and considers the implications of their different approaches to uncertainty quantification.

## 1.4 Gaussian Processes versus Neural Networks

Gaussian Processes and neural networks are both flexible machine learning approaches that can be used to learn non-linear relationships between input variables and a continuous output. Both methods have been applied extensively to scientific problems, and neither is universally preferable. The suitability of either approach depends on the characteristics of the data and the requirements of the problem. For the application considered in this thesis, an important consideration is that the model should provide not only a prediction of the observable but also a quantified uncertainty associated with that prediction.

### 1.4.1 Neural Networks

Neural networks are flexible parametric models that construct a mapping between input and output variables through a sequence of interconnected layers. Each layer applies a transformation to its inputs using a set of learned parameters, typically consisting of weights and biases, followed by a non-linear activation function. By combining multiple layers, a neural network can represent highly non-linear relationships between its inputs and outputs [18].

A simple neural network consists of an input layer, one or more hidden layers and an output layer. The input layer receives the variables describing the system, while the hidden layers transform these inputs into increasingly complex representations. The output layer produces the quantity predicted by the network. An example of this structure is shown in Figure 1.1.

During training, the parameters of the network are adjusted to minimise a chosen loss function evaluated using the available training data. The flexibility of the network is determined by factors such as the number of layers, the number of parameters and the choice of activation functions. This flexibility has made neural networks particularly successful for problems involving complex relationships between high-dimensional inputs and outputs. They are consequently widely used in scientific applications, including experimental particle physics analysis, simulation and reconstruction [3, 4].

For a conventional neural network regression model, the output is typically a single predicted value for a given input. The network therefore provides a deterministic mapping once its param-

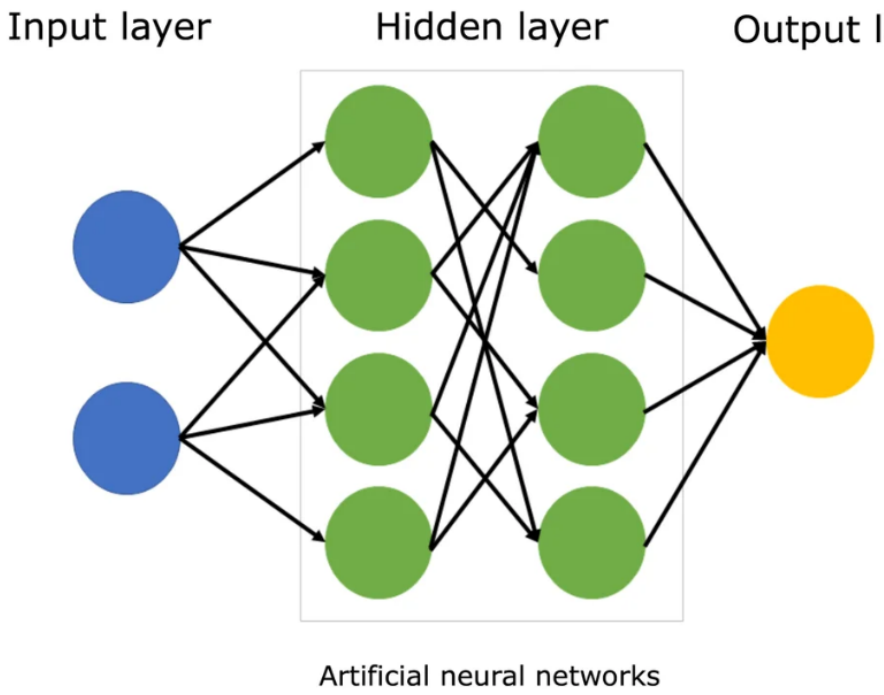



Figure 1.1: Example of a simple neural network showing the input, hidden and output layers and the connections between them. Taken from [19].

eters have been trained. Uncertainty can nevertheless be incorporated into neural network models. For example, probabilistic output models can be used to predict a distribution rather than a single value, while ensemble methods and Bayesian neural networks can provide estimates of predictive or parameter uncertainty. These approaches can be effective, but they introduce additional modelling assumptions or inference procedures beyond those required for a standard deterministic neural network.

### 1.4.2 Comparison with Gaussian Processes

Gaussian Processes take a fundamentally different approach to function approximation. Rather than representing the unknown relationship using a fixed set of fitted parameters, a GP defines a probability distribution over possible functions. The covariance kernel determines how strongly function values at different input locations are related. Once the GP is conditioned on observations, this distribution produces a predictive distribution at any new input location.

The difference in the representation of uncertainty is particularly important. In a standard neural network regression model, the network produces a point prediction – additional techniques are

required if an uncertainty is desired. In contrast, predictive uncertainty is an intrinsic component of the standard GP framework. The GP provides a predictive mean and variance, with the uncertainty determined by the covariance structure and the information provided by the observations. Consequently, uncertainty is produced as part of the regression procedure rather than being added as a separate modelling stage.

The two approaches also differ in how assumptions about the underlying function are incorporated. Neural networks can represent highly complex relationships through their architecture and learned parameters, with the architecture, activation functions and training procedure influencing the resulting model. GPs instead express assumptions about the behaviour of the function primarily through the choice of covariance kernel. The kernel specifies how strongly information at one input location influences predictions at another and therefore controls properties such as the smoothness and correlation structure of the predicted function. The choice of kernel is consequently an important modelling assumption when applying a GP.

This distinction does not mean that GPs are inherently more suitable than neural networks. Neural networks are capable of representing extremely complex functions and can be advantageous when large and diverse datasets are available or when the problem involves very high-dimensional inputs. They can also incorporate physical constraints and can be combined with probabilistic methods when uncertainty quantification is required. Conversely, the standard GP formulation has computational limitations as the number of training observations becomes large, since inference involves operations on the covariance matrix. The choice between the two approaches must therefore be based on the requirements of the particular problem.

For the application considered in this thesis, the principal requirement is not simply to obtain an accurate continuous representation of the measured observable – the model must also provide a quantified uncertainty across the input space. This uncertainty is required to assess how strongly the available experimental measurements constrain the resulting prediction and to distinguish regions where the model is well constrained from those where the available information provides weaker constraints.

The probabilistic formulation of a GP provides this uncertainty as an intrinsic component of the regression model. This makes it particularly appropriate for the objective of constructing a continuous representation of an experimentally measured observable while retaining information about the reliability of the prediction. A neural network could also be extended to provide uncertainty estimates, but doing so would require an additional probabilistic, ensemble or Bayesian

framework. For the present application, the direct incorporation of predictive uncertainty into the GP framework therefore provides a strong motivation for its use.

Gaussian Processes are therefore adopted in this work primarily because their probabilistic formulation provides predictive uncertainty as an intrinsic component of the regression model. The characteristics of the experimental data considered in this thesis, including their sparse and irregular distribution across the relevant kinematic space, introduce additional challenges for the application of ML methods. These challenges, and the motivation for applying the GP framework to sparse experimental data, are considered in the following section.

## 1.5 Application to Hadron Physics

The applications of Gaussian Processes discussed in Section 1.3 demonstrate their use in a range of scientific regression and uncertainty quantification problems. In particular, applications to experimental measurements show that GPs can be used to construct continuous representations of physical observables from discrete observations while retaining information about the uncertainty of the prediction. These characteristics make GP regression relevant to the analysis of experimental data in hadron physics.

Hadron physics seeks to understand the structure and interactions of strongly interacting particles, including mesons and baryons. Experimental studies of hadron reactions provide measurements of physical observables as functions of the kinematic variables describing the reaction. These observables are therefore naturally represented as functions over a kinematic space. However, experiments do not provide measurements at every possible point in this space. Instead, measurements are obtained at a finite set of discrete kinematic points, determined by the experimental setup, available beam energies, detector acceptance and the feasibility of extracting the observable.

A continuous representation of these measurements is useful for several aspects of hadron physics analysis. Physical models and reaction analyses may require an estimate of an observable at kinematic points where no direct measurement is available. Similarly, measurements from different experiments may cover different regions of the same kinematic space, making it useful to combine them into a common representation. A regression model can provide a means of connecting these individual measurements and estimating the observable throughout

the region of interest.

The use of a ML regression method for this purpose is not intended to replace the underlying physical description of the reaction. Rather, it provides a data-driven representation of the measured observable that can be used alongside established theoretical and experimental methods. This is particularly useful when the primary objective is to represent the available experimental information without imposing a detailed theoretical model for the functional dependence of the observable on its kinematic variables.

Gaussian Processes are well suited to this type of probabilistic regression because they provide both an estimate of the observable and a quantified predictive uncertainty. The resulting model can therefore represent not only the measured behaviour but also the degree to which the available experimental information constrains the prediction throughout the kinematic space. This is particularly important for hadron physics data, where the experimental coverage is often non-uniform and the available measurements can be separated by substantial regions of kinematic space.

### 1.5.1 The Issue of Sparse Experimental Data

The principal challenge in applying this approach to hadron physics measurements is the sparsity of the available experimental data. Measurements of an observable are generally available only at a finite number of kinematic points, and these points are not necessarily distributed uniformly throughout the accessible kinematic region. Some regions may contain many measurements, while others may contain relatively little experimental information or no measurements at all.

This issue becomes increasingly important when the observable depends on more than one kinematic variable. A collection of measurements that provides reasonable coverage in each variable individually may nevertheless provide relatively poor coverage of the full multidimensional space. The number of measurements required to populate a multidimensional space uniformly also increases rapidly with the number of dimensions. Consequently, experimental datasets that are adequate for studying individual kinematic dependencies can remain sparse when considered as samples of the full observable.

Sparse coverage affects the interpretation of any regression model constructed from the data.

In regions containing measurements, the available observations provide direct constraints on the predicted observable. Between these measurements, the model must infer the behaviour of the observable from the surrounding data and the assumptions incorporated into the regression method. In regions far from available measurements, the prediction is increasingly dependent on those assumptions rather than on direct experimental information.

For this reason, a smooth interpolation alone is insufficient for the present application. A regression model should also provide information about how strongly each prediction is supported by the available measurements. Gaussian Processes provide this information through their predictive uncertainty. The uncertainty can vary across the kinematic space according to the locations and uncertainties of the experimental measurements, allowing regions with strong experimental constraints to be distinguished from regions where the available information is limited.

The sparse nature of the data does not, however, constitute an intrinsic advantage of Gaussian Processes. A GP remains dependent on assumptions encoded through its covariance function and hyperparameters, and these assumptions become particularly important where experimental constraints are weak. Applying GPs to sparse hadron physics data therefore requires careful consideration of the covariance structure, the treatment of experimental uncertainties and the methods used to determine the model hyperparameters.

The combination of multidimensional kinematic dependence, irregular experimental coverage and measurement uncertainty therefore presents a non-trivial regression problem. The objective of this thesis is to investigate how Gaussian Process regression can be applied to this problem to construct continuous representations of hadron physics observables while providing reliable estimates of the associated predictive uncertainty.

The physical observables and reaction processes relevant to this problem are introduced in the following chapter. The subsequent chapters then develop the statistical and Gaussian Process methodology required to construct and validate the resulting representations.

## 1.6 Aims and Thesis Structure

### 1.6.1 Aims and Original Contributions

The aim of this thesis is to investigate the application of Gaussian Process regression to sparse experimental data in hadron physics. In particular, the work considers the construction of continuous representations of experimentally measured observables across multidimensional kinematic spaces, while retaining a quantified uncertainty associated with the resulting predictions.

A central requirement of the work is therefore not only to obtain predictions of the measured observables, but also to assess the reliability of the associated predictive uncertainties. The methodology developed in this thesis investigates the construction and optimisation of Gaussian Process models for this purpose, including the treatment of multidimensional experimental data, the determination of appropriate kernel hyperparameters and the assessment of the resulting uncertainty estimates.

The work also considers the practical challenges introduced by sparse and irregularly distributed experimental measurements. Particular attention is given to how the available data constrain the GP across the kinematic space and how the predictive uncertainty behaves in regions with limited experimental information. Methods for assessing the calibration of the resulting uncertainties are therefore developed alongside the regression methodology.

Except where explicitly stated otherwise, all computational code developed for this thesis was written by the author using standard `NumPy` and `SciPy` libraries [20, 21]. Although established Gaussian Process implementations are available, a custom implementation was chosen to provide direct control over the components of the GP framework, including covariance construction, posterior evaluation, uncertainty calculation and hyperparameter optimisation. This was particularly important during the development of the empirical coverage-based hyperparameter methodology, which requires repeated evaluation of the GP posterior within a custom optimisation procedure. The custom implementation therefore allowed the framework to be developed and modified without being constrained by the interfaces or optimisation procedures of existing GP libraries. All of the code used in the main body of this thesis is available at the public GitHub repository [22].

The original contributions of this thesis include the development of a Gaussian Process framework for sparse multidimensional hadron physics observables, a methodology for determining the model hyperparameters in the presence of sparse data, and procedures for validating the resulting predictive uncertainties. The framework is first tested using controlled pseudodata studies before being applied to experimental polarisation observables and kaon scattering data. The resulting approach provides a data-driven means of constructing continuous representations of experimental observables while retaining information about the reliability of the predictions.

### 1.6.2 Thesis Structure

The remainder of this thesis is organised as follows. Chapter 2 introduces the hadron physics background required for the subsequent analysis. The chapter discusses the quark model and the development of Quantum Chromodynamics, hadron spectroscopy, reaction and scattering theory, reaction kinematics and coordinate systems and polarisation observables. It then describes the structure and limitations of the experimental data used in this work, including the combination of measurements from different experimental datasets. The chapter concludes by placing these measurements within the broader process through which experimental data are converted into physical interpretation.

Chapter 3 introduces the statistical and data-science foundations required for the methodology developed in this thesis. Probability theory, Bayesian inference, regression and uncertainty quantification are introduced before the Gaussian Process framework is developed mathematically. The chapter then considers covariance kernels, hyperparameter optimisation and multidimensional optimisation methods, followed by the statistical measures used to compare distributions. Interpolation, extrapolation and convex hull methods are also introduced where required by the subsequent analysis.

Chapter 4 develops the Gaussian Process framework used in this thesis. The experimental datasets and GP model specification are introduced before the methods used to optimise the model for sparse datasets are developed. The optimisation procedure, treatment of the kinematic domain and construction of the convex hull are then described. The chapter concludes by summarising the role of the individual components within the complete GP framework.

Chapter 5 validates the resulting framework using controlled pseudodata studies. The construc-

tion of pseudodata is described before the calibration of the GP predictive uncertainties and the ability of the model to recover known functional behaviour are investigated.

Chapter 6 presents the application of the validated framework to experimental hadron physics data. The first part considers two-dimensional polarisation observables, while the second applies the framework to kaon scattering data and considers both individual experimental datasets and their combination. The resulting probability surfaces and their physical interpretation are discussed.

Finally, Chapter 7 summarises the main findings and original contributions of the thesis and discusses potential directions for future work, including extensions to higher-dimensional observables, physics-informed covariance kernels and the use of the GP framework as an emulator.

# Chapter 2

# Hadron Physics and Reaction Theory

## 2.1 Introduction to Hadron Physics

The quest to understand the fundamental constituents of matter and the forces governing their interactions has long driven the development of physics. Hadron physics is concerned with the study of composite particles formed from quarks and gluons, collectively known as hadrons. Hadrons are broadly divided into baryons, which are described as states containing three valence quarks, and mesons, which are described predominantly as quark-antiquark states. Their observed properties arise from the interactions between their constituent quarks and gluons and from the underlying dynamics of the strong interaction.

Understanding the structure and interactions of hadrons requires experimental information over a range of energies and reaction channels. One of the primary tools for obtaining this information is scattering, in which an initial state is prepared and the resulting particles are measured. By studying quantities such as cross-sections, angular distributions, energy dependence and polarisation observables, information about the underlying interaction and the intermediate states contributing to a reaction can be extracted. Scattering experiments therefore provide an important connection between the theoretical description of the strong interaction and experimentally accessible properties of hadrons.

A particularly important regime is that of low-energy hadron physics, where the strong interac-

tion becomes non-perturbative. In this regime, the interaction cannot generally be treated using straightforward perturbative expansions, and the observed properties of hadrons emerge from complicated many-body dynamics. Understanding how the fundamental quark and gluon degrees of freedom give rise to the observed spectrum and properties of hadrons therefore remains an important problem in contemporary physics.

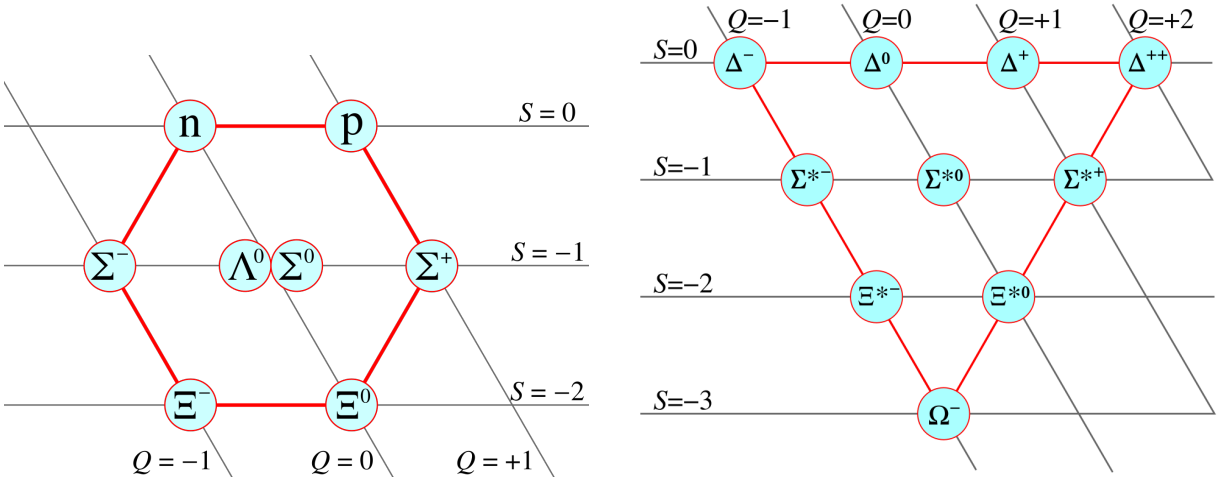
The study of hadrons has therefore developed through increasingly fundamental descriptions, from phenomenological classifications based on flavour symmetries to the quark model and ultimately to Quantum Chromodynamics (QCD). The development of this description and its implications for our understanding of hadrons are discussed in the following section.

## 2.2 The Quark Model and the Development of QCD

Before the development of a gauge theory for the strong force, the quark model was introduced as a phenomenological framework [23]. The classification of hadrons was first achieved through the use of $\mathrm{SU}(3)$ flavour symmetry, which arranged hadrons into multiplets according to their quantum numbers. The baryon octet and decuplet and pseudoscalar and vector meson nonets were systematically organised within this symmetry framework by Gell-Mann [24]. This structure, often referred to as the "Eightfold Way", revealed similarities in hadron masses and properties, suggesting an underlying symmetry in the strong force.

The interpretation of these multiplets in terms of more fundamental constituents – quarks – was subsequently proposed independently by Gell-Mann and Zweig [25, 26]. In this picture, the observed hadrons correspond to specific combinations of the three light quark flavours, up ($u$), down ($d$) and strange ($s$), and their corresponding antiquarks. These quarks form a flavour triplet under $\mathrm{SU}(3)$, and combinations of them give rise to the different hadron multiplets. The members of a multiplet can be represented in terms of their quantum numbers, with the baryon and meson diagrams commonly arranged according to electric charge, $Q$, and strangeness, $S$. The electric charge is determined by the charges of the constituent quarks, with $Q_u = +2/3$ and $Q_d = Q_s = -1/3$, while strangeness, $S$, is a quantum number associated with strange quark content, with a strange quark carrying $S = -1$ and an anti-strange quark carrying $S = +1$. Consequently, the positions of the hadrons within the diagrams reflect their underlying quark composition and the flavour symmetry structure.

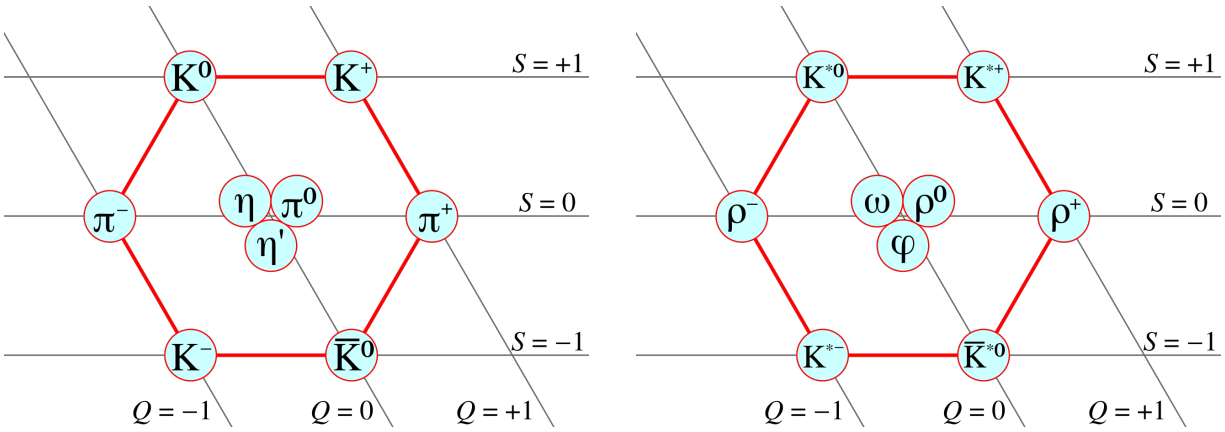
The baryon octet contains the lowest lying spin-$1/2$ baryons, while the baryon decuplet contains a set of low-lying spin-$3/2$ baryons. Here, the spin of a particle refers to its intrinsic angular momentum, with the values $1/2$ and $3/2$ denoting the corresponding spin quantum number. In the quark model, baryons are described as three quark states, whose intrinsic spins combine to give the total spin of the baryon. Similarly, the pseudoscalar and vector mesons form nonets consisting predominantly of quark-antiquark states. The pseudoscalar mesons have spin 0, while the vector mesons have spin 1. The resulting quark interpretations of these baryon and meson multiplets are illustrated in Figures 2.1 and 2.2, respectively.




(a) Baryon Octet.

(b) Baryon Decuplet.

Figure 2.1: The baryon octet (left) and decuplet (right), showing the organisation of the light baryon states according to their strangeness, $S$, and electric charge, $Q$. The horizontal axis denotes the electric charge and the vertical axis denotes the strangeness of each state. The position of each baryon therefore reflects its quantum numbers and, in the quark model, its underlying combination of up ($u$), down ($d$) and strange ($s$) quarks. The octet consists of spin-$1/2$ baryons, while the decuplet consists of spin-$3/2$ baryons, where spin denotes the intrinsic angular momentum quantum number of the particle. The octet and decuplet classification was described by Gell-Mann [24], with their quark interpretation subsequently proposed independently by Gell-Mann and Zweig [25, 26]. Taken from [27].




(a) Pseudoscalar Meson Nonet. (b) Vector Meson Nonet.

Figure 2.2: Pseudoscalar (left) and vector (right) meson nonets, showing the organisation of the light meson states according to their strangeness, $S$, and electric charge, $Q$. Unlike baryons, which are described as three quark states, mesons are described predominantly as quark-antiquark states. The pseudoscalar nonet contains spin-0 mesons, while the vector nonet contains spin-1 mesons, where spin denotes the intrinsic angular momentum quantum number of the particle. The position of each meson in the diagrams reflects its quantum numbers and its underlying combination of up ($u$), down ($d$) and strange ($s$) quarks and their corresponding anti-quarks. The nonet classification was described by Gell-Mann [24], with the quark interpretation subsequently proposed independently by Gell-Mann and Zweig [25, 26]. Taken from [27].

Despite its striking success, the quark model is not a complete theory of the strong force. It provides a classification scheme and an intuitive picture of hadrons in terms of constituent quarks, but it does not specify the underlying dynamics responsible for binding them. In particular, it offers no fundamental explanation for confinement – why quarks are never observed as free particles [28]. Additionally, while the model incorporates an approximate $\mathrm{SU}(3)$ flavour symmetry, this symmetry is global and phenomenological in nature: it does not explain either its explicit breaking or the dynamical generation of constituent quark masses from more fundamental degrees of freedom. More generally, the quark model lacks a consistent relativistic framework, limiting its predictive power for processes involving high energies or large momentum transfer [29]. In particular, it provides no mechanism for describing interactions in terms of quantum fields and is not formulated as a fully relativistic quantum field theory consistent with principles such as locality and Lorentz invariance. These limitations become especially significant in regimes where relativistic effects and processes such as particle creation and annihilation play an essential role.

Taken together, these shortcomings indicate that a purely symmetry based description is insufficient. What is required instead is a dynamical theory in which the interactions between quarks arise from an underlying principle. The key insight is that promoting internal symmetries to lo-

cal (gauge) symmetries naturally introduces interaction fields and tightly constrains their form. In this framework, the strong force is described by a non-Abelian gauge theory based on SU(3) colour symmetry, in which quarks carry colour charge and interact via the exchange of gluons [30]. This construction not only provides a consistent relativistic quantum field theory, but also leads to qualitatively new phenomena – such as asymptotic freedom at high energies and confinement at low energies – that are absent in the original quark model. This theory is QCD. The formulation of QCD as the gauge theory of the strong force was established following the discovery of asymptotic freedom independently by Gross and Wilczek and by Politzer [31, 32].

QCD is firmly established as the theory that best describes the strong force, with its dynamics encoded in the QCD Lagrangian, and its predictions have been extensively validated in high-energy experiments. In the full quantum theory, physical observables are defined through the QCD path integral, in which expectation values are obtained by summing over all possible quark and gluon field configurations weighted by the action – the spacetime integral of the QCD Lagrangian. This formalism provides a complete, first principles description of the strong force, but in practice it is only analytically tractable in certain limits. In high-energy regimes, ($Q \gg 1$GeV), the strong coupling constant $\alpha_s$ becomes small due to asymptotic freedom, allowing Perturbative Quantum Chromodynamics (pQCD) – organised as expansions in powers of $\alpha_s$ – to produce precise predictions for processes involving large momentum transfer. This framework successfully accounts for the behaviour of quarks and gluons at short distances, enabling accurate calculations for phenomena such as deep inelastic scattering [33].

However, at energy scales characteristic of hadron physics ($Q \sim 1$GeV), the coupling becomes large and pQCD becomes increasingly unreliable. This behaviour is illustrated in Figure 2.3. The theory enters a non-perturbative regime in which key emergent phenomena, including confinement and dynamical chiral symmetry breaking, dominate the dynamics. These effects prevent a straightforward analytic treatment and obscure a direct connection between the fundamental QCD Lagrangian and the observed spectrum of hadrons. Developing a controlled theoretical description of this low-energy domain therefore remains one of the central challenges in hadron physics.

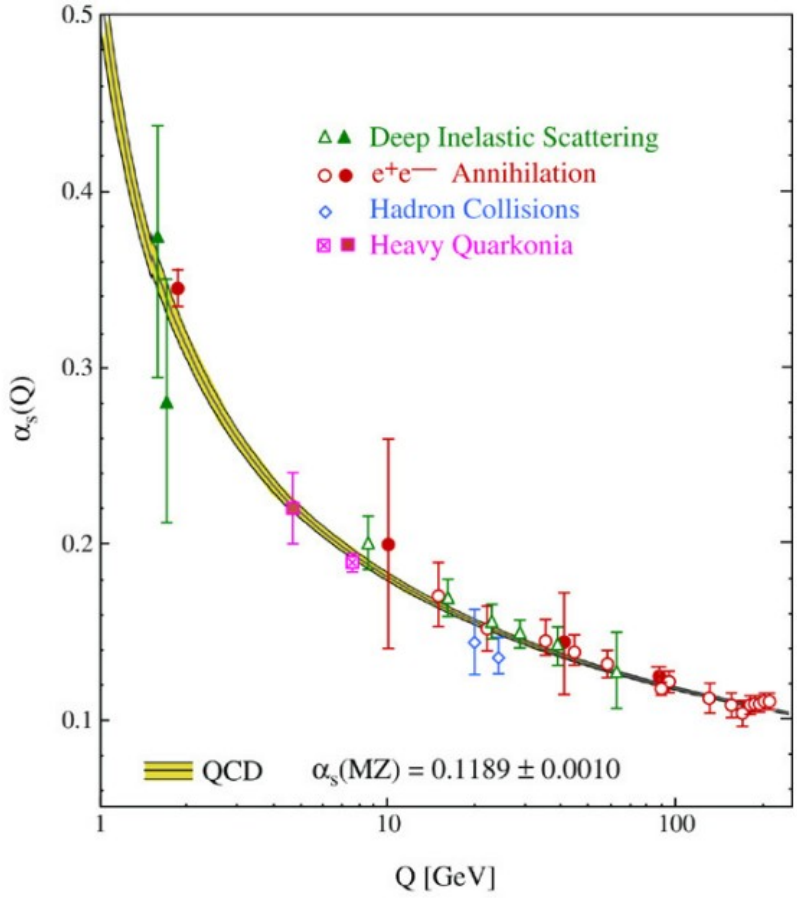



Figure 2.3: Summary of experimental measurements of the QCD coupling constant $\alpha_s(Q)$ as a function of the energy scale $Q$. The measurements, obtained from a range of different processes and experiments, demonstrate the running of the strong coupling, with $\alpha_s$ decreasing as the energy scale increases. The measurements are compared with the QCD prediction, showing good agreement across the range of energy scales probed. The increase of $\alpha_s$ towards lower energies illustrates the transition towards the non-perturbative regime, where perturbative QCD becomes increasingly unreliable. Taken from [34].

This is where non-perturbative methods become essential. Lattice QCD offers a first principles approach by discretising spacetime into a finite grid and evaluating the QCD path integral numerically. The advent of large scale Monte Carlo simulations has enabled computations of hadron observables such as masses, decay constants and form factors directly from the theory. Within lattice QCD, the behaviour of quark fields is governed by the discretised Dirac operator in a background of gluon fields. The eigenvalues and eigenvectors of this operator – often referred to as quark modes – encode important information about the structure of the QCD vacuum. In particular, the low-lying quark modes are closely associated with long distance physics and play a central role in dynamical chiral symmetry breaking. They also contribute significantly to hadron correlation functions, which are used to extract particle spectra in lattice simulations [35]. Figure 2.4 illustrates an example of such a lattice QCD spectroscopy calculation, showing

the extracted excited nucleon spectrum organised according to spin and parity quantum numbers. The pattern of states closely follows the multiplet structure predicted by the $\mathrm{SU}(6) \otimes \mathrm{O}(3)$ quark model, demonstrating that lattice QCD is capable of reproducing the structure of baryon excitations directly from the underlying theory [36].

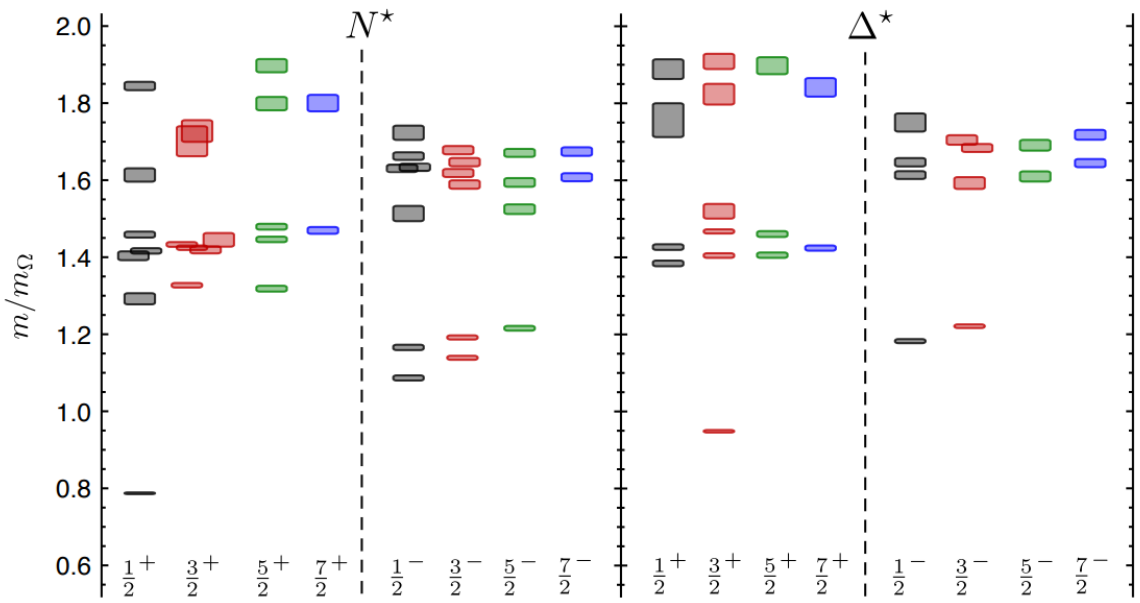



Figure 2.4: Excited nucleon and $\Delta$ baryon spectrum obtained from a lattice QCD calculation, with the extracted states organised according to their spin-parity quantum numbers $J^P$. Each level represents an extracted energy eigenstate, with the vertical position indicating its mass relative to the $\Omega$ baryon mass. The states are grouped according to their corresponding $J^P$ quantum numbers, illustrating the rich spectrum of ground and excited nucleon and $\Delta$ states that can be obtained from lattice QCD. The observed pattern of states can also be compared with the multiplet structure expected from quark model descriptions of baryons. Taken from [36].

In this way, lattice QCD serves as a bridge between the fundamental quark-gluon dynamics of QCD and the effective descriptions provided by the quark model proposed by Gell-Mann and Zweig. It allows tests of how the intuitive picture of hadrons as bound states of constituent quarks emerges from the underlying theory and quantifies the extent to which this picture remains valid in different regimes [37].

Despite this, the results from lattice QCD are obtained under controlled approximations – such as finite lattice spacing, limited volumes and unphysical quark masses – that require careful extrapolation to the physical limit. Moreover, many observables of phenomenological interest remain challenging to access directly within the lattice framework. As a consequence, theoretical predictions must utilise high precision experimental data. In the absence of a fully predictive

theoretical framework at low energies, these measurements provide an important means of accessing hadron structure by constraining phenomenological models, guiding the development of lattice QCD approaches and enabling the extraction of physical quantities that cannot be calculated from first principles alone.

## 2.3 Hadron Spectroscopy

The quark model provides a useful framework for organising the known hadrons in terms of their constituent quarks and quantum numbers. However, the hadron spectrum extends well beyond the lowest lying states considered in the multiplet classifications described in Figures 2.1 and 2.2. Hadrons can be produced in a range of excited states, corresponding to different combinations of quark spin, orbital angular momentum and radial excitation. The systematic study of these states and their properties is known as hadron spectroscopy. By measuring the masses, widths, quantum numbers and decay properties of excited hadrons, spectroscopy provides information about the underlying dynamics responsible for the formation of bound states and resonances [35, 38].

In the quark model description, excited hadrons arise from different configurations of the constituent quarks. For baryons, the three constituent quarks can combine with different total spin and orbital angular momentum configurations, producing states with different total angular momentum and parity, $J^P$. Similarly, mesons can occupy different orbital and radial excitations, giving rise to a spectrum of states with different quantum numbers. In an idealised quark model picture, these states can be organised into multiplets according to the underlying flavour, spin and spatial symmetries. Comparing the predicted spectrum with experimentally observed states therefore provides a means of testing the extent to which the quark model description captures the dynamics of the strong interaction [35].

The interpretation of excited hadrons is complicated by the fact that they are generally unstable. Unlike stable particles, excited hadrons can decay into lighter hadrons through the strong interaction. They are therefore identified through their effects on scattering and production processes rather than being observed as stable particles. A resonance corresponds to an unstable state associated with a pole of the scattering amplitude in the complex energy plane. The position of such a pole provides information about the resonance energy and width, while its couplings to different reaction channels describe how strongly the state interacts with those channels [35,

38]. Resonance properties are therefore extracted indirectly from experimental measurements through the analysis of the underlying reaction amplitudes. In a scattering process, a reaction threshold is the minimum energy required for a particular final state to be produced. The opening of a new reaction channel at its threshold can therefore significantly influence the behaviour of the scattering amplitude.

The $\Lambda(1405)$ provides an important example of this behaviour. It is an excited strange baryon located close to the $\bar{K}N$ threshold and has long been recognised as an unusual state within the hadron spectrum. A simple isolated three quark interpretation does not fully account for its observed properties. Instead, its behaviour is strongly influenced by its coupling to meson-baryon channels, particularly $\bar{K}N$ and $\pi\Sigma$ [35, 39]. The close proximity of the resonance to the $\bar{K}N$ threshold means that the opening of this channel occurs close to the energy of the $\Lambda(1405)$, making its properties particularly sensitive to the coupled-channel dynamics.

The $\Lambda(1405)$ has been studied using a range of theoretical models and experimental data, leading to variation in the extracted pole positions. Figure 2.5 illustrates this spread for the $\Lambda(1405)$ and related resonances. The variation reflects the dependence of the extracted pole positions on the theoretical description and experimental constraints used in each analysis.

Several analyses, including lattice QCD calculations, have provided evidence for a two-pole structure of the $\Lambda(1405)$, with two nearby poles contributing to the observed structure and exhibiting different couplings to the available reaction channels [39–44]. Together, these studies indicate that the observed properties of the $\Lambda(1405)$ are not readily described by a single isolated resonance and are sensitive to the interactions between the relevant reaction channels. The $\Lambda(1405)$ therefore illustrates several of the broader challenges in hadron spectroscopy. The interpretation of an observed resonance depends on the available reaction channels, the precision and kinematic coverage of the experimental measurements, and the theoretical framework used to analyse the corresponding amplitudes. A complete understanding of such states consequently requires information from multiple reactions and observables, together with a consistent treatment of their underlying reaction dynamics.

The observed spectrum does not always correspond straightforwardly to the states expected from a simple constituent quark picture. In particular, the number and properties of experimentally established excited baryons have historically differed from predictions based on quark model calculations. This discrepancy has motivated the study of so-called “missing resonances”, referring to states predicted by theoretical models but not yet clearly identified experimentally [35,

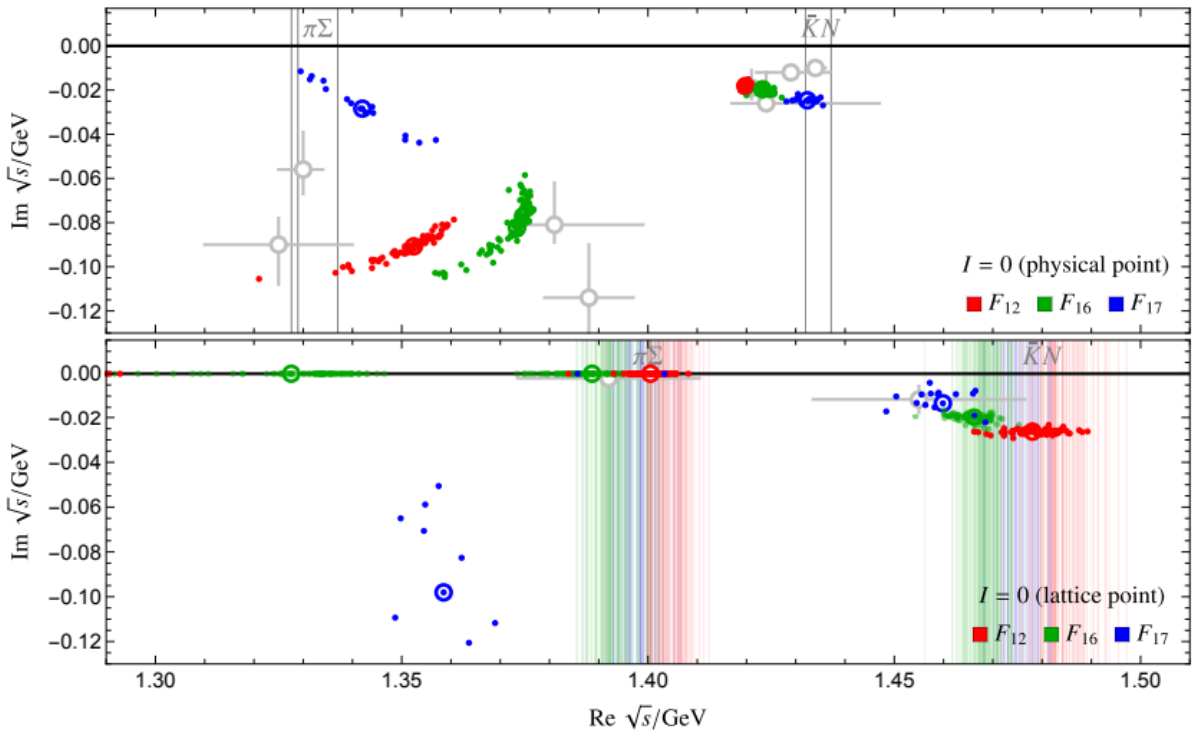



Figure 2.5: Pole positions of the $\Lambda(1405)$ and related resonances in the complex energy plane, showing results obtained from a range of theoretical models and experimental data inputs. The different colours correspond to the different models and analyses considered, illustrating the variation in the extracted pole positions. The real part of the pole position corresponds to the resonance energy, while the imaginary part is related to its width. The spread in the extracted pole positions highlights the sensitivity of the $\Lambda(1405)$ pole structure to the underlying model and data constraints. Taken from [40].

38, 45]. The identification of such states is complicated by their potentially weak couplings to commonly studied reaction channels, overlapping resonances and the presence of substantial non-resonant contributions. Consequently, the absence of a clearly observed structure in a particular reaction does not necessarily imply that the corresponding state does not exist.

The missing resonance problem is also connected to the limitations of simplified descriptions of hadrons as isolated constituent quark states. QCD permits more complicated configurations involving multiple quarks and gluons, and experimental studies have identified candidates whose properties are difficult to accommodate within the conventional quark model classification. Examples include tetraquark states containing four quarks and pentaquark states containing five quarks [46, 47]. A number of such candidates have been observed, particularly in the heavy quark sector, providing evidence that the hadron spectrum may contain states with more complicated internal structures than those represented by the simplest quark-antiquark and three quark configurations. These developments have motivated continued investigation of the role of multi-quark configurations and hadron interactions in the formation of excited states.

More generally, hadron spectroscopy relies on experimental measurements of quantities such as cross sections and polarisation observables to constrain the amplitudes governing hadron reactions. These measurements provide complementary information about the underlying dynamics and can be used to determine the properties of resonances and other structures in the hadron spectrum [35]. The reaction and scattering framework connecting these observables to the underlying interaction is introduced in the following section.

## 2.4 Reaction and Scattering Theory

The properties of excited hadrons cannot generally be measured directly because unstable states decay into lighter particles. Instead, information about their properties is obtained through reactions in which the relevant particles are produced and detected. Scattering and reaction experiments therefore provide a connection between the underlying strong interaction dynamics and experimentally measurable quantities such as cross sections, angular distributions and polarisation observables.

A reaction can be represented schematically as

$$a + b \to c + d, \tag{2.1}$$

where $a$ and $b$ denote the initial state particles and $c$ and $d$ denote the final state particles. The dynamics of the reaction are described by a reaction amplitude, which contains information about the possible interaction mechanisms connecting the initial and final states. Observable quantities are constructed from these amplitudes and therefore provide indirect information about the underlying interaction. For example, a differential cross-section describes the probability of producing a particular final state as a function of the reaction kinematics.

For reactions involving excited hadrons, the reaction amplitude can contain contributions from both resonant and non-resonant processes. Resonant contributions are associated with unstable states that appear as poles of the scattering amplitude, while non-resonant contributions arise from other allowed interaction mechanisms. These contributions can interfere, meaning that the measured observables depend not only on the individual contributions but also on their relative phases. Consequently, the properties of an individual resonance cannot generally be determined from an isolated feature in a single observable. Instead, measurements of multiple observables

across the relevant kinematic region are used to constrain the underlying reaction amplitudes.

A common approach to extracting information about reaction amplitudes is Partial Wave Analysis (PWA), in which the reaction amplitude is decomposed into contributions with different angular momentum and quantum number content. The energy and angular dependence of the measured observables can then be used to constrain these contributions and, subsequently, extract information about the resonances contributing to the reaction. The quality of such an analysis therefore depends on the availability and precision of experimental measurements across the relevant kinematic region [48].

A further complication arises when several reaction channels are accessible at the same energy. In this situation, the channels can be connected through transitions between their respective initial and final states, meaning that the dynamics in one channel can influence observables in another. Such processes are described using a coupled-channel framework, in which the relevant amplitudes are considered simultaneously. This is particularly important for resonances located close to a reaction threshold or strongly coupled to multiple channels. In the strangeness $S = -1$ sector, for example, the $\bar{K}N$ and $\pi\Sigma$ channels play an important role in the description of the $\Lambda(1405)$ [35, 39]. The proximity of the $\Lambda(1405)$ to the $\bar{K}N$ threshold means that its properties are sensitive to the dynamics of these coupled channels. Consequently, measurements of different reactions and final states provide complementary constraints on the underlying amplitudes.

The reaction framework therefore provides the connection between the experimental observables considered in this thesis and the amplitude analyses used to extract information about hadron interactions and resonances. The kinematic variables and coordinate conventions used to describe these reactions are introduced in the following section.

## 2.5 Reaction Kinematics and Coordinate Systems

The kinematics of a reaction describe the energies, momenta and directions of the particles involved. For a two body reaction, these quantities can be described particularly simply in the centre-of-mass frame, where the total three-momentum of the initial state is zero. In this frame, the incoming particles have equal and opposite momenta, while conservation of momentum requires the outgoing particles to have equal and opposite momenta. The kinematic configuration of a two body reaction is therefore determined by the available energy together with the direction

of the outgoing particles.

For the reactions considered in this thesis, the incident energy provides one of the principal variables describing the reaction. In photoproduction, the incident photon energy determines the total energy available to produce the final state particles. The measured observables can therefore be considered as functions of the reaction energy and the angular configuration of the final state. The experimental datasets used in this thesis consist of measurements at discrete points in this kinematic space, with an observable and its associated experimental uncertainty provided at each measured point.

The kinematics of the pseudoscalar meson photoproduction reaction considered in this thesis are illustrated in Figure 2.6. The incoming photon and proton form the initial state, while the outgoing kaon and $\Sigma^+$ form the final state. The direction of the outgoing $\Sigma^+$ is fixed by the two body kinematics and provides the basis for defining the angular variables used in the polarisation measurement. The subsequent weak decay of the $\Sigma^+$ provides access to information about its polarisation [49].

The reaction geometry is defined with respect to the incident photon direction and the direction of the outgoing particles. The plane containing the incident photon and the outgoing particles defines the reaction plane. A coordinate system can then be defined relative to this plane and the direction of the outgoing recoil hyperon. This provides a consistent basis in which the polarisation of the produced hyperon can be described by its components along three orthogonal directions [49].

In addition to the reaction kinematics, the polarisation state of the incident photon provides an additional degree of freedom in photoproduction experiments. For a linearly polarised photon beam, the electric field vector of the photon oscillates preferentially along a particular direction transverse to its direction of propagation. The degree of linear polarisation is denoted by $P_L^\gamma$, while the orientation of the polarisation direction is described by the beam polarisation angle $\alpha$. The angle $\alpha$ specifies the orientation of the photon polarisation relative to the reference direction defined by the reaction geometry.

The angular position of the outgoing hyperon around the incident beam direction is described by the azimuthal angle $\phi$. The relative orientation of the beam polarisation and the recoil direction is therefore described by the combination $\alpha - \phi$. This relative angle is important for the dependence of the measured polarisation observables on the orientation of the linearly polarised

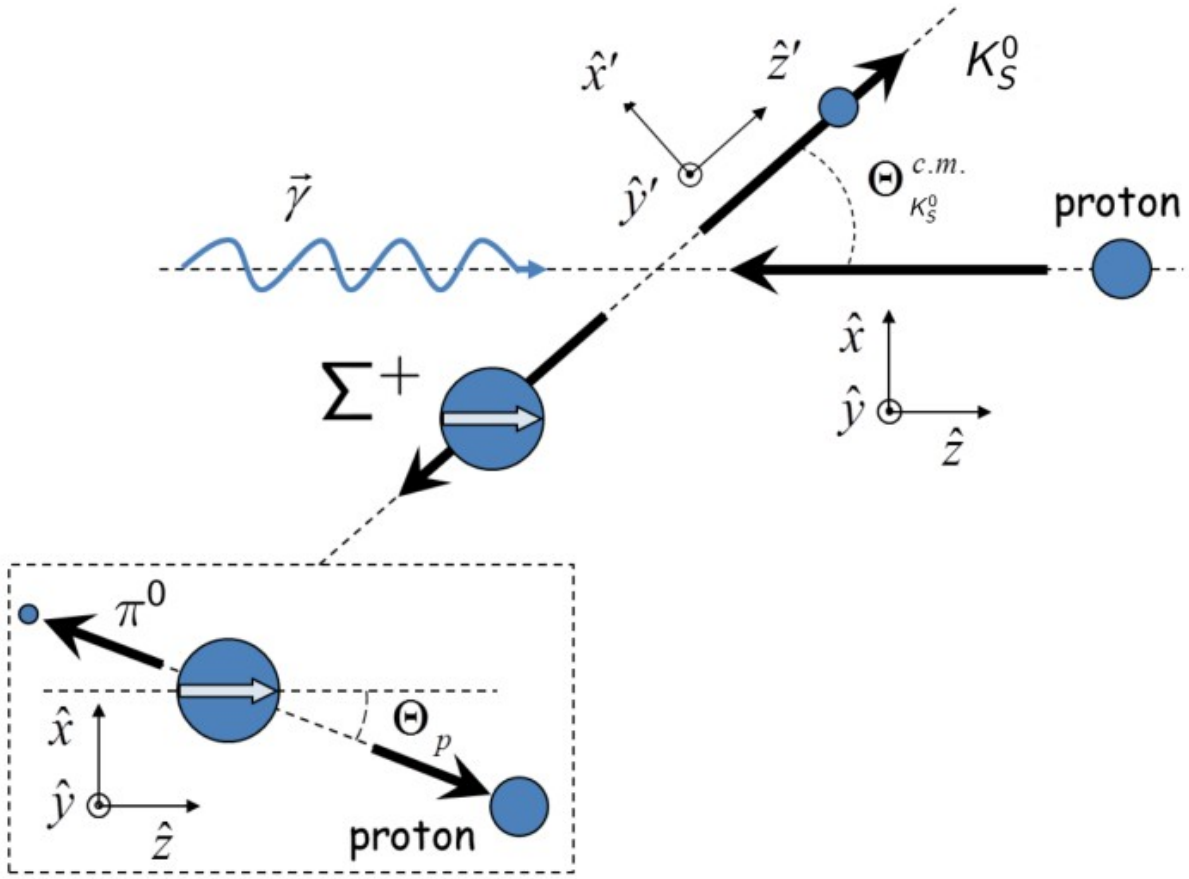



Figure 2.6: Kinematics of the $\gamma p \to K^0\Sigma^+$ reaction. The incoming photon and proton form the initial state, while the outgoing kaon and $\Sigma^+$ form the final state. Taken from [49].

photon beam.

The polarisation of the recoil $\Sigma^+$ can be accessed through its weak decay, $\Sigma^+ \to p\pi^0$. Since the decay is parity violating, the angular distribution of the decay products depends on the polarisation of the parent hyperon. The measured decay distribution can therefore be used to determine the components of the recoil polarisation in the chosen coordinate system. These components are denoted by $P_x^R$, $P_y^R$ and $P_z^R$, with the precise definition of the corresponding axes given by the reaction geometry.

The coordinate system and polarisation components provide the geometrical basis for defining the polarisation observables considered in this thesis. In particular, measurements involving different beam and recoil polarisation configurations provide information beyond that contained in the unpolarised cross section. The resulting observables and their relationship to the underlying reaction amplitudes are discussed in the following section.

The second reaction considered in this thesis is $\bar{K}N$ scattering, in which an antikaon interacts with a nucleon and can produce a number of different final states. Different reaction channels have different thresholds, such that a particular final state becomes accessible only above the corresponding threshold energy. The reaction is therefore characterised by the incident energy together with the final state channel.

The two reactions considered in this thesis consequently provide examples of different kinematic descriptions relevant to hadron physics. Pseudoscalar meson photoproduction involves energy and angular variables, while $\bar{K}N$ scattering can be described in terms of the incident energy for a given reaction channel. These variables provide the basis for describing the observables discussed in the following section.

## 2.6 Polarisation Observables

Polarisation observables provide information about reaction dynamics that cannot be obtained from the unpolarised cross section alone. In pseudoscalar meson photoproduction, the reaction is described by four independent complex amplitudes, subject to the constraints imposed by parity conservation and Lorentz invariance [50]. The observables measured experimentally are constructed from bilinear combinations of these amplitudes. While the unpolarised differential cross section provides information about combinations of their magnitudes, polarisation observables provide additional sensitivity to their relative magnitudes and phases [51]. Measurements of multiple polarisation observables therefore provide complementary constraints on the underlying reaction amplitudes.

Polarisation observables are generally expressed as dimensionless asymmetries that quantify differences in reaction rates for different polarisation configurations. An asymmetry can be defined generically in terms of two event yields, $N_+$ and $N_-$, as

$$\hat{a} = \frac{N_+ - N_-}{N_+ + N_-}, \tag{2.2}$$

where the two yields correspond to the relevant polarisation dependent configurations. This definition gives the asymmetry a physical range of $-1 \leq \hat{a} \leq 1$, with the limiting values corresponding to complete dominance of one configuration. The specific polarisation observables considered below are obtained from particular combinations of beam, target and recoil polarisa-

tion states [51–53].

The polarisation observables considered in this thesis are the photon beam asymmetry $\Sigma$, recoil polarisation $P$, target asymmetry $T$, and the beam-recoil observables $O_x$ and $O_z$. For a linearly polarised photon beam, their contributions to the measured cross-section can be expressed as

$$\frac{d\sigma}{d\Omega} = 1 + PP_y^R - \left(\Sigma + TP_y^R\right) P_L^\gamma \cos 2(\alpha - \phi) + \left(O_x P_x^R + O_z P_z^R\right) P_L^\gamma \sin 2(\alpha - \phi), \tag{2.3}$$

where the quantities describing the reaction geometry and recoil polarisation are defined in Section 2.5 [54]. The different terms in this expression correspond to different combinations of beam, target and recoil polarisation information and therefore provide distinct constraints on the reaction amplitudes.

The photon beam asymmetry, $\Sigma$, is one of the simplest polarisation observables. It measures the dependence of the cross-section on the linear polarisation of the incoming photon, specifically the difference in production rates when the photon polarisation is parallel versus perpendicular to the reaction plane defined by the photon and kaon momenta [55]. $\Sigma$ is also sensitive to the relative contributions of natural and unnatural parity exchange mechanisms. Values of $\Sigma$ near $+1(-1)$ indicate dominance of natural (unnatural) parity exchange, while intermediate values show a mixture of both contributions. Because it can be measured with a polarised photon beam and an unpolarised target, $\Sigma$ is often one of the first polarisation observables extracted in experimental programs.

Another key observable is the recoil polarisation, $P$, which describes the polarisation of the $\Sigma^+$ hyperon along the axis perpendicular to the reaction plane. $P$ can even be non-zero when the incoming photon and the target proton are unpolarised, as it arises purely from the interference among the complex amplitudes in the reaction. The measurement of $P$ relies on the asymmetry in the weak decay of the $\Sigma^+$, allowing a direct probe of the spin transfer from the initial system to the hyperon [56]. Because $P$ is sensitive to the imaginary parts of amplitude products, it provides complementary information to $\Sigma$, which primarily probes real interference terms.

The target asymmetry, $T$, characterises the sensitivity of the cross-section to transverse polarisation effects and encodes interference patterns that are distinct from those accessible through $P$ or $\Sigma$, providing an additional constraint on the underlying amplitudes. Together, the single polarisation observables $\Sigma$, $P$ and $T$ allow a partial characterisation of the reaction amplitude, probing the interference between different spin-flip and non-spin-flip components and offering insight into the underlying resonance contributions. Although the target is unpolarised, $T$ can

still be accessed through its coupling to the recoil polarisation in the $\cos 2(\alpha - \phi)$ term of the cross-section. The self-analysing weak decay of the $\Sigma^+$ allows the recoil polarisation component $P_y^R$ to be determined from the angular distribution of the decay products which, in combination with the known beam polarisation, enables the extraction of $T$ without the need for a polarised target [49].

Double polarisation observables further increase the information available from this reaction. In particular, the beam recoil observables $O_x$ and $O_z$ measure the correlation between the linear polarisation of the incoming photon and the polarisation of the recoiling hyperon along the $x$ and $z$ axes respectively. These observables are especially valuable because they provide access to the relative phases between amplitudes that are not probed by single polarisation observables. For example, $O_x$ is sensitive to interference between amplitudes contributing to transverse hyperon polarisation, while $O_z$ probes correlations along the photon momentum direction.

Taken together, the observables $\Sigma$, $P$, $T$, $O_x$ and $O_z$ probe different combinations of the underlying complex amplitudes. Their simultaneous measurement therefore provides substantially more information about the reaction than the unpolarised cross section alone. In particular, the sensitivity of polarisation observables to different interference terms helps distinguish amplitude solutions that produce similar cross sections but different polarisation behaviour. This reduces ambiguities in amplitude analyses and provides stronger constraints on partial wave decompositions.

The observables are also sensitive to interference between different reaction mechanisms. Contributions from processes such as $s$-channel nucleon resonances, $t$-channel meson exchange and $u$-channel hyperon exchange can interfere constructively or destructively, producing characteristic energy and angular dependences in the measured observables [57]. Consequently, measurements of $\Sigma$, $P$, $T$, $O_x$ and $O_z$ can be used to constrain models of the reaction dynamics and investigate contributions from excited hadron states [50].

## 2.7 Structure and Limitations of Experimental Data

The theoretical descriptions discussed in the preceding sections are ultimately constrained by experimental measurements. An experiment does not provide continuous knowledge of an observable across all possible kinematic configurations; instead, measurements are obtained over

a finite kinematic range and are typically reported at discrete points or within finite bins. The accessible region is further restricted by the experimental acceptance and by the resolution with which the relevant kinematic quantities can be reconstructed. Experimental datasets therefore provide a finite sampling of the underlying physical dependence rather than a continuous description of the observable.

The available measurements also span several generations of experimental facilities. Modern experiments can provide measurements over substantially larger kinematic ranges and with greater statistical precision than many earlier experiments. Earlier measurements, however, remain an important component of modern hadron physics datasets. Many experiments from the 1950s through to the 1970s relied on detectors such as bubble chambers (see Figure 2.7), in which particle trajectories were recorded photographically. Although these experiments operated at comparatively low event rates and required photographic images to be processed to extract the relevant kinematic information, they provided measurements that remain useful for constraining regions of kinematic space and reaction channels that may have limited coverage from more recent experiments [58]. Some historical experiments also used beam types, target configurations or reaction conditions that are no longer readily available, meaning that their measurements cannot necessarily be reproduced at contemporary facilities. Measurements obtained using different experimental technologies must therefore often be considered together when constructing a world dataset [35, 59].

An important consequence of the experimental measurement process is that the reconstructed kinematic quantities have finite precision. Momentum, angular and energy resolutions introduce uncertainties into the determination of the kinematic configuration of each event. These effects can also influence the measured value of an observable, for example through events being reconstructed in neighbouring bins or through a broadening of the measured distribution. Experimental analyses must therefore account for detector effects when extracting physical observables from measured event distributions. The use of finite binning introduces an additional trade-off: finer bins provide greater resolution of the kinematic dependence but contain fewer events and can therefore have larger statistical uncertainties, while wider bins improve statistical precision at the expense of kinematic resolution.

The resulting measurements generally provide a sparse and non-uniform sampling of the available kinematic space. Measurements tend to be concentrated in regions where the experimental acceptance, event yield and analysis sensitivity are sufficient to obtain reliable results, while other regions may contain relatively few measurements or no direct measurements at all. For a photoproduction reaction, for example, an observable may be measured at a finite set of centre-

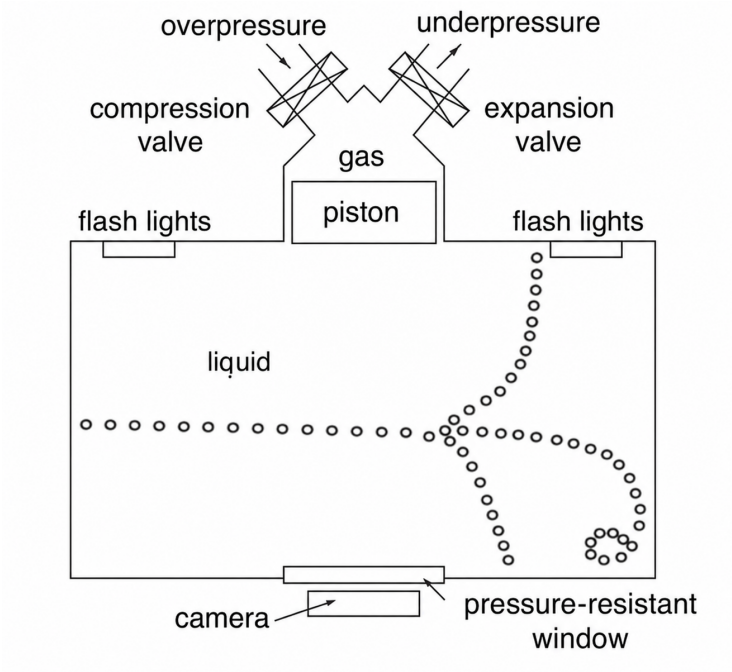



Figure 2.7: Schematic construction of a bubble chamber, illustrating the recording of charged particle trajectories as bubble tracks within a superheated liquid. Bubble chambers were widely used in early hadron physics experiments, and measurements obtained from these experiments continue to form part of modern world datasets. Taken from [60].

of-mass energies and scattering angles rather than as a continuous function of these variables. The distribution of measurements across this space therefore contains information not only about the observable itself, but also about the regions in which experimental constraints are available [35].

This incomplete coverage is particularly relevant for reaction amplitude and spectroscopy analyses. The energy and angular dependence of measured observables provides constraints on the reaction amplitudes and on the contributions of different reaction mechanisms and excited hadron states. Where measurements are widely separated or absent over substantial regions of kinematic space, however, the available data provide less direct information about the behaviour of the underlying reaction. Different descriptions can consequently provide comparable agreement with the measured points while predicting different behaviour in regions where no measurements are available. This makes the distribution of the experimental measurements an important consideration when interpreting reaction amplitude analyses and extracting information about the hadron spectrum [35, 59].

The experimental data are also heterogeneous. Measurements of the same or related observables may originate from different experiments and experimental configurations, each with its

own kinematic coverage, binning scheme, detector response and uncertainty structure. Combining such measurements can provide a more complete description of a reaction, but the resulting dataset may contain measurements with substantially different densities and uncertainty scales. In addition, measurements are associated with both statistical and systematic uncertainties. Statistical uncertainties arise from the finite number of events contributing to a measurement, while systematic uncertainties can originate from sources including detector calibration, acceptance corrections, particle identification, luminosity and beam properties. Systematic effects may also introduce correlations between measurements, meaning that their uncertainties cannot always be treated as independent. This is discussed in more detail in the following subsection.

Consequently, experimental hadron physics datasets represent finite, sparse and heterogeneous samples of the underlying physical processes. The available measurements provide constraints only over limited regions of kinematic space and are subject to both statistical and systematic uncertainties. These characteristics must be taken into account when using experimental observables to constrain reaction amplitudes and study the hadron spectrum, since the physical dependence between measured points and in regions without direct measurements must ultimately be inferred from the available experimental information.

### 2.7.1 Combining Experimental Datasets

An important consideration in the construction of world datasets is how measurements from different facilities should be combined. In practice, experiments may be weighted according to their quoted statistical uncertainties, estimated systematic uncertainties or other assessments of experimental quality. However, quoted uncertainties are not always directly comparable between experiments, particularly when systematic effects are treated differently or are incompletely reported. As a result, the relative influence assigned to different datasets in a global fit can depend on assumptions concerning the treatment of uncertainties and the quality of the available measurements. This issue is encountered in large scale partial wave analyses such as those performed by SAID [61] and MAID [62], where extensive datasets from different facilities must be reconciled within a single framework. While such procedures are necessary for practical global analyses, the resulting world dataset therefore reflects not only the underlying experimental measurements but also the statistical and methodological assumptions used to combine them.

Normalisation shifts are a frequently used approach for addressing differences between exper-

imental datasets [63]. In a global fit, each experimental dataset may be assigned an overall scale factor that is optimised simultaneously with the physical model parameters. This allows relative normalisation adjustments between experiments and can improve the agreement between datasets. Such adjustments can be appropriate when the dominant discrepancy between experiments is associated with an overall normalisation uncertainty. However, a normalisation factor cannot in general account for differences in the energy dependence, local structure or uncertainty behaviour of two datasets. If the underlying measurements contain more complex discrepancies, allowing additional normalisation parameters may reduce the apparent disagreement without establishing whether the datasets are statistically compatible. The treatment of normalisation uncertainties must therefore be considered alongside the other sources of experimental uncertainty.

From a Bayesian standpoint, hierarchical modelling provides a framework in which differences between experiments can be represented explicitly. In such approaches, quantities such as relative normalisation factors or uncertainty rescaling parameters can be treated as additional parameters with assigned prior distributions, allowing uncertainty associated with the relationship between datasets to be propagated through the subsequent analysis [64]. This provides a principled mechanism for representing uncertainty in the combination of experiments, but the resulting inference remains dependent on the assumed hierarchical structure and prior distributions. The treatment of correlations and the choice of which experimental effects should be modelled therefore remain important considerations.

Global averaging procedures, such as those reviewed by the Particle Data Group [38], similarly encounter the problem of combining measurements that are not fully consistent. Although systematic criteria can be applied to assess dataset quality and consistency, the resulting averages may require decisions concerning outlier treatment, uncertainty inflation and correlations between measurements. Such procedures are useful when a single representative value is required, but reducing several measurements to a single estimate can also obscure the structure of any disagreement between the constituent datasets.

These considerations illustrate a general limitation of representing heterogeneous experimental measurements by a single effective dataset. When measurements from different experiments are statistically compatible, their combination can provide a more precise and comprehensive constraint on the underlying physical observable. When substantial differences are present, however, the resulting representation depends on how those differences are treated and may conceal information about the disagreement between the original measurements. This is particularly relevant when the combined dataset is subsequently used to constrain an interpolation or

theoretical model. A reliable treatment of experimental data must therefore distinguish between uncertainty arising from limited measurements of an underlying physical quantity and variation between measurements that may indicate unresolved experimental inconsistencies.

For hadron physics reactions, this issue is particularly important because historical measurements from different experiments may provide the only available constraints over significant regions of the kinematic space. Discarding measurements that disagree requires an external judgement about which measurements should be retained, while forcing all measurements into a single representation can suppress potentially important information about their differences. Retaining information about the individual measurements and their associated uncertainties can therefore provide a more transparent representation of the available experimental evidence.

## 2.8 From Experimental Measurements to Physics Interpretation

The preceding sections have introduced the theoretical framework of hadron physics, the description of reactions and scattering, and the nature of the experimental measurements used to investigate these processes. These concepts can be brought together by considering the general pathway from an experimental measurement to the interpretation of the underlying physics.

Although the precise procedure depends on the reaction and experimental technique, the overall process can be broadly summarised as:

1. **Experimental measurement** – particles produced in a reaction are detected and their measurable properties are recorded.
2. **Reaction reconstruction** – the measured quantities are used to reconstruct the kinematics and identify the reaction being studied.
3. **Determination of physical observables** – reconstructed events are used to determine quantities such as cross sections and polarisation observables.
4. **Reaction amplitude analysis** – the measured observables are used to constrain the amplitudes describing the underlying reaction dynamics.

5. **Resonance and spectroscopy analysis** – the resulting amplitudes are analysed to determine the properties and contributions of excited hadrons.

6. **Physics interpretation** – the extracted information is compared with theoretical descriptions of hadrons and the strong interaction.

### 2.8.1 Experimental measurement

A hadron physics experiment begins with an interaction between an initial state and a target, producing a set of final state particles. These particles are detected using a combination of detector subsystems, each designed to measure different properties of the particles. The CLAS12 detector at Jefferson Laboratory (shown in Figure 2.8) provides a representative example of a modern large acceptance spectrometer used to study electron and photon induced reactions and combines several complementary detector systems. The principal subsystems relevant to the measurement and reconstruction of the final state particles are [65]:

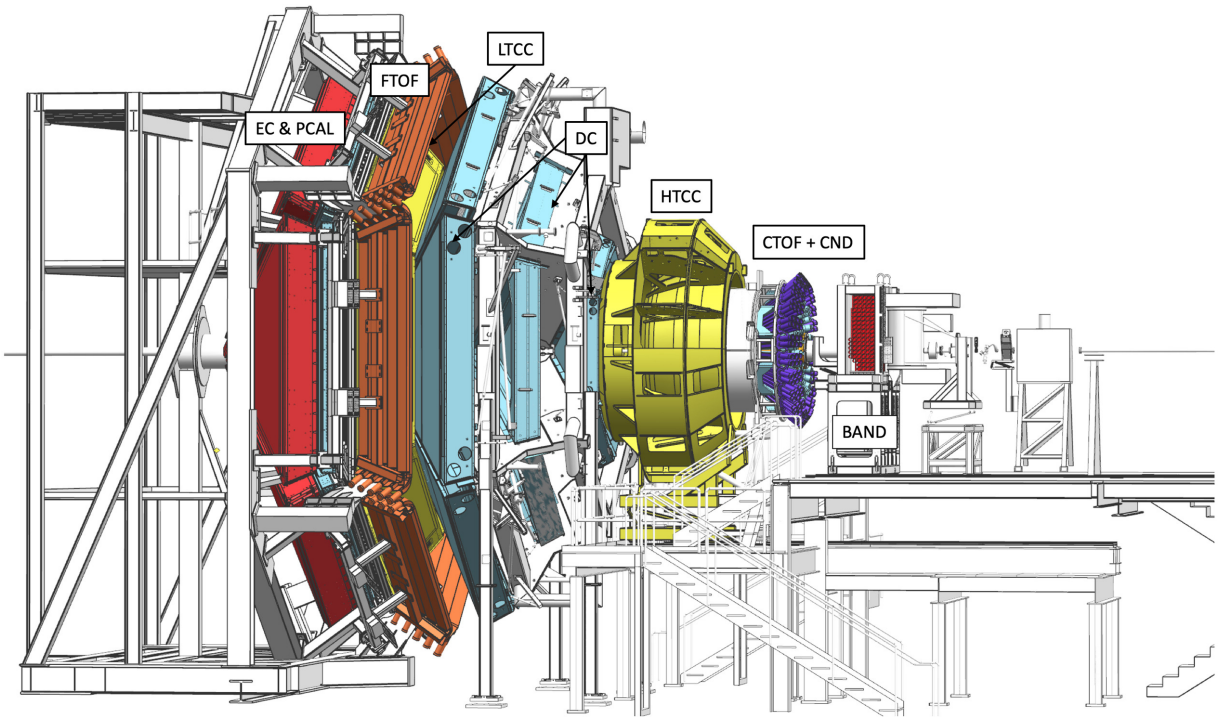



Figure 2.8: Schematic overview of the CLAS12 detector at Jefferson Laboratory, showing the principal detector subsystems used for particle tracking, energy measurement and particle identification. Taken from [65].

- **Electromagnetic Calorimeter (EC) and Pre-shower Calorimeter (PCAL):** measure the energy deposited by electromagnetic showers produced by electrons and photons.

PCAL provides finer segmentation and acts as a pre-shower detector, improving the reconstruction of electromagnetic particles and aiding particle identification.

- **Forward Time-of-Flight (FTOF):** measures the time taken by charged particles to travel from the target to the forward detector. Combined with the reconstructed momentum and path length, the measured time-of-flight provides information for particle identification, including the separation of pions and kaons over a substantial momentum range.
- **Low-Threshold Cherenkov Counter (LTCC):** detects Cherenkov radiation produced by sufficiently relativistic charged particles. Its response provides additional particle identification information, particularly for distinguishing pions from kaons at higher momenta.
- **Drift Chambers (DC):** provide tracking information for charged particles in the forward detector. The positions of ionisation signals in the gas-filled chambers are used to reconstruct particle trajectories, which can then be combined with the magnetic field to determine their momenta.
- **High-Threshold Cherenkov Counter (HTCC):** provides particle identification information for highly relativistic charged particles through the detection of Cherenkov radiation. In particular, it provides an important means of identifying scattered electrons in CLAS12.
- **Central Time-of-Flight (CTOF) and Central Neutron Detector (CND):** provide time-of-flight and neutral particle detection capabilities in the central detector region. CTOF measures the timing of charged particles and contributes to their identification, while the CND provides detection of neutrons that do not leave tracks in the charged particle tracking system.
- **Backward Angle Neutron Detector (BAND):** detects neutrons emitted at very backward angles. Its scintillator bars provide timing and energy deposition information that can be used to identify and reconstruct backward-going neutrons.

These detector systems provide complementary measurements of the particles produced in the reaction. Tracking detectors determine the trajectories of charged particles, calorimeters measure energy deposition, and timing and Cherenkov detectors provide information for particle identification. Combining information from multiple subsystems therefore allows the properties of individual final state particles to be reconstructed.

### 2.8.2 Reaction reconstruction

The information recorded by the detector is used to reconstruct the kinematic properties of each event. Measurements from the different detector subsystems are combined to determine quantities such as particle momenta, energies and directions, while the response of the detector is used to identify the particles involved in the reaction. Conservation laws and the known properties of the reaction provide additional constraints on the reconstruction.

The reconstructed events can then be selected according to the reaction of interest. Depending on the experimental setup and reaction, this can involve identifying the relevant final state particles and applying kinematic or event selection criteria. The resulting event sample provides the basis for determining the physical observables associated with the reaction.

### 2.8.3 Determination of physical observables

The reconstructed events are used to determine experimentally accessible physical observables. Depending on the reaction, these can include total and differential cross sections, angular distributions and polarisation observables. As discussed in Section 2.6, polarisation observables provide information about the reaction dynamics that is complementary to that obtained from unpolarised cross sections.

The resulting measurements provide experimental constraints on the underlying reaction. However, these constraints are available only over the kinematic coverage of the experiment and are associated with the statistical and systematic uncertainties discussed in Section 2.7.

### 2.8.4 Reaction amplitude analysis

The measured observables are related to the underlying reaction amplitudes that describe the dynamics connecting the initial and final states. As discussed in Section 2.4, these amplitudes can contain contributions from different reaction mechanisms, including resonant and non-resonant processes. Since these contributions can interfere, measurements of multiple observables and

reaction channels can provide complementary constraints on the amplitudes.

The reaction amplitudes can be constrained by fitting theoretical or phenomenological models to the measured observables. Such fits determine the parameters or coefficients of the reaction model that provide the best description of the available experimental data. The quality of the resulting description therefore depends on both the experimental measurements and their uncertainties, as well as on the kinematic coverage and range of observables available.

Partial wave and coupled-channel analyses [66–68] provide approaches for using experimental measurements to constrain the energy and angular dependence of reaction amplitudes. In a partial wave analysis, the reaction amplitude is decomposed into contributions with different angular momentum and quantum number content, with the corresponding coefficients constrained by fits to the experimental observables. The resulting amplitudes provide a description of the reaction that can subsequently be used to investigate the properties of the contributing hadron states.

### 2.8.5 Resonance and spectroscopy analysis

The constrained reaction amplitudes can then be analysed to investigate excited hadrons and their properties. Resonances are associated with poles of the scattering amplitude, whose positions and couplings provide information about their energies, widths and interactions with different reaction channels.

The interpretation of these states can become particularly challenging when several reaction channels are coupled or when resonances occur close to reaction thresholds. Measurements from multiple reactions and observables can therefore provide complementary constraints on the same underlying physics.

### 2.8.6 Physics interpretation

The extracted information can finally be considered in the context of theoretical descriptions of hadrons and the strong interaction. The observed properties of excited states can be compared

with predictions from approaches such as constituent quark models, lattice QCD and reaction theory calculations.

Together, these stages establish the connection between experimental measurements and the interpretation of hadron interactions and the hadron spectrum. Experimental measurements provide constraints on the physical process, reaction theory provides the framework for relating those measurements to the underlying amplitudes, and spectroscopy provides the means of interpreting those amplitudes in terms of the properties of hadron states.

# Chapter 3

# Statistical and Data-Science Foundations

## 3.1 Probability Theory

Probability theory provides the mathematical framework for quantifying uncertainty and describing the statistical behaviour of experimental measurements. In scientific applications, probability provides a means of representing uncertainty arising from factors such as finite measurement resolution, statistical fluctuations, experimental noise and limited data. This provides the foundation for the statistical inference and uncertainty quantification methods used throughout this thesis.

Experimental measurements can be modelled as random variables, representing quantities whose observed values are subject to statistical fluctuations and measurement uncertainty. A random variable may be discrete or continuous and is characterised by an associated probability distribution. For a continuous random variable $X$, this distribution is described by a Probability Density Function (PDF), $p(X)$, where the integral over an interval gives the probability that the variable takes a value within that interval. The probability distribution therefore provides a complete statistical description of the possible outcomes of a measurement and allows quantities such as the mean and variance to be calculated.

The expectation value of a random variable provides its mean or average value,

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} X p(X)\,\mathrm{d}X, \tag{3.1}$$

while the variance quantifies the spread of the distribution about its mean,

$$\mathrm{Var}(X) = \mathbb{E}\left[(X - \mathbb{E}[X])^2\right]. \tag{3.2}$$

The standard deviation is given by the square root of the variance and provides a commonly used measure of uncertainty in experimental measurements.

In many physical applications, measurements are well approximated by Gaussian distributions due to the central limit theorem [69], particularly when uncertainties arise from the accumulation of many small, approximately independent effects. The Gaussian distribution is therefore frequently used to represent measurement uncertainties. Other probability distributions are also important in physics; for example, Poisson distributions are commonly used to describe counting experiments, where the measured quantity is the discrete number of events observed within a given interval [70]. The Poisson distribution is particularly appropriate when events occur independently with a well-defined expected rate, and its variance is equal to its mean. For sufficiently large event counts, the Poisson distribution approaches a Gaussian distribution, while for small event counts its discrete nature and asymmetry can become important. The appropriate probability distribution therefore depends on the underlying measurement process and the sources of uncertainty being considered.

When multiple quantities are measured simultaneously, a multivariate probability distribution is required. Such distributions allow correlations between measurements to be represented in addition to the uncertainty associated with each individual observable. These correlations can be particularly important in scattering experiments, where kinematic constraints, common systematic uncertainties and shared experimental conditions can lead to non-trivial covariance structures.

For two random variables $X$ and $Y$, conditional probability describes the distribution of one variable given knowledge of the other. This is written as

$$p(X|Y), \tag{3.3}$$

and describes the probability distribution of $X$ when the value of $Y$ is known. Conditional probability provides the basis for statistical inference, where information about observed quantities is used to constrain unknown quantities.

The concepts introduced here provide the probabilistic foundation for the Bayesian inference, regression and uncertainty quantification methods developed in the following sections. In particular, representing experimental measurements as probability distributions allows both the uncertainties in the available data and the uncertainties in subsequent predictions to be treated within a consistent statistical framework.

## 3.2 Bayesian Inference

Bayesian inference provides a framework for updating knowledge about unknown quantities as new experimental data become available. In contrast to a purely probabilistic description of measurements, Bayesian inference provides a mechanism for combining prior information with evidence from observed data to obtain a posterior distribution for the quantities of interest. This is particularly useful in scientific applications where available data may be sparse or subject to substantial uncertainty.

For model parameters $\theta$ and observed data $D$, Bayes' theorem [71] relates the posterior probability distribution to the likelihood and prior distribution, and is often written in proportional form as

$$p(\theta|D) \propto p(D|\theta)p(\theta). \tag{3.4}$$

The likelihood describes how probable the observed data are for a given set of model parameters. It is determined by the assumed statistical model for the measurements and therefore depends on the experimental uncertainties and correlations. For example, when measurements are assumed to have Gaussian uncertainties, the likelihood can be written in terms of the residual between the model prediction and the observed data, weighted by the experimental covariance matrix. For a vector of observations $y$ and model predictions $f(\theta)$, this takes the form

$$p(y|\theta) \propto \exp\left[-\frac{1}{2}\left(y - f(\theta)\right)^T C^{-1}\left(y - f(\theta)\right)\right], \tag{3.5}$$

where $C$ is the covariance matrix describing the experimental uncertainties and correlations [72].

The prior distribution represents information about the model parameters before considering the observed data. In physics applications, prior information may arise from theoretical consid-

erations, previous measurements or physical constraints. For example, a prior may restrict a parameter to a physically meaningful range or encode an expected scale for a quantity. The use of a prior is particularly relevant when data are sparse, since the available measurements may not provide sufficient information to constrain all model parameters independently. However, the choice of prior should be justified carefully, as it can introduce model dependence into the inference to varying degrees. A relatively weak prior may impose only general physical constraints, while a more informative prior may incorporate assumptions about the expected behaviour of the system. In particular, assuming a specific functional form for a model, or imposing conditions on the expected posterior distribution, can substantially influence the resulting inference. The balance between incorporating prior knowledge and minimising unnecessary model dependence is therefore an important consideration when performing inference from sparse experimental data.

The experimental setup primarily enters the likelihood through the measurement uncertainties, correlations and detector response rather than directly determining the prior distribution. However, knowledge of the experimental conditions can indirectly motivate prior assumptions when it provides relevant physical constraints on the quantities being inferred. It is therefore important to distinguish prior information about the underlying physical model from the statistical uncertainty associated with the measurement process.

The posterior distribution combines the prior information with the information provided by the experimental data. Unlike a single best fit estimate, it provides a probability distribution over the allowed values of the parameters, allowing both parameter estimates and their associated uncertainties to be obtained. Predictions for new observations can similarly be obtained by propagating uncertainty in the model parameters through the posterior distribution.

### 3.2.1 Bayesian and Frequentist Inference

Bayesian and frequentist statistics provide two different frameworks for interpreting probability and performing statistical inference. In the frequentist approach, model parameters are treated as fixed but unknown quantities, while probability is associated with the repeated outcomes of an experiment. Inference is therefore typically based on quantities such as maximum likelihood estimates, confidence intervals and hypothesis tests, whose interpretation is defined in terms of repeated sampling from the underlying statistical model.

In the Bayesian approach, probability is instead used to represent uncertainty about unknown quantities. Model parameters are therefore described by probability distributions, with the posterior distribution representing the updated state of knowledge after observing the data. This allows prior information and experimental evidence to be incorporated within a single framework [64].

The distinction is particularly important when interpreting uncertainty intervals. A frequentist confidence interval is constructed such that, under repeated experiments, a specified proportion of intervals generated by the procedure would contain the true parameter value. It does not strictly assign a probability to the parameter itself. A Bayesian credible interval, in contrast, is obtained directly from the posterior distribution and can be interpreted as containing a specified probability mass of the parameter given the observed data and the assumed prior.

Neither framework is universally superior, and both provide useful approaches to statistical inference. The Bayesian formulation is particularly appropriate for the present work because the objective is not only to estimate a function from experimental measurements, but also to propagate uncertainty and quantify the probability of possible function values throughout the kinematic region. This becomes especially important when experimental measurements are sparse and unevenly distributed, since prior assumptions and the statistical structure of the model can provide information between measured points.

## 3.3 Regression

Regression provides a framework for learning functional relationships between measured observables and underlying input variables. In hadron physics, this often involves reconstructing a continuous function – such as a cross-section or polarisation observable – from a finite and potentially sparse set of experimental measurements. Formally, if the regression problem follows Gaussian statistics then it may be expressed as modelling observations $y_i$ obtained at inputs $x_i$ according to

$$y_i = f(x_i) + \varepsilon_i, \tag{3.6}$$

where $f(x)$ represents the underlying latent physical function and $\varepsilon_i$ represents measurement noise and other experimental uncertainties.

A key distinction in regression methodology is that between deterministic and probabilistic approaches [73]. In a deterministic approach, the objective is to identify a single function that provides an estimate of the relationship between the inputs and observations. For example, a parametric model may assume a functional form $f(x;\theta)$ characterised by a finite set of parameters $\theta$, which are determined by minimising an appropriate loss function. A least-squares objective is commonly used when the measurement uncertainties are assumed to be Gaussian, while a maximum likelihood approach can incorporate a more general model for the measurement uncertainties. The resulting model provides a single preferred estimate of the underlying function. Although uncertainties on the fitted parameters can subsequently be estimated and propagated to the fitted function, the regression remains centred on a single best fit model rather than explicitly treating the function itself as a random quantity.

In sparse data regimes, such an approach can be restrictive. Multiple functional behaviours may be statistically consistent with the available measurements, particularly in regions where data are absent or experimental uncertainties are large. A deterministic fit must nevertheless select one function to represent the relationship between the measured points. Consequently, uncertainty associated with the choice of functional behaviour may not be fully represented by uncertainties obtained from the fitted parameters alone. This distinction becomes increasingly important when predictions are required away from measured points.

Probabilistic regression addresses this issue by assigning a probability distribution to the possible functional relationships between the inputs and observations. Rather than identifying only a single preferred function, the regression model describes a distribution over functions that is conditioned on the available data and any prior assumptions. Predictions at new input values are therefore probability distributions rather than single values, allowing both the expected prediction and its associated uncertainty to be quantified. In hadron physics, Markov Chain Monte Carlo (MCMC) methods [74] provide an established and widely used methodology for sampling from posterior distributions in probabilistic analyses. MCMC methods generate samples from the posterior distribution of the model parameters, allowing the corresponding range of physically plausible predictions to be explored.

A MCMC method generates samples from a probability distribution by constructing a Markov chain whose stationary distribution is the target distribution. The defining property of a Markov chain is that the probability of transitioning to the next state depends only on the current state, rather than the full history of the chain. By generating a sufficiently long sequence of such states, the distribution of the samples approaches the desired posterior distribution, allowing quantities such as parameter uncertainties and correlations to be estimated without explicitly evaluating

the normalisation of the posterior. Importantly, the resulting posterior distribution need not be Gaussian. The samples therefore contain information about the full shape of the distribution, including asymmetry, correlations and multiple modes, rather than only its mean and variance.

The distinction between deterministic and probabilistic regression is separate from the distinction between parametric and non-parametric approaches. Parametric regression assumes that the underlying function can be described using a finite number of parameters, whereas more flexible or non-parametric approaches do not require a fixed functional form with a predetermined number of parameters. A probabilistic formulation can be applied to either type of model. This distinction is particularly relevant for sparse experimental data, where imposing a specific functional form may introduce substantial model dependence.

All regression methods rely on assumptions about the relationship between the observations and the underlying function. These include assumptions about the noise structure – such as independence, Gaussian-like behaviour or the presence of correlated systematic effects – as well as assumptions about the behaviour of the underlying function. In parametric models, these assumptions may be encoded explicitly through the chosen functional form. More flexible approaches can instead encode structural assumptions such as smoothness, continuity or characteristic length scales.

In the context of hadron reaction analysis[1], such assumptions are not merely technical choices but can incorporate physical knowledge. Constraints arising from kinematics, symmetry principles, positivity conditions and known asymptotic behaviour can restrict the space of physically admissible functions. A well-constructed regression framework must therefore balance flexibility with physically motivated constraints or regularisation. This balance is particularly important in sparse regimes, where the available data may be insufficient to determine the functional behaviour uniquely and different modelling assumptions can lead to different predictions in poorly constrained regions.

Probabilistic function-space methods provide a natural framework for addressing this ambiguity. By placing a probability distribution over possible functions, they allow uncertainty associated with both the observations and the inferred functional behaviour to be represented explicitly. Methods that model correlations between predictions at different input points are particularly useful in this context, since information from measured regions can be propagated to nearby

[1] Here hadron reaction is taken to mean any type of physics reaction involving hadrons, e.g. scattering, decays, production, etc.

unmeasured regions in a controlled manner. Gaussian Processes provide one such approach and are introduced in Section 3.5.

## 3.4 Uncertainty Quantification

A central objective of probabilistic modelling is not only to provide accurate mean predictions, but to characterise uncertainty in a manner that is both statistically consistent and physically interpretable. In hadron physics – where datasets may be sparse and systematics correlated – uncertainty quantification plays a defining role in determining the reliability of theoretical conclusions.

It is useful to distinguish between uncertainty associated with observed data and uncertainty associated with model predictions [75]. Observable uncertainty refers to the statistical and systematic errors attached to experimental measurements. These may arise from counting statistics, detector resolution, acceptance corrections, normalisation uncertainties or correlated systematic effects. Such uncertainties are typically provided in the form of covariance matrices that encode both variances and inter-point correlations. They describe variability in measured observables conditional on the true underlying physical quantity.

Predictive uncertainty, by contrast, concerns uncertainty in the inferred underlying function or in future observations conditioned on the available data [76]. Even if measurements were noise-free, predictive uncertainty would remain in regions where data are limited or absent. This component reflects incomplete knowledge of the physical behaviour being modelled. Within a Bayesian framework, predictive uncertainty arises from posterior uncertainty over model parameters or functions, together with the assumed observational noise model. A coherent regression framework must therefore propagate experimental covariance information through the inference procedure and combine it with structural model uncertainty to produce physically meaningful uncertainty bands, as is the case in GP regression.

The predictive distribution formalises this concept. For a new input value, it specifies the full probability distribution of either the latent function value or a future observable, conditional on the observed dataset. Rather than providing a single number, the predictive distribution encodes the range of plausible outcomes and their associated probabilities. In GP regression, this distribution is Gaussian and characterised by a predictive mean and covariance matrix. The pre-

dictive mean represents the best estimate given the data and prior assumptions, while the predictive variance quantifies the residual uncertainty. Both quantities are obtained from the posterior distribution after conditioning on the observed data and optimising the hyperparameters.

This predictive variance typically contains two conceptually distinct contributions [77]. The first is epistemic uncertainty, arising from limited information about the underlying function and reducible through the acquisition of additional data. The second is aleatoric uncertainty[2], arising from intrinsic measurement noise and irreducible without improvements in experimental precision. Distinguishing between these components clarifies how uncertainty evolves as new information is incorporated and is particularly relevant in sparse regimes, where epistemic uncertainty may dominate outside measured regions.

Quantifying predictive uncertainty is not sufficient in itself; its reliability must be assessed. Validation tools are therefore required to evaluate whether predicted uncertainties are statistically consistent with observed deviations. One standard diagnostic in experimental physics is the residual distribution, which is often referred to as the pull distribution in physics. For validation data, the pull is defined as the residual between prediction $(y_{pred})$ and observation $(y_{obs})$, normalised by the predicted standard deviation $(\delta y_{pred})$. This can be expressed as:

$$\text{pull} = \frac{y_{obs} - y_{pred}}{\delta y_{pred}}. \tag{3.7}$$

If uncertainties are correctly estimated and residuals approximately Gaussian, the pull distribution should be centred at zero with unit variance. A broadened pull distribution indicates underestimated uncertainties, while a narrowed distribution indicates overestimation. A systematic shift away from zero reveals bias in the predictive model.

Closely related is the concept of empirical coverage [78]. A predictive interval constructed at nominal credibility level $\alpha$ is said to be calibrated if the true values fall within that interval with frequency approximately $\alpha$ under repeated sampling. Under-coverage signals overconfident predictions, whereas over-coverage indicates conservative but potentially uninformative uncertainty estimates. In sparse data regimes, ensuring appropriate coverage is particularly important, as models may otherwise extrapolate with unjustified certainty in poorly constrained regions. Figure 3.1 shows example empirical coverage curves comparing predictive calibration. The well-calibrated model (blue) closely follows the diagonal (ideal $y = x$), while the overcon-

[2]Often called statistical uncertainty in physics.

fident model (orange) falls below the line and the underconfident model (green) lies above it. Each curve is computed over discrete coverage levels, giving a step-like appearance.

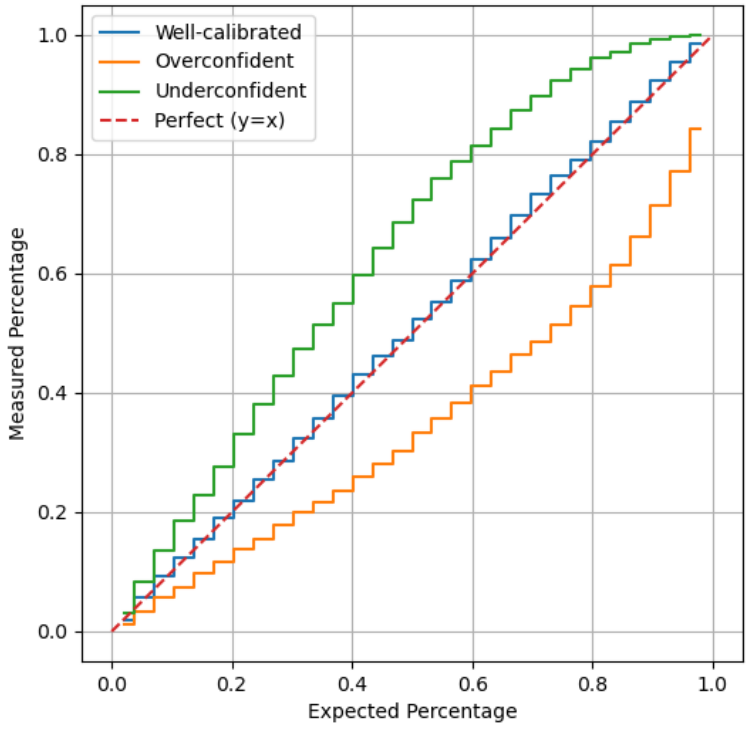



Figure 3.1: Example Empirical Coverage Curves.

For Gaussian predictive distributions, the nominal coverage associated with a given number of standard deviations can be determined from the Cumulative Distribution Function (CDF) of the standard normal distribution [79] (this coverage is equivalent to the $x$-axis of Figure 3.1). The CDF describes the probability that a random variable takes a value less than or equal to a given value. For a standard normal distribution, it therefore gives the cumulative probability up to a specified number of standard deviations from the mean. The CDF, denoted by $\Phi(z)$, gives the probability that a standard normal variable takes a value less than or equal to $z$ and is defined as

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-t^2/2}\, dt, \tag{3.8}$$

where $t$ is a dummy integration variable and $z$ is the upper limit of the integral. The probability of a value $x$ lying within $k$ standard deviations of the predicted mean is therefore

$$P(-z \leq x \leq z) = \Phi(z)\Phi(-z), \tag{3.9}$$

where $z$ is the chosen number of standard deviations defining the interval. This provides the expected coverage for a Gaussian predictive distribution at each chosen interval. For example,

intervals of $1\sigma$ and $1.96\sigma$ correspond to approximately 68.3% and 95.0% coverage, respectively. These values provide the nominal coverage levels against which empirical coverage can be compared when assessing the calibration of the predictive uncertainties [80].

Finally, the quality of uncertainty quantification involves balancing calibration with sharpness. Calibration refers to statistical consistency between predicted and observed variability, while sharpness measures the concentration of predictive distributions independently of the realised outcomes. A model that assigns excessively broad intervals will achieve nominal coverage but provide little useful information. Conversely, a model that produces overly narrow intervals may appear precise but fail to achieve correct coverage. The goal is therefore to obtain predictions that are as sharp as possible while remaining well calibrated.

In the modelling of hadron observables, this balance is essential. Meaningful uncertainty bands must reflect genuine physical ignorance without artificially inflating variability or masking systematic bias. A principled probabilistic framework provides the tools to assess and maintain this balance, ensuring that uncertainty estimates are both statistically consistent and physically informative.

## 3.5 Gaussian Processes

Gaussian Processes (GPs) provide a probabilistic framework for regression in which the object of inference is a function rather than a finite set of parameters. They may be understood as a natural extension of multivariate Gaussian distributions to infinite-dimensional function spaces. Whereas a multivariate Gaussian defines a joint distribution over a finite collection of random variables, a GP defines a joint distribution over function values evaluated at any finite set of input points.

Formally, a GP is defined as a collection of random variables such that any finite subset follows a multivariate normal distribution. A function $f(x)$ is said to follow a GP if, for any inputs $x_1, \ldots, x_n$, the vector $(f(x_1), \ldots, f(x_n))$ is multivariate Gaussian distributed. This is written as

$$f(\vec{x}) \sim \mathcal{N}(m(\vec{x}), \kappa(\vec{x}, \vec{x})), \tag{3.10}$$

where $m(\vec{x})$ is the mean function and $\kappa(\vec{x}, \vec{x})$ is the covariance function, or kernel [6].

This definition means that a GP is completely specified by its first and second moments. The mean function determines the expected value of the function at each input point, while the covariance function determines the relationship between function values at different input points. The GP therefore provides a probabilistic distribution over possible functions rather than specifying a single functional form.

The mean function is defined as

$$m(\vec{x}) = \mathbb{E}[f(\vec{x})], \tag{3.11}$$

and represents the prior expectation of the function at each input point. In many applications, the mean function is chosen to be zero. This does not prevent the posterior from taking non-zero values, since the observed data modify the prior expectation through the covariance structure.

The covariance function is defined as

$$\kappa(\vec{x}_i, \vec{x}_j) = \mathrm{Cov}(f(\vec{x}_i), f(\vec{x}_j)), \tag{3.12}$$

and describes the covariance between function values at two input points. The covariance therefore determines how strongly knowledge of the function at one point is related to its behaviour at another. The properties of the covariance function determine the class of functions that can be represented by the GP. Covariance functions and their properties are discussed in Section 3.6.

GP regression proceeds by conditioning the prior process on observed data. Suppose that measurements $y = (y_1, \ldots, y_n)$ are observed at input points $X = (\vec{x}_1, \ldots, \vec{x}_n)$, where each $\vec{x}_i$ represents a point in the input or kinematic space. The corresponding values are collected into the vector

$$y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}. \tag{3.13}$$

To make predictions at new input points, consider a second set of points

$$X_* = (\vec{x}_1^*, \ldots, \vec{x}_m^*), \tag{3.14}$$

where the values of the observable are to be predicted. These predicted values are collected into

$$y_* = \begin{bmatrix} y_1^* \\ \vdots \\ y_m^* \end{bmatrix}. \tag{3.15}$$

The defining property of a GP is that the values of the underlying function at any finite collection of input points follow a multivariate Gaussian distribution. Consequently, the values at the combined set of observed and prediction points, $(X, X_*)$, form a joint multivariate Gaussian distribution. The mean vectors associated with the observed and prediction points are denoted by

$$M = m(X), \qquad M_* = m(X_*), \tag{3.16}$$

and the covariance matrices are constructed from the kernel according to

$$K_f = \kappa(X, X), \tag{3.17}$$

$$K_* = \kappa(X, X_*), \tag{3.18}$$

$$K_{**} = \kappa(X_*, X_*). \tag{3.19}$$

In practice, the observed values are subject to experimental uncertainties. These can be incorporated by assuming that the measured values are related to the underlying function through additive Gaussian noise,

$$y = f(X) + \varepsilon, \qquad \varepsilon \sim \mathcal{N}(0, \sigma^2), \tag{3.20}$$

where $\sigma^2$ is the covariance matrix of the experimental uncertainties. Since the sum of Gaussian random variables is itself Gaussian distributed, the inclusion of experimental noise modifies the covariance of the observed values without changing the covariance between the observed and prediction points. The covariance matrix associated with the observed data therefore becomes

$$K = K_f + \sigma^2 = \kappa(X, X) + \sigma^2. \tag{3.21}$$

The resulting joint distribution of the observed values and the predicted values can consequently be written as

$$\begin{bmatrix} y \\ y_* \end{bmatrix} \sim \mathcal{N}\left( \begin{bmatrix} M \\ M_* \end{bmatrix}, \begin{bmatrix} K & K_* \\ K_*^T & K_{**} \end{bmatrix} \right). \tag{3.22}$$

This result follows from two properties of Gaussian distributions. First, the definition of a GP ensures that function values at any finite set of input points are jointly Gaussian. Second, the experimental uncertainties are assumed to be Gaussian and independent of the underlying function. Adding this Gaussian noise to the observed function values therefore preserves the multivariate Gaussian form of the joint distribution.

The predictive distribution is obtained by conditioning the joint distribution in Equation 3.22 on the observed values $y$ [6]. The joint probability density corresponding to this distribution is

$$p(y, y_*) \propto \exp\left[-\frac{1}{2}\begin{bmatrix} y-M \\ y_*-M_* \end{bmatrix}^T \begin{bmatrix} K & K_* \\ K_*^T & K_{**} \end{bmatrix}^{-1} \begin{bmatrix} y-M \\ y_*-M_* \end{bmatrix}\right]. \tag{3.23}$$

For fixed observed values $y$, the conditional distribution $p(y_*|y)$ is obtained from the joint distribution by treating $y$ as constant. The exponent of the joint Gaussian can then be expanded into terms involving $y$ and $y_*$. Terms that depend only on $y$ contribute only to the normalisation of the conditional distribution and can therefore be omitted. The terms that depend on $y_*$ can be written as

$$p(y_*|y) \propto \exp\left[-\frac{1}{2}\left(y_*^T K_{**}^{-1} y_* - 2y_*^T K_{**}^{-1} K_*^T K^{-1}(y-M)\right)\right]. \tag{3.24}$$

Completing the square in $y_*$ gives

$$p(y_*|y) \propto \exp\left[-\frac{1}{2}\left(y_*-\mu_*\right)^T \Sigma_*^{-1}\left(y_*-\mu_*\right)\right], \tag{3.25}$$

where

$$\mu_* = M_* + K_*^T K^{-1}(y-M), \tag{3.26}$$

$$\Sigma_* = K_{**} - K_*^T K^{-1} K_*. \tag{3.27}$$

This expression has the form of a multivariate Gaussian distribution and can therefore be written as

$$p(y_*|y) \sim \mathcal{N}(\mu_*, \Sigma_*), \tag{3.28}$$

where $\mu_*$ and $\Sigma_*$ are the predictive mean and covariance, respectively.

The predictive mean in Equation 3.26 consists of the prior mean $M_*$ together with a correction determined by the difference between the observed values $y$ and their prior mean $M$. The cross-covariance matrix $K_*$ determines how information from the observed points is transferred to the prediction points, while $K^{-1}$ accounts for the covariance of the observations, including their experimental uncertainties.

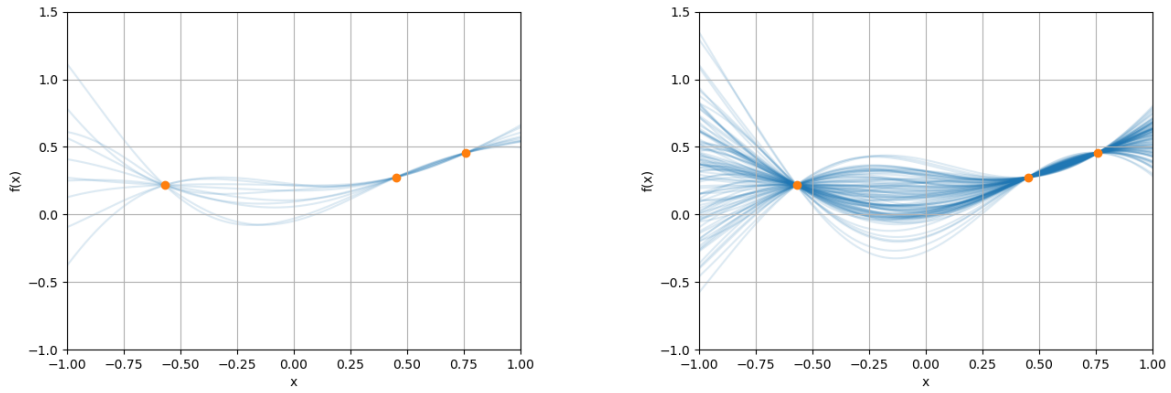
The predictive covariance in Equation 3.27 represents the remaining uncertainty in the predicted values after conditioning on the observations. The second term in this expression represents the reduction in uncertainty resulting from the information provided by the observed data. The magnitude of this reduction depends on the covariance between the prediction and observed points. Prediction points that are strongly correlated with the observed data are more strongly constrained and consequently have smaller predictive uncertainties. Conversely, prediction points that are weakly correlated with the observations retain a larger proportion of their prior uncertainty. In regions sufficiently far from the observed data, the cross-covariance $K_*$ becomes small and the predictive distribution approaches the prior distribution. This provides a natural mechanism for propagating both the information contained in the measurements and the associated uncertainty into regions where no direct measurements are available.

The GP framework can therefore be interpreted as a Bayesian approach to regression. In the terminology introduced in Section 3.2, the GP defines a prior distribution over functions through its mean and covariance functions. The observed data then provide the likelihood through the assumed measurement model, and conditioning the prior distribution on these observations produces the posterior distribution over functions. In this sense, the predictive distribution derived above is the functional analogue of the posterior distribution in Equation 3.4. Rather than obtaining a posterior distribution over a finite set of model parameters, GP regression obtains a posterior distribution over possible functions. This interpretation is consistent with the Bayesian formulation of GPs as non-parametric models, in which the prior is placed directly over functions rather than over the parameters of a prescribed functional form [6].

The Bayesian nature of the GP is particularly relevant when predicting at unmeasured points. The prior encodes assumptions about the behaviour and correlations of the underlying function, while the observed data update these assumptions through the likelihood. The predictive mean and covariance therefore describe the posterior state of knowledge about the function after conditioning on the available measurements. This provides a natural framework for propagating both experimental information and uncertainty into regions where measurements are sparse or unavailable, which is a central motivation for the GP framework developed in this thesis.

The probabilistic nature of the GP can also be illustrated by drawing samples from the posterior distribution over functions. Unlike samples drawn from the prior, posterior samples are conditioned on the observed measurements and therefore reflect both the information contained in the data and the assumptions encoded by the covariance function. Figure 3.2 shows posterior samples drawn from the same GP posterior using 10 and 100 samples. For this illustrative example, the Radial Basis Function (RBF) kernel is used to define the covariance function (this is

discussed further in the next section), while the observed values are assumed to have no associated experimental uncertainty, such that the posterior samples are constrained to pass through the observed points.




(a) 10 samples.

(b) 100 samples.

Figure 3.2: Posterior GP samples illustrating the range of functions consistent with the observed data. Each curve represents a possible function passing between the observed points, with the ensemble of curves representing the functions permitted by the posterior distribution. Increasing the number of samples provides a denser representation of this range of possible functions.

With a small number of samples, the individual functions illustrate the range of behaviours permitted by the posterior distribution, while increasing the number of samples provides a progressively clearer representation of the possible functions. The spread of the sampled functions therefore provides a visual representation of the predictive uncertainty. These samples provide a direct representation of the distribution over possible functions, while the predictive mean and covariance provide a compact statistical description of the same distribution. Figure 3.3 shows the posterior mean together with uncertainty bands derived from the predictive covariance. The sampled functions can be interpreted as individual realisations from this posterior distribution, fluctuating around the predictive mean with a spread determined by the local predictive variance.

The preceding discussion establishes the basic framework of GP regression. A prior distribution over functions is specified through a mean function and covariance function, observations are incorporated by conditioning this prior on the available data, and the resulting posterior provides predictions together with associated uncertainties. The covariance function is particularly important because it determines the statistical relationship between function values at different input points and therefore controls the types of functions that can be represented by the GP. Covariance functions, commonly referred to as kernels, and their properties are discussed in the Section 3.6.

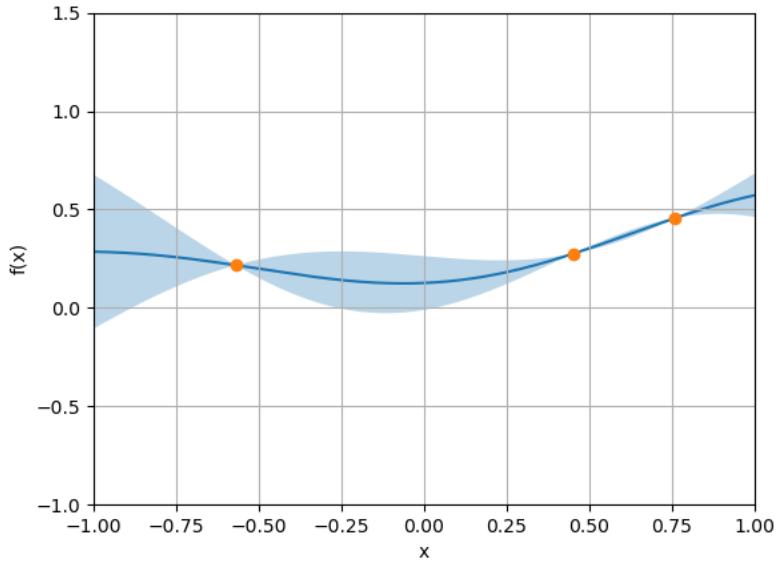



Figure 3.3: Example GP fit showing the posterior mean and uncertainty band. The predictive distribution summarises the range of functions represented by the posterior samples shown in Figure 3.2.

### 3.5.1 Cholesky Decomposition

Explicitly computing $K^{-1}$ of the predictive GP mean in Equation 3.26 is computationally expensive and can introduce unnecessary numerical instability. A standard alternative is to factorise the covariance matrix using a Cholesky decomposition [6]; this factorises a symmetric positive-definite matrix into the product of a lower triangular matrix and its transpose, allowing linear systems involving the covariance matrix to be solved without explicitly computing its inverse (these properties of the covariance matrix are discussed further in the following section). The Cholesky decomposition expresses the covariance matrix as

$$K = LL^T, \tag{3.29}$$

where $L$ is a lower triangular matrix. This factorisation allows the required linear systems involving $K$ to be solved using forward and backward substitution, avoiding the explicit calculation of $K^{-1}$ and providing a more numerically stable and computationally efficient approach.

The predictive mean requires the quantity $K^{-1}(y-M)$. Rather than finding the inverse explicitly, $\alpha$ is defined as the solution to $K\alpha = (y-M)$. Using the Cholesky decomposition, this system is solved through two triangular linear systems: $Lz = (y-M)$ is solved for $z$ and the result used to solve $L^T\alpha = z$ for $\alpha$. Since triangular systems can be solved efficiently using forward

and backward substitution, this approach is both computationally cheaper and numerically more stable than explicitly computing $K^{-1}$. The posterior mean can then be written as

$$\mu_* = M_* + K_*^T \alpha. \tag{3.30}$$

This formulation is mathematically equivalent to Equation 3.26 whilst avoiding explicit inversion of the covariance matrix. A similar approach is used to evaluate the posterior covariance, as shown in Equation 3.27. Defining $v = L^{-1}K_*$ where $K_*$ is the covariance matrix between the known inputs and the prediction points, the predictive covariance becomes

$$\Sigma_* = K_{**} - v^T v, \tag{3.31}$$

where $\Sigma_*$ is the covariance matrix of the posterior predictions and $K_{**}$ is the covariance matrix evaluated between the prediction points. The matrix $v$ therefore contains the contribution from the covariance between the training and prediction points after solving the corresponding triangular system. This expression is equivalent to Equation 3.27 but evaluates the required operations using the Cholesky factorisation rather than explicitly calculating $K^{-1}$.

## 3.6 GP Kernels

The covariance function, commonly referred to as the kernel, is a central component of a GP. It specifies the covariance between function values at different input points,

$$\kappa(\vec{a}, \vec{b}) = \mathrm{Cov}(f(\vec{a}), f(\vec{b})), \tag{3.32}$$

and therefore determines how strongly observations at one location influence predictions at another. The kernel also encodes assumptions about the properties of the underlying function, such as its smoothness, characteristic scale of variation, periodicity and amplitude.

For a covariance function to define a valid GP, the resulting covariance matrix must be symmetric and positive semi-definite. A matrix is symmetric if it is equal to its transpose,

$$K = K^T. \tag{3.33}$$

For a covariance matrix, this follows from the symmetry of the covariance between two random

variables, such that the covariance between $f(\vec{a})$ and $f(\vec{b})$ is the same as that between $f(\vec{b})$ and $f(\vec{a})$.

The covariance matrix must also be positive semi-definite. This requires that

$$\vec{c}^T K \vec{c} \geq 0 \tag{3.34}$$

for every vector $\vec{c}$. Positive semi-definiteness ensures that the covariance matrix represents physically and mathematically valid variances for arbitrary linear combinations of the function values. In particular, for any linear combination

$$z = \vec{c}^T \vec{f}, \tag{3.35}$$

its variance is given by

$$\mathrm{Var}(z) = \vec{c}^T K \vec{c}, \tag{3.36}$$

which cannot be negative. These conditions ensure that the kernel defines a valid covariance matrix and consequently a valid multivariate Gaussian distribution for every finite collection of function values [6].

The behaviour of a kernel can be understood in terms of the covariance between two input points. When two points have a large covariance, their function values are strongly correlated and knowledge of the function at one point provides substantial information about the other. Conversely, when the covariance is small, the two function values are weakly correlated and the function is able to vary more independently between them. The rate at which the covariance decreases with distance therefore determines the characteristic scale over which information is propagated through the input space.

Different kernel choices correspond to different assumptions about the regularity and structure of the underlying function. Several commonly used kernels are described below.

### 3.6.1 Radial Basis Function Kernel

The squared exponential, or radial basis function (RBF), kernel is one of the most widely used covariance functions for modelling smooth functions [6]. In its general form, it is defined as

$$\kappa\left(\vec{a},\vec{b}\right) = \sigma_f^2 \exp\left(-\frac{1}{2}\sum_{i=1}^{d}\frac{(a_i - b_i)^2}{\ell_i^2}\right), \tag{3.37}$$

where $d$ denotes the number of input dimensions, $\ell_i$ is the characteristic length scale associated with the $i$-th dimension, and $\sigma_f^2$ is the amplitude. The kernel is stationary, meaning that the covariance depends only on the separation between the input points rather than their absolute positions. The covariance decreases smoothly with increasing separation, such that nearby points are strongly correlated while points separated by distances large compared with their characteristic length scales become increasingly uncorrelated.

The characteristic length scale $\ell_i$ determines the rate at which correlations decay along the corresponding input dimension. A small length scale allows the function to vary over shorter distances, while a large length scale produces correlations over greater distances and therefore favours more slowly varying behaviour. In multiple dimensions, assigning a separate length scale to each input dimension allows the kernel to represent different characteristic scales of variation in different directions through the input space. The amplitude $\sigma_f^2$ controls the overall scale of the covariance and therefore the typical magnitude of variations about the mean function.

An important property of the RBF kernel is the smoothness of the functions it produces. Functions drawn from a GP with an RBF covariance function are infinitely differentiable, meaning that derivatives of all orders exist and are continuous [6]. Consequently, the functions represented by this kernel cannot contain discontinuities, sharp corners, or other non-differentiable features. The RBF kernel therefore imposes a strong prior assumption that the underlying function varies smoothly throughout the input space. The length scales determine the characteristic distance over which this smooth variation occurs, while the amplitude determines its overall scale.

The RBF kernel consequently provides a relatively simple covariance model in which the hyperparameters control the scale and smoothness of the functions represented by the GP. Alternative kernel choices modify these assumptions and can therefore represent functions with different

degrees of smoothness or other structural properties.

### 3.6.2 Matérn kernel

A second commonly used choice is the Matérn kernel [81]. Unlike the RBF kernel, which imposes infinite differentiability, the Matérn family provides a direct means of controlling the smoothness of the functions represented by the GP. The Matérn covariance function can be written as

$$\kappa_\nu\left(\vec{a},\vec{b}\right) = \sigma_f^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\sqrt{2\nu}\sqrt{\sum_{i=1}^{d}\frac{(a_i-b_i)^2}{\ell_i^2}}\right)^{\nu} K_\nu\left(\sqrt{2\nu}\sqrt{\sum_{i=1}^{d}\frac{(a_i-b_i)^2}{\ell_i^2}}\right), \tag{3.38}$$

where $\sigma_f^2$ is the amplitude, $\ell_i$ is the characteristic length scale associated with the $i$-th input dimension, $\nu$ is the smoothness parameter, $\Gamma$ is the Gamma function, and $K_\nu$ is a modified Bessel function of the second kind.

Similar to the RBF kernel, the length scales $\ell_i$ determine the rate at which correlations decrease with separation along each input dimension. A small $\ell_i$ allows functions to vary over shorter distances in the corresponding dimension, while a large $\ell_i$ produces correlations over greater distances and therefore favours more slowly varying behaviour. The amplitude $\sigma_f^2$ determines the overall variance scale of the function, while $\nu$ controls its smoothness. Smaller values of $\nu$ produce functions with less smooth and more rapidly varying behaviour, whereas increasing $\nu$ results in progressively smoother functions.

The smoothness parameter provides an important distinction between the Matérn and RBF kernels. For finite values of $\nu$, functions drawn from a Matérn GP have a finite degree of differentiability, with the number of mean square derivatives increasing as $\nu$ increases [6]. Thus, unlike the RBF kernel, the Matérn kernel can represent functions that are smooth but not infinitely differentiable. In the limit $\nu \to \infty$, the Matérn kernel approaches the RBF kernel, recovering its infinitely differentiable behaviour [6].

For certain commonly used values of $\nu$, the Matérn kernel has simpler analytical forms. For

example, when $\nu = 3/2$,

$$\kappa_{3/2}\left(\vec{a},\vec{b}\right) = \sigma_f^2\left(1+\sqrt{3}r\right)\exp\left(-\sqrt{3}r\right), \tag{3.39}$$

while for $\nu = 5/2$,

$$\kappa_{5/2}\left(\vec{a},\vec{b}\right) = \sigma_f^2\left(1+\sqrt{5}r+\frac{5r^2}{3}\right)\exp\left(-\sqrt{5}r\right), \tag{3.40}$$

where

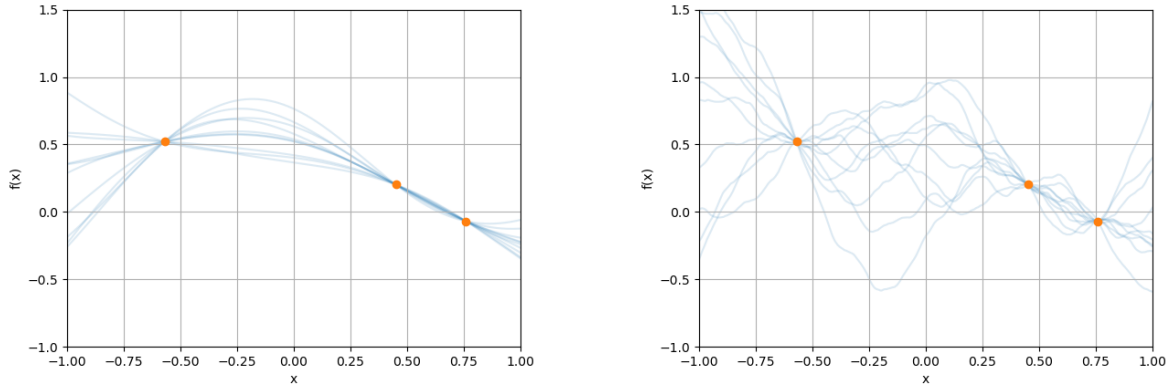
$$r = \sqrt{\sum_{i=1}^{d}\frac{(a_i-b_i)^2}{\ell_i^2}}. \tag{3.41}$$

### 3.6.3 Comparison of RBF and Matérn Kernels

The choice between the RBF and Matérn kernels therefore represents a trade-off in the assumptions made about the smoothness of the underlying function. The RBF kernel provides the strongest smoothness assumption, with functions that are infinitely differentiable, but this may be unnecessarily restrictive when the underlying physical behaviour contains more rapid or less regular variations. The Matérn family relaxes this assumption by allowing the degree of smoothness to be controlled through $\nu$, providing greater flexibility at the cost of introducing an additional hyperparameter and allowing less smooth functions. Consequently, while the RBF kernel may be appropriate when a highly smooth underlying function is expected, Matérn kernels can provide a more flexible description when the degree of smoothness is not known a priori.

Figure 3.4 illustrates the effect of the kernel choice by showing posterior samples generated using the RBF and Matérn kernels. The two sets of samples are conditioned on the same observed data, but the different covariance functions impose different assumptions on the smoothness of the underlying function. The resulting functions therefore exhibit different behaviour despite being constrained by the same measurements.

Other kernel families can encode different types of structure. For example, periodic kernels can represent functions with repeating behaviour, while linear kernels describe covariance associated with linear relationships between input variables. Kernel combinations can also be constructed to represent more complex forms of behaviour, allowing multiple structural assumptions to be




(a) RBF kernel.

(b) Matérn kernel.

Figure 3.4: Posterior GP samples generated using different covariance functions and conditioned on the same observed data. The different kernels impose different assumptions about the smoothness of the underlying function, resulting in different sets of possible functions. The RBF kernel produces infinitely differentiable functions and therefore favours highly smooth behaviour, while the Matérn kernel allows less smooth functions depending on the choice of $\nu$. The figure illustrates how these differing smoothness assumptions influence the range of functions that can be represented while remaining consistent with the same observations.

incorporated into a single covariance function.

The choice of kernel therefore determines the prior distribution over functions and consequently influences the behaviour of the resulting GP. Different kernels can produce substantially different function spaces even when conditioned on the same observations. Kernel selection is consequently an important modelling consideration, particularly when the available data provide limited information about the underlying functional behaviour.

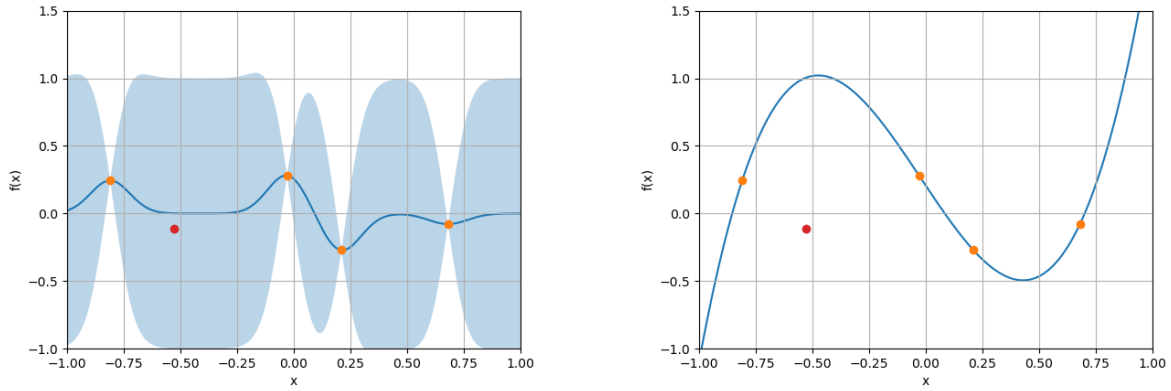
## 3.7 Hyperparameter Optimisation

The kernel functions introduced in Section 3.6 contain parameters that determine the properties of the functions represented by the GP. These parameters are referred to as hyperparameters and are generally not known *a priori*. The hyperparameters are collected into a vector

$$\theta = (\theta_1, \theta_2, \ldots, \theta_p), \tag{3.42}$$

where $p$ is the total number of hyperparameters.

As an example, consider the RBF kernel introduced in Equation 3.37. The characteristic length scale $\ell$ determines how rapidly correlations decay as a function of distance in input space. A small length scale allows the GP to vary over shorter distances and therefore provides greater local flexibility, while a large length scale produces correlations over greater distances and favours smoother, more slowly varying behaviour. If $\ell$ is too small, the GP may exhibit rapid variations between neighbouring datapoints and become overly sensitive to individual measurements, as illustrated in Figure 3.5a. Conversely, if $\ell$ is too large, the GP may smooth out genuine variation in the underlying function, resulting in biased predictions between measurements, as illustrated in Figure 3.5b. These examples demonstrate how the choice of hyperparameters can substantially influence the behaviour of the resulting GP.




(a) Length scale $\ell$ too small.

(b) Length scale $\ell$ too large.

Figure 3.5: Example GP fits illustrating the effect of an inappropriate choice of the RBF kernel length scale $\ell$. In both cases, the GP is conditioned on the same set of observed datapoints, but different values of $\ell$ are used. For a length scale that is too small (left), the GP is able to vary rapidly between neighbouring datapoints, resulting in increased local variation and sensitivity to individual measurements. For a length scale that is too large (right), the stronger correlations over larger distances impose greater smoothness on the posterior mean, potentially suppressing genuine variation in the underlying function. The examples demonstrate that the length scale controls the balance between local flexibility and smoothness and therefore has a significant influence on the resulting GP prediction.

The hyperparameters must therefore be determined using an appropriate optimisation procedure. The standard approach to determining these hyperparameters is to maximise the marginal likelihood of the observed data [6].

The marginal likelihood is obtained by integrating over all possible values of the latent function. For the observed data considered in Section 3.5, this can be written as

$$P(y \mid X, \theta) = \int P(y \mid f(X)) P(f(X) \mid X, \theta), df(X). \tag{3.43}$$

Here, $P(y \mid f(X))$ describes the probability of obtaining the observed data for a given underlying function, while $P(f(X) \mid X, \theta)$ describes the prior probability distribution of the latent function for a given set of input points and hyperparameters. From the GP formulation introduced in Section 3.5, the latent function is distributed according to

$$p(f(X) \mid X, \theta) \sim \mathcal{N}(M, K_f), \tag{3.44}$$

where $M = m(X)$ is the mean vector and $K_f = \kappa(X,X)$ is the covariance matrix of the latent function. The likelihood of the observed data for a given latent function is therefore

$$P(y \mid f(X)) = \mathcal{N}(f(X), \Sigma), \tag{3.45}$$

where $\Sigma = \sigma^2 I$ represents the covariance matrix formed from experimental uncertainties. Since the observed values are given by

$$y = f(X) + \Sigma, \tag{3.46}$$

the marginalisation over $f(X)$ combines the uncertainty in the latent function with the experimental uncertainty. As established in Section 3.5, the resulting covariance matrix is

$$K = K_f + \Sigma. \tag{3.47}$$

The marginal distribution of the observed data is therefore

$$y \sim \mathcal{N}(M, K). \tag{3.48}$$

Solving the integral in Equation 3.43 gives the marginal probability density

$$P(y \mid X, \theta) = \frac{1}{\sqrt{(2\pi)^n |K|}} \exp\left[-\frac{1}{2}(y-M)^T K^{-1} (y-M)\right]. \tag{3.49}$$

Taking the negative logarithm gives the negative marginal log-likelihood,

$$-\log P(y \mid X, \theta) = \frac{1}{2}(y-M)^T K^{-1}(y-M) + \frac{1}{2}\log|K| + \frac{n}{2}\log(2\pi). \tag{3.50}$$

The optimal hyperparameters can then be obtained by minimising this quantity,

$$\theta^* = \arg,\min_{\theta} \left[-\log P(y \mid X, \theta)\right]. \tag{3.51}$$

### 3.7.1 Issues with Sparse Datasets

The marginal likelihood provides a principled method for determining the hyperparameters of a GP. However, maximising the marginal likelihood does not necessarily ensure that the resulting GP has the properties required when modelling sparse experimental data. In particular, when the available measurements provide limited coverage of the underlying function, the marginal likelihood may provide only weak constraints on some hyperparameters. Different combinations of hyperparameters can consequently produce similarly probable descriptions of the observed data, while giving substantially different behaviour between the measured points. This is particularly relevant for length scale parameters, which determine the characteristic distance over which the GP covariance decreases and therefore control the smoothness of the resulting function.

The behaviour of the marginal likelihood can be illustrated using the $O_z$ polarisation observable for the $\gamma p \to K^0\Sigma^+$ reaction [82]. The dataset contains approximately 21 measurements covering the full angular range, $\cos\theta = -1$ to $+1$, over an energy range of approximately 800 MeV. Figure 3.6 shows the negative marginal log-likelihood as a function of the angular and energy length scale parameters for this dataset.

The negative log-likelihood surface demonstrates two features that are important for sparse datasets. First, the optimum occurs at a relatively small angular length scale. Such a value permits the GP to vary rapidly with angle between the available measurements, despite the comparatively smooth behaviour expected of the observable over the measured angular range. Second, the contours of the surface are approximately flat in the energy direction, forming an extended ridge across a broad range of energy length scales. This indicates that a substantial range of energy length scales provides similarly good descriptions of the available measurements, meaning that the energy dependence is poorly constrained by the marginal likelihood.

This behaviour arises because the marginal likelihood evaluates how well a particular probabilistic model describes the measurements that are available. When those measurements sparsely sample the underlying physical dependence, there may be insufficient information to distinguish between different behaviours in regions between the measured points. Consequently, a hyperparameter set can provide a good marginal likelihood while producing undesirable interpolation behaviour or poorly calibrated predictive uncertainties in regions where measurements are sparse.

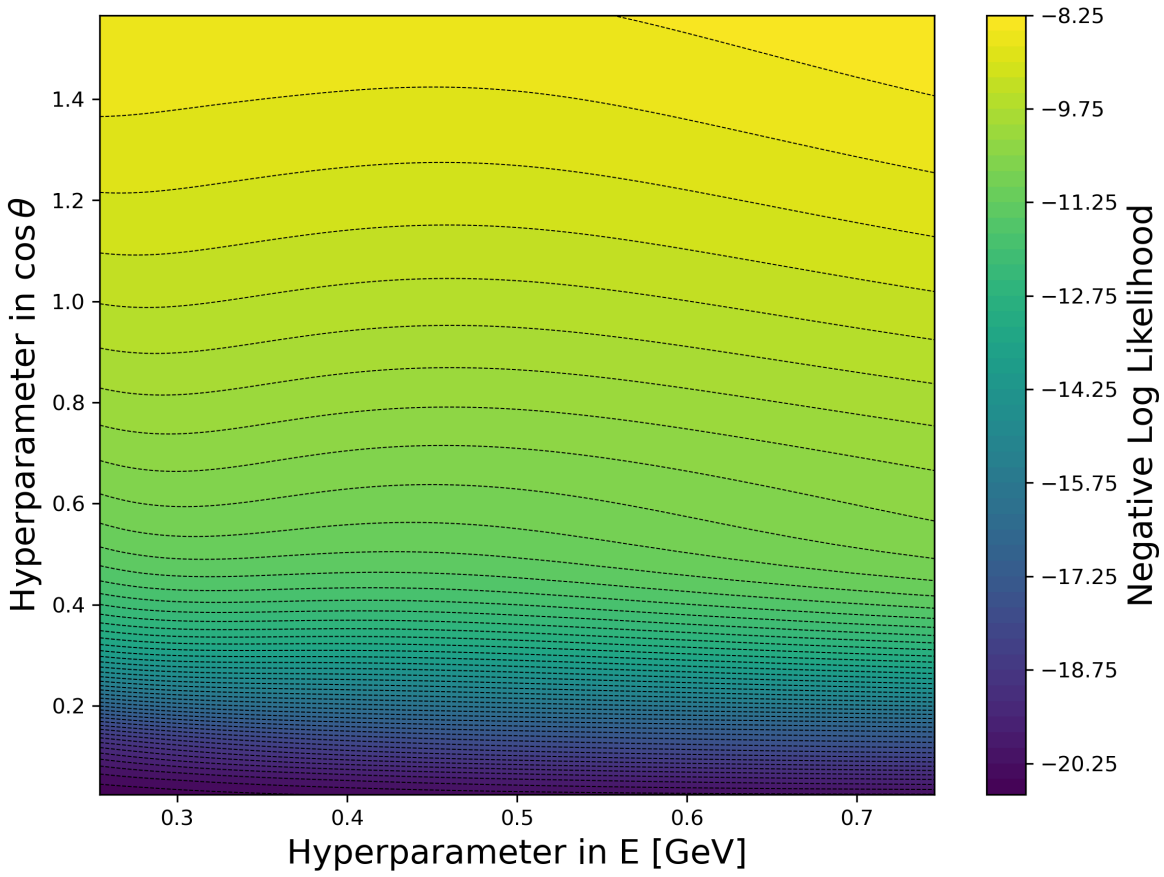



Figure 3.6: Negative marginal log-likelihood surface calculated from Equation 3.51 for the $O_z$ polarisation observable [82]. The approximately flat contours in the energy direction indicate that the energy length scale is poorly constrained by the marginal likelihood, while the minimum occurs at a relatively small angular length scale.

This distinction is important for the application considered in this thesis. The objective is not simply to determine the GP that provides the most probable description of the measured points, but to construct an interpolation that provides a statistically reliable representation of the observable between those measurements. The predictive uncertainties must also provide a meaningful indication of the confidence in the interpolated function. These requirements are not directly encoded in the marginal likelihood alone.

The example in Figure 3.6 therefore demonstrates that, for sparse experimental datasets, marginal-likelihood optimisation may not sufficiently constrain the hyperparameters relevant to the desired interpolation. Rather than replacing the marginal likelihood entirely, this motivates the inclusion of additional criteria that directly assess properties such as predictive uncertainty calibration and the behaviour of the GP between measured points. A hyperparameter optimisation procedure incorporating these requirements is developed in Chapter 4.

## 3.8 Finding a Multidimensional Optimum

Many optimisation problems require the minimum of a function that depends on multiple variables. The objective of the optimisation is to identify the point at which the function takes its minimum value.

In one dimension, this problem can be visualised as finding the lowest point of a curve. In multiple dimensions, the objective function instead defines a landscape over the parameter space, with each point corresponding to a particular combination of the variables. The resulting landscape can contain multiple local minima, flat regions and strongly correlated directions, making it difficult to identify the global minimum through a simple systematic search. An optimisation method must therefore balance exploration of the available space with convergence towards regions where the objective function is small.

One approach to exploring a multidimensional space is to generate an initial set of points that provides broad coverage of the allowed region. Latin Hypercube Sampling (LHS) is one method for achieving this [83]. If $m$ samples are required, each dimension of the search space is divided into $m$ equally sized intervals, with one random sample selected from each interval. The samples in each dimension are then independently permuted before being combined to form the final multidimensional sample points. Consequently, each interval is sampled exactly once in every dimension. This provides more uniform marginal coverage of the search space than purely random sampling and can be useful when generating an initial population of candidate solutions. LHS can easily be implemented by using the `smt.sampling_methods.LHS` package [84].

Two approaches commonly used to explore multidimensional spaces are Markov Chain Monte Carlo (MCMC) and Particle Swarm Optimisation (PSO). Although both methods can explore complex multidimensional landscapes, they have different underlying purposes. MCMC is fundamentally a sampling technique used to characterise a probability distribution (and is commonly used in hadron physics in this context), whereas PSO is an optimisation algorithm designed to locate a minimum of an objective function.

### 3.8.1 Markov Chain Monte Carlo

As discussed in Section 3.3, MCMC is a sampling method used to approximate complex probability distributions by generating a sequence of parameter values that form a Markov chain. This process introduces a stochastic exploration of the parameter space, allowing the chain to move through regions of different probability density while favouring regions of higher probability. After an initial burn-in period, the samples can be treated as draws from the target distribution, provided that the chain has converged.

For a $D$-dimensional variable $\vec{x}$, this produces a sequence of samples $\vec{x}^{(1)}, \vec{x}^{(2)}, \ldots, \vec{x}^{(N)}$, where each sample represents a point in the multidimensional space. The target distribution can be related to an objective function. For example, a distribution can be defined as

$$P(\vec{x}) \propto \exp\left[-L(\vec{x})\right]. \tag{3.52}$$

Under this definition, regions where $L(\vec{x})$ is small have a greater probability of being sampled. The resulting samples therefore provide information about the structure of the objective landscape, including regions of high probability, local minima and correlations between dimensions.

In practical applications, the number of samples required for convergence is not necessarily known in advance. Instead, convergence diagnostics can be evaluated periodically during sampling and used to determine whether the chains have sufficiently explored the target distribution. This allows the sampling to be terminated once the convergence requirements are satisfied, rather than requiring a fixed number of iterations to be chosen in advance. Two commonly used diagnostics for assessing MCMC convergence are the Gelman-Rubin statistic [64] and the integrated autocorrelation time [85].

#### Gelman-Rubin Statistic

The Gelman-Rubin statistic, $\hat{R}$, assesses convergence by comparing the variation within individual chains with the variation between different chains [64]. Consider $m$ chains, each containing $n$ samples. The mean of chain $i$ is denoted by $\bar{\theta}_i$, while the overall mean of all chains is denoted

by $\bar{\theta}$. The variance within the chains is estimated from

$$W = \frac{1}{m}\sum_{i=1}^{m} s_i^2, \tag{3.53}$$

where $s_i^2$ is the variance of the samples in the $i$th chain. The between-chain variance is given by

$$B = \frac{n}{m-1}\sum_{i=1}^{m}\left(\bar{\theta}_i - \bar{\theta}\right)^2. \tag{3.54}$$

These two estimates are combined to give an estimate of the marginal posterior variance,

$$\hat{V} = \frac{n-1}{n}W + \frac{1}{n}B. \tag{3.55}$$

The Gelman-Rubin statistic is then defined as

$$\hat{R} = \sqrt{\frac{\hat{V}}{W}}. \tag{3.56}$$

When $\hat{R} \approx 1$, the within-chain and between-chain variance estimates are in agreement, indicating that the chains have mixed well and are sampling from the same distribution. Values of $\hat{R}$ significantly greater than unity indicate that the chains have not yet converged. The statistic therefore provides a measure of consistency between independently initialised chains.

**Integrated Autocorrelation Time**

While $\hat{R}$ assesses agreement between chains, the integrated autocorrelation time provides information about the statistical independence of samples within a chain. Successive MCMC samples are generally correlated, meaning that a chain containing a large number of samples does not necessarily contain the same amount of information as an equivalent number of independent samples.

For a chain containing $n$ samples, the autocovariance at lag $t$, where $t$ denotes the separation

between samples in the chain, can be written as

$$C(t) = \frac{1}{n-t} \sum_{s=1}^{n-t} \left(\theta_s - \bar{\theta}\right) \left(\theta_{s+t} - \bar{\theta}\right). \tag{3.57}$$

The corresponding normalised autocorrelation function is

$$\rho(t) = \frac{C(t)}{C(0)}. \tag{3.58}$$

The integrated autocorrelation time is then defined as

$$\tau = 1 + 2 \sum_{t=1}^{\infty} \rho(t). \tag{3.59}$$

The value of $\tau$ provides an estimate of the number of consecutive correlated samples required to obtain approximately one independent sample. Consequently, larger values of $\tau$ indicate stronger correlations within the chain and a lower effective number of independent samples.

Rather than requiring $\tau$ to reach a particular value, its stability can be used as a convergence criterion. If successive estimates of the integrated autocorrelation time are denoted by $\tau_{\text{old}}$ and $\tau_{\text{new}}$, a relative stability measure can be defined as

$$\tau_{\text{stability}} = \frac{|\tau_{\text{old}} - \tau_{\text{new}}|}{\tau_{\text{new}}}. \tag{3.60}$$

A small value of $\tau_{\text{stability}}$ indicates that the estimated autocorrelation time has become stable as the chains grow, suggesting that additional sampling is unlikely to substantially alter the estimate.

The Gelman-Rubin statistic and integrated autocorrelation time therefore provide complementary convergence information. $\hat{R}$ assesses whether independently initialised chains are sampling consistently from the same distribution, while $\tau_{\text{stability}}$ assesses whether the estimated autocorrelation structure has stabilised. Requiring both measures to indicate convergence provides a more robust stopping criterion than relying on either diagnostic alone.

### Identifying Multimodal Solutions

The probability distribution obtained from an MCMC calculation may contain multiple modes corresponding to distinct regions of the multidimensional parameter space. This is particularly important when using MCMC samples to identify an optimum, as the mean of the samples does not necessarily correspond to a minimum of the objective function. For example, in Figure 3.7, two modes are separated in parameter space, such that their mean may lie between the modes in a region of relatively low probability.

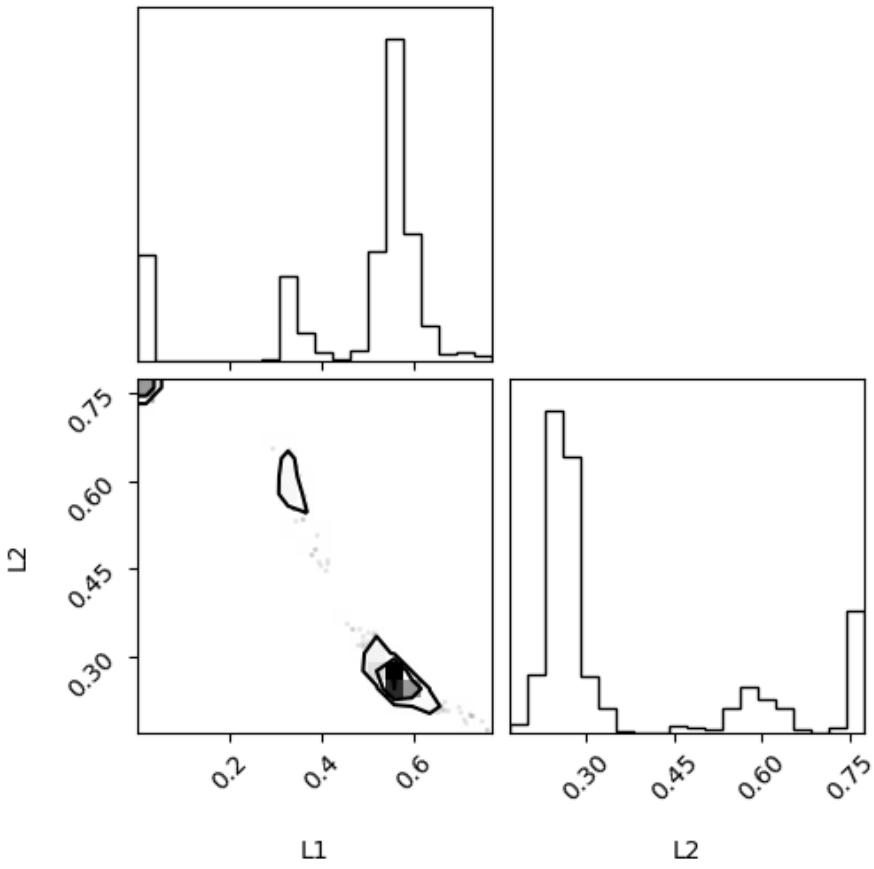



Figure 3.7: Example two-dimensional corner plot illustrating the multimodal nature of the probability distribution obtained from the MCMC calculation. The presence of distinct modes indicates multiple preferred regions of the parameter space, such that the mean of all sampled points may lie between the modes and not correspond to an optimum.

The sampled points can instead be separated into clusters representing distinct regions of the parameter space. $K$-means clustering, implemented using `sklearn.cluster.KMeans` [86], provides one approach for achieving this by partitioning the MCMC samples into $K$ clusters. Each sample is assigned to the cluster whose centre is closest according to the Euclidean dis-

tance, after which the cluster centres are updated by taking the mean of the assigned samples. This process is repeated until the cluster centres have stabilised or a maximum number of iterations is reached.

The appropriate value of $K$ is not necessarily known beforehand. As $K$-means requires $K \geq 2$, the possibility that the sampled distribution contains only a single mode is considered first. The 1D distribution of the samples can be examined using `scipy.signal.find_peaks` [21] to identify distinct local maxima. If a single peak is identified, the samples do not require further clustering and the corresponding region can be used directly for subsequent optimisation. If multiple peaks are identified, $K$-means clustering is performed for a range of possible values of $K$.

The quality of the resulting clustering is assessed using the silhouette score, originally proposed by Rousseeuw [87]. For a sample $x$, the silhouette score is defined as

$$S(x) = \frac{b(x) - a(x)}{\max\left(a(x), b(x)\right)}, \tag{3.61}$$

where $a(x)$ is the mean distance between $x$ and the other samples within its own cluster, while $b(x)$ is the mean distance between $x$ and the samples in the nearest neighbouring cluster. The overall silhouette score is obtained by taking the mean of $S(x)$ over all samples. A larger value indicates that the samples are, on average, better separated into distinct clusters.

The silhouette score is evaluated for multiple values of $K$, with the value producing the largest mean score selected as the preferred number of clusters. The resulting cluster centres then provide representative points within the distinct regions of the parameter space. Each cluster centre is used as an initial guess in `scipy.optimize.minimize` [21], allowing a local optimisation of the objective function to be performed from each identified region. The minimum obtained across these local optimisations is then selected as the final optimum.

This procedure allows the MCMC calculation to serve as a global exploration of the multidimensional objective function landscape while avoiding the assumption that its samples represent a single approximately Gaussian region. Distinct modes can therefore be identified and subsequently refined using direct local optimisation, reducing the likelihood that an optimum is missed because the MCMC samples occupy multiple separated regions of parameter space.

### 3.8.2 Particle Swarm Optimisation

PSO is a population based optimisation algorithm in which a collection of candidate solutions, referred to as particles, simultaneously explores the multidimensional search space [88]. Each particle represents one possible solution and is described by a position vector $\vec{x}_i(t)$, where $i$ identifies the particle and $t$ denotes the iteration of the optimisation. Each particle also has an associated velocity vector $\vec{v}_i(t)$, which determines the change in its position between successive iterations.

The objective function is evaluated at the position of every particle. During the optimisation, each particle records the position at which it has previously obtained its lowest objective function value. This is referred to as the particle's personal best position and is denoted by $\vec{p}_i(t)$. The swarm also records the best position found by any particle, referred to as the global best position and denoted by $\vec{g}(t)$. The corresponding objective function values are $L(\vec{p}_i(t))$ and $L(\vec{g}(t))$, respectively.

The particle velocities are updated at each iteration according to three components: inertia, cognitive attraction and social attraction [89],

$$\vec{v}_i(t+1) = \omega\vec{v}_i(t) + c_1\vec{r}_1\left[\vec{p}_i(t) - \vec{x}_i(t)\right] + c_2\vec{r}_2\left[\vec{g}(t) - \vec{x}_i(t)\right], \tag{3.62}$$

where $\omega$ is the inertia coefficient, $c_1$ is the cognitive coefficient, $c_2$ is the social coefficient, and $\vec{r}_1$ and $\vec{r}_2$ are stochastic vectors.

The three terms in Equation 3.62 have distinct roles:

- **Inertia:** The term $\omega\vec{v}_i(t)$ retains part of the particle's previous velocity. This allows particles to maintain their existing direction of motion and promotes exploration of the search space.
- **Cognitive:** The term $c_1\vec{r}_1[\vec{p}_i(t) - \vec{x}_i(t)]$ attracts a particle towards its own best known position. This allows each particle to use information from its individual search history.
- **Social:** The term $c_2\vec{r}_2[\vec{g}(t) - \vec{x}_i(t)]$ attracts a particle towards the best position found by the swarm. This allows information discovered by one particle to influence the movement of all other particles and promotes convergence towards favourable regions.

The relative contributions of these three components are controlled by the coefficients $\omega$, $c_1$ and $c_2$. The inertia coefficient determines the influence of the particle's previous motion, while the cognitive and social coefficients determine the strength of the attraction towards the particle's own best position and the swarm's global best position, respectively.

Following the velocity update, the position of each particle is updated according to

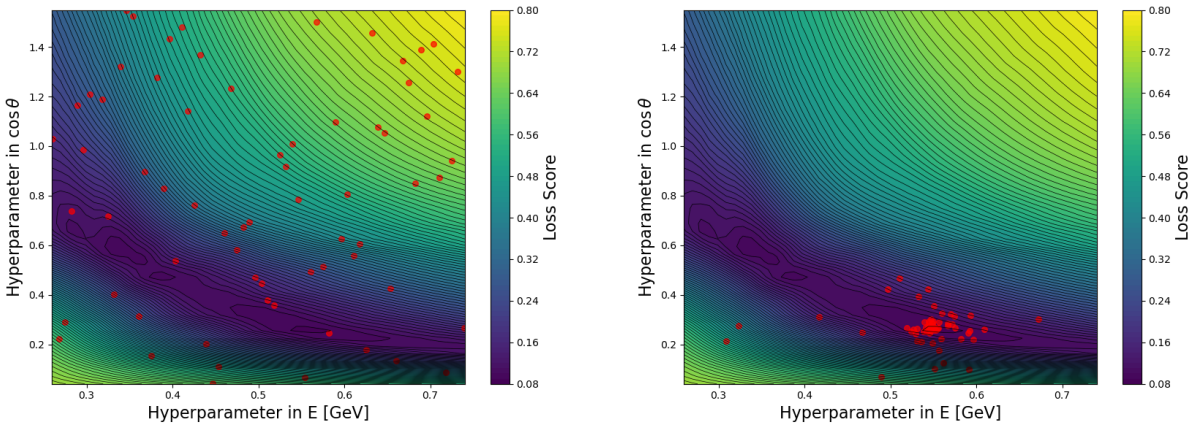
$$\vec{x}_i(t+1) = \vec{x}_i(t) + \vec{v}_i(t+1). \tag{3.63}$$

The objective function is then evaluated at the new position, and the personal and global best positions are updated if an improved solution has been found. Repeating this process causes the swarm to explore the search space while progressively concentrating around regions associated with low objective function values. An example of this evolution is shown in Figure 3.8.

In practical optimisation problems, the search space is often restricted to a finite region. Each dimension can therefore be assigned lower and upper bounds, such that $x_{j,\min} \leq x_{i,j}(t) \leq x_{j,\max}$, where $x_{i,j}$ is the $j$th component of the position vector of particle $i$. Similarly, limits can be placed on the velocity to prevent particles from making excessively large movements between iterations.

The optimisation is repeated until a specified convergence criterion or maximum number of iterations is reached. The global best position at the end of the optimisation provides the candidate solution corresponding to the lowest objective function value encountered by the swarm.

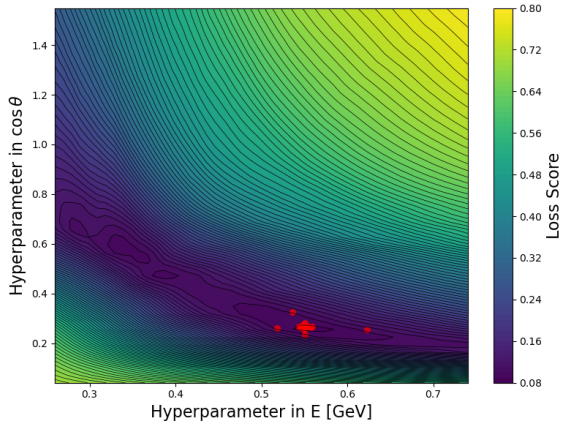
The principal distinction between MCMC and PSO is therefore their intended use. MCMC generates samples from a probability distribution and can be used to characterise the structure, uncertainties and correlations of a multidimensional space, while also allowing preferred or minimum regions to be identified. In contrast, PSO uses a population of candidate solutions to search directly for a minimum. Both methods can explore complex multidimensional landscapes without requiring the objective function to be analytically differentiable, making them useful for problems where conventional gradient based approaches may be unsuitable.




(a) Iteration 0.




(b) Iteration 50.




(c) Iteration 100.

Figure 3.8: Evolution of the particle positions during a Particle Swarm Optimisation PSO. The particles initially explore a broad region of the parameter space before progressively concentrating around regions associated with lower objective function values. The changing distribution illustrates the transition from broad exploration towards convergence on a favourable region of the search space.

## 3.9 Comparing CDFs

Several standard measures can be used to quantify the difference between two CDFs. These measures reduce the difference between two curves to a single scalar quantity, allowing the agreement between the corresponding probability distributions to be quantified. Different measures emphasise different aspects of the discrepancy between the curves. In particular, the Kolmogorov-Smirnov (KS) distance, the Cramér-von Mises (CvM) statistic, the Anderson-Darling (AD) statistic and the Wasserstein distance provide several established approaches for comparing one-dimensional distributions through their CDFs [90, 91]. These measures are introduced here in terms of two generic CDFs, denoted by $F_1(c)$ and $F_2(c)$.

### 3.9.1 Kolmogorov-Smirnov Distance

The Kolmogorov-Smirnov (KS) distance provides a simple measure of the largest difference between two CDFs. It is defined as

$$D_{\mathrm{KS}} = \sup_{c} |F_1(c) - F_2(c)|, \tag{3.64}$$

where sup denotes the supremum, or least upper bound, of the quantity over all values of $c$. In other words, $D_{\mathrm{KS}}$ is the smallest value that is greater than or equal to $|F_1(c) - F_2(c)|$ for every value of $c$. For continuous functions over a finite range, the supremum corresponds to the maximum vertical separation between the two CDF curves.

The KS distance therefore depends only on the point at which the two curves differ most. A large discrepancy confined to a small region can produce a large KS distance, while differences distributed across a broad range may produce the same value provided that their maximum separation is the same. Consequently, the KS distance provides a measure of the largest local discrepancy between two distributions rather than their overall agreement across the full range. The KS statistic is commonly used for goodness-of-fit and two-sample comparisons because of this direct interpretation in terms of the maximum separation between cumulative distributions [91].

### 3.9.2 Cramér-von Mises Distance

The Cramér-von Mises (CvM) approach instead considers the discrepancy between the two CDFs across their entire range. A general form can be written as

$$D_{\mathrm{CvM}} = \int [F_1(c) - F_2(c)]^2 \, w(c), dc, \tag{3.65}$$

where $w(c)$ is a weighting function that determines the relative importance assigned to discrepancies at different values of $c$. In other words, the weighting function controls how strongly a particular region of the CDF contributes to the overall measure. A larger value of $w(c)$ gives greater importance to differences between the two CDFs in that region, while a smaller value reduces their contribution. For a uniform weighting, $w(c) = 1$, all values of $c$ contribute equally

and the expression reduces to

$$D_{\mathrm{CvM}} = \int [F_1(c) - F_2(c)]^2 \, dc. \tag{3.66}$$

Unlike the KS distance, which considers only the largest difference, the CvM measure incorporates discrepancies throughout the entire range of $c$. The squared difference also means that larger discrepancies contribute disproportionately to the resulting value. Consequently, two pairs of CDFs with the same maximum separation can produce different CvM distances if the discrepancies are distributed differently across the range. The CvM statistic is therefore sensitive to the overall agreement between the two curves rather than only their largest local difference [91].

### 3.9.3 Anderson-Darling Distance

The Anderson-Darling (AD) approach extends the CvM concept by introducing a weighting function that gives greater importance to discrepancies in the tails of the distributions. A commonly used form is

$$D_{\mathrm{AD}} = \int \frac{[F_1(c) - F_2(c)]^2}{F_2(c)\,[1 - F_2(c)]}, dF_2(c). \tag{3.67}$$

The weighting factor

$$w(c) = \frac{1}{F_2(c)\,[1 - F_2(c)]} \tag{3.68}$$

increases as $F_2(c)$ approaches either zero or one. Consequently, differences between the two CDFs near the lower and upper tails receive greater weight than differences near the centre of the distribution. The AD measure is therefore particularly sensitive to discrepancies in the tails, making it useful when agreement in the extreme regions of a distribution is of particular interest [91].

The distinction between the CvM and AD approaches is therefore primarily the weighting applied to different regions of the distribution. With approximately uniform weighting, the CvM measure considers discrepancies across the distribution more evenly, whereas the AD measure explicitly increases the contribution from the tails. These different weighting schemes allow the same pair of CDFs to be assessed with different sensitivities to where the discrepancy occurs.

### 3.9.4 Wasserstein Distance

The Wasserstein distance provides another standard approach for comparing probability distributions. It is based on the concept of optimal transport, where the difference between two distributions is quantified by considering the minimum amount of "work" required to transform one distribution into the other [90]. This can be illustrated using the analogy of moving piles of dirt. If two piles represent two probability distributions, one distribution can be transformed into the other by moving probability mass between different locations. The work required to perform this transformation depends on both the amount of probability mass moved and the distance over which it is moved. The Wasserstein distance corresponds to the minimum total work required.

For one-dimensional probability distributions, the first-order Wasserstein distance has a particularly simple representation in terms of their corresponding CDFs,

$$W = \int |F_1(c) - F_2(c)|\, dc. \tag{3.69}$$

Thus, for 1D distributions, the Wasserstein distance is equal to the total area between the two CDF curves. Unlike the KS distance, which considers only the maximum separation, the Wasserstein distance incorporates the discrepancy across the entire range of $c$. Unlike the CvM and AD measures, which use the squared difference between the CDFs, the Wasserstein distance uses the absolute difference. The resulting quantity also has a direct interpretation in terms of the distance over which probability mass must be transported.

If the CDFs are evaluated on a uniformly spaced grid, Equation 3.69 can be approximated by the corresponding Riemann sum,

$$W \approx \sum_k |F_1(c_k) - F_2(c_k)|\, \delta c, \tag{3.70}$$

where $c_k$ denotes the $k$-th grid point and $\delta c$ is the spacing between adjacent points. This form provides a direct way of evaluating the Wasserstein distance when the CDFs are available only at discrete values of $c$.

### 3.9.5 Comparison of CDF Measures

The four measures described above quantify differences between two CDFs in different ways. The KS distance is determined by the largest vertical separation between the curves and is therefore sensitive to the maximum local discrepancy. The CvM measure considers the squared discrepancy across the full range, providing a measure of overall agreement. The AD measure uses a similar integrated squared discrepancy but places greater emphasis on the tails of the distributions. The Wasserstein distance integrates the absolute separation between the curves and, in one dimension, has the additional interpretation of the minimum transport cost between the two distributions.

These differences are illustrated schematically by considering two CDFs with discrepancies occurring in different regions. A metric based only on the maximum separation, such as the KS distance, does not distinguish between discrepancies with the same maximum magnitude but different extents. Measures based on an integrated difference, such as the CvM and Wasserstein distances, account for the extent of the discrepancy across the full range. The AD measure additionally changes the relative importance of the discrepancy according to its location, giving greater weight to the tails. The appropriate measure therefore depends on the properties of the distributions being compared and the aspect of their agreement that is of interest.

## 3.10 Interpolation and Extrapolation

A central consideration when applying a regression model to sparse data is whether predictions are made within the region supported by the observations or beyond it. This distinction is commonly described in terms of interpolation and extrapolation. Interpolation refers to prediction within the region of the input space for which observations are available, whereas extrapolation refers to prediction outside the region supported by the observed data. The distinction is particularly important for sparse datasets, where apparently small extensions beyond the measured region can correspond to substantial distances from the nearest observations.

For a GP, the distinction between interpolation and extrapolation is also related to the covariance structure specified by the kernel. For the RBF kernel introduced as Equation 3.37, the covariance

between two points decreases exponentially with their separation,

$$\kappa(x,x_)=\sigma_f^2\exp\left(-\frac{(x-x_)^2}{2\ell^2}\right), \tag{3.71}$$

where $\ell$ is the characteristic length scale. Consequently, two points separated by a distance much greater than the length scale,

$$|x-x_*|\gg\ell, \tag{3.72}$$

have negligible covariance and are effectively uncorrelated under the kernel.

This behaviour has a direct consequence for the GP predictive distribution. Predictions made sufficiently far from the observed data have increasingly little dependence on the observations and instead approach the prior distribution. For a zero prior mean and a unit prior variance, the predictive mean therefore approaches zero while the predictive variance approaches unity as the distance from the observed data becomes large. This behaviour is illustrated in Figure 3.9.

The length scale therefore provides a measure of the spatial range over which observations can influence a prediction. However, interpolation and extrapolation should not be regarded as being defined solely by the kernel length scale. Rather, the geometric extent of the observed dataset provides a natural first definition of the region in which predictions are supported by measurements. In one dimension this region is straightforward to identify, while in higher dimensions its geometry becomes more complicated.

For a one-dimensional dataset, the region spanned by the observations is given by the closed interval

$$[x_{\min},x_{\max}], \tag{3.73}$$

where $x_{\min}$ and $x_{\max}$ are the minimum and maximum observed input values. Predictions within this interval are geometrically interpolative, while predictions outside it are geometrically extrapolative.

In higher dimensional spaces, the component wise minimum and maximum values define an axis-aligned hyper-rectangle. Although this provides a simple extension of the one-dimensional interval, it can contain substantial regions in which no observations are present. Consequently, the component-wise bounds do not necessarily provide an appropriate representation of the region supported by the data.

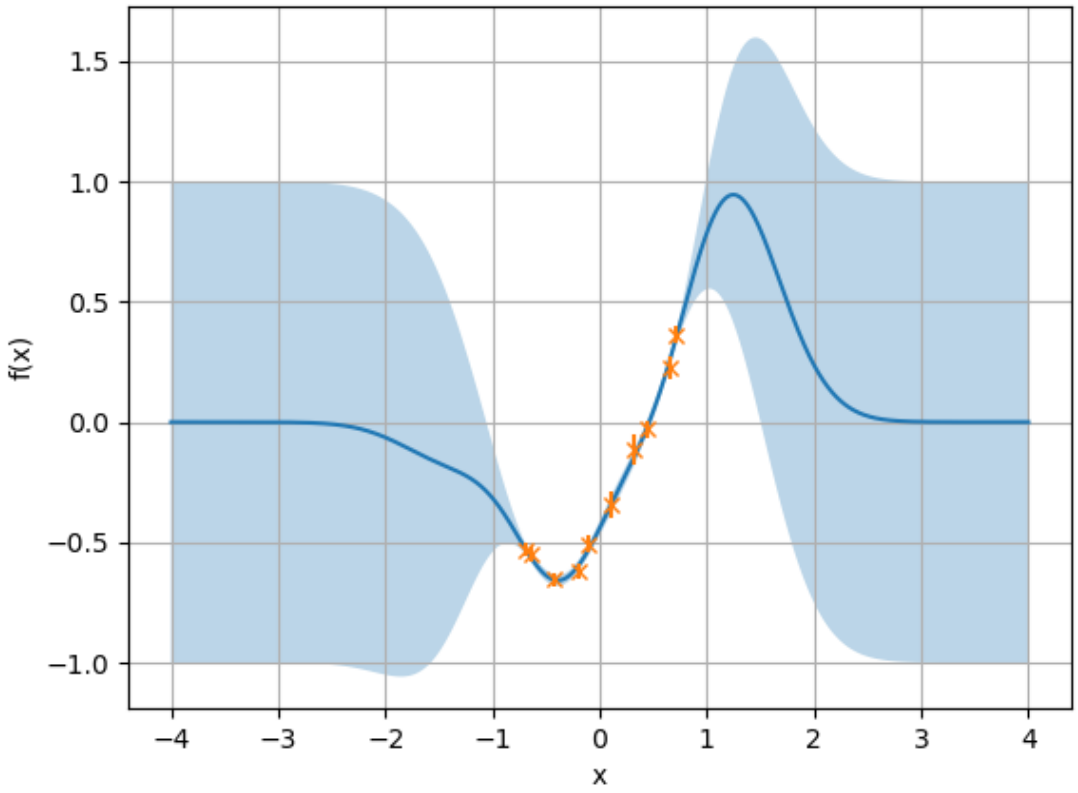



Figure 3.9: Example GP extrapolation plot using an RBF kernel. As the prediction point moves increasingly far from the observed data relative to the kernel length scale, $\ell$, the covariance between the prediction point and the observations approaches zero. Consequently, the predictive mean approaches the prior mean of zero and the predictive variance approaches the prior variance of unity.

A more general geometric definition can be obtained using the convex hull of the observed points. The convex hull provides the smallest convex set containing all observations and therefore provides a natural generalisation of the one-dimensional interval to higher-dimensional input spaces. The mathematical definition and properties of the convex hull are discussed in the following section.

## 3.11 Convex Hulls

The convex hull is the smallest convex set containing a given set of points. A set $S \subset \mathbb{R}^D$ is convex if, for any two points $a, b \in S$, the entire line segment connecting them is also contained

within $S$ [92]. This can be expressed mathematically as

$$\{\theta a + (1-\theta)b, |, \theta \in [0,1]\} \subset S. \tag{3.74}$$

The expression

$$\theta a + (1-\theta)b \tag{3.75}$$

is a convex combination of the points $a$ and $b$. The coefficients are non-negative and sum to unity, ensuring that the resulting point lies on the line segment connecting $a$ and $b$. Convexity therefore requires that a set contain the complete line segment between every pair of its points.

The convex hull generalises this idea from pairs of points to any finite collection of points. To illustrate this, consider three points $x_1, x_2, x_3 \in X$. First, a convex combination of two points can be formed as

$$y = \alpha x_1 + (1-\alpha)x_2, \qquad \alpha \in [0,1]. \tag{3.76}$$

By the definition of convexity, $y$ lies on the line segment connecting $x_1$ and $x_2$ and is therefore contained within the convex hull of these points. A convex combination of $y$ with the third point $x_3$ can then be formed:

$$z = \beta y + (1-\beta)x_3, \qquad \beta \in [0,1]. \tag{3.77}$$

Substituting the expression for $y$ gives

$$\begin{aligned} z &= \beta\left(\alpha x_1 + (1-\alpha)x_2\right) + (1-\beta)x_3 \\ &= (\beta\alpha)x_1 + (\beta(1-\alpha))\,x_2 + (1-\beta)x_3. \end{aligned}$$

The coefficients in this expression are all non-negative and sum to unity:

$$\begin{aligned} &\beta\alpha + \beta(1-\alpha) + (1-\beta) \\ =&\beta(\alpha + 1 - \alpha) + (1-\beta) \\ =&\beta + (1-\beta) \\ =&1. \end{aligned}$$

Thus, the point $z$ can be expressed as a weighted sum of all three points, with non-negative weights whose sum is unity. This demonstrates how repeated convex combinations naturally lead to the more general definition of a convex hull.

For a finite set of observations

$$X = x_1, \ldots, x_N \subset \mathbb{R}^D, \tag{3.78}$$

the convex hull is therefore defined as

$$\mathrm{conv}(X) = \left\{ \sum_{i=1}^{N} \lambda_i x_i, \middle| , \lambda_i \geq 0, ; \sum_{i=1}^{N} \lambda_i = 1 \right\}. \tag{3.79}$$

Each point within the convex hull can therefore be expressed as a convex combination of the observed points. The coefficients $\lambda_i$ determine the contribution of each observed point to the resulting position, while the conditions

$$\lambda_i \geq 0, \qquad \sum_{i=1}^{N} \lambda_i = 1 \tag{3.80}$$

ensure that the resulting point remains within the geometric envelope defined by the observations.

In one dimension, the convex hull reduces to the interval between the minimum and maximum observed values,

$$\mathrm{conv}(X) = [x_{\min}, x_{\max}], \tag{3.81}$$

recovering the usual definition of the region spanned by a one-dimensional dataset. In two or more dimensions, the convex hull instead forms a polygon, polytope or higher-dimensional convex body enclosing the observed points. An example of a two-dimensional convex hull is shown in Figure 3.10.

The convex hull therefore provides a geometry based definition of the region spanned by a set of observations. It avoids the inclusion of regions that are contained within the component-wise bounds of the dataset but lie outside the actual geometric extent of the measured points. This makes the convex hull a useful basis for defining a prediction domain for multidimensional datasets.

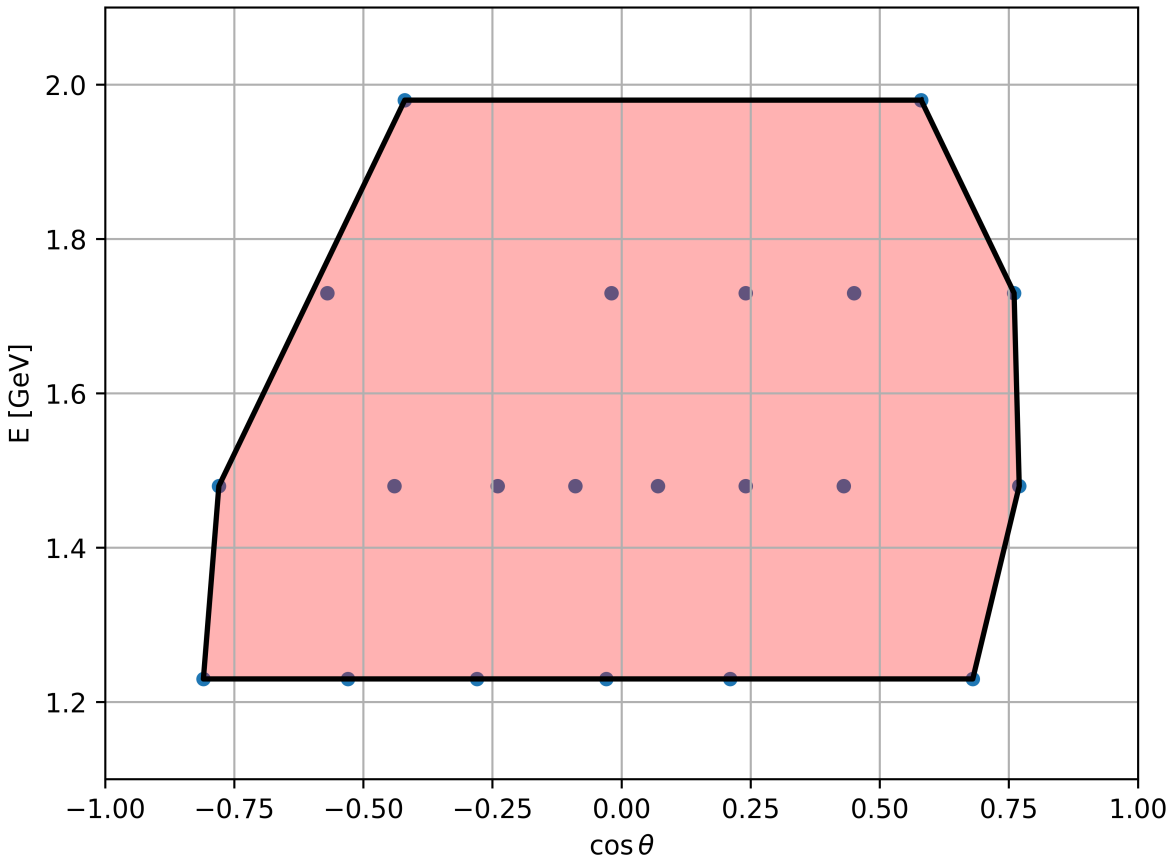



Figure 3.10: 2D convex hull constructed from polarisation observables measured as a function of $\cos\theta$ and energy in [82]. The convex hull (here shaded in red) defines the smallest convex region containing the measured data points, providing a geometrically permissible interpolation domain in the two-dimensional input space.

# Chapter 4

# Gaussian Process Framework for Hadron Data

## 4.1 Overview of the GP Framework

Chapter 2 established that experimental hadron physics data provide sparse and non-uniform measurements of observables across kinematic spaces, while Chapter 3 introduced the statistical foundations required to model such data using GPs. This chapter describes the specific GP framework developed in this thesis for interpolating experimental hadron physics observables and quantifying the associated predictive uncertainty.

The framework is designed to address the requirements associated with sparse experimental measurements. In particular, the hyperparameters of the GP must be selected not only to provide a suitable description of the available measurements, but also to produce reliable predictive uncertainties and physically sensible behaviour between measured points. The interpolation region must also be distinguished from regions where the model is required to extrapolate beyond the available experimental constraints.

The complete framework consists of several stages. The experimental measurements are first represented in a form suitable for the GP, with the kinematic variables forming the input space and the measured observable providing the target quantity. The associated experimental uncer-

tainties are incorporated into the model, after which the GP is specified using an appropriate mean function and covariance kernel. The kernel hyperparameters are then determined using the optimisation procedure developed in this thesis. Finally, the region of kinematic space in which the GP is constrained by the available measurements is identified, allowing predictions to be distinguished from extrapolations.

The following sections describe each stage of the framework, including the representation of the experimental data, specification of the GP, selection of the covariance kernel and identification of the interpolation region. Particular attention is given to the development of two methodological components: the optimisation of the GP hyperparameters using a combination of statistical and physics-motivated criteria, and the construction of a discretised convex hull to identify the region of kinematic space supported by the available measurements. The computational implementation of the framework and the complete analysis workflow are also described.

## 4.2 Experimental Data

The experimental measurements considered in this thesis are used directly as the constraints on the GP. Each measurement consists of a set of kinematic variables, a measured observable and an associated experimental uncertainty. The dimensionality and density of the datasets differ between the two applications considered.

For the polarisation observables, the measurements are described by two kinematic variables: the reaction energy and $\cos\theta$, where $\theta$ is the scattering angle. The measurements are distributed across four discrete energy bins spanning approximately 1.2–2.0 GeV. Each energy bin contains between 2 and 8 measurements at different values of $\cos\theta$, resulting in sparse and non-uniform sampling of the two-dimensional kinematic space [82].

The $\bar{K}N$ scattering data are 1D, with the incident energy providing the kinematic variable. The data span approximately 0–300 MeV and comprise measurements from a range of individual experiments. Each experiment contains between 3 and 8 measurements, resulting in particularly sparse sampling of the underlying energy dependence [93–108][1].

[1] It is interesting to note here that [102] is an unpublished technical report but is listed as a reference in several published papers [109, 110].

No interpolation or smoothing is applied to the experimental measurements before they are supplied to the GP. The resulting measured values, kinematic coordinates and associated uncertainties are used directly in the GP calculations.

## 4.3 GP Model Specification

The GP framework used in this thesis adopts a zero mean function,

$$m(X) = 0. \tag{4.1}$$

The use of a zero mean function is a standard choice in GP regression, particularly when the primary information about the underlying function is provided by the observed data rather than by a well-motivated parametric mean function [6]. The mean function represents the prior expectation of the latent function before the experimental measurements are taken. Setting this function to zero therefore does not imply that the physical observable is expected to be zero. Instead, it specifies that no predetermined functional dependence is imposed on the observable in the absence of data.

This choice is particularly appropriate for the applications considered in this thesis. The aim is to interpolate experimentally measured observables without imposing a specific functional form motivated by a phenomenological model. Introducing a non-zero parametric mean function would require additional assumptions about the expected behaviour of the observable and could influence the resulting interpolation, particularly in regions where the available measurements provide weak constraints. The zero mean function instead provides a neutral prior mean, with the experimental measurements determining the posterior prediction in the regions where the data provide information. This also avoids introducing additional parameters associated with the specification and optimisation of a mean function.

With this choice, and after applying the Cholesky decomposition described in Section 3.5.1, the predictive mean is therefore given by

$$\mu_* = K_*^T \alpha, \tag{4.2}$$

while the predictive covariance is

$$\Sigma_* = K_{**} - v^T v. \tag{4.3}$$

These are the expressions used to calculate the predictive mean and uncertainty throughout the framework.

The covariance function used throughout this work is the RBF kernel introduced in Section 3.6 as Equation 3.37. The choice of covariance function is motivated by the nature of the experimental hadron physics data considered in this thesis. As discussed in Section 2.7, these datasets provide discrete and often sparse measurements of physical observables across the underlying kinematic space, while the observables themselves are expected to vary continuously with the kinematic variables. Consequently, the interpolation should provide a smooth representation between the measured points without introducing local variations that are not supported by the available experimental information.

The RBF kernel provides a suitable prior for this purpose because functions drawn from an RBF GP are infinitely differentiable. As discussed in Section 3.6, this represents a strong assumption about the smoothness of the underlying function: not only the observable itself, but derivatives of all orders are continuous. For the applications considered here, this provides a desirable constraint on the behaviour of the posterior between measurements. In regions where experimental data are sparse, the posterior is consequently encouraged to vary smoothly according to the characteristic length scales rather than developing rapid changes between individual measurements that are not directly supported by the data.

The limited number of measurements in the experimental datasets considered here means that there can be substantial freedom in the behaviour of a flexible interpolating function between neighbouring data points. A kernel with a finite degree of differentiability, such as a Matérn kernel with finite $\nu$, would permit less smooth functions and could therefore accommodate more local variation. While this flexibility can be advantageous when such behaviour is physically expected, it is not required for the interpolation considered here. The infinitely differentiable functions produced by the RBF kernel instead provide a consistent smoothness prior across the kinematic domain.

The choice of the RBF kernel therefore represents a deliberate balance between flexibility and the assumptions required to obtain a physically sensible interpolation. The kernel provides sufficient flexibility through the independent length scales to accommodate different rates of variation across the kinematic dimensions, while its infinite differentiability prevents the posterior from developing unnecessarily irregular behaviour in regions that are weakly constrained by the experimental data. This is particularly important for the sparse datasets considered in this

thesis, where the posterior behaviour between measurements is determined substantially by the covariance structure.

The amplitude $\sigma_f^2$ appearing in the general form of Equation 3.37 is not included as a free hyperparameter in the implementation used here. Omitting this parameter reduces the number of quantities that must be determined from the sparse experimental datasets. The length scales are therefore the principal hyperparameters of the covariance function and are determined using the optimisation procedure developed in the following section. The RBF kernel used here is therefore of the form

$$\kappa\left(\vec{a},\vec{b}\right) = \exp\left(-\frac{1}{2}\sum_{i=1}^{d}\frac{(a_i-b_i)^2}{\ell_i^2}\right). \tag{4.4}$$

## 4.4 Hyperparameter Optimisation for Sparse Datasets

As established in Section 3.7, the standard marginal-likelihood approach does not sufficiently constrain the GP hyperparameters for the sparse datasets considered in this thesis. A new hyperparameter optimisation procedure is therefore developed specifically for this setting. Rather than relying on the likelihood of the observed measurements alone, the procedure evaluates the statistical behaviour of the resulting GP interpolation and uses this to determine suitable values of the RBF kernel length scales.

The optimisation is based on the predictive mean and uncertainty produced by the GP. For a given set of hyperparameters (length scales), the predictions are compared with the observed measurements over a range of uncertainty intervals. This provides a direct assessment of whether the predictive uncertainties are appropriately calibrated. The resulting discrepancy from the expected statistical behaviour is then quantified using the Wasserstein distance introduced in Section 3.9. Physics-informed bounds and a boundary penalty are additionally incorporated to restrict the optimisation to a reasonable region of hyperparameter space.

The optimisation procedure consists of the following stages:

1. **Evaluate empirical coverage:** For a given set of kernel length scales, the GP predictive mean and uncertainty are used to determine the fraction of observed datapoints contained within a range of predictive uncertainty intervals.

2. **Calculate the Wasserstein distance:** The empirical coverage is compared with the nominal coverage expected for a Gaussian predictive distribution. The discrepancy between the two coverage curves is quantified using the one-dimensional Wasserstein distance.

3. **Apply physics-informed bounds:** Bounds are imposed on each kernel length scale based on the distribution of the available input data. These bounds restrict the optimisation to a region in which the length scales remain related to the characteristic spacing and extent of the measured data.

4. **Apply the boundary penalty:** A small penalty is introduced for length scales approaching the imposed boundaries. This discourages the optimisation from accumulating at the limits of the allowed parameter space while still permitting boundary solutions when they provide the best overall description of the data.

5. **Optimise the combined loss:** The Wasserstein distance and boundary penalty are combined to form the final loss function. The optimal length scales are obtained by minimising this loss over the allowed hyperparameter space.

The individual components of this procedure are described in the following subsections.

### 4.4.1 Empirical Coverage

For a given set of GP hyperparameters, the predictive mean and uncertainty can be used to assess the statistical behaviour of the resulting model. In particular, the predictive uncertainty can be evaluated through its empirical coverage. The statistical definition of empirical coverage was introduced in Section 3.4; here, it is used as a criterion for assessing different choices of the kernel hyperparameters.

Since the GP provides both a predictive mean and standard deviation, the fraction of measured datapoints that fall within a specified number of predictive standard deviations can be calculated and compared with the corresponding nominal coverage. For a given value of $c_k$, where $c_k$ represents the number of predictive standard deviations defining the interval, the inclusion of the $i$-th datapoint within this interval is defined by

$$S_{k,i} = \begin{cases} 1, & \text{if } \left|y_{\text{fit},i} - y_{\text{data},i}\right| \leq c_k e_{\text{fit},i}, \\ 0, & \text{otherwise.} \end{cases} \tag{4.5}$$

Here, $y_{\text{fit},i}$ and $e_{\text{fit},i}$ denote the GP predictive mean and standard deviation at the input location of the $i$-th datapoint, respectively, while $y_{\text{data},i}$ is the corresponding experimental measurement.

Using this indicator, the measured coverage at each value of $c_k$ is calculated as

$$F_{\text{meas},k} = \frac{1}{n} \sum_{i=1}^{n} S_{k,i}, \tag{4.6}$$

where $n$ is the total number of datapoints. The resulting values of $F_{\text{meas},k}$ give the fraction of measurements contained within the corresponding predictive uncertainty interval.

The measured coverage can also be interpreted as an empirical CDF. It gives the fraction of absolute standardised residuals, defined by the difference between the measured value and the GP predictive mean normalised by the predictive standard deviation, that are less than or equal to $c_k$. It is therefore the empirical CDF of the absolute standardised residuals, evaluated at $c_k$. This interpretation allows the measured coverage to be compared directly with the predicted coverage as two one-dimensional CDFs.

To assess the behaviour of the predictive uncertainties over a range of interval sizes, values of $c_k$ between 0 and 3 are considered. For a correctly calibrated Gaussian predictive distribution, the standardised residuals are expected to follow a standard normal distribution. Since the coverage interval is defined by the absolute deviation from the predictive mean, the relevant random variable is the absolute value of a standard normal variable. The corresponding predicted coverage is therefore the probability that a standard normal random variable lies within $c_k$ standard deviations of its mean.

The predicted coverage can therefore be written as

$$F_{\text{pred},k} = \Phi(c_k) - \Phi(-c_k), \tag{4.7}$$

where $\Phi$ is the standard normal CDF. This quantity gives the probability that a standard normal random variable lies within $c_k$ standard deviations of its mean. Equivalently, $F_{\text{pred},k}$ is the CDF of the absolute value of a standard normal random variable, evaluated at $c_k$. It therefore represents the expected fraction of measurements contained within $c_k$ predictive standard deviations for a correctly calibrated Gaussian predictive distribution.

The empirical coverage plot is constructed by plotting the nominal coverage, $F_{\text{pred},k}$, on the

horizontal axis and the corresponding measured coverage, $F_{\text{meas},k}$, on the vertical axis. The two quantities therefore represent the expected and measured fractions of datapoints contained within the corresponding predictive uncertainty intervals. Since both quantities are CDFs evaluated at the same value of $c_k$, perfect calibration corresponds to $F_{\text{meas},k} = F_{\text{pred},k}$ which gives the diagonal line in the empirical coverage plot. An example is shown in Figure 4.1.

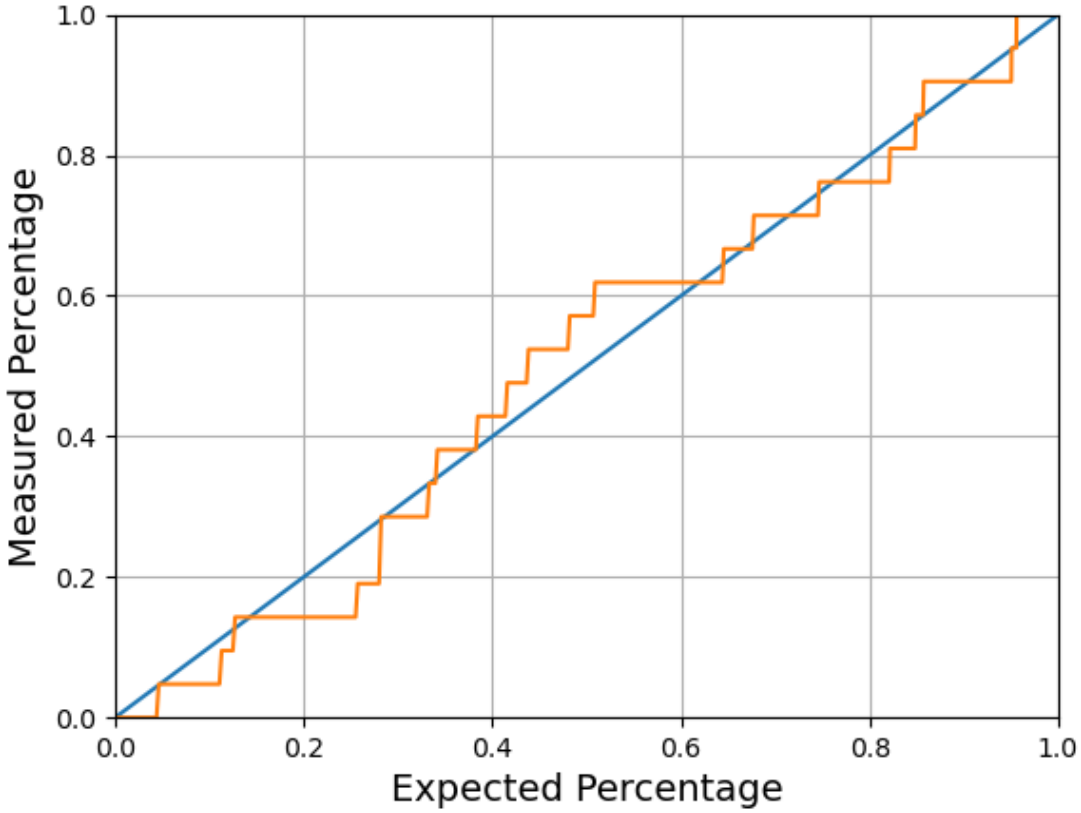



Figure 4.1: Example empirical coverage plot for a specified set of GP hyperparameters. The nominal coverage, $F_{\text{pred}}$, expected for a Gaussian predictive distribution is shown on the horizontal axis, while the measured coverage, $F_{\text{meas}}$, obtained from the GP predictive mean and uncertainty is shown on the vertical axis. The diagonal line represents perfect calibration, for which $F_{\text{meas}} = F_{\text{pred}}$. The nominal coverage is calculated from the standard normal CDF using Equation 4.7. During the optimisation procedure, the hyperparameters are varied and the resulting measured coverage curves are evaluated to minimise their discrepancy from the nominal expectation.

For a well-calibrated predictive distribution, the measured coverage should closely follow the diagonal line over the full range of nominal coverage. If the predictive uncertainties are too small, the measured coverage will tend to fall below the nominal coverage, indicating under-coverage. Conversely, excessively large uncertainties result in over-coverage, with the measured curve lying above the nominal expectation.

The hyperparameters are therefore varied during the optimisation procedure, with the coverage calculation repeated for each set of values considered. Each set of hyperparameters produces a corresponding set of measured and nominal coverage values, allowing the statistical behaviour of the resulting GP to be assessed. The discrepancy between the measured and nominal coverage is then quantified using the Wasserstein distance, providing the scalar loss used in the subsequent optimisation.

### 4.4.2 Wasserstein Distance

The empirical coverage procedure described in Section 4.4.1 provides a measured coverage curve, $F_{\text{meas}}$, which is the empirical CDF of the absolute standardised residuals. This can be compared with the nominal coverage, $F_{\text{pred}}$, which is the corresponding CDF expected for a correctly calibrated Gaussian predictive distribution. A scalar measure of the discrepancy between these two CDFs is therefore required in order to compare different choices of the GP hyperparameters during the optimisation procedure. Several measures for comparing CDFs were introduced in Section 3.9, including the KS, CvM, AD and Wasserstein distances.

For the present application, the Wasserstein distance is selected because it provides an integrated measure of the discrepancy between the measured and nominal coverage across the full range of interval sizes. This is desirable because the optimisation aims to obtain predictive uncertainties that exhibit appropriate coverage behaviour across a range of uncertainty intervals, rather than minimising only the largest deviation at a particular coverage level.

The KS distance considers only the maximum separation between the two curves and therefore does not account for the extent of discrepancies across the remainder of the coverage range. The CvM statistic incorporates discrepancies across the full range but uses a squared difference, giving greater weight to larger deviations. The AD statistic similarly uses a squared difference while additionally weighting the tails of the distributions. These properties are not required for the present application, where the aim is to measure the overall discrepancy from the expected coverage behaviour without preferentially weighting either the largest deviations or the tails. The Wasserstein distance, introduced in Section 3.9, instead uses the absolute difference between the two CDFs and incorporates their separation across the full range. It therefore provides a suitable scalar measure for the coverage-based optimisation used here.

For the coverage curves considered in this work, the Wasserstein distance is evaluated using the measured and nominal coverage values generated for each value of $c_k$. Using the discrete form introduced in Equation 3.70, the distance is calculated as

$$W \approx \sum_k \left| F_{\text{meas},k} - F_{\text{pred},k} \right| \delta c, \tag{4.8}$$

where $F_{\text{meas},k}$ is the measured coverage obtained from Equation 4.6, $F_{\text{pred},k}$ is the corresponding nominal coverage obtained from Equation 4.7, and $\delta c$ is the spacing between adjacent values of $c_k$.

The resulting value of $W$ therefore quantifies the overall discrepancy between the measured and expected coverage behaviour. A value of $W = 0$ corresponds to perfect agreement between the two coverage CDFs, while larger values indicate increasing deviations from the expected coverage. In terms of the empirical coverage plot, the Wasserstein distance is therefore related to the accumulated separation between the measured coverage curve and the diagonal line representing perfect calibration.

For each set of GP hyperparameters, the coverage calculation is repeated and a corresponding value of $W$ is obtained. The Wasserstein distance therefore provides the primary statistical component of the loss function used to determine the optimal kernel length scales. However, the statistical coverage criterion alone does not impose constraints on the physical scale of the resulting length scales. Physics-informed bounds and a boundary penalty are therefore introduced in the following subsection.

### 4.4.3 Physics-Informed Boundaries

When performing an optimisation, it is useful to impose suitable boundaries on the parameter space to ensure that the resulting optimum remains physically reasonable. For the GP hyperparameters considered in this work, the available experimental data provide a natural basis for defining these boundaries. In particular, the characteristic length scales should remain related to the spatial scales over which the experimental data provide constraints.

For each input (kinematic) dimension $i \in \{1, \ldots, D\}$, let $\{x_{i,j}\}_{j=1}^{N}$ denote the values of that dimension sorted in ascending order. The spacings between consecutive datapoints are then

defined as

$$\delta x_{i,j} = x_{i,j+1} - x_{i,j}, \qquad j = 1, \ldots, N-1. \tag{4.9}$$

The lower bound $L_i$ is defined as the smallest non-zero spacing between unique datapoints in the corresponding input dimension. This provides a minimum characteristic scale that is resolved by the available data and prevents the optimisation from selecting length scales smaller than the smallest non-zero separation between measured points. If no non-zero spacing exists, corresponding to all values in that dimension being identical, the lower bound is set to zero.

The upper bound $U_i$ is defined by the total range of the available data in the corresponding input dimension. This ensures that the maximum allowed length scale remains related to the extent of the region over which experimental measurements are available. The bounds are therefore determined directly from the distribution of the measured input coordinates rather than being selected as arbitrary numerical limits. For each kinematic dimension, they are given by

$$\begin{aligned} L_i &= \begin{cases} \min\limits_j \delta x_{i,j}, & \text{if } \exists\, j \text{ such that } \delta x_{i,j} > 0, \\ 0, & \text{otherwise,} \end{cases} \\ U_i &= x_{i,N} - x_{i,1}. \end{aligned} \tag{4.10}$$

These bounds define the allowed region of hyperparameter space over which the subsequent optimisation is performed. They are described as physics-informed because they are constructed from the kinematic scales resolved by the experimental measurements and therefore constrain the GP to length scales relevant to the physical region represented by the data. They should not, however, be interpreted as fundamental physical limits on the possible variation of the underlying observable. Rather, they provide data-informed safeguards against selecting characteristic scales that are poorly supported by the available measurements.

### 4.4.4 Combined Loss Function

The Wasserstein distance introduced in Section 4.4.2 provides a measure of the statistical agreement between the measured and nominal coverage curves. The physics-informed bounds introduced in Section 4.4.3 additionally restrict the allowed hyperparameter space. These constraints alone, however, do not prevent the optimum from occurring directly on one of the imposed

boundaries.

An optimum located on a boundary is not necessarily invalid, but it indicates that the optimisation would prefer a value outside, or at the extreme edge, of the allowed parameter space. An interior optimum is therefore preferable where possible, as it indicates that the selected hyperparameters are determined primarily by the structure of the data and the statistical objective rather than by the imposed limits. An interior solution also provides stronger evidence that the physics-informed boundaries are acting as safeguards rather than actively determining the result.

To encourage interior solutions, a boundary proximity penalty is introduced into the loss function. For each length scale $\ell_i$, a score is defined to quantify its proximity to the corresponding lower and upper bounds, $L_i$ and $U_i$. The score is zero when the length scale lies on either boundary and increases linearly towards the midpoint between the bounds, where it reaches a maximum value of 0.5. This is defined as

$$S_i = \min\left(\frac{\ell_i - L_i}{U_i - L_i}, \frac{U_i - \ell_i}{U_i - L_i}\right). \tag{4.11}$$

The minimum score across all length scales is then used,

$$S = \min_i(S_i). \tag{4.12}$$

This ensures that the penalty is determined by the length scale closest to one of its boundaries. Consequently, the penalty reflects the most boundary constrained dimension rather than averaging the proximity of the individual length scales.

The boundary penalty is then defined as

$$P(\ell) = 1 - 2S. \tag{4.13}$$

The resulting penalty lies between zero and one for length scales within their allowed bounds. A value of $P(\ell) = 1$ corresponds to at least one length scale lying directly on a boundary, while $P(\ell) = 0$ corresponds to all length scales lying at the midpoint of their respective allowed ranges. Intermediate values describe the degree of proximity to the nearest boundary. An illustrative one-dimensional example of this penalty is shown in Figure 4.2, where the lower and upper bounds are $L = 0$ and $U = 1$, respectively.

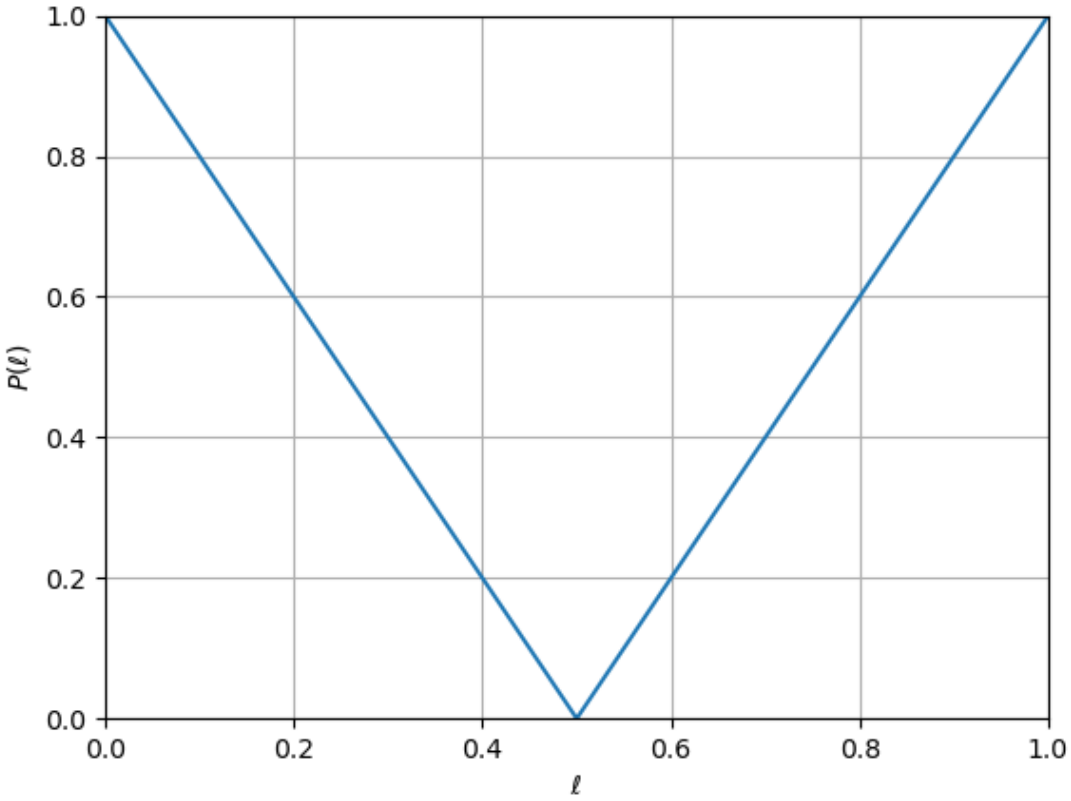



Figure 4.2: Example one-dimensional boundary penalty $P(\ell)$ for lower and upper bounds $L = 0$ and $U = 1$. The penalty is zero at the midpoint of the allowed interval and increases linearly towards either boundary, where it reaches its maximum value of one.

The boundary penalty is incorporated into the Wasserstein loss with a weighting of 1%. The resulting combined loss is therefore

$$L_{\text{total}} = W + 0.01P(\ell), \tag{4.14}$$

where $W$ is the Wasserstein distance between the measured and nominal coverage curves and $P(\ell)$ is the boundary proximity penalty.

The small weighting of the boundary term ensures that the optimisation remains dominated by the primary Wasserstein loss. Consequently, the penalty does not substantially distort the loss surface or override a statistically meaningful minimum in favour of an interior solution. Instead, it provides a small preference for solutions away from the imposed boundaries when competing solutions have similar Wasserstein losses. In particular, a boundary solution is not excluded: if the Wasserstein loss is sufficiently lower at a boundary than elsewhere in the allowed parameter

space, the 1% penalty is insufficient to prevent the boundary solution from being selected.

The combined loss therefore balances two complementary requirements. The Wasserstein term determines how well the predictive uncertainties reproduce the expected statistical coverage, while the boundary penalty discourages the optimisation from selecting length scales at the limits of the physically informed parameter space when this provides no substantial statistical advantage. The resulting objective can consequently be used to identify hyperparameters that provide appropriate uncertainty calibration while remaining away from artificially imposed boundaries where possible. An example of the resulting combined loss surface for the $O_z$ polarisation observable is shown in Figure 4.3, where the minimum is well constrained within the allowed hyperparameter space.

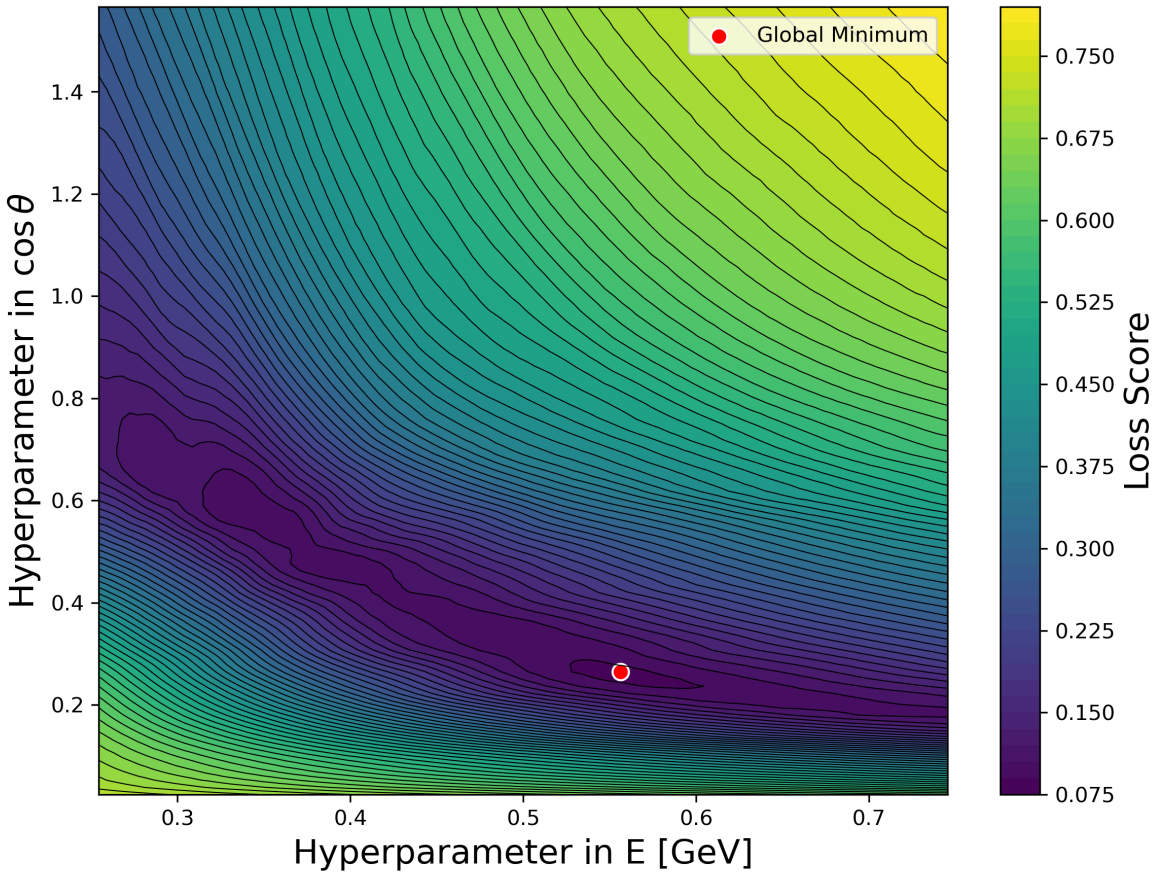



Figure 4.3: Combined loss surface calculated for the $O_z$ polarisation observable [82] as a function of the two RBF kernel length scales. The surface is evaluated using the Wasserstein distance between the measured and nominal coverage curves together with the boundary proximity penalty defined in Equation 4.14. The minimum corresponds to the combination of length scales that provides the lowest overall loss and therefore represents the optimal hyperparameter choice for this dataset. The relatively well-constrained minimum provides a distinct preferred region within the allowed hyperparameter space, in contrast to the poorly constrained behaviour observed for the marginal-likelihood in Figure 3.6.

## 4.5 Optimisation Strategy

The hyperparameters of the GP are determined by minimising the combined loss function described in Section 4.4.4. The optimisation is performed using PSO, with each particle representing a candidate set of kernel length scales. The general principles and mathematical formulation of PSO were introduced in Section 3.8.

PSO was selected as the primary optimisation method because it is specifically designed to search for an optimum of an objective function and does not require the loss function to be differentiable. This is well suited to the present application, where the loss incorporates empirical coverage and therefore does not have a simple analytical form. Each particle is evaluated by constructing the corresponding GP, calculating its predictive mean and uncertainty, and evaluating the empirical coverage, Wasserstein distance and boundary penalty. These quantities are combined according to Equation 4.14 to provide the loss associated with the particle.

Although PSO is well suited to this optimisation problem, MCMC is a commonly used approach in hadron physics analyses for exploring multidimensional parameter spaces and identifying regions of favourable solutions. It was therefore important to assess whether the choice of optimisation method affected the resulting solutions. Both MCMC and PSO were applied to the same set of pseudodatasets, described in more detail in Chapter 5, and evaluated using the same combined loss function (Equation 4.14). The comparison showed that PSO consistently obtained lower loss values while also requiring less computational time than MCMC for the pseudodatasets considered. Based on these results, PSO was adopted as the primary optimisation method used throughout this work. A detailed comparison of the two approaches, including the resulting loss values and computational runtimes, is provided in Appendix A.

The velocity of each particle is updated according to the PSO Equation 3.62 introduced in Section 3.8,

$$\vec{v}_i(t+1) = \omega\vec{v}_i(t) + c_1\vec{r}_1\left[\vec{p}_i(t) - \vec{x}_i(t)\right] + c_2\vec{r}_2\left[\vec{g}(t) - \vec{x}_i(t)\right], \tag{4.15}$$

where the three terms correspond to the inertia, cognitive and social components, respectively. The weighting of these components determines the balance between continued exploration of the parameter space and convergence towards previously identified favourable solutions.

For the implementation used in this work, the inertia coefficient is varied with iteration according

to

$$\omega(t) = \max\left(0.4, 0.9 - 0.002t\right). \tag{4.16}$$

The relatively large initial value of $\omega$ allows particles to retain a substantial fraction of their previous velocity, promoting broad exploration of the hyperparameter space during the early stages of the optimisation. As the optimisation progresses, the decreasing inertia coefficient reduces the influence of the previous velocity and shifts the swarm towards more local exploration around favourable regions. The lower bound of $\omega = 0.4$ ensures that the inertia contribution does not vanish completely, retaining some capacity for particles to move away from localised regions of the search space.

The cognitive and social coefficients are both set to

$$c_1 = c_2 = 1.4. \tag{4.17}$$

The equal weighting of the cognitive and social components provides a balance between the tendency of particles to return towards their own best-known positions and the tendency of the swarm to move towards the best solution identified globally. This allows individual particles to continue exploring the parameter space while maintaining convergence towards regions of low loss.

The stochastic vectors appearing in Equation 3.62 are independently sampled according to

$$\vec{r}_1, \vec{r}_2 \sim \mathscr{U}(0,1)^D, \tag{4.18}$$

where $D$ is the dimensionality of the hyperparameter space. The multiplication between these stochastic vectors and the displacement vectors in Equation 3.62 is performed component-wise. Thus, for each dimension $j$,

$$\left[\vec{r}_1 \left(\vec{p}_i - \vec{x}_i\right)\right]_j = r_{1,j}\left(p_{i,j} - x_{i,j}\right), \tag{4.19}$$

with an analogous expression for the social component. The stochastic vectors therefore independently scale the magnitude of the cognitive and social contributions in each dimension rather than performing a vector dot product.

Although the components of $\vec{r}_1$ and $\vec{r}_2$ are restricted to positive values, the resulting velocity contributions are not restricted to positive directions. The direction in each dimension is determined by the corresponding displacement between the current particle position and the target

position. For example, if $x_{i,j} > p_{i,j}$, then $p_{i,j} - x_{i,j} < 0$, giving a negative cognitive contribution in that dimension. Conversely, if $x_{i,j} < p_{i,j}$, the contribution is positive. The stochastic vectors therefore introduce randomness into the magnitude of the movement, while the relative positions of the particle and its target determine its direction. This allows particles to move towards favourable solutions regardless of whether the required change in a given hyperparameter is positive or negative.

The particle velocities are subsequently clipped component-wise according to

$$-v_{\max,j} \leq v_{i,j} \leq v_{\max,j}, \tag{4.20}$$

where $v_{\max,j}$ is the maximum allowed velocity in dimension $j$. In the implementation used here, this value is set equal to the total kinematic range of the corresponding input dimension,

$$v_{\max,j} = U_j - L_j, \tag{4.21}$$

where $L_j$ and $U_j$ are the lower and upper bounds of that dimension defined in Equation 4.10. This limits the size of individual particle movements relative to the scale of the search space, reducing excessively large updates and improving numerical stability. The particle positions are similarly constrained to the physics-informed bounds defined in Equation 4.10.

To reduce the likelihood of stagnation or premature convergence, a soft-restart mechanism is also implemented [111]. If the global best solution does not improve for a specified number of iterations, defined as the patience parameter, a small stochastic perturbation is applied to the particle positions. In the implementation used here, the patience is set to 100 iterations. This redistributes the swarm within the allowed parameter space while retaining the information accumulated by the optimisation, providing an opportunity to escape a local minimum or other stagnant region of the loss surface.

The optimisation is continued until convergence or until the maximum number of iterations is reached, with the latter set to 500 iterations in the present implementation. The global best particle position at the end of the optimisation is taken as the optimal set of kernel length scales and is subsequently used for the final GP interpolation.

## 4.6 Convex Hull Construction

As discussed in Sections 3.10 and 3.11, the convex hull provides a geometric definition of the region spanned by a set of observations and can therefore be used to distinguish interpolation from extrapolation. In the framework developed in this thesis, this region is represented by a discretised set of kinematic points at which the GP predictions can subsequently be evaluated.

For the one-dimensional $\bar{K}N$ scattering datasets, the interpolation region is simply the interval between the minimum and maximum measured incident energies. For the two-dimensional polarisation observable datasets, the interpolation region is defined by the convex hull of the measurements in the $(E, \cos\theta)$ input space. Where the experimental measurements are provided in bins, the outer bin edges are incorporated into the construction so that the resulting region reflects the finite kinematic extent of the measurements rather than only their central values.

The continuous convex hull itself is calculated using the `scipy.spatial.ConvexHull` implementation [21]. However, the resulting geometric representation is not directly suitable for the subsequent GP calculations, which require a discretised set of points with specified resolutions in each kinematic dimension. A dedicated procedure is therefore used to generate the discretised boundary and interior of the convex hull.

For a set of points in $n$ dimensions, the convex hull algorithm identifies the boundary facets of the hull and returns the vertices defining the simplicial representation of these facets. These vertices provide the starting point for constructing the discretised boundary. For each pair of vertices defining an edge of a facet, the points along the edge are generated according to the specified discretisation in each kinematic dimension.

Let $\vec{a}$ and $\vec{b}$ denote two vertices defining an edge, and let $\vec{r}$ denote the corresponding discretisation resolutions. The dimension containing the largest number of discretisation intervals along the edge is selected as the incrementing dimension,

$$j^* = \arg\max_j \left( \frac{|a_j - b_j|}{r_j} \right), \tag{4.22}$$

where $j = 1, \ldots, n$. If multiple dimensions attain the same maximum value, one is selected arbitrarily.

Assuming, without loss of generality, that $a_{j^*} < b_{j^*}$, let $\{x_{i,j^*}\}$ denote the discretised coordinates between the two vertices along the incrementing dimension. The corresponding parameter along the edge is

$$t_i = \frac{a_{j^*} - x_{i,j^*}}{a_{j^*} - b_{j^*}}, \qquad t_i \in [0,1]. \tag{4.23}$$

The remaining coordinates are then obtained from the parametric equation of the straight line connecting the two vertices,

$$x_{i,j} = a_j + t_i(b_j - a_j), \qquad j \neq j^*. \tag{4.24}$$

The resulting coordinates are mapped onto the specified discretisation grid according to

$$x_{i,j} \mapsto r_j \,\mathrm{round}\left(\frac{x_{i,j}}{r_j}\right). \tag{4.25}$$

Repeating this procedure for each edge of each facet produces a discretised representation of the boundary. The points generated from different facets are then combined and duplicate points are removed before the interior of the convex hull is populated.

To populate the interior, the discretised boundary points are ordered according to their kinematic coordinates. A dimension in which the boundary exhibits variation is selected as the incrementing dimension. For a set of boundary points $e$, this is chosen as the highest dimension for which

$$\max_k e_{k,j} - \min_k e_{k,j} \neq 0. \tag{4.26}$$

The corresponding discretisation step is then defined by

$$\vec{s}_j = \begin{cases} r_j, & j = j', \\ 0, & \text{otherwise.} \end{cases} \tag{4.27}$$

Consecutive boundary points are examined to identify segments that differ only in the selected incrementing dimension. Two points $e_k$ and $e_{k+1}$ satisfy this condition when

$$e_{k,j} = e_{k+1,j}, \qquad \forall j \neq j'. \tag{4.28}$$

When this condition is satisfied, the points between the two boundary coordinates are populated

according to

$$S = \left\{ x_i = e_k + n\vec{s} \;\middle|\; x_{i,j'} \leq e_{k+1,j'}, \quad n \in \mathbb{N} \right\}. \tag{4.29}$$

Applying this procedure across the boundary generates the discretised points contained within the convex hull.

A limitation of this procedure arises from the ordering of the boundary points. The standard ordering used to identify consecutive points does not necessarily place all geometrically adjacent points next to one another. Consequently, two points that define the endpoints of a valid discretised segment can be separated by other points in the ordering, causing intermediate points to be omitted.

For example, consider a standard pyramid with vertices $[1,1,1]$, $[1,5,1]$, $[5,1,1]$, $[5,5,1]$ and $[3,3,3]$ and unit resolution in each dimension. The expected number of points in the discretised convex hull can be computed by summing the number of points in each horizontal "layer" of the pyramid: $1^2+3^2+5^2 = 35$, as shown in Figure 4.4. However, the algorithm returns only 31 points; specifically, the points $[2,3,2]$, $[4,3,2]$, $[3,2,2]$ and $[3,4,2]$ – highlighted in red in Figure 4.4 – are missing from the returned convex hull.

The missing points arise because the points defining the endpoints of the corresponding discretised segments are not adjacent under the standard ordering. For example, the point $[3,4,2]$ lies on the facet defined by the vertices $[1,5,1]$, $[5,5,1]$ and $[3,3,3]$. The points $[2,4,2]$ and $[4,4,2]$ define the endpoints of a discretised segment containing $[3,4,2]$, but they are separated in the standard ordering. Consequently, the condition in Equation 4.28 is never applied to this pair and the intermediate point is not generated.

To overcome this ordering dependence, the construction is repeated under cyclic permutations of the kinematic dimensions. For a set of $n$ dimensions, the coordinates are cyclically permuted according to

$$[1,2,\ldots,n-1,n] \to [n,1,2,\ldots,n-2,n-1]. \tag{4.30}$$

The points are then re-sorted in the permuted coordinate system and the discretisation procedure is repeated. After each iteration, the generated points are mapped back to the original coordinate system.

Repeating the construction for each cyclic permutation changes which geometrically related points are adjacent under the ordering. Points that are separated under one ordering can therefore

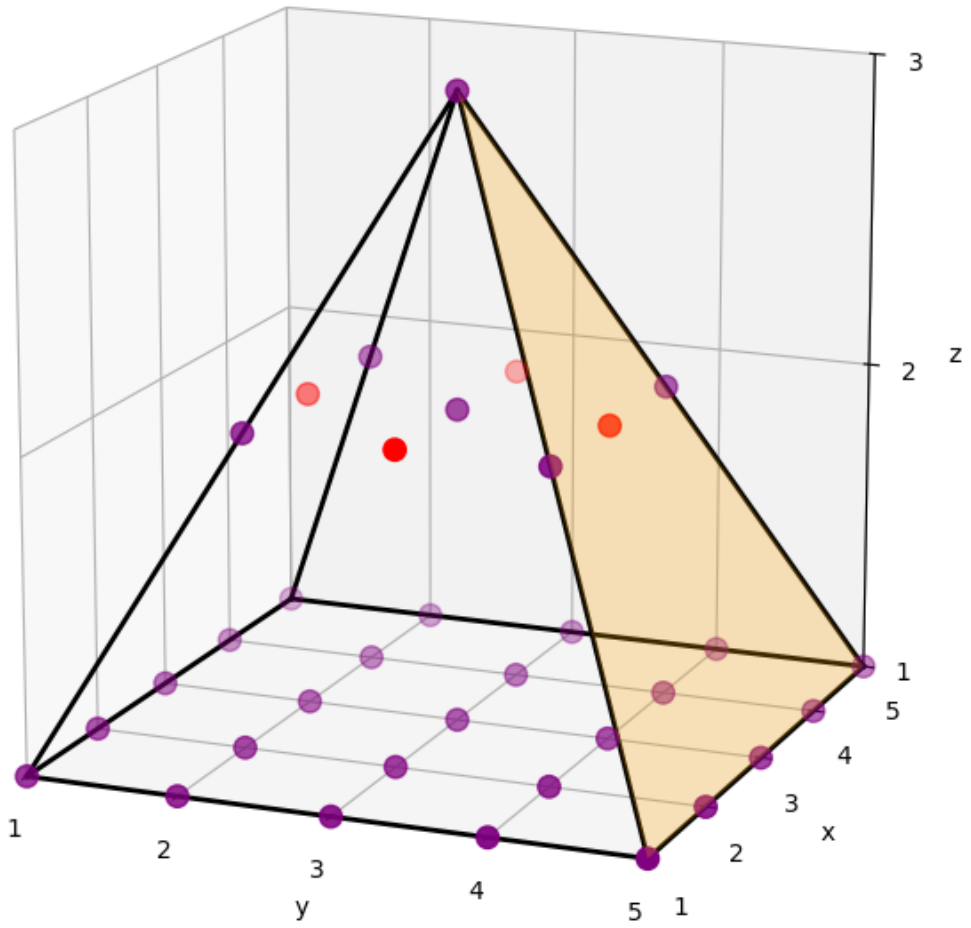



Figure 4.4: Example three-dimensional pyramidal convex hull used to test the discretisation algorithm. The geometry contains 35 points at unit resolution, providing a simple test case for assessing whether the discretised construction recovers the complete convex hull.

become adjacent under another. The resulting point sets are combined and duplicate points are removed, ensuring that relevant discretised segments are captured across the different coordinate orderings.

The effect of the ordering problem and its resolution are illustrated by considering the facet containing the missing point $[3,4,2]$. A two-dimensional projection of this facet along the $y$ axis is shown in Figure 4.5. The cyclic permutation procedure causes the relevant boundary points to become adjacent under an alternative ordering, allowing the previously omitted point to be generated.

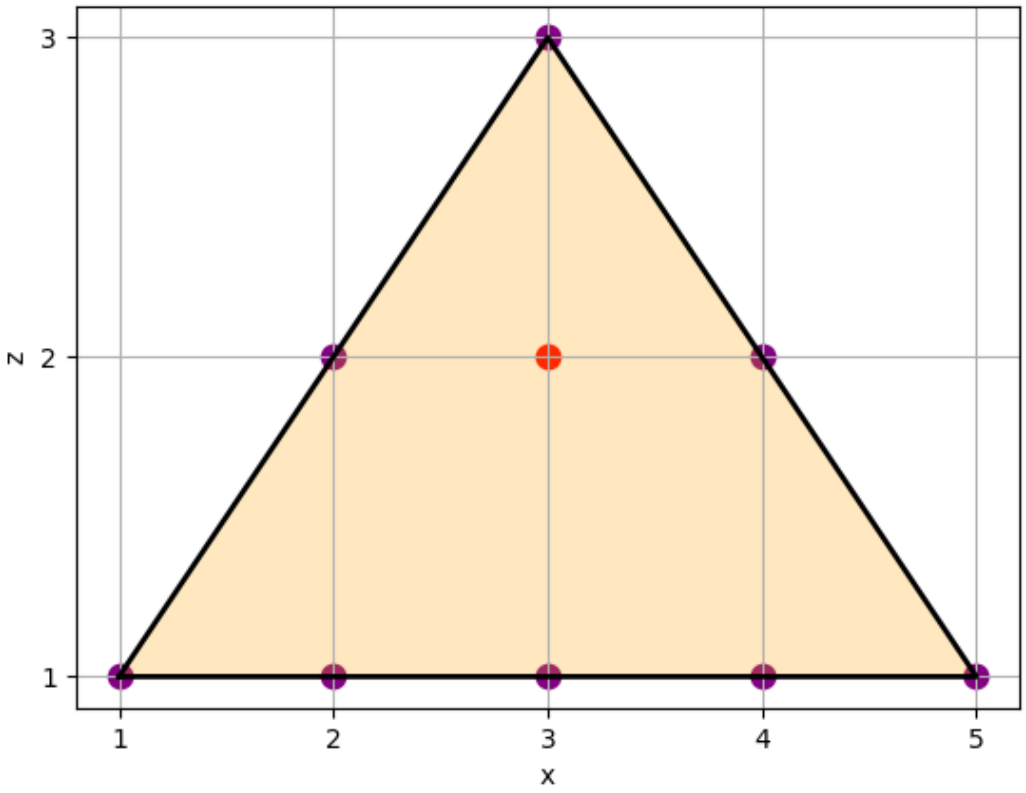



Figure 4.5: Two-dimensional projection of the pyramidal facet along the $y$ axis. The missing point $[3,4,2]$ lies between the boundary points $[2,4,2]$ and $[4,4,2]$. The example illustrates how the standard ordering can prevent valid neighbouring points from being identified by the discretisation procedure.

Applying the construction under all cyclic permutations successfully recovers the missing interior points. For the pyramidal test geometry, the resulting discretised convex hull contains the expected 35 points. This provides a simple validation of the construction procedure and demonstrates that the ordering dependence of the initial implementation can be removed without changing the underlying geometric convex hull.

### 4.6.1 Additional Binning Information

The final stage of the construction incorporates information about the finite kinematic bins associated with the experimental measurements. Although the convex hull construction described above is a general procedure that can be applied to datasets in arbitrary dimensions, the treatment of binning introduced in this section is designed specifically for the two-dimensional polarisa-

tion observable datasets considered in this thesis, and can be applied more generally to datasets with a similar kinematic structure.

The inclusion of this information is motivated by the way in which experimental measurements are obtained and reported. As discussed in Section 2.7, a measured value of a polarisation observable is typically associated with a finite range of reaction energies and scattering angles rather than a single infinitesimally precise point in kinematic space. The quoted central kinematic coordinates therefore provide a representative location for a measurement that contains information from a finite region surrounding that point. If only the central values were used when constructing the convex hull, the resulting interpolation region could be artificially smaller than the region actually probed by the experiment. This would cause predictions close to the outermost measurements to be classified as extrapolations even when they remain within the kinematic range covered by the corresponding experimental bins.

Incorporating the bin boundaries therefore provides a more physically representative definition of the region constrained by the measurements. The intention is not to treat every point within a bin as an independent measurement, nor to introduce additional experimental data that were not measured. Instead, the bin information is used to determine the geometric extent of the region from which the reported measurements originate. The resulting convex hull consequently reflects both the locations of the measured central values and the finite kinematic coverage associated with those measurements.

For the polarisation observables considered in this work, the measurements are binned in both reaction energy and $\cos\theta$. Where explicit bin edge information is available, the outer edges of the measured bins are incorporated directly into the initial point set used to construct the convex hull. In this way, the geometric boundary of the interpolation region extends to the experimentally covered kinematic range rather than terminating at the central values of the outermost measurements.

For the angular coordinate $\cos\theta$, only the outer edges of the leftmost and rightmost measured bins are required when defining the boundary. Let $a_1$ and $a_m$ denote the mean $\cos\theta$ values of the leftmost and rightmost bins at a given energy. When explicit bin edge information is unavailable, the outer edges are estimated from the separation between neighbouring measured angular values. This assumes that the measured angular points are representative of a sequence of neighbouring bins, such that the distance between adjacent central values provides an estimate of the corresponding bin spacing. The estimated outer edges are constrained to the physical

interval $\cos\theta \in [-1,1]$:

$$\Delta a_1 = \max\left(a_1 - \frac{|a_2 - a_1|}{2}, -1\right), \tag{4.31}$$

$$\Delta a_m = \min\left(a_m + \frac{|a_m - a_{m-1}|}{2}, 1\right). \tag{4.32}$$

This prescription reflects the expected structure of approximately contiguous angular bins: the boundary of the outermost bin is taken to lie approximately halfway between its central value and the neighbouring bin centre, unless this would extend beyond the physical angular range. The restrictions at $\cos\theta = \pm 1$ ensure that the estimated bin extent remains within the physically accessible scattering angle range.

The energy bin widths are determined using an analogous approach. The measured energy values are treated as representative central values of neighbouring energy bins, such that their separations provide an estimate of the local bin spacing. For the lowest and highest measured energies, the half-width is taken from the distance to the single adjacent energy value. For intermediate energies, the bin width is determined from the mean of the neighbouring half-separations. Thus, for measured energy $E_i$, the estimated bin width is

$$\Delta E_i = \begin{cases} \dfrac{|E_2 - E_1|}{2}, & i = 1, \\ \dfrac{|E_n - E_{n-1}|}{2}, & i = n, \\ \dfrac{1}{2}\left(\dfrac{|E_i - E_{i-1}|}{2} + \dfrac{|E_{i+1} - E_i|}{2}\right), & 2 \le i \le n-1. \end{cases} \tag{4.33}$$

The corresponding estimated bin edge coordinates are then given by

$$E_i^{\pm} = E_i \pm \Delta E_i. \tag{4.34}$$

The use of neighbouring measurements to estimate the bin extent is motivated by the expected organisation of the experimental data into adjacent kinematic bins. For a regularly or approximately regularly binned dataset, the separation between neighbouring central values provides a natural estimate of the distance between the corresponding bin centres. Taking half of this separation therefore gives an estimate of the distance from a central value to the boundary between neighbouring bins. For intermediate energy values, averaging the two neighbouring half-separations allows the prescription to accommodate variations in bin width while maintaining a

continuous estimate of the local kinematic coverage.

For each energy level containing more than one angular measurement, the resulting angular and energy edges provide additional boundary points required to construct the convex hull. These points are incorporated before the convex hull calculation so that the resulting discretised domain accounts for the finite kinematic coverage of the experimental bins. Importantly, these additional points do not represent additional measurements of the observable; they only extend the geometric domain to account for the region of kinematic space associated with the existing measurements.

When explicit binning information is unavailable, the above prescription therefore provides an estimate of the outer bin boundaries based on the observed spacing of the measured kinematic coordinates. This is a practical approximation to the likely bin structure rather than a reconstruction of the exact experimental binning. Its use is justified where the data exhibit an ordered and approximately contiguous bin structure, as is the case for the polarisation observable datasets considered here. Cases in which this assumption cannot be reliably made, such as a dataset containing only a single measured energy or only one angular point at each energy, require the bin boundaries to be determined separately from the available experimental information.

The convex hull construction itself remains independent of the treatment of binning and can therefore be applied without these additional steps to datasets for which measurements are represented by well-defined point coordinates. The binning prescription developed here is instead an application specific extension intended to incorporate information that is particularly relevant for the polarisation observable data used in this work. It is therefore applicable most directly to datasets with a similar structure, namely measurements organised over discrete energy and angular bins.

This binning prescription was developed specifically for the two-dimensional pseudoscalar meson photoproduction data considered in this thesis [82]. An example of the resulting discretised convex hull, including the additional kinematic extent associated with the experimental binning, is shown in Figure 4.6.

The resulting set of discretised points defines the kinematic domain used for the subsequent GP interpolation. The construction is therefore applied to the input coordinates rather than to the measured observable itself, ensuring that the resulting domain represents the region of kinematic space supported by the available experimental measurements and their associated binning.

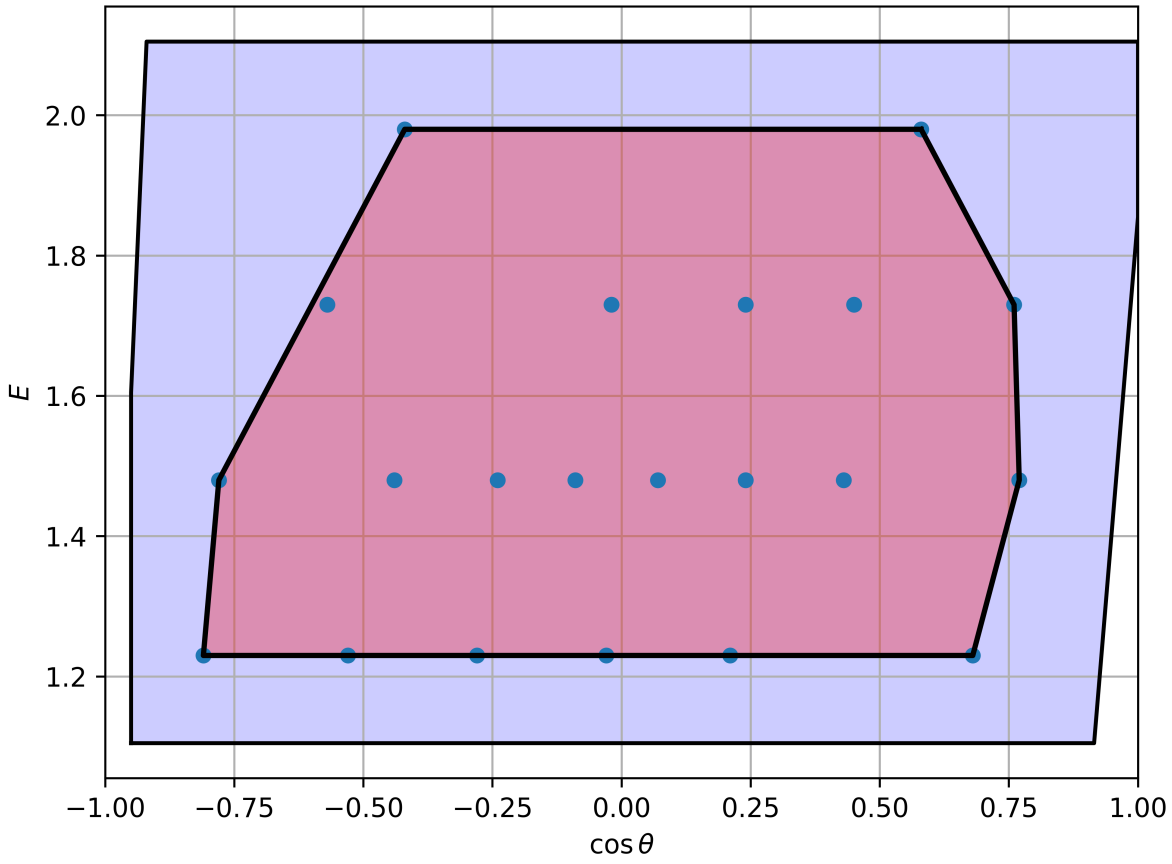



Figure 4.6: Two-dimensional discretised convex hull for the polarisation observables measured in [82], including the additional kinematic extent associated with the experimental binning. The original convex hull is shaded in red and is defined by the central measurement coordinates, while the extended region is shaded in blue and incorporates the estimated boundaries of the experimental bins, providing a representation of the full kinematic region covered by the measurements.

### 4.6.2 Computational Considerations

An alternative approach is to use a Delaunay triangulation to determine which points in a discretised parameter space grid lie within the convex hull [112]. However, this requires the full grid to be generated before the points inside the hull can be identified. The grid must therefore contain every possible combination of the discretised coordinates, including points that ultimately fall outside the convex hull and are not required for the final interpolation domain.

This becomes increasingly demanding as the dimensionality or resolution of the parameter space increases. Adding an additional dimension introduces another set of coordinate values that must be combined with those of the existing dimensions, while increasing the resolution similarly in-

creases the number of possible coordinate combinations. Consequently, the number of points in the complete grid grows rapidly with both dimensionality and resolution. Since the coordinates of all of these points must be stored before the point-in-hull classification can be performed, the associated memory requirements can become a significant limitation.

The issue is particularly relevant when the convex hull occupies only a relatively small fraction of the parameter space. In this case, a large number of the generated grid points lie outside the convex hull and are subsequently discarded. The computational resources required to generate and store these points are therefore used for points that do not contribute to the final interpolation domain.

The algorithm developed in this work takes a different approach by constructing the discretised convex hull domain directly from the facets of the continuous convex hull. Rather than first generating the complete parameter-space grid and then determining which points are contained within the hull, the algorithm constructs the boundary of the hull and subsequently fills the region enclosed by this boundary. Only points required to represent the discretised convex hull are therefore generated and retained.

This approach reduces the amount of unnecessary computation and memory associated with constructing the interpolation domain. The advantage is not that the convex hull calculation itself is eliminated, since the continuous convex hull must still be determined, but rather that the potentially large Cartesian grid used for point-in-hull classification does not need to be materialised. The resulting construction is therefore better matched to the specific purpose of this calculation, where the required output is the discretised subset of parameter space contained within the convex hull.

Although the applications considered in this thesis involve only 1D and 2D kinematic spaces, the algorithm was designed as a general multidimensional procedure. This was motivated by the possibility of applying the construction to higher-dimensional parameter spaces, such as those encountered in the DVCS analysis discussed later in Section 7.4.1. In such cases, the rapid growth in the number of points in a complete Cartesian grid makes avoiding its explicit construction increasingly important.

## 4.7 Role of the Gaussian Process Framework

The experimental data considered in this thesis form a finite and non-uniform sampling of the underlying physical observables, as discussed in Section 2.7. Measurements are available only at discrete kinematic points or within finite bins, with the density and uncertainties of the measurements varying across the kinematic space. These limitations are particularly relevant when the measured observables are used in subsequent reaction amplitude and spectroscopy analyses, where information across a broader kinematic region may be required.

The GP framework developed in this thesis is intended to address this gap between the discrete nature of the experimental measurements and the more continuous kinematic dependence required by subsequent analyses. The framework is positioned after the determination of physical observables and before their use in reaction amplitude analyses within the broader physics analysis pipeline described in Section 2.8. Its purpose is to provide an interpolated representation of the measured observables between existing measurements, together with an estimate of the associated predictive uncertainty.

This can be particularly useful for Partial Wave Analyses (PWAs), which use experimental observables across a range of energies and angles to constrain the reaction amplitudes. Broad and sufficiently dense experimental coverage provides stronger constraints on the kinematic dependence of these amplitudes. However, the experimental data available for a given reaction may contain substantial gaps in their energy or angular coverage. An interpolation between existing measurements can provide estimates at kinematic locations where no direct measurement is available, allowing the information contained in the surrounding measurements to contribute more continuously to a subsequent analysis.

The benefit of such an interpolation is not limited to filling individual gaps between measurements. Measurements from different experiments may have different kinematic coverage, meaning that combining them can produce an irregular collection of measurements across the relevant parameter space. A continuous representation can provide a common description of these measurements and allow observables to be evaluated at kinematic locations where individual experiments provide no direct result. This can be useful when comparing experimental information with theoretical calculations or when constructing the inputs required for subsequent amplitude analyses.

An important requirement is that the interpolation should retain information about how strongly the available measurements constrain the predicted observable. A prediction made between closely spaced measurements should generally be better constrained than one made in a region separated from the nearest experimental data. The predictive uncertainty provided by the GP therefore gives additional information beyond the interpolated central value, allowing the degree of experimental constraint to be carried through to subsequent analyses. The convex hull construction described in Section 4.6 further provides a geometric definition of the region supported by the available measurements, preventing regions requiring extrapolation from being treated as equally constrained.

The framework does not introduce additional experimental measurements. Instead, it provides a statistical representation of the information already contained in the available data and estimates the behaviour between measured points. This distinction is important because interpolation cannot compensate for genuinely missing experimental coverage. Where the available measurements provide insufficient information to constrain an observable, the appropriate solution remains additional experimental measurements rather than treating model predictions as experimental constraints.

The use of a GP is therefore motivated by the particular structure of the experimental data and by the requirements of their subsequent use. The framework provides a means of making the information contained in sparse measurements more readily usable across the kinematic region they support, while retaining an explicit representation of predictive uncertainty and distinguishing interpolation from extrapolation.

The standard GP formulation alone, however, does not guarantee that the resulting interpolation has the statistical properties required for this application. As demonstrated in Section 3.7.1, marginal-likelihood optimisation can leave the kernel hyperparameters weakly constrained for sparse datasets, potentially resulting in undesirable behaviour between measurements or poorly calibrated predictive uncertainties. The customised optimisation procedure developed in this chapter is therefore used to select the hyperparameters according to the requirements of the interpolation rather than relying solely on the standard marginal likelihood. The complete implementation and analysis workflow are described in the following sections.

## 4.8 Summary of the GP Framework

This chapter has presented the Gaussian Process framework developed in this thesis for interpolating sparse experimental hadron physics data and quantifying the associated predictive uncertainty. The experimental measurements used as inputs to the framework were described in Section 4.2, including the different kinematic dimensionalities and sampling densities of the datasets considered. The specification of the GP, including the choice of a zero mean function and the use of the RBF covariance kernel with independent characteristic length scales, was presented in Section 4.3.

A customised hyperparameter optimisation procedure was developed to address the requirements of the sparse datasets considered in this work. The overall procedure was introduced in Section 4.4, with empirical coverage used to assess the calibration of the predictive uncertainties in Section 4.4.1. The Wasserstein distance was selected to quantify the discrepancy between the measured and nominal coverage curves, as described in Section 4.4.2. Physics-informed bounds on the kernel length scales and a boundary proximity penalty were then incorporated into the combined loss function, as described in Sections 4.4.3 and 4.4.4. The resulting loss is minimised using PSO, with the optimisation strategy and its implementation described in Section 4.5. The selection of PSO was supported by a comparison with MCMC on identical pseudodatasets, with the detailed comparison provided in Appendix A.

The construction of the region over which the GP interpolation is performed was described in Section 4.6. The 1D $\bar{K}N$ datasets use the range of the measured kinematic variable, while the 2D polarisation observable datasets use a convex hull constructed from the available measurements. The treatment of finite experimental binning and its incorporation into the interpolation region was described in Section 4.6.1. The computational considerations motivating the direct construction of the discretised convex hull, rather than generating a complete parameter-space grid, were discussed in Section 4.6.2.

The role of the resulting framework within the broader physics analysis was considered in Section 4.7. The framework provides a means of making the information contained in sparse experimental measurements more readily usable across the kinematic region supported by those measurements, while retaining an explicit representation of predictive uncertainty and distinguishing interpolation from extrapolation.

Taken together, the methodology provides a GP framework in which the model specification, hyperparameter optimisation and interpolation domain are designed specifically for the characteristics of sparse experimental hadron physics data. The resulting framework provides statistically controlled interpolations without introducing additional experimental measurements, while retaining information about the uncertainty associated with predictions between measured points. Its performance and limitations are assessed using pseudodata and experimental datasets in Chapter 5.

# Chapter 5

# Validation of GP Framework

## 5.1 Construction of Pseudodata

In order to validate the methodology presented here, in particular the hyperparameter optimisation for sparse datasets, pseudodata, whose true underlying function is known, can be tested. Here we base the pseudodata on the two-dimensional polarisation observables of the $\gamma p \to K^0\Sigma^+$ reaction discussed in Section 4.2, where the sparse and non-uniform kinematic coverage of the experimental measurements was introduced. These observables are typically measured as functions of their energy (usually expressed in GeV) and scattering angle (usually expressed as $\cos\theta$).

Based on the general formalism of pseudoscalar meson photoproduction [50], the angular dependence of polarisation observables can be expressed as sums of Legendre polynomials, reflecting the underlying partial wave expansion of the reaction amplitudes. The energy dependence of observables is commonly associated with resonant structures, which can be parameterised using Breit-Wigner or, more generally, Cauchy-like distributions that reflect the pole structure of the underlying amplitudes [50]. However, for the purpose of validating the proposed methodology, such physically motivated line shapes are not required. Instead, we adopt a Gaussian parameterisation for the energy-dependent part of the pseudodata. While Gaussians do not reproduce the exact analytic properties of resonance dynamics, they provide a smooth, well-controlled approximation that is sufficient for constructing test functions with known ground truth. This

choice allows us to generate stable, differentiable pseudodata surfaces while avoiding additional complications arising from resonance interference effects, threshold behaviour or long tails inherent to Breit-Wigner forms. The Gaussian choice therefore serves as a convenient and computationally robust approximation to more realistic energy dependences for the purposes of this validation.

The 2D pseudodata surface is therefore of the form

$$f(w,\cos\theta) = \sum_{l=0}^{L} c_l g_l(w) P_l(\cos\theta), \tag{5.1}$$

where $c_l \sim \mathscr{U}(-1,1)$ are randomly sampled coefficients, $g_l(w)$ are Gaussian functions of the form

$$g_l(w) = \exp\left[-\frac{(w-\mu_l)^2}{2\sigma_l^2}\right], \tag{5.2}$$

and $P_l(\cos\theta)$ is the $l$-th order Legendre polynomial. Here, $\mathscr{U}(a,b)$ denotes a continuous uniform distribution between $a$ and $b$. For simplicity, each Legendre polynomial $P_l(\cos\theta)$ is associated with a single energy-dependent Gaussian $g_l(w)$. This avoids additional coupling between angular and energy components, making the pseudodata surface easier to construct and interpret while still capturing the essential 2D structure needed to validate the hyperparameter optimisation procedure. Here we set $L = 3$ so that each pseudodata surface has a total of 12 parameters.

In order to mimic the data from $\gamma p \to K^0\Sigma^+$, we generate the pseudodata surface across the angular range $-0.85 \leq \cos\theta \leq +0.85$ and an energy range of $w_{\min} = 1.2$GeV to $w_{\max} = 2$GeV. These angular limits are chosen to mimic the incomplete angular coverage resulting from detector acceptance. To ensure that the underlying true function is mathematically interesting in this region, we set $\mu_l \sim \mathscr{U}(1.4,1.8)$ and $\sigma_l^2 \sim \mathscr{U}(0.25,0.75)$. It is also important to note that these polarisation observables are asymmetries and are therefore confined to $[-1,1]$. If $\max|f(w,\cos\theta)| > 1$, the entire function is divided by $\max|f(w,\cos\theta)|$, ensuring that the observable values remain within this physical range.

The experimental data are sparse and are measured at a small number of discrete energy values, with only a limited number of angular measurements at each energy. To mimic this structure, the energy range $[w_{\min}, w_{\max}]$ is divided into $n$ equally sized bins, such that the width of each bin is

$$\Delta w = \frac{w_{\max} - w_{\min}}{n}. \tag{5.3}$$

One random energy value is then sampled from each bin, giving the set of discrete energy values

$$w_i \sim \mathscr{U}\left(w_{\min} + (i-1)\Delta w, w_{\min} + i\Delta w\right), \qquad i = 1, \ldots, n. \tag{5.4}$$

A similar procedure is then used for $\cos\theta$, with the angular range divided into $m_i$ equally sized bins at each energy value and one random value sampled from each bin. To mimic the sparsity of the experimental data, we fix $n = 4$ energy values and independently sample the number of angular measurements at each energy according to

$$m_i \sim \mathscr{U}(4, 5, \ldots, 10), \qquad i = 1, \ldots, n, \tag{5.5}$$

where $\mathscr{U}(4,5,\ldots,10)$ denotes a discrete uniform distribution over the integers from 4 to 10. Thus, the pseudodata contain four energy values with between four and ten angular measurements at each energy, mimicking the sparse and non-uniform sampling of the experimental dataset.

The pseudodata are modelled on data taken by the CLAS collaboration which, for this experiment [82], used a linearly polarised photon beam, for which each event is assigned a binary label (positive or negative outcome) corresponding to the sign of the polarisation-dependent asymmetry.

For a given kinematic point, the outcome of a single measurement can therefore be treated as a Bernoulli random variable. The probability of observing the positive or negative outcome ($P_+, P_-$, respectively) is determined by the asymmetry value obtained from the generated surface, denoted by $\hat{a}$. This quantity represents the underlying expectation value of the observable at that kinematic point and corresponds to the ensemble-averaged asymmetry that would be obtained in the limit of infinite statistics. In an experimental setting, however, only a finite number of events are observed, and the measured asymmetry therefore fluctuates about this expectation value due to statistical noise. In the pseudodata generation, this effect is modelled by sampling outcomes according to the probabilities $P_+$ and $P_-$ at each kinematic point, given by

$$P_\pm = \frac{1 \pm \hat{a}}{2}. \tag{5.6}$$

This form ensures that the probabilities are properly normalised, $P_+ + P_- = 1$, and that they remain bounded between 0 and 1 provided $|\hat{a}| \leq 1$, as required for a physical polarisation observable.

As this is a counting experiment, the total number of events $N$ is expected to follow Poisson statistics:

$$N \sim \text{Poisson}(\lambda), \tag{5.7}$$

where $\lambda$ is the effective count for the experiment. In the pseudodata construction, $\lambda \sim \mathscr{U}(200, 1000)$ to provide a range of event counts representative of the experimental setting. For a given $N$, the number of positive outcomes $N_+$ follows a binomial distribution

$$N_+ \sim \text{Binomial}\left(N, \frac{1+\hat{a}}{2}\right). \tag{5.8}$$

Thus, $N_- = N - N_+$. We can define an unbiased estimator for the asymmetry as

$$z = \frac{N_+ - N_-}{N}. \tag{5.9}$$

The statistical uncertainty of this measured asymmetry comes from the finite number of events detected and the probabilistic nature of polarisation measurements. Using standard results from binomial statistics [70], the variance of $z$ is given by

$$\text{Var}(z) = \frac{1-\hat{a}^2}{N}. \tag{5.10}$$

Although $\hat{a}$ is known during pseudodata generation, experimental analyses do not have access to the true underlying asymmetry. Therefore, to mimic a realistic measurement, $\hat{a}$ is replaced by the measured value $z$, yielding the estimated statistical uncertainty

$$\sigma_z \approx \sqrt{\frac{1-z^2}{N}}. \tag{5.11}$$

This captures the essential statistical behaviour of polarisation asymmetry measurements in the Poisson-limited regime. The uncertainty decreases as $1/\sqrt{N}$ with increasing statistics, while the binomial variance decreases as $|\hat{a}|$ approaches unity.

We now have a suitable sparse pseudodataset with a known underlying function. We can therefore apply the hyperparameter optimisation procedure developed in Chapter 4 within the convex hull of these points, as described in Sections 4.4 and 4.6.

After generating the GP surface, points are generated from the posterior mean in order to assess

its suitability. If the original input consists of $p$ pseudodata points, the same number of points are generated from the GP posterior mean. These points are also required to have comparable coverage of the kinematic space, defined by the convex hull of the original data. To achieve this, candidate points are generated using Latin Hypercube Sampling [83] to obtain a uniform, space-filling distribution (discussed in Section 3.8). These points are rounded to the experimental resolution, and the sampling density is increased until the number of valid points is sufficient to match the number of original pseudodata points. If more points are generated than required, a random subset is selected. This provides datasets with the same sample size and comparable kinematic coverage, allowing subsequent analyses of the original and GP-derived datasets to be compared directly.

In total, 100 independent pseudodata surfaces and corresponding datasets are generated to provide an ensemble for the validation studies presented in the following sections.

## 5.2 Uncertainty Calibration Test

We first verify that both the generated pseudodata and sampled GP datapoints exhibit the expected scatter about the underlying true functions given in Equation 5.1. The pull (otherwise known as the residual) of the datapoints can be found using:

$$\text{pull} = \frac{y_{\text{func}} - y_{\text{eval}}}{\delta y_{\text{eval}}}, \tag{5.12}$$

where $y_{\text{eval}}$ denotes either the pseudodata mean or the corresponding GP value, depending on the dataset under consideration, and $\delta y_{\text{eval}}$ is the associated estimated uncertainty. In all cases, $y_{\text{func}}$ represents the value of the known underlying function at the same kinematic point. If the uncertainties are correctly estimated and the deviations from the underlying function are approximately Gaussian, the pull distribution is expected to have zero mean and unit variance. Under this assumption, a point satisfies

$$|\text{pull}| \leq k \quad \Longleftrightarrow \quad y_{\text{func}} \in [y_{\text{eval}} - k\delta y_{\text{eval}}, y_{\text{eval}} + k\delta y_{\text{eval}}], \tag{5.13}$$

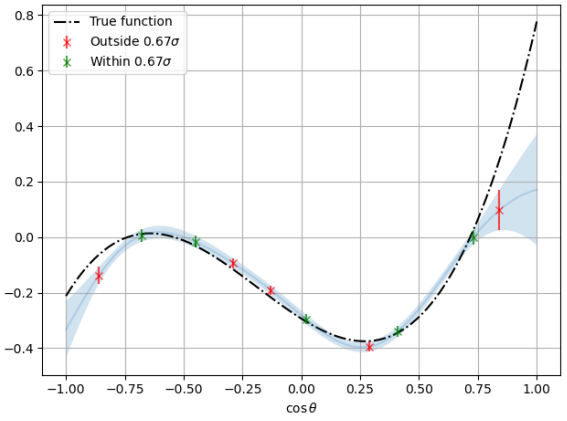
where $k$ specifies the number of standard deviations considered. Consequently, the fraction of points satisfying this condition provides a measure of the empirical coverage of the corresponding uncertainty interval.

For a standard normal distribution, the expected fraction of points within an interval of width $k\sigma$ is given by

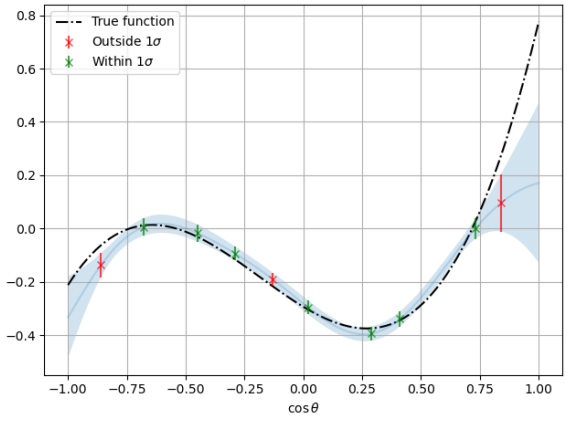
$$F_{\text{pred}}(k) = \Phi(k) - \Phi(-k), \tag{5.14}$$

where $\Phi$ is the cumulative distribution function of the standard normal distribution, as discussed in Section 3.4. For example, the expected coverage is approximately 50.0%, 68.3% and 95.0% for $0.67\sigma$, $1\sigma$ and $1.96\sigma$, respectively. The empirical coverage is calculated by evaluating the pull at each pseudodata or GP point and determining the fraction satisfying $|\text{pull}| \leq k$ for a range of values of $k$. This is similar to the procedure used to generate the empirical coverage plot introduced in Section 4.4.1, where the measured coverage is compared with the nominal Gaussian expectation over a range of uncertainty levels. In the present case, however, the underlying function is known by construction, allowing the coverage of the pseudodata and GP uncertainties to be assessed directly. This procedure is repeated independently for each of the 100 generated pseudodata surfaces, and the resulting coverage fractions are averaged over the ensemble to provide a measure of the typical uncertainty calibration for each dataset. Agreement between the measured and expected coverage indicates that the estimated uncertainties provide an appropriate representation of the deviations from the known underlying function, while systematic over- or under-coverage indicates that the uncertainties are respectively too large or too small.

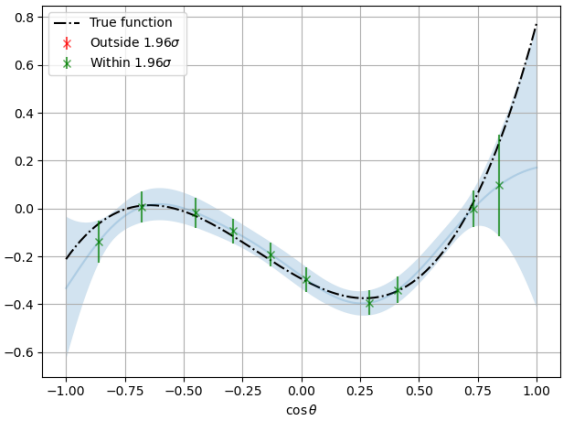
An example of this procedure is shown in Figure 5.1. The known underlying function (with functional form given in Equation 5.1) is shown by the black dash dotted curve, while the evaluated values are shown as individual points. Green points satisfy the coverage criterion for the specified uncertainty interval, while red points lie outside the corresponding interval. The uncertainty bands are scaled according to the value of $k$ considered in each panel, illustrating how the fraction of points contained within the interval changes as the confidence level is increased.




(a) $0.67\sigma$




(b) $1\sigma$




(c) $1.96\sigma$

Figure 5.1: Example of the uncertainty coverage test for three confidence intervals. The black dash dotted curve represents the known underlying function given by Equation 5.1, while the points represent the corresponding evaluated values. Green points satisfy the coverage criterion $|\text{pull}| \leq k$ and are therefore contained within the corresponding $k\sigma$ uncertainty interval, while red points fall outside this interval. The uncertainty bands around the evaluated values are scaled by the value of $k$ shown in each panel. The panels show the coverage for $0.67\sigma$, $1\sigma$ and $1.96\sigma$, for which a standard normal distribution would contain approximately 50.0%, 68.3% and 95.0% of points, respectively. The figure therefore provides a visual illustration of how the empirical coverage criterion is applied to the pseudodata and corresponding GP points.

The results are summarised in Table 5.1. The pseudodata reproduce the expected Gaussian coverage very closely, with measured fractions of 49.6%, 68.2% and 95.3% for the $0.67\sigma$, $1\sigma$ and $1.96\sigma$ intervals, respectively, compared with expected values of 50.0%, 68.3% and 95.0%. The small differences from the expected values are consistent with statistical fluctuations arising from the finite number of generated points. This provides a validation of the uncertainty-generation procedure used to construct the pseudodata and demonstrates that the statistical fluctuations introduced during the pseudodata construction produce uncertainties with the expected

coverage properties.

| $\sigma$ Band | Expected Percentage of Points (%) | Measured Percentage of Known Points (%) | Measured Percentage of GP Points (%) |
|---|---|---|---|
| $0.67\sigma$ | 50.0 | 49.6 | 63.8 |
| $1\sigma$ | 68.3 | 68.2 | 78.9 |
| $1.96\sigma$ | 95.0 | 95.3 | 97.4 |

Table 5.1: Expected and measured coverage fractions for the pseudodata and corresponding GP points. The expected values are calculated assuming a standard normal distribution of pulls. The measured values are averaged over the ensemble of 100 independently generated pseudodata surfaces.

The points generated from the GP exhibit higher coverage than the Gaussian expectation, particularly at the lower uncertainty levels. At $0.67\sigma$, 63.8% of the points are contained within the uncertainty interval, compared with an expected 50.0%, while at $1\sigma$ the corresponding coverage is 78.9%, compared with 68.3%. At $1.96\sigma$, however, the measured coverage of 97.4% is much closer to the expected value of 95.0%. This indicates a degree of overcoverage in the GP uncertainties, particularly for the narrower uncertainty intervals, meaning that the predicted uncertainties are somewhat conservative in this regime. The observed behaviour may be associated with the smoothing and correlation structure of the GP, which can reduce the point-to-point variation of the posterior mean relative to the underlying statistical fluctuations in the pseudodata. The overcoverage therefore does not indicate underestimated uncertainties; instead, the GP uncertainty intervals are sufficiently broad that the known function is contained more frequently than expected under the nominal Gaussian model. The closer agreement at $1.96\sigma$ is particularly relevant for subsequent analyses, where larger uncertainty intervals provide a useful measure of the confidence associated with the interpolated function.

Overall, the coverage test demonstrates that the uncertainties produced by the GP do not systematically underestimate the uncertainty associated with the interpolated surface for the pseudodata considered here. While the GP exhibits some overcoverage at lower confidence levels, the effect becomes substantially smaller at $1.96\sigma$, where the measured coverage is close to the Gaussian expectation. The pseudodata therefore provide a validation of the statistical construction, while the comparison with the GP points demonstrates that the interpolation does not introduce substantial underestimation of the associated uncertainty. Having established this aspect of the GP performance, the next test considers a complementary requirement: whether the interpolation

preserves the underlying functional information contained within the sparse pseudodata. This is assessed in the following section by testing the recovery of the known functional form.

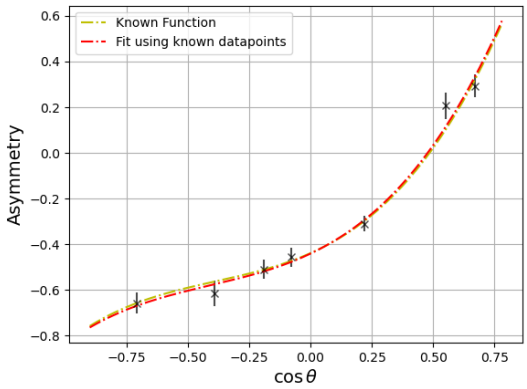
## 5.3 Recovery of Functional Form

A further requirement of the GP representation is that it should retain the information contained in the original measurements in a form suitable for subsequent physics analysis. As discussed in Section 2.8, experimental measurements are ultimately used to constrain models of the underlying reaction dynamics. This commonly involves fitting a chosen functional form or theoretical model to the available measurements and extracting the corresponding parameters. The ability to perform such a fit is therefore an important test of whether the GP interpolation preserves the relevant information contained within the sparse dataset.

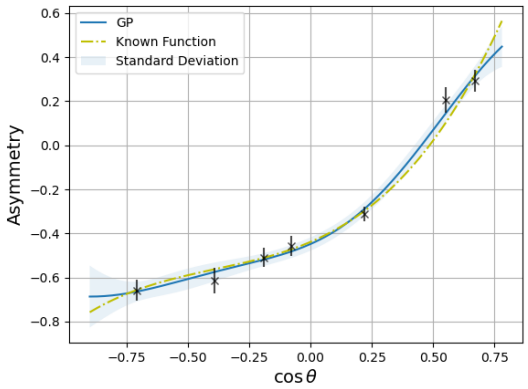
To test this, we compare fits performed directly using the original pseudodata with equivalent fits performed using datapoints sampled from the GP surface. An illustrative comparison of the resulting fitted surfaces is shown through their one-dimensional projections in Figure 5.2. These projections are intended only to provide a visual example of the similarity between the two fits; the quantitative comparison is performed using the fitted coefficients.

The known functional form of the underlying two-dimensional surface, given by Equation 5.1, is first fitted directly to the pseudodata using a least-squares method implemented with the `iminuit` software package [113]. An example of the resulting fit is shown in Figure 5.2a. A GP is then fitted to the same pseudodata, producing the interpolated surface shown in Figure 5.2b. As described in Section 5.1, a set of datapoints is subsequently sampled from the GP surface with the same number of points and comparable kinematic coverage to the original pseudodata. The same least-squares fitting procedure is then applied to these sampled GP datapoints using the original functional form.

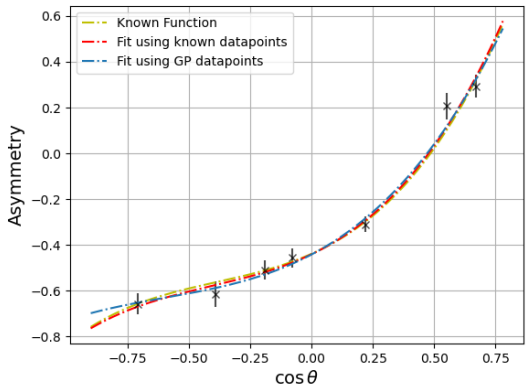
The quantitative comparison is performed using the fitted coefficients across the ensemble of 100 independently generated pseudodata surfaces. For each generated surface, the functional form is fitted independently to both the original pseudodata and the corresponding sampled GP datapoints. The resulting fitted coefficients can then be compared through their pulls, defined using Equation 5.12. For a given coefficient, the pull measures the difference between the value obtained from a particular fit and the corresponding true value used to generate the pseudodata,




(a) Fit to the functional form using the pseudodata.




(b) GP fit to the pseudodata.




(c) Fits to the functional form using the pseudodata and sampled GP datapoints.

Figure 5.2: Illustrative comparison of fits to the known functional form for a generated two-dimensional surface. The fit performed using the original pseudodata is compared with the equivalent fit performed using datapoints sampled from the GP surface. The results are shown through one-dimensional projections of the fitted surfaces and are intended to illustrate their overall similarity. The quantitative comparison between the two approaches is instead performed using the pulls of the fitted coefficients shown in Table 5.2.

relative to the estimated uncertainty on the fitted coefficient. If the fitting procedure provides an appropriate description of the coefficient uncertainties, these pulls are expected to be approximately Gaussian distributed.

We therefore model the pull distribution of each fitted coefficient using a Gaussian distribution. The fitted mean and variance of this Gaussian provide a compact description of the distribution and allow the behaviour of the coefficients obtained from the original pseudodata and sampled GP datapoints to be compared quantitatively. In particular, similar variances indicate that the two datasets provide comparable constraints on the fitted coefficients, while substantial differences in the fitted means or variances could indicate that the interpolation has altered the information

available to the subsequent functional form fit.

The fitted Gaussian means and variances for each coefficient are summarised in Table 5.2 and shown in Figure B.1 for both the original pseudodata and the sampled GP datapoints. The two datasets produce broadly comparable pull distributions across all fitted coefficients. The variances are generally of similar magnitude, with the GP-based fits showing some reduction in variance for several coefficients. This behaviour is consistent with the mild over-coverage observed in the uncertainty calibration test, where the GP uncertainties were found to be somewhat conservative at lower confidence levels. Importantly, the differences between the two sets of fitted coefficients do not exhibit a consistent pattern across the different coefficients that would indicate a significant loss or systematic distortion of information through the interpolation.

| | Pseudodata | | GP Datapoints | |
|---|---|---|---|---|
| Coefficient | Mean | Variance | Mean | Variance |
| $c_0$ | -0.037 | 0.746 | -0.150 | 0.978 |
| $\mu_0$ | 0.111 | 0.731 | -0.202 | 0.684 |
| $\sigma_0^2$ | -0.058 | 0.636 | -0.235 | 0.822 |
| $c_1$ | -0.098 | 0.902 | 0.021 | 0.828 |
| $\mu_1$ | -0.012 | 0.658 | -0.023 | 0.723 |
| $\sigma_1^2$ | -0.146 | 0.630 | -0.060 | 0.444 |
| $c_2$ | 0.093 | 0.706 | -0.007 | 0.448 |
| $\mu_2$ | -0.008 | 0.837 | -0.060 | 0.650 |
| $\sigma_2^2$ | -0.047 | 0.623 | -0.187 | 0.655 |
| $c_3$ | -0.008 | 1.013 | 0.042 | 0.776 |
| $\mu_3$ | -0.115 | 0.685 | -0.057 | 0.721 |
| $\sigma_3^2$ | -0.208 | 0.635 | -0.311 | 0.837 |

Table 5.2: Means and variances of the Gaussian distributions fitted to the pull distributions of the fitted coefficients obtained using the original pseudodata and sampled GP datapoints.

The fitted coefficients obtained from the sampled GP datapoints are therefore consistent with those obtained from the original pseudodata. The fitted variances are generally of comparable magnitude, while the differences in the fitted means do not show a consistent systematic trend across the coefficients. These results indicate that the interpolation has retained the information required for the subsequent functional form fit. The small differences observed between the two approaches are expected when comparing fits to finite datasets and do not indicate a significant distortion of the underlying functional information.

An additional advantage of the GP representation is that it provides a continuous description of the observable within the kinematic region supported by the experimental data, rather than restricting subsequent analyses to the discrete locations at which measurements are available. This allows the observable to be evaluated at kinematic points where no direct experimental measurement exists, together with an associated estimate of the interpolation uncertainty. Such a representation may be particularly useful when comparing experimental information with theoretical models that are only defined or applicable over specific kinematic regions, since the GP can provide estimates at the required points without introducing an additional functional parameterisation. The interpolation does not provide information beyond that supported by the experimental data, and its reliability therefore depends on the available kinematic coverage and the associated uncertainties.

Overall, this test demonstrates that the GP surface preserves sufficient information from the sparse pseudodata to support a subsequent functional form fit. This is an important property for the intended physics application, where an interpolation is not an endpoint in itself but forms an intermediate representation between sparse experimental measurements and the extraction of information about the underlying reaction dynamics. The agreement between the coefficient pull distributions obtained from the original pseudodata and sampled GP datapoints provides evidence that the GP can act as a suitable surrogate for the original measurements while retaining the information required for downstream physics analyses.

## 5.4 Summary

This chapter has validated the GP framework developed in Chapter 4 using pseudodata for which the underlying functional form is known. The construction of these pseudodata, described in Section 5.1, was designed to reproduce the sparse and non-uniform kinematic coverage characteristic of experimental datasets. The uncertainty calibration test in Section 5.2 demonstrated that the uncertainties associated with both the pseudodata and the GP interpolation exhibit the expected coverage behaviour. In particular, the GP showed some overcoverage at lower confidence levels but no evidence of systematic underestimation of the interpolation uncertainty. The recovery of functional form in Section 5.3 then demonstrated that fitting the known functional dependence to datapoints sampled from the GP produces results comparable to those obtained from the original pseudodata, indicating that the interpolation retains the information required for subsequent physics analysis.

Together, these tests provide validation of the principal requirements established for the framework in Chapter 4: that the GP should provide statistically meaningful uncertainties while retaining the information contained within sparse measurements. The results therefore support the use of the framework as an intermediate representation between the discrete experimental measurements described in Section 2.7 and subsequent physics analyses. In addition, the continuous representation provided by the GP allows the observable to be evaluated at unmeasured kinematic points within the region supported by the data, as discussed in Section 4.7. The validation presented here therefore provides the basis for applying the framework to the experimental datasets and physics observables considered in Chapter 6.

# Chapter 6

# Results and Discussion

## 6.1 Application to Polarisation Observables

The first application of the GP framework is to the sparse two-dimensional polarisation observable data for the $\gamma p \to K^0\Sigma^+$ reaction. The experimental measurements used in this analysis are taken from [82] and provide discrete samples of the observables over photon energy and $\cos\theta$, with limited and non-uniform coverage across the kinematic region. The purpose of this application is to demonstrate how the framework developed in Chapter 4 can be used to construct continuous representations of the measured observables while retaining associated uncertainty information. The resulting GP surfaces for $\Sigma$, $P$, $T$, $O_x$ and $O_z$ are presented below.

### 6.1.1 GP Fits to 2D Polarisation Observables

The GP framework was applied independently to the five polarisation observables $\Sigma$, $P$, $T$, $O_x$ and $O_z$. Each observable is treated as a function of photon energy and $\cos\theta$, with the experimental measurements providing the constraints on the resulting two-dimensional surface. No interpolation or smoothing is applied to the experimental data prior to the GP fit.

The resulting fits are shown in Figures 6.1–6.5. In each figure, the lower panel shows the two-

dimensional GP surface as a function of photon energy and $\cos\theta$. The four panels above show one-dimensional projections of the same surface at photon energies where experimental measurements are available. These projections allow the fitted dependence on $\cos\theta$ to be compared directly with the experimental data, while the two-dimensional surface provides the interpolation between the measured kinematic points.

The fitted surfaces exhibit different levels of structure between the five observables. The $\Sigma$ observable shows a comparatively smooth dependence across the kinematic region (Figure 6.1), while smaller local structures are visible in the $P$ (Figure 6.2) and $O_x$ (Figure 6.4) projections. More pronounced structure is present in the $T$ (Figure 6.3) and $O_z$ (Figure 6.5) projections. In particular, the projection at 1.98 GeV for $T$ exhibits a pronounced local maximum at positive $\cos\theta$, while the corresponding projection for $O_z$ exhibits a pronounced local minimum in a similar region of $\cos\theta$.

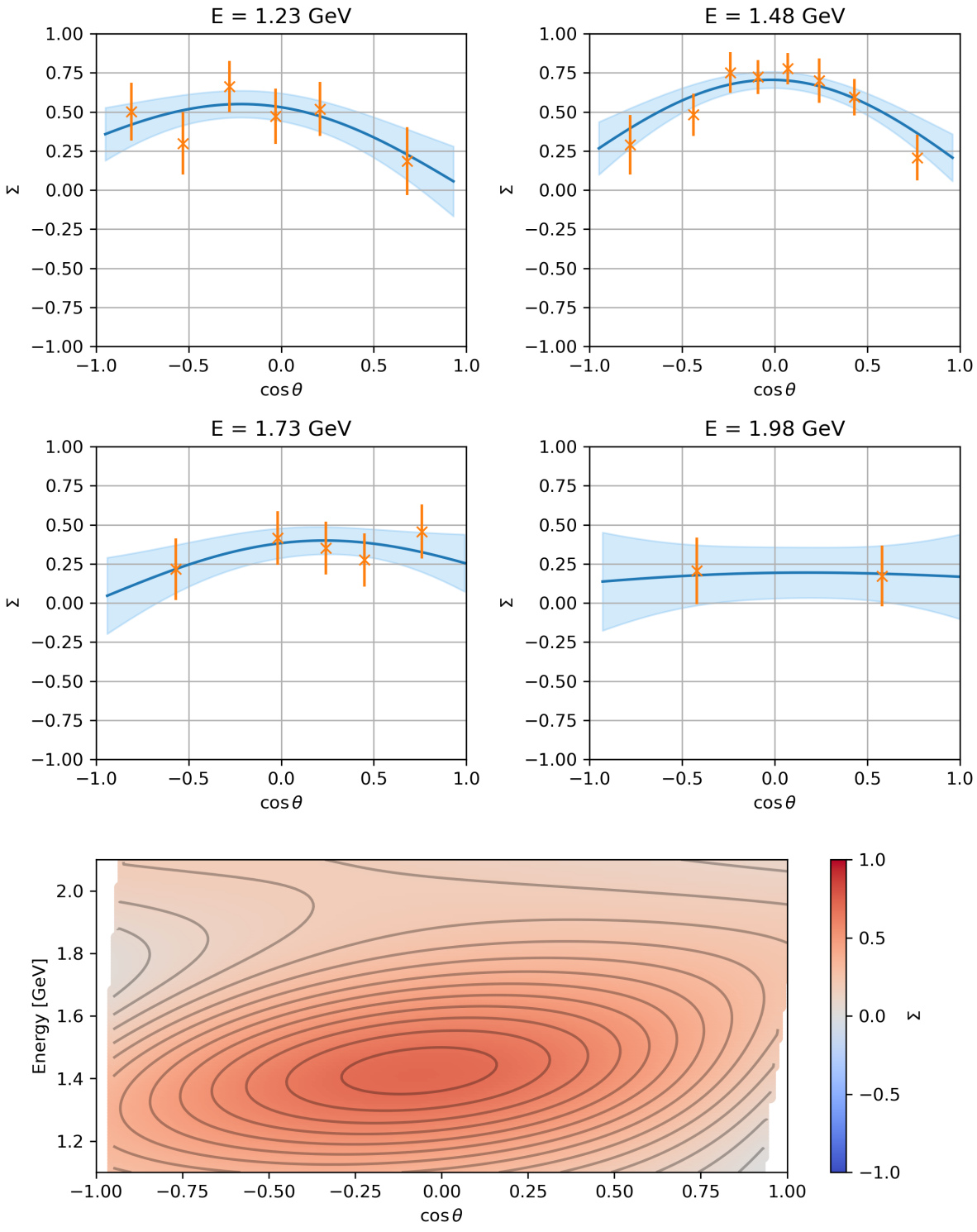



Figure 6.1: GP fit to the Σ polarisation observable for the $\gamma p \to K^0\Sigma^+$ reaction. The lower panel shows the two-dimensional GP surface as a function of photon energy and $\cos\theta$, while the upper panels show one-dimensional projections of the fitted surface at selected photon energies. The experimental measurements are shown for comparison.

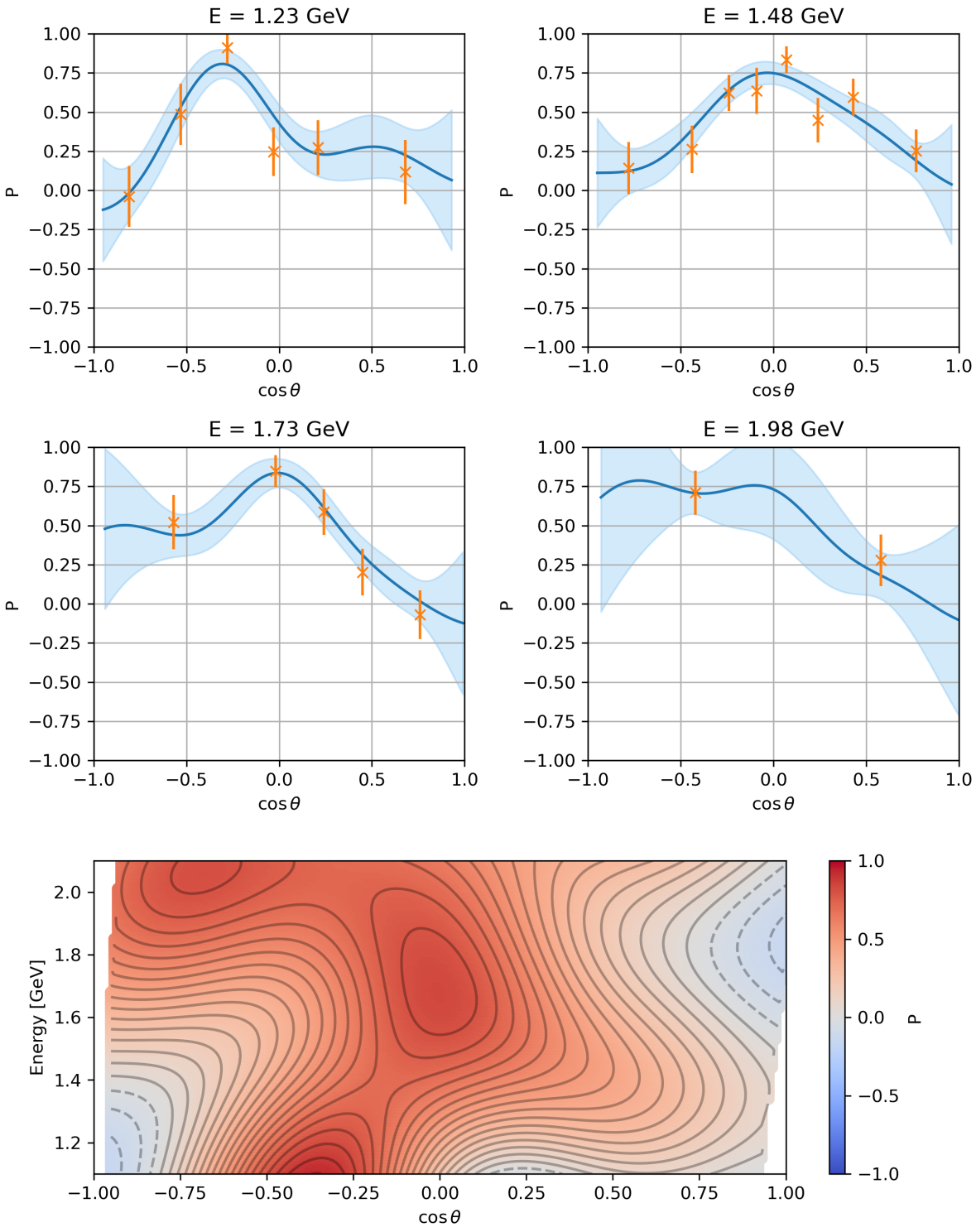



Figure 6.2: GP fit to the $P$ polarisation observable for the $\gamma p \to K^0\Sigma^+$ reaction. The lower panel shows the two-dimensional GP surface as a function of photon energy and $\cos\theta$, while the upper panels show one-dimensional projections of the fitted surface at selected photon energies. The experimental measurements are shown for comparison.

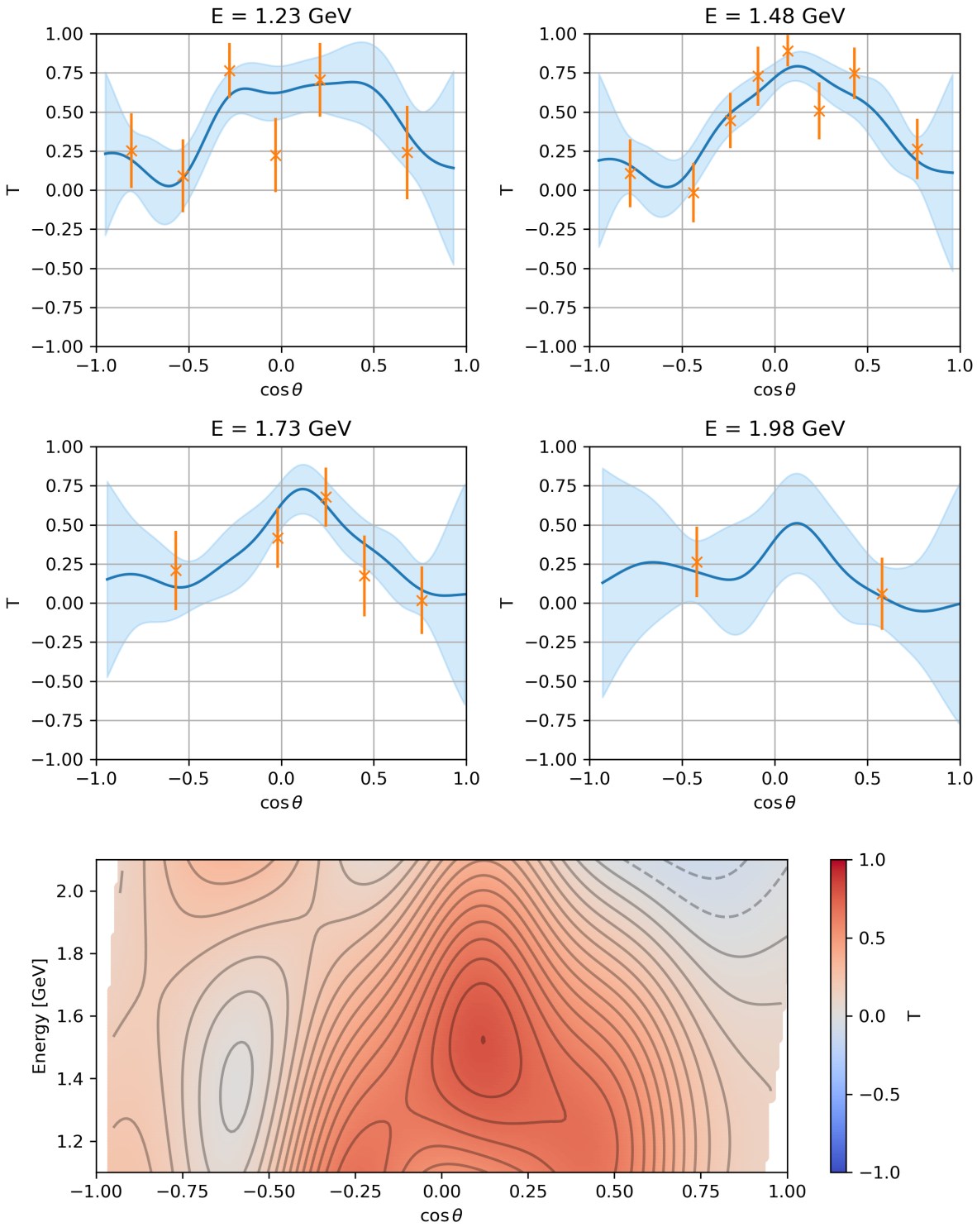



Figure 6.3: GP fit to the $T$ polarisation observable for the $\gamma p \to K^0\Sigma^+$ reaction. The lower panel shows the two-dimensional GP surface as a function of photon energy and $\cos\theta$, while the upper panels show one-dimensional projections of the fitted surface at selected photon energies. The experimental measurements are shown for comparison. The projection at 1.98 GeV exhibits a pronounced local maximum at positive $\cos\theta$.

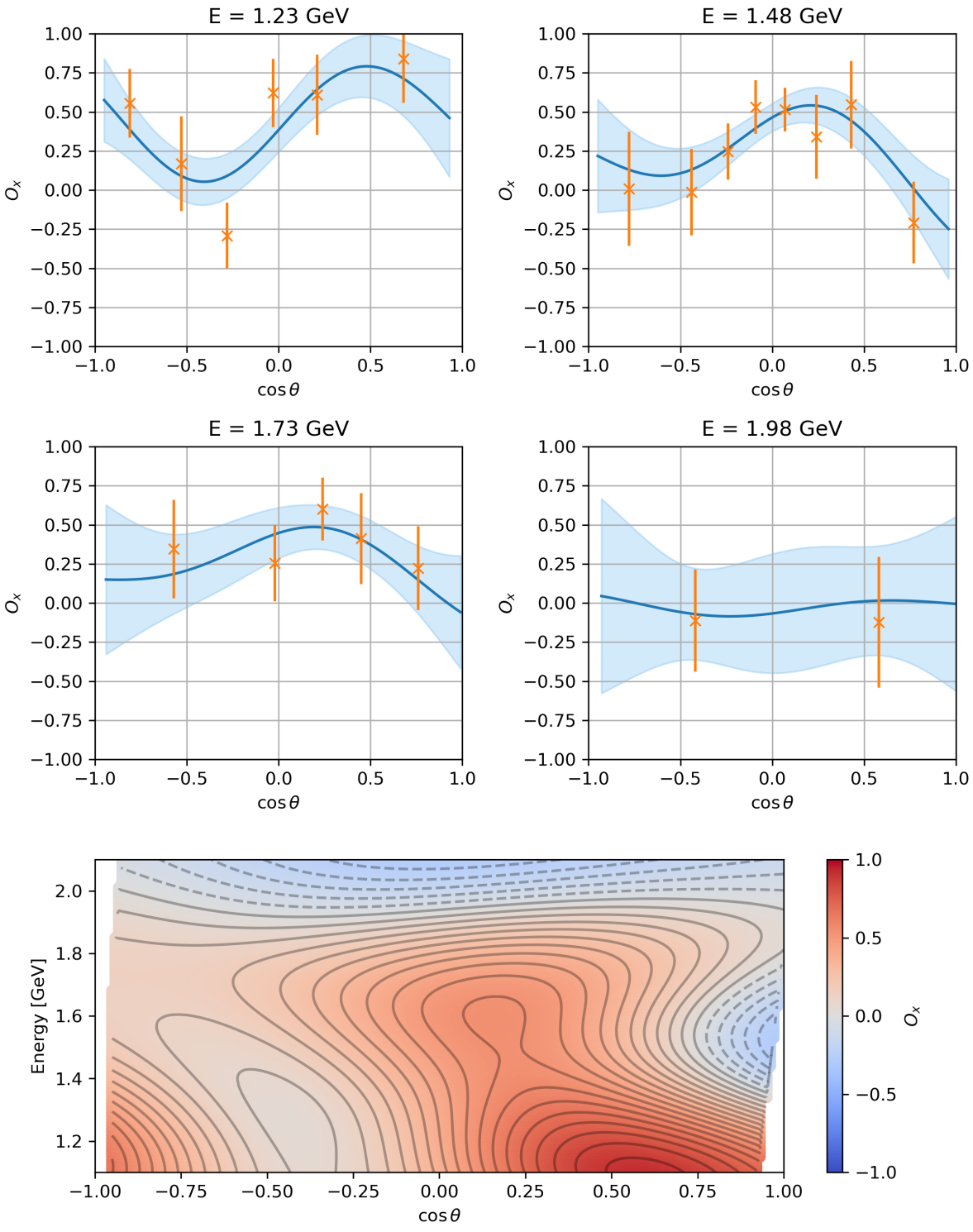



Figure 6.4: GP fit to the $O_x$ polarisation observable for the $\gamma p \to K^0\Sigma^+$ reaction. The lower panel shows the two-dimensional GP surface as a function of photon energy and $\cos\theta$, while the upper panels show one-dimensional projections of the fitted surface at selected photon energies. The experimental measurements are shown for comparison.

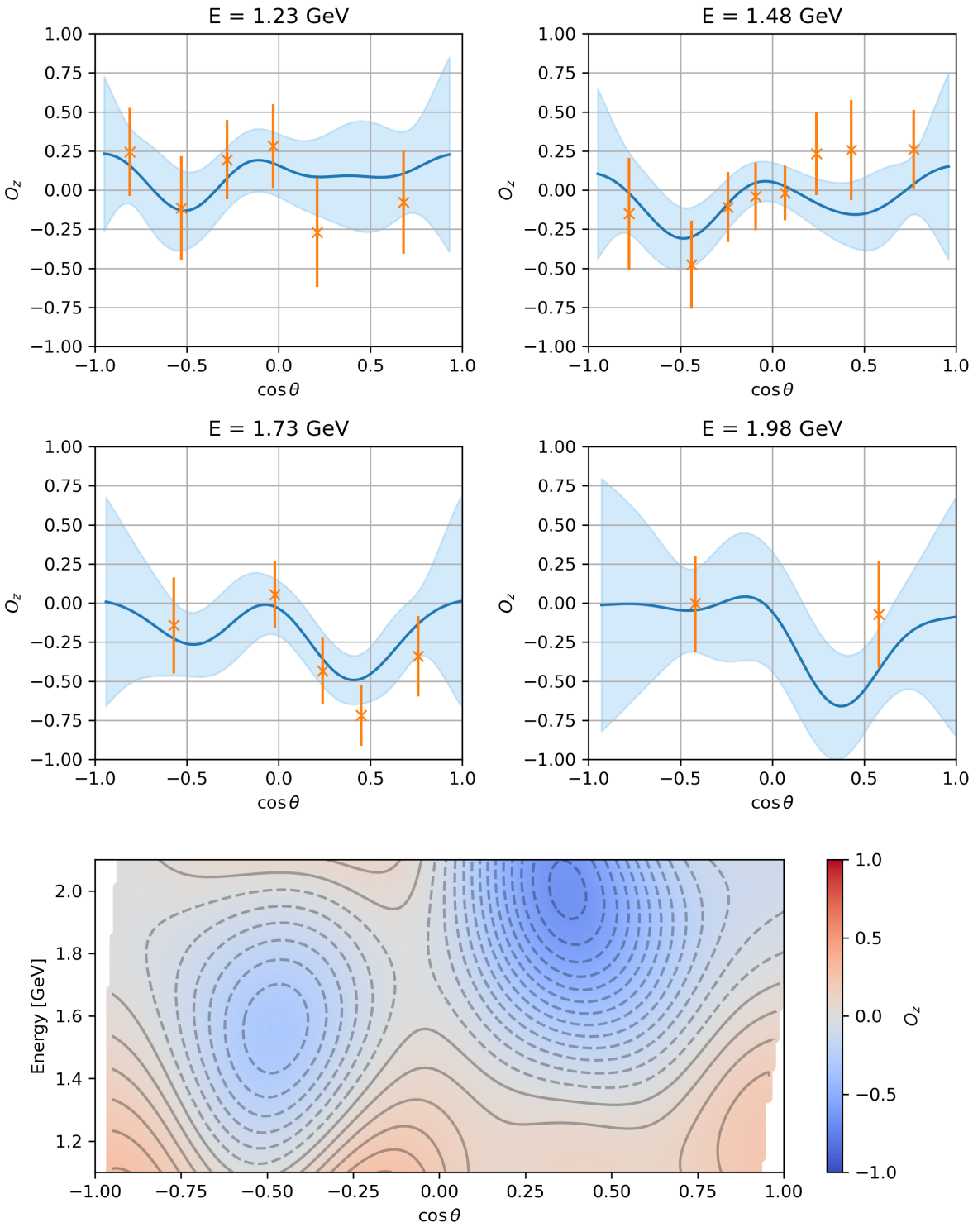



Figure 6.5: GP fit to the $O_z$ polarisation observable for the $\gamma p \to K^0\Sigma^+$ reaction. The lower panel shows the two-dimensional GP surface as a function of photon energy and $\cos\theta$, while the upper panels show one-dimensional projections of the fitted surface at selected photon energies. The experimental measurements are shown for comparison. The projection at 1.98 GeV exhibits a pronounced local minimum at positive $\cos\theta$.

### 6.1.2 Results and Interpretation

A key result of the two-dimensional treatment is that the dependence on $\cos\theta$ at a given energy is constrained by measurements across the surrounding energy–$\cos\theta$ region. The GP covariance correlates nearby points in both input dimensions, so measurements at neighbouring energies can contribute to the prediction at an energy where the experimental coverage is limited. The extent of this correlation is determined by the learned energy and angular length scales, allowing the fitted surface to incorporate information across both dimensions rather than treating each energy independently. This is particularly important for the present data, where only a small number of measurements are available at each photon energy. The resulting interpolation is therefore constrained not only by measurements at the energy being considered, but also by the structure of the observable across the surrounding kinematic region.

This behaviour is particularly evident in the projections at 1.98 GeV for $T$ and $O_z$. The $T$ observable exhibits a pronounced local maximum at positive $\cos\theta$, while $O_z$ exhibits a pronounced local minimum in a similar region of $\cos\theta$. These structures are obtained by evaluating the fitted two-dimensional surfaces at $E = 1.98$ GeV and therefore reflect the correlated energy–$\cos\theta$ dependence learned from the full dataset. The projections demonstrate that the two-dimensional treatment can retain non-trivial dependence on $\cos\theta$ even when only a small number of measurements are available at the energy being considered.

This is important because the experimental coverage at individual energies is sparse and non-uniform. If the dependence on $\cos\theta$ were instead fitted independently at each photon energy, only the measurements available at that energy would constrain the resulting function, and the information supplied by measurements at neighbouring energies would be discarded. In such a treatment, the pronounced structures observed in the $T$ and $O_z$ projections could be difficult or impossible to recover if the available measurements at 1.98 GeV do not provide sufficient information to resolve them. The smaller structures present in the $P$ and $O_x$ projections are likewise constrained by the surrounding two-dimensional data, while the comparatively smooth behaviour of $\Sigma$ indicates that the same framework can also represent observables for which there is less local structure. The differences between the fitted observables demonstrate that the GP does not impose a common functional shape, but instead adapts the interpolated surface to the constraints provided by each dataset.

The uncertainty associated with the fitted surface provides an indication of how strongly the interpolation is constrained by the experimental data. Regions with measurements are generally

more tightly constrained, while the uncertainty increases where the kinematic coverage becomes sparse. Consequently, structures inferred in poorly sampled regions should be interpreted with greater caution than those occurring where measurements provide stronger constraints. The spatial variation of the predictive uncertainty also provides information that could be useful when planning future measurements. Regions in which the predictive uncertainty is large indicate areas that are comparatively weakly constrained by the available multidimensional dataset and could therefore be considered when planning additional measurements. In this sense, the GP uncertainty can provide not only an assessment of the current interpolation, but also an indication of where additional measurements may be most informative. At the same time, the fitted surfaces provide a continuous representation of all five observables at arbitrary kinematic points within the region supported by the data, rather than restricting subsequent analyses to the discrete locations at which experimental measurements were reported. The associated uncertainties are retained when evaluating the GP at these points, allowing both the interpolated observable and the reliability of that interpolation to be considered together.

The results therefore demonstrate the principal benefit of applying the framework to sparse two-dimensional experimental data: the GP reproduces the available measurements while providing a continuous representation across the measured kinematic region, with information from neighbouring energies contributing to the dependence on $\cos\theta$. This provides a suitable representation of the sparse experimental information for subsequent physics analyses.

## 6.2 Application to Kaon Scattering Data

The second application of the GP framework considers total cross-section measurements for $K^- p$ scattering. Unlike the polarisation observable application in Section 6.1, these data are one-dimensional functions of the centre-of-mass energy and comprise measurements collected from multiple independent experiments over several decades. The analysis considers the six possible final states, $K^- p$, $\bar{K}^0 n$, $\Lambda\pi^0$, $\Sigma^0\pi^0$, $\Sigma^+\pi^-$ and $\Sigma^-\pi^+$, using the experimental measurements compiled in the available world dataset [93–108]. These channels form a coupled system in the strangeness $S = -1$ sector and are particularly relevant to studies of the $\Lambda(1405)$, whose properties are strongly connected to the coupled $\bar{K}N$ and $\pi\Sigma$ dynamics [39, 41–43]. The analysis therefore provides a test of the GP framework in a setting where the available measurements are not only sparse, but also heterogeneous and potentially inconsistent between experiments.

### 6.2.1 World Dataset

The first analysis considers the complete world dataset, with the measurements from the available experiments combined for each scattering channel. Applying the GP framework to these combined datasets tests whether the measurements can be represented by a single continuous function of energy. If the measurements provide statistically compatible constraints on a common underlying cross-section, the GP should produce a smooth interpolation through the available data. Conversely, substantial disagreement between experiments may result in competing constraints that cannot be accommodated naturally by a single smooth function.

The resulting GP fits are shown in Figure 6.6. In most channels, the fitted functions exhibit pronounced oscillatory behaviour between adjacent experimental measurements. These oscillations are particularly evident at lower energies, where the measurements from different experiments show substantial variation. The behaviour is considerably less pronounced in the channels for which the available measurements are provided by two non-overlapping experiments, where there is less direct competition between measurements over the same energy range.

The observed oscillations provide evidence that the measurements contributing to the world dataset cannot be represented adequately by a single smooth underlying function within the assumptions of the GP model. This interpretation is supported by the validation studies in Chapter 5, where the same framework was shown to recover known functional behaviour from sparse pseudodata while producing statistically meaningful uncertainties. The behaviour observed here is therefore qualitatively different from that obtained for the controlled validation datasets and from the smooth interpolation obtained for the polarisation observables in Section 6.1. The result instead indicates substantial inter-experimental tension within the world dataset.

When incompatible measurements are supplied simultaneously to a GP, the fitted function is required to accommodate all of the available constraints. If the measurements differ in their energy dependence or local structure, this can result in rapid variations between neighbouring data points as the model attempts to accommodate conflicting observations. The resulting interpolation should therefore not be interpreted as a physical prediction of the cross-section in these regions, since the behaviour reflects the incompatibility of the input measurements rather than a well-constrained underlying dependence. The world dataset result consequently demonstrates the importance of considering the consistency of the constituent experiments before interpreting a combined interpolation.

The result motivates separating the world dataset into its constituent experiments. This allows the differences between individual measurements to be examined directly and distinguishes variation intrinsic to the experimental data from behaviour introduced when the measurements are forced into a single interpolated function.

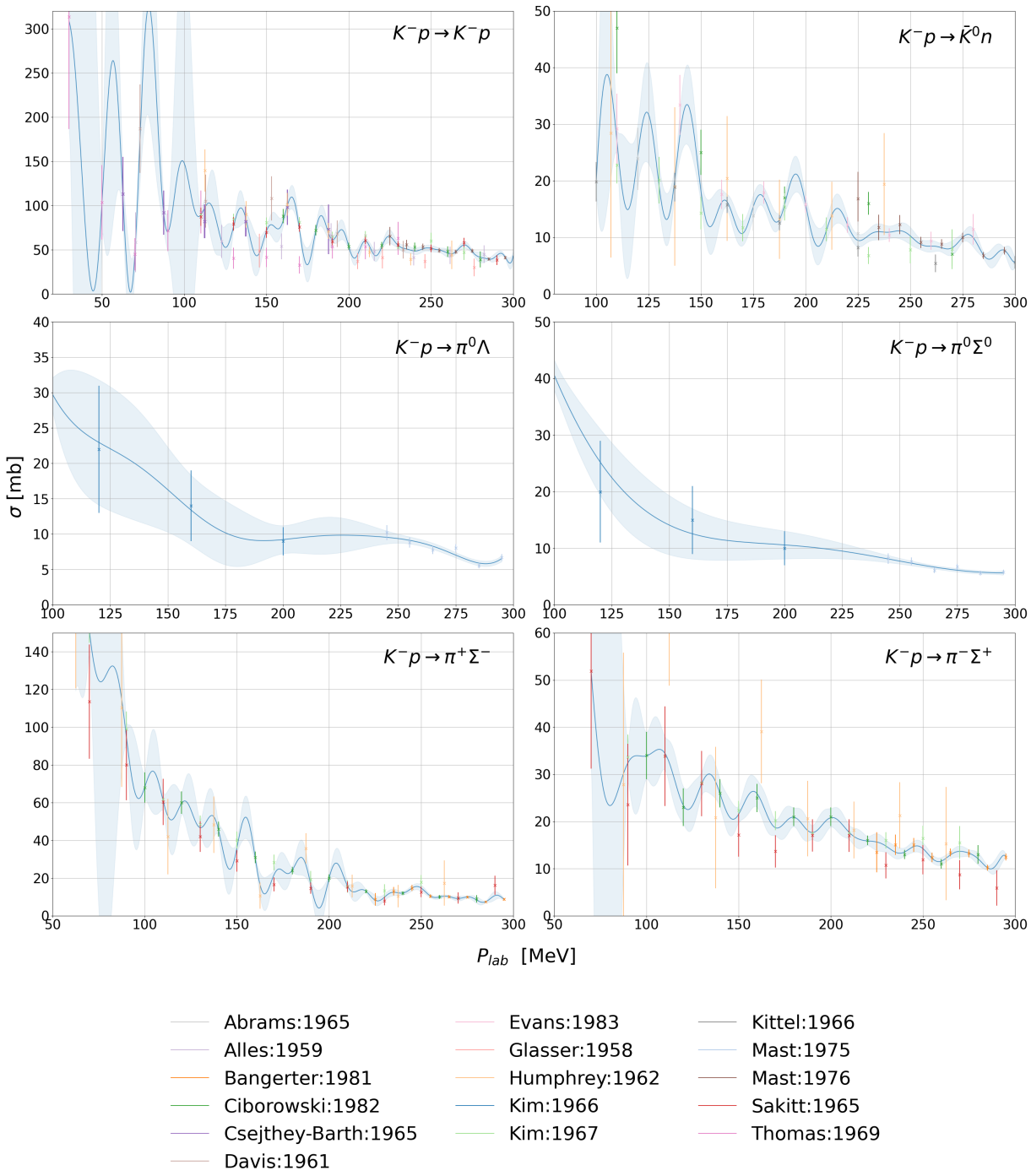



Figure 6.6: GP fits to the combined world datasets for the different $K^- p$ scattering final states. The curves show the posterior mean and one standard deviation uncertainty band obtained when measurements from all available experiments are fitted simultaneously within each channel. The pronounced oscillatory behaviour between neighbouring measurements in several channels, particularly at lower energies, indicates substantial tension between the constituent experimental datasets and demonstrates the difficulty of representing the world data by a single smooth function.

### 6.2.2 Individual Experimental Datasets

The individual components of the world dataset were therefore fitted separately using the same GP framework. The resulting fits are shown in Figure 6.7, with the mean prediction and one standard deviation uncertainty band shown for each experiment. A substantial variation in the measured cross-sections is evident between experiments, particularly at lower energies. The differences extend beyond an overall displacement or normalisation factor, with the experiments exhibiting different central values and local energy dependence. This demonstrates that the oscillatory behaviour observed in the combined world dataset is associated with genuine differences between the constituent experimental measurements rather than being produced solely by the interpolation procedure.

The individual fits also illustrate the effect of the amount of available experimental information on the resulting uncertainty. Some datasets are associated with substantially broader GP uncertainty bands than others. For example, the Abrams dataset for $K^- p \to \bar{K}^0 n$ contains only three measurements in the relevant energy region [93], providing relatively little information with which to constrain the underlying energy dependence. A wide range of functions can consequently remain compatible with these measurements under the GP prior, resulting in a larger posterior uncertainty. The increased uncertainty is therefore a reflection of the limited constraining power of the dataset rather than an indication of reduced reliability of the GP methodology itself.

The comparison between the individual fits also demonstrates why replacing the constituent measurements with a single effective curve can be problematic. Where several experiments provide overlapping measurements, their individual GP predictions can occupy substantially different regions of the cross-section–energy plane. Representing these measurements by a single central value would necessarily discard information about the extent and structure of this disagreement. In particular, the variation between experiments cannot in general be characterised solely by a single Gaussian uncertainty about a common mean when the underlying predictions exhibit distinct or separated structures.

The individual fits therefore provide the basis for constructing a representation that retains the information contained in the separate experimental measurements rather than enforcing their agreement. Such a representation can be formed by combining the probability densities associated with the individual GP predictions at each energy.

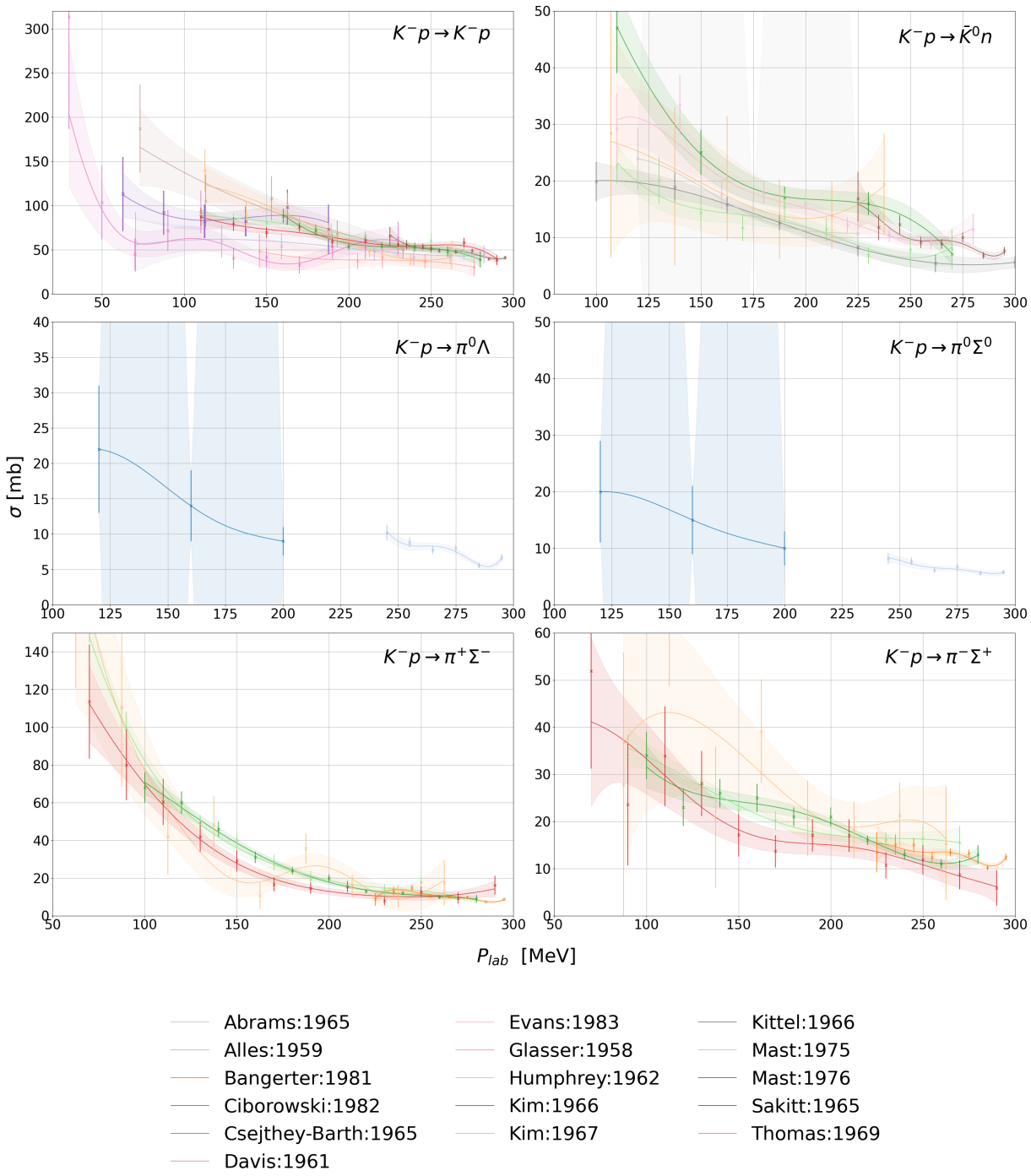



Figure 6.7: Individual GP fits to the experimental datasets contributing to the $K^-p$ scattering world dataset. The curves show the posterior mean and one standard deviation uncertainty band for each individual experiment. The substantial variation between experiments, particularly at lower energies, demonstrates differences in both the central values and local energy dependence of the measurements. The differing uncertainty bands also reflect the varying constraining power of the individual datasets.

### 6.2.3 Combined Probability Surfaces

The individual GP fits provide probabilistic descriptions of the cross-section for each experimental dataset. These can be combined by representing the prediction from each GP at a given energy as a Gaussian distribution and summing the resulting distributions. This construction is equivalent to a Gaussian mixture model [114] and produces a probability density over cross-section and energy while retaining the information from the individual experiments.

For a given energy $E$, the prediction from experiment $i$ is represented by

$$y_i(E) \sim \mathcal{N}\left(\mu_i(E), \sigma_i^2(E)\right), \tag{6.1}$$

where $\mu_i(E)$ and $\sigma_i(E)$ are the mean and standard deviation of the corresponding GP prediction. The combined probability density is then constructed as an equally weighted Gaussian mixture,

$$p(y \mid E) = \frac{1}{N}\sum_{i=1}^{N} \mathcal{N}\left(y; \mu_i(E), \sigma_i^2(E)\right), \tag{6.2}$$

where $N$ is the number of experimental datasets being combined.

The quantity $p(y \mid E)$ is a probability density in the cross-section $y$. The probability associated with a finite interval of cross-section can be approximated by

$$P(y \mid E) \approx p(y \mid E)\delta y, \tag{6.3}$$

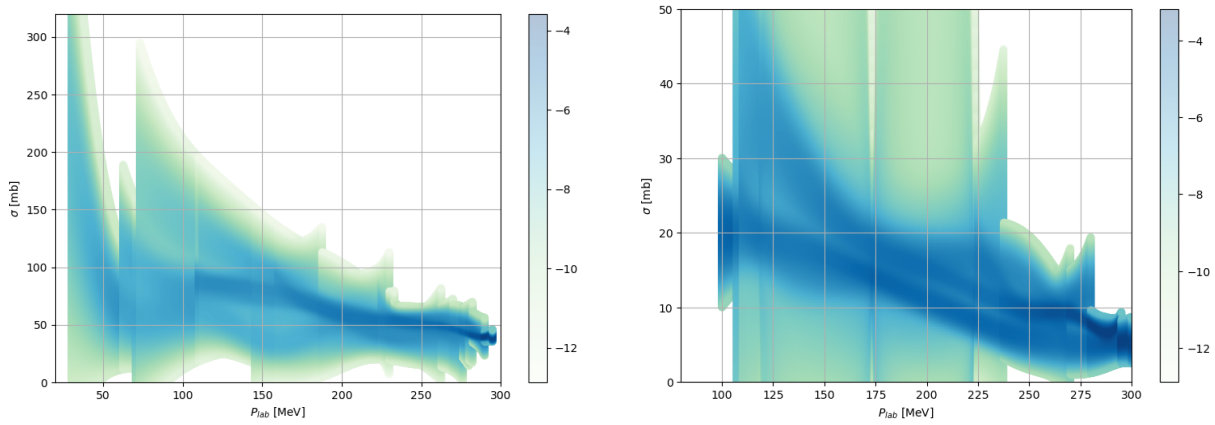
where $\delta y$ is the width of the cross-section interval being considered. Thus, multiplication by $\delta y$ converts the probability density into an approximate probability associated with a finite interval. This does not introduce additional information into the probability distribution, but provides a representation that is more directly related to the probability of a theoretical prediction lying within a specified range of the observable.

The resulting probability distribution depends on both the energy and the cross-section and can therefore be represented as a two-dimensional probability surface over the $(E, y)$ plane. At a fixed energy, a slice through this surface corresponds to the combined probability distribution of the cross-section at that energy, while variation along the energy axis describes how this distribution changes across the kinematic range. This representation is particularly useful when comparing theoretical predictions with the available experimental constraints, since a predicted

cross-section can be located within the surface and its degree of experimental support assessed.

In contrast to the dataset combination approaches discussed in Section 2.7, no additional weighting is assigned to the individual experiments in this construction. The relative quality and reliability of the datasets cannot be established from the GP predictions alone, and assigning weights based on quoted uncertainties, normalisation adjustments or other measures of experimental quality would introduce additional assumptions into the combined representation. In particular, such a weighting scheme could artificially favour agreement with one experiment while suppressing disagreement with another. Equal weighting is therefore adopted so that each experimental dataset contributes equally to the probability surface, without making an a priori judgement about which experiment provides the more reliable constraint. This allows both agreement and disagreement between the datasets to remain visible in the resulting probability density.

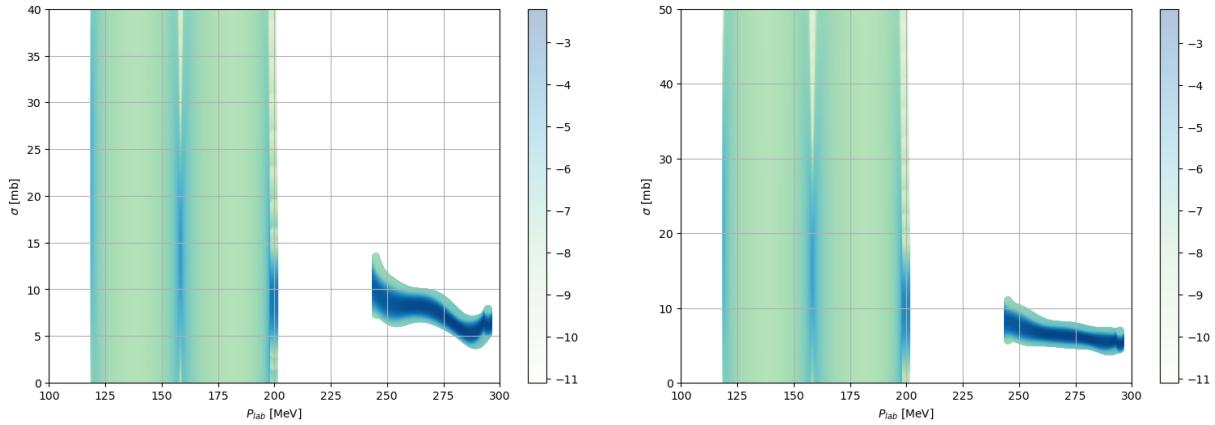
An experiment with a larger GP uncertainty therefore produces a broader Gaussian component, while a more precisely constrained prediction produces a narrower component. Repeating this construction over the energy range produces the probability surfaces shown in Figure 6.8. The surfaces display the logarithm of the probability associated with the finite cross-section interval, $\log[P(y \mid E)]$. Taking the logarithm allows the large dynamic range of the probability to be represented while retaining the relative structure of the experimentally supported regions.




(a) $K^- p \to K^- p$.



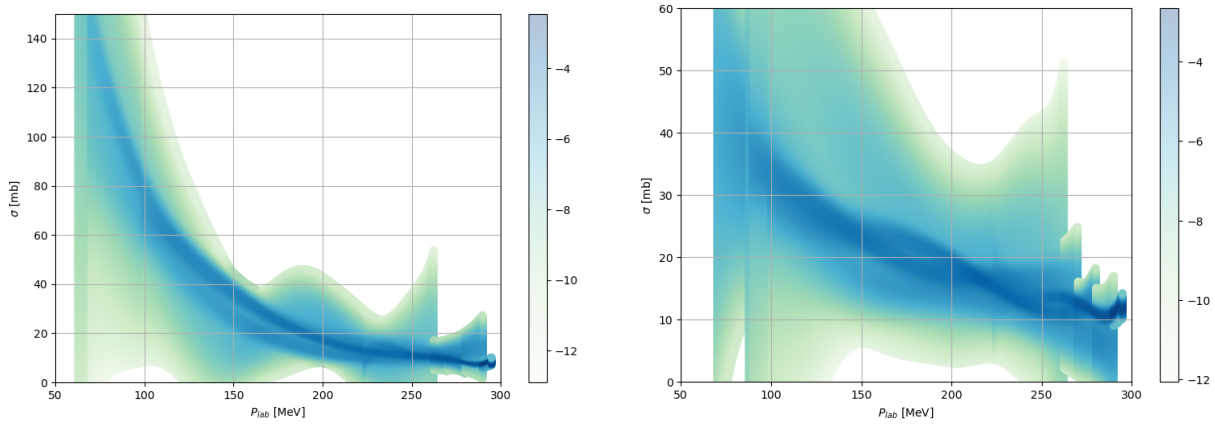

(b) $K^- p \to \bar{K}^0 n$.




(c) $K^- p \to \pi^0 \Lambda$.




(d) $K^- p \to \pi^0 \Sigma^0$.




(e) $K^- p \to \pi^+ \Sigma^-$.




(f) $K^- p \to \pi^- \Sigma^+$.

Figure 6.8: Combined probability surfaces constructed from the individual experimental GP predictions for the six $K^- p$ scattering final states. At each energy, the individual GP predictions are represented by Gaussian components whose sum is equivalent to a Gaussian mixture model. The resulting surfaces retain the uncertainties associated with the individual experiments and multimodal structure arising from disagreement between their measurements. The double-ridged structure visible in the $K^- p \to \bar{K}^0 n$ channel at lower energies illustrates the preservation of distinct regions of experimental support.

### 6.2.4 Interpretation of the Results

The probability surfaces provide a representation of the experimental information that differs from combining all measurements into a single fitted curve. Where the individual experiments are consistent, the probability density is concentrated around a common cross-section, providing a strong constraint on the observable. Where the experiments disagree, the separate regions of support remain visible. The resulting representation therefore shows the extent to which the available measurements can be reconciled without requiring them to be reduced to a single central value.

This distinction is particularly relevant at lower energies, where several of the experimental datasets exhibit substantial differences. In the $K^- p \to \bar{K}^0 n$ channel, for example, the double-ridged structure in Figure 6.8b indicates that the available measurements support two distinct regions of cross-section. A conventional combination into a single effective dataset could instead produce an intermediate value between these regions, potentially obscuring the disagreement between the experiments. The probability surface retains both regions and therefore provides a more direct representation of the experimental evidence available in this region.

The same behaviour can be seen to different extents in the other channels. Regions where the experimental datasets overlap produce concentrated probability density, whereas differences in their central values or energy dependence lead to broader or separated structures. The resulting surfaces therefore provide more information than the posterior mean of a GP fitted to the complete world dataset. In the latter case, disagreement between experiments can manifest as rapid variations in the fitted function as the GP attempts to satisfy competing constraints. The probability surfaces instead separate the contributions from the individual experiments, making it possible to distinguish the uncertainty associated with an individual measurement from the variation between different measurements.

This is particularly useful when considering the coupled nature of the $K^- p$ scattering system. The six final states considered in this analysis are coupled through the underlying reaction dynamics, so constraints from the different channels enter jointly when constructing a coupled-channel description. The low-energy region is especially important for studies of the $\Lambda(1405)$, whose properties are closely connected to the coupled $\bar{K}N$ and $\pi\Sigma$ channels [39, 41–43]. Preserving the experimental constraints separately therefore allows the structure of the available measurements in each channel to be considered when constructing a coupled-channel description, rather than requiring the experimental input to be reduced to a single curve beforehand.

The double-ridged structure observed in the $K^- p \to \bar{K}^0 n$ channel should not itself be interpreted as evidence for two physical states. The probability surface represents the distribution of experimental constraints, and the presence of multiple regions of probability can result from disagreement between measurements rather than from multiple physical contributions to the underlying cross-section. Nevertheless, retaining such structure is valuable because it prevents an experimental discrepancy from being removed before the data are considered within a coupled-channel analysis. A subsequent physical model can then be compared with the full experimental representation and determine whether the observed structure can be accounted for by the underlying reaction dynamics.

The probability surfaces also provide a useful way of identifying where the experimental information is comparatively well constrained and where greater uncertainty remains. A concentrated probability density indicates that the available experiments provide similar constraints, while broad or separated regions indicate that the inferred cross-section depends more strongly on which experimental dataset is considered. This distinction is important when using the experimental information to constrain theoretical models, since a model prediction lying within a broad region of probability should not be interpreted in the same way as one lying within a tightly concentrated region. The regions of disagreement may also provide useful guidance for future experimental measurements. Where independent datasets give substantially different constraints, additional measurements in the corresponding kinematic region could help determine whether the difference reflects statistical fluctuations, systematic effects or genuine variation in the underlying observable. The probability surfaces therefore provide not only a representation of the current experimental constraints, but also an indication of regions where further measurements could be particularly informative.

The approach therefore differs from conventional dataset combination procedures in its treatment of experimental disagreement. Weighted averages, normalisation adjustments and global fits can be appropriate when the constituent measurements are sufficiently consistent and a single best estimate is required. However, when the measurements exhibit substantial differences, these approaches necessarily require assumptions about how the disagreement should be treated. The present approach postpones this decision by retaining the separate experimental constraints. This is particularly useful when there is insufficient independent information to determine whether the observed differences arise from statistical fluctuations, systematic effects or other experimental considerations.

An alternative approach to the treatment of differences between experiments would be to extend the GP model to include experiment-dependent nuisance parameters. For example, an additive

model could take the form

$$y_i(x) = f(x) + c_i + \varepsilon_i, \tag{6.4}$$

where $c_i$ represents an experiment-specific offset with an associated prior distribution. Such a model provides a natural way of accounting for systematic differences that can be represented by a constant shift between datasets. However, the differences observed between the experimental datasets considered here are not restricted to such offsets, with differences also present in their local energy dependence. An experiment-dependent constant would therefore not fully capture the variation observed between the datasets. More generally, introducing such parameters would require assumptions about the form and magnitude of the inter-experimental differences. The present approach instead avoids imposing such a model and retains the separate experimental constraints for subsequent analysis.

A limitation of the present interpretation is that the probability surfaces cannot determine the origin of the disagreement between experiments. The GP propagates the uncertainties supplied with the measurements, but systematic effects that are not included in these uncertainties cannot be identified through the interpolation. Correlations between measurements arising from common experimental sources are likewise not explicitly represented. This limitation enters at the experimental measurement stage of the physics pipeline described in Section 2.7, before the observables are subsequently used in reaction amplitude and coupled-channel analyses. The probability surfaces should therefore be interpreted as a representation of the available experimental constraints rather than as a statistical resolution of the differences between experiments.

Overall, the results demonstrate that the GP framework can be used not only to interpolate individual sparse experimental datasets, but also to construct a probabilistic representation of heterogeneous measurements without first forcing them into agreement. The resulting surfaces preserve both the uncertainty of the individual datasets and the differences between their predictions. For the $K^- p$ scattering data considered here, this provides a useful representation of the experimental input to subsequent coupled-channel analyses, particularly in regions where the available measurements do not define a unique underlying cross-section.

# Chapter 7

# Conclusions and Future Work

## 7.1 Summary of the Thesis

This thesis has developed a Gaussian Process (GP) framework for constructing continuous, uncertainty-aware representations of sparse experimental data in hadron physics. The work was motivated by the limited and irregular kinematic coverage of many hadron physics measurements, where a useful representation must account for both the available experimental constraints and the uncertainty associated with predictions between measured points.

The framework addresses three main challenges: determining suitable GP hyperparameters for sparse datasets, defining the region of kinematic space supported by the available measurements, and combining information from multiple experimental datasets while retaining their individual statistical structure. These aspects were investigated using pseudodata and subsequently demonstrated using experimental polarisation observables and kaon scattering cross-sections.

## 7.2 Principal Findings and Original Contributions

A central contribution of this thesis is the development of a hyperparameter optimisation procedure designed for the sparse datasets considered here. Rather than relying on marginal-likelihood maximisation, the procedure evaluates the statistical behaviour of the GP predictive uncertainties through empirical coverage. The Wasserstein distance is used to quantify the agreement between the predicted and expected distributions, providing a loss function that can be directly incorporated into the optimisation. Combined with Particle Swarm Optimisation (PSO) and physics-informed parameter bounds, this provides a practical strategy for determining suitable GP hyperparameters.

The validation studies using pseudodata demonstrate that the resulting predictive uncertainties can be assessed against known statistical expectations. The pseudodata also provide a controlled environment in which the ability of the GP to recover known functional behaviour can be investigated independently of the complexities of experimental measurements. These studies therefore provide validation of both the uncertainty calibration procedure and the broader GP framework.

A second contribution is the development of a discretised multidimensional convex hull procedure for defining the region of kinematic space supported by the experimental data. This provides a means of distinguishing interpolation from extrapolation and prevents regions outside the experimental coverage from being treated in the same way as regions constrained by measurements. The discretised construction allows the resolution to be controlled independently in each dimension and permits experimental binning information to be incorporated into the representation.

The framework was applied to the two-dimensional polarisation observables $\Sigma$, $P$, $T$, $O_x$ and $O_z$ for the $\gamma p \to K^0\Sigma^+$ reaction. The resulting GP surfaces provide continuous representations of the experimental measurements across the energy–angle plane while retaining quantified predictive uncertainties. The covariance structure allows measurements at neighbouring energies and angles to contribute to predictions across the kinematic space. This allows non-trivial angular structures to be represented even where only a limited number of measurements are available at an individual energy.

The polarisation observable application also demonstrates the value of retaining the predictive uncertainty alongside the interpolated observable. Regions with stronger experimental con-

straints generally have smaller uncertainties, while sparsely sampled regions retain larger uncertainties. The resulting surfaces therefore provide information not only about the estimated observable but also about the degree to which that estimate is constrained by the available measurements.

A further contribution is the development of a method for combining independent GP representations of multiple experimental datasets into probability surfaces. Rather than first reducing the measurements to a single averaged dataset, the individual experimental constraints are retained throughout the GP regression and subsequently combined. This allows non-trivial features of the individual datasets to remain visible in the resulting probability distribution.

The application to kaon scattering demonstrates this approach in a setting where different experimental datasets can impose distinct constraints on the observable. In particular, the resulting probability surfaces can contain separated regions of high probability, as observed for the $K^- p \to \bar{K}^0 n$ channel near the $\Lambda(1405)$ region. Such structures would not necessarily be apparent in a representation based only on an averaged best fit trajectory. The probability surface approach therefore provides a means of retaining information about both agreement and tension between experimental datasets.

Taken together, these contributions establish a framework for representing sparse hadron physics data probabilistically while retaining quantified uncertainties and the statistical structure of the available measurements. The framework has been demonstrated for two-dimensional polarisation observables and one-dimensional kaon scattering cross-sections, providing a basis for its application to other experimental and theoretical problems.

## 7.3 Conclusions

The results of this thesis demonstrate the feasibility of using Gaussian Processes to construct continuous, uncertainty-aware representations of sparse experimental data in hadron physics. The combination of empirical uncertainty calibration, multidimensional definition of the data-supported region and probabilistic treatment of multiple datasets addresses several practical challenges associated with applying GP regression to these data.

The resulting representations provide a connection between discrete experimental measurements and the continuous or probabilistic information required by subsequent physical analyses. The GP uncertainty provides an explicit indication of how strongly different regions are constrained and can therefore also identify areas where additional experimental measurements could provide stronger constraints.

The work does not replace the underlying physical models or subsequent phenomenological analyses. Instead, it provides a statistical representation of the experimental information that can be used alongside them. In particular, the probability surfaces constructed from multiple datasets retain information that may otherwise be lost when experimental measurements are reduced to a single averaged representation.

The framework developed in this thesis therefore provides a foundation for further investigation of GP methods in hadron physics. The principal extensions considered in this work concern higher-dimensional observables, physics-informed covariance functions and the use of GPs as emulators of computationally expensive theoretical calculations.

## 7.4 Future Work

The GP framework developed in this thesis has been demonstrated using two-dimensional pseudoscalar meson photoproduction observables and one-dimensional kaon scattering cross-sections. Several extensions would provide opportunities to investigate the broader applicability of the methodology.

### 7.4.1 Higher Dimensional Observables

The existing GP framework can be extended beyond the two-dimensional pseudoscalar meson photoproduction studies to investigate processes characterised by higher-dimensional kinematic dependence. Although the current implementation has been tested using, at most, two kinematic variables, the underlying GP methodology is not inherently restricted to two dimensions. Higher-dimensional applications would nevertheless provide an important test of the scalability

of the framework, particularly because experimental measurements generally become increasingly sparse as the dimensionality of the kinematic space increases.

One particularly relevant example is Deeply Virtual Compton Scattering (DVCS), an exclusive lepton-hadron scattering process in which an incoming electron exchanges a highly virtual photon with a proton, resulting in the emission of a real photon while leaving the proton intact. As illustrated in Figure 7.1, the process is described in terms of two scattering planes: the lepton plane defined by the incoming and outgoing electron momenta $(k, k')$ and the hadron plane associated with the virtual and real photon momenta $(q_1, q_2)$. The figure also defines the azimuthal angle $\phi$ between these planes, together with the target spin orientation $S$, which play an important role in the angular dependence of DVCS observables. DVCS is of considerable interest in hadron physics because it provides access to the proton's Generalised Parton Distribution (GPD), which encode correlated information about the longitudinal momentum and spatial structure of quarks and gluons inside the proton. DVCS is typically characterised by five kinematic dimensions: the squared four-momentum transfer of the exchanged photon $Q^2$, which sets the interaction energy scale; the Bjorken scaling variable $x_B$, related to the momentum fraction carried by the struck parton; the squared momentum transfer to the proton $t$, which probes the transverse spatial structure of the target; the azimuthal angle $\phi$ between the lepton and hadron scattering planes; and the beam energy, which determines the overall collision kinematics and accessible phasespace region [115].

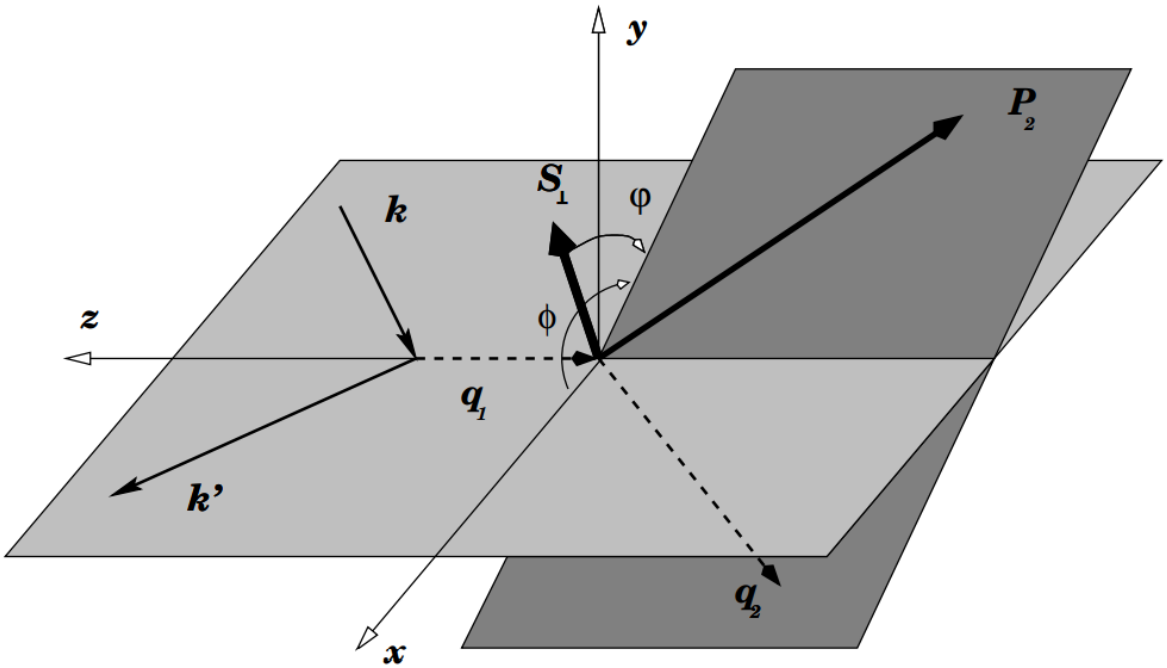



Figure 7.1: Kinematics of DVCS. Taken from [115].

Extending the framework to such a problem would provide a significantly more demanding test of the methodology. As the dimensionality increases, the volume of the accessible phase space grows rapidly, while the number of measurements required to obtain uniform coverage increases correspondingly. A dataset that provides reasonable coverage in each individual variable may therefore remain sparse when considered in the full multidimensional space. The GP must consequently make use of correlations between the available measurements while providing an uncertainty that reflects the degree to which each region is constrained.

The choice of kernel, sampling strategy and hyperparameter optimisation procedure would become increasingly important in higher-dimensional applications. The RBF kernel used in the present work provides a useful baseline, but its performance in a higher-dimensional and potentially more structured physical system would need to be assessed. The predictive uncertainty could also be used to identify regions where the available data provide weak constraints and where additional experimental or theoretical sampling would be most valuable.

Applying the framework to a higher-dimensional process such as DVCS would therefore provide an opportunity to investigate its scalability beyond the low-dimensional applications considered in this thesis.

### 7.4.2 Physics Informed Kernels

A natural extension of the framework is the investigation of physics-informed choices of covariance kernel. The RBF kernel provides a useful baseline for the applications considered in this thesis because it imposes a smooth correlation structure across the kinematic space. However, it also assumes a particular form of smoothness, and alternative covariance structures may be more appropriate for physical systems exhibiting different types of behaviour.

As discussed in Section 3.6, the Matérn class of kernels provides a means of controlling the smoothness of the underlying function through an additional parameter. Rational quadratic kernels provide another possible extension by representing a range of characteristic length scales, while periodic kernels may be useful where periodic behaviour is known to occur.

Composite kernels can also be constructed from combinations of simpler covariance functions. Such constructions could allow different types of behaviour to be represented simultaneously,

for example a smooth overall dependence together with shorter scale structure.

More generally, physics-informed kernel design could incorporate known symmetries, conservation laws or invariances directly into the covariance function. This would provide a systematic means of introducing established physical knowledge into the GP prior while retaining the probabilistic uncertainty quantification of the framework.

Future work could therefore compare alternative and physics-informed kernels using both predictive performance and uncertainty calibration. This would establish whether additional physical structure can improve the representation of sparse experimental data without artificially reducing the uncertainty in regions where experimental constraints remain weak.

### 7.4.3 Emulator

The GP framework developed in this thesis could also be extended to construct emulators of computationally expensive theoretical models. In theoretical hadron physics, repeated evaluation of a reaction model across a large parameter space can require substantial computational resources, particularly when the model contains several free parameters or when the calculation is embedded within a statistical inference procedure. A GP emulator could provide a computationally efficient approximation to such calculations while retaining an estimate of the emulator uncertainty.

In this application, the GP inputs would include both the kinematic variables and the parameters defining the theoretical model. For example, a theoretical model with three free parameters evaluated across two kinematic dimensions would result in a five-dimensional GP input space. The GP would learn the dependence of the theoretical prediction on both the model parameters and the kinematic variables.

The emulator would be trained using a finite set of full theoretical calculations selected to cover the relevant model-parameter and kinematic space. Once trained, it could replace repeated evaluations of the full theoretical model during parameter inference, reducing the computational cost of exploring the model parameter space.

The uncertainty provided by the emulator could also be used to identify regions where additional full theoretical calculations are required. New calculations could then be added to the training set and the emulator retrained, providing an adaptive refinement procedure in which computational resources are concentrated in regions where the emulator is less well constrained.

This would provide a natural extension of the uncertainty-aware methodology developed in this thesis from experimental interpolation to theoretical emulation. It could be particularly valuable for analyses in which large numbers of model evaluations are required, such as Bayesian parameter inference or coupled-channel phenomenology.

# Appendix A

# PSO vs MCMC

As discussed in Section 4.5, PSO was selected as the primary hyperparameter optimisation method after comparison with MCMC. Both methods were applied to the same ensemble of pseudodatasets and evaluated using the combined loss function of Equation 4.14. This appendix presents the results of this comparison.

The MCMC procedure was considered converged when both convergence criteria defined in Section 3.8 were satisfied. In particular, the Gelman–Rubin statistic, defined in Equation 3.56, was required to satisfy $\hat{R} < 1.1$, while the relative change in the integrated autocorrelation time, defined in Equation 3.60, was required to satisfy $\tau_{\text{stability}} < 0.1$. The $\hat{R} < 1.1$ threshold was chosen as the convergence criterion used for the comparison presented here. Although more stringent thresholds, such as $\hat{R} < 1.01$, are commonly used in modern applications, stricter criteria generally increase the computational cost of the MCMC procedure [116, 117]. The $\tau_{\text{stability}} < 0.1$ criterion requires successive estimates of the integrated autocorrelation time to differ by less than 10%, providing a practical measure of when the autocorrelation estimate has stabilised.

## A.1 Validation of MCMC Optimisation

The MCMC optimisation procedure described in Section 3.8 was applied to the same pseudo-datasets used to validate the PSO-based framework in Chapter 5. The same coefficients and initial "known" datapoints were used, allowing the resulting GP surfaces to be compared directly.

The uncertainty calibration and recovery-of-functional-form results obtained using MCMC are summarised in Tables A.1 and A.2, respectively. The corresponding pull distributions are shown in Figure B.2 in Appendix B.

| $\sigma$ Band | Expected Percentage of Points (%) | Measured Percentage of Known Points (%) | Measured Percentage of GP Points (%) |
|---|---|---|---|
| $0.67\sigma$ | 50.0 | 49.6 | 55.2 |
| $1\sigma$ | 68.3 | 68.2 | 74.8 |
| $1.96\sigma$ | 95.0 | 95.3 | 97.1 |

Table A.1: Expected and measured coverage fractions for the pseudodata and corresponding GP points using MCMC optimisation.

The results show that the GP surfaces obtained using MCMC exhibit similar coverage behaviour to those obtained using PSO. The coverage is somewhat higher than the nominal values at $0.67\sigma$ and $1\sigma$, while remaining close to the expected value at $1.96\sigma$. These results indicate that the uncertainty calibration obtained using MCMC is sufficient for the pseudodata validation.

The recovery of the underlying functional form provides a second comparison between the original pseudodata and the GP representation. The fitted coefficient pulls are summarised in Table A.2, with the full pull distributions provided in Appendix B.

| | Pseudodata | | GP Datapoints | |
|---|---|---|---|---|
| Coefficient | Mean | Variance | Mean | Variance |
| $c_0$ | -0.037 | 0.746 | -0.101 | 0.792 |
| $\mu_0$ | 0.111 | 0.731 | -0.023 | 0.862 |
| $\sigma_0^2$ | -0.058 | 0.636 | -0.078 | 0.627 |
| $c_1$ | -0.098 | 0.902 | -0.141 | 0.716 |
| $\mu_1$ | -0.012 | 0.658 | 0.013 | 0.577 |
| $\sigma_1^2$ | -0.146 | 0.630 | -0.015 | 0.748 |
| $c_2$ | 0.093 | 0.706 | 0.015 | 0.484 |
| $\mu_2$ | -0.008 | 0.837 | 0.025 | 0.645 |
| $\sigma_2^2$ | -0.047 | 0.623 | -0.198 | 0.635 |
| $c_3$ | -0.008 | 1.013 | 0.029 | 0.525 |
| $\mu_3$ | -0.115 | 0.685 | -0.043 | 0.783 |
| $\sigma_3^2$ | -0.208 | 0.635 | -0.099 | 0.885 |

Table A.2: Means and variances of the Gaussian distributions fitted to the pull distributions of the fitted coefficients obtained using the original pseudodata and sampled GP datapoints, with MCMC optimisation.

The coefficient pull distributions obtained using MCMC are broadly comparable between the original pseudodata and the sampled GP datapoints. Although individual coefficients show differences in the fitted variances, there is no consistent indication of a substantial change in the information available for the subsequent functional-form fit. Thus, as with PSO, the MCMC-optimised GP surfaces satisfy the requirements of the pseudodata validation.

## A.2 Comparison of PSO and MCMC

Since both optimisation methods satisfy the pseudodata validation tests, the choice between them is determined by their performance in the optimisation itself. The loss scores and computational runtimes obtained from the two methods are therefore compared below.

The loss scores of Equation 4.14 for PSO and MCMC are shown in Figure A.1 and summarised in Table A.3. Since both methods were applied to the same pseudodatasets, the loss obtained for

each dataset can be compared directly.

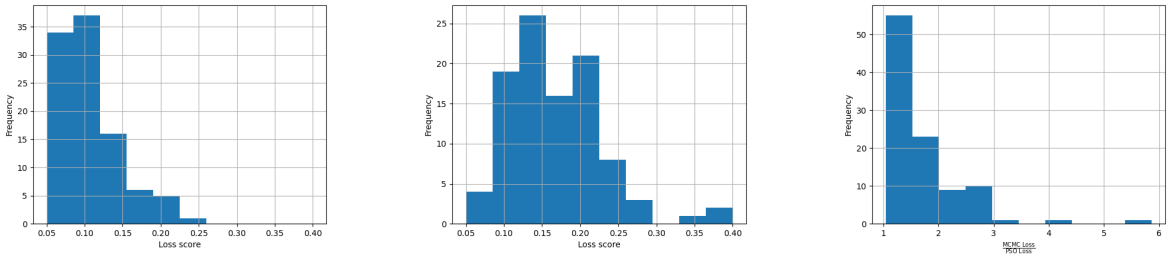
| | PSO Loss | MCMC Loss | $\frac{\text{MCMC Loss}}{\text{PSO Loss}}$ |
|---|---|---|---|
| Mean | 0.106 | 0.167 | 1.675 |
| Std Dev | 0.040 | 0.062 | 0.683 |
| Min | 0.049 | 0.077 | 1.039 |
| Max | 0.240 | 0.394 | 5.867 |

Table A.3: Summary statistics of the loss scores obtained using PSO, MCMC, and their ratio across the pseudodatasets considered.

The ratio of the MCMC loss to the corresponding PSO loss is shown in the final column of Table A.3 and in Figure A.1. The results show that

$$\min\left(\frac{\text{MCMC Loss}}{\text{PSO Loss}}\right) > 1. \tag{A.1}$$

Thus, for every pseudodataset considered, MCMC produced a higher loss than PSO. The ratio reached a maximum of approximately 5.9, although approximately 75% of the ratios were below 2. The mean ratio was 1.675, indicating that the MCMC loss was, on average, approximately 68% higher than the corresponding PSO loss.




(a) PSO (b) MCMC (c) Normalised

Figure A.1: Comparison of the optimisation loss scores obtained using PSO and MCMC. The first two panels show the loss values obtained by each optimisation method, while the third shows the MCMC loss normalised by the corresponding PSO loss.

The computational runtimes for PSO and MCMC, measured using the `time` package [118], are shown in Figure A.2 and summarised in Table A.4. As with the loss scores, the runtimes can be compared directly because both methods were applied to the same pseudodatasets.

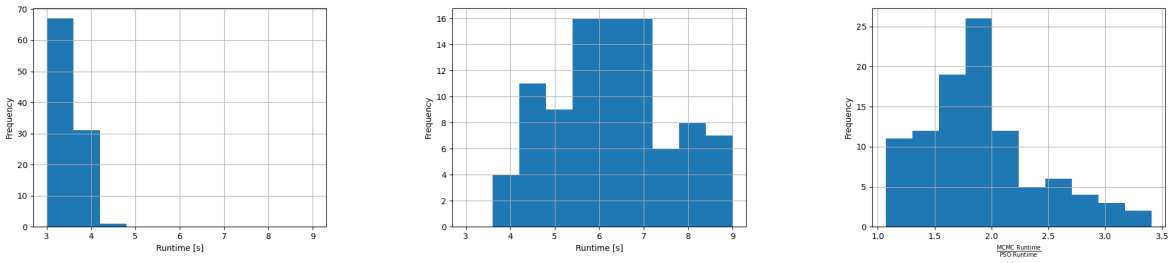
| | PSO Runtime (s) | MCMC Runtime (s) | $\frac{\text{MCMC Runtime}}{\text{PSO Runtime}}$ |
|---|---|---|---|
| Mean | 3.474 | 6.577 | 1.904 |
| Std Dev | 0.277 | 1.711 | 0.519 |
| Min | 2.970 | 3.910 | 1.068 |
| Max | 4.207 | 11.549 | 3.410 |

Table A.4: Summary statistics of the computational runtimes obtained using PSO, MCMC, and their ratio across the pseudodatasets considered.

The ratio of the MCMC runtime to the corresponding PSO runtime is shown in the final column of Table A.4 and in Figure A.2. The results show that

$$\min\left(\frac{\text{MCMC Runtime}}{\text{PSO Runtime}}\right) > 1. \tag{A.2}$$

Thus, MCMC required more computational time than PSO for every pseudodataset considered. The maximum runtime ratio was approximately 3.4, while approximately 65% of the ratios were below 2. The mean ratio was 1.904, indicating that MCMC required approximately 90% more computational time on average.




(a) PSO (b) MCMC (c) Normalised

Figure A.2: Comparison of the computational runtimes obtained using PSO and MCMC. The first two panels show the runtimes obtained by each optimisation method, while the third shows the MCMC runtime normalised by the corresponding PSO runtime.

The comparison therefore indicates that PSO provides a more effective optimisation method for the loss function considered here. Both methods were capable of producing GP surfaces that satisfied the pseudodata validation tests, but PSO consistently obtained lower loss values while also requiring less computational time. Given the specific optimisation problem addressed in

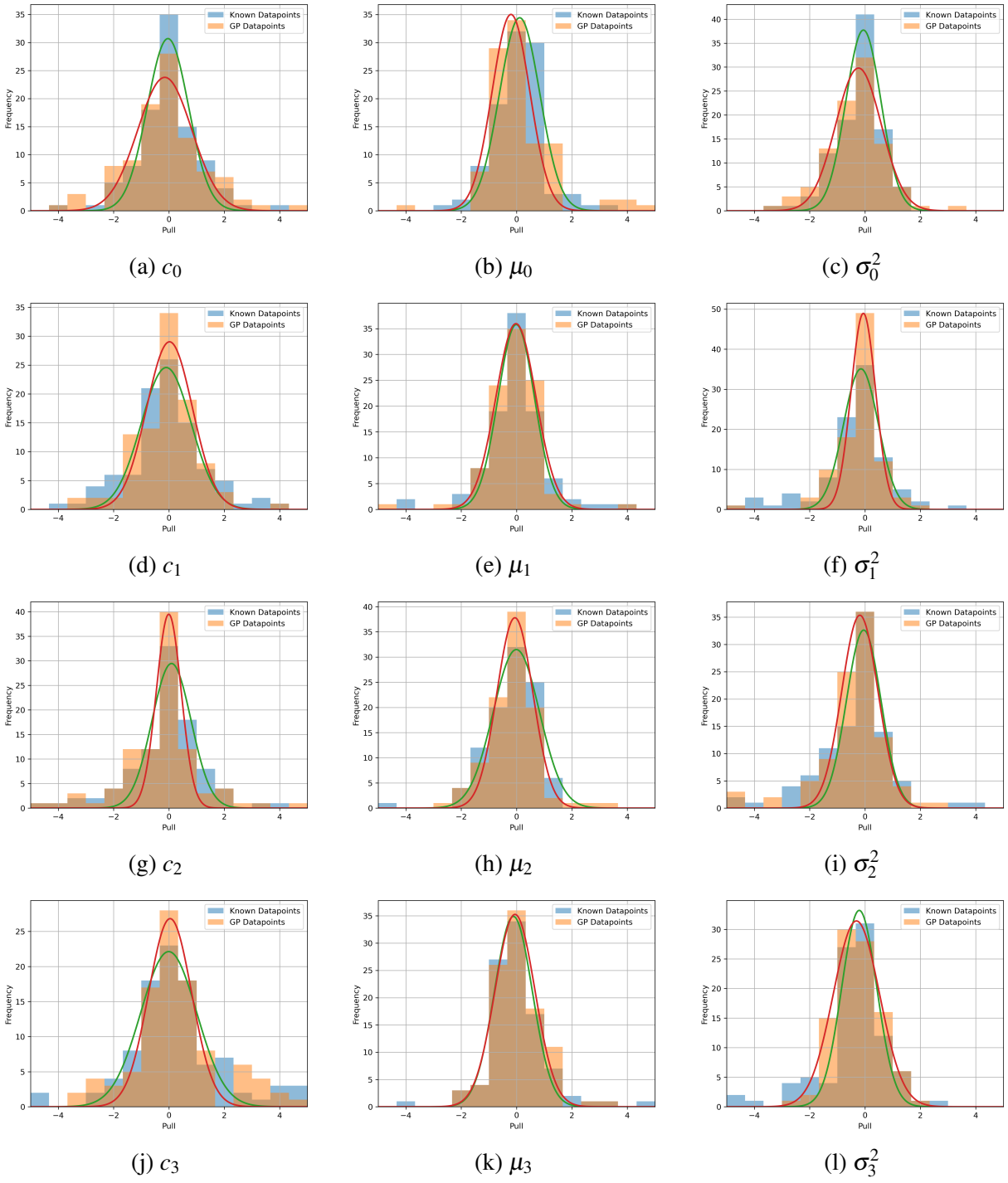
this work, these results demonstrate that PSO provides the more suitable approach and it was therefore selected as the primary optimisation method used throughout this work.

# Appendix B

# Pull Distributions of Fitted Coefficients

This appendix presents the full pull distributions of the fitted coefficients used in the recovery-of-functional-form validation described in Section 5.3. The coefficients were obtained by fitting the functional form given in Equation 5.1 to both the original pseudodata and datapoints sampled from GP surfaces. The pulls were calculated using Equation 5.12 over the ensemble of 100 independently generated pseudodata surfaces. The corresponding fitted Gaussian means and variances are summarised in Tables 5.2 and A.2 for the PSO and MCMC optimisation methods, respectively.

The first figure shows the pull distributions obtained when the GP hyperparameters were optimised using PSO, while the second shows the corresponding distributions when MCMC was used. The latter provides a comparison with the alternative optimisation approach discussed in Section 4.5. The full distributions are included here to provide the underlying results from which the summary statistics presented in the main text and the MCMC comparison in Appendix A were obtained.



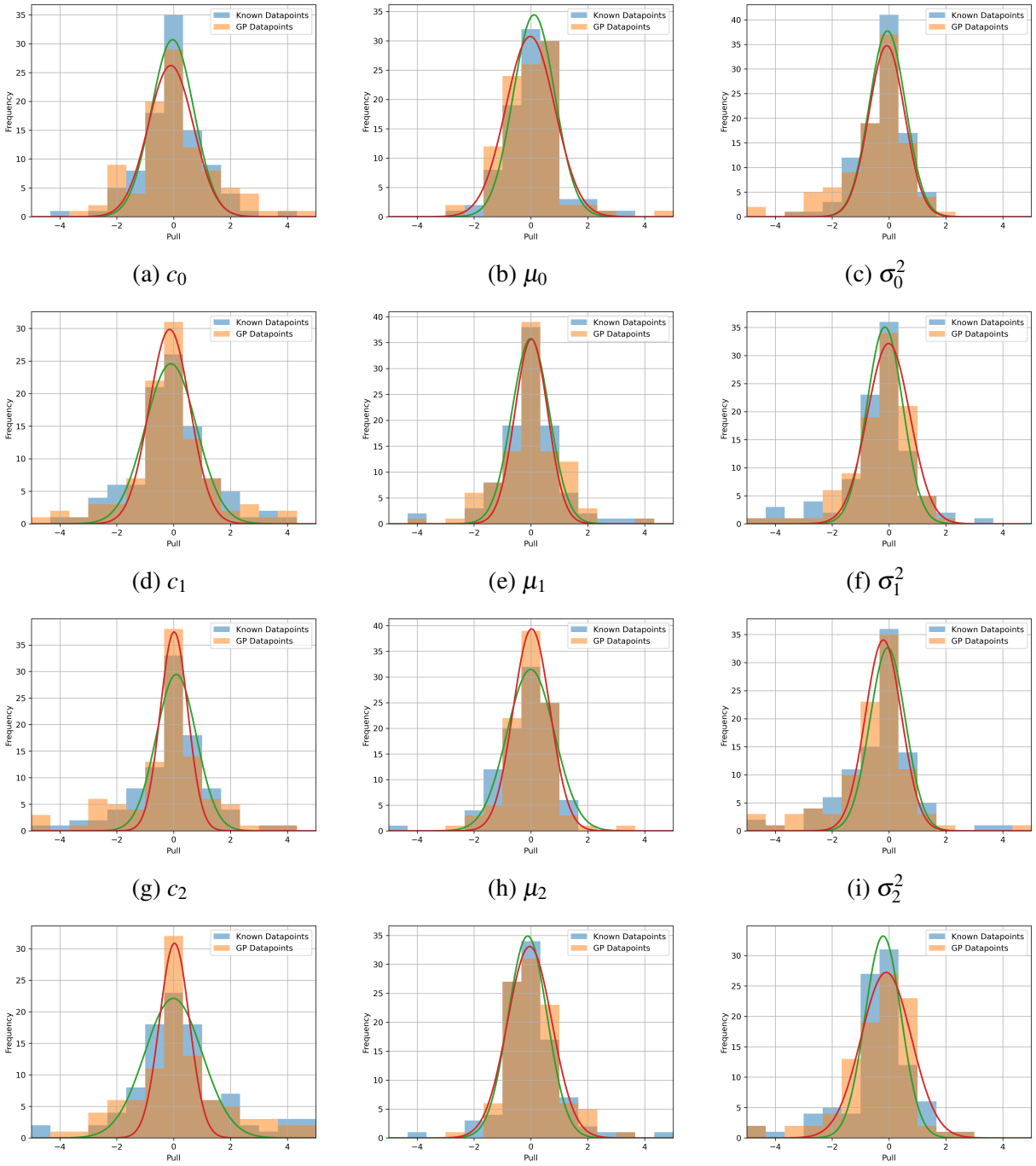

(a) $c_0$ (b) $\mu_0$ (c) $\sigma_0^2$

(d) $c_1$ (e) $\mu_1$ (f) $\sigma_1^2$

(g) $c_2$ (h) $\mu_2$ (i) $\sigma_2^2$

(j) $c_3$ (k) $\mu_3$ (l) $\sigma_3^2$

Figure B.1: Pull distributions of the fitted coefficients obtained using PSO, corresponding to the summary statistics in Table 5.2.




(a) $c_0$ (b) $\mu_0$ (c) $\sigma_0^2$

(d) $c_1$ (e) $\mu_1$ (f) $\sigma_1^2$

(g) $c_2$ (h) $\mu_2$ (i) $\sigma_2^2$

(j) $c_3$ (k) $\mu_3$ (l) $\sigma_3^2$

Figure B.2: Pull distributions of the fitted coefficients obtained using MCMC, corresponding to the summary statistics in Table A.2.

# Appendix C

# List of Recurring Acronyms

**AD** Anderson-Darling

**CDF** Cumulative Distribution Function

**CvM** Cramér-von Mises

**DVCS** Deeply Virtual Compton Scattering

**GP** Gaussian Process

**GPD** Generalised Parton Distribution

**KS** Kolmogorov-Smirnov

**LHS** Latin Hypercube Sampling

**MCMC** Markov Chain Monte Carlo

**ML** Machine Learning

**PDF** Probability Density Function

**pQCD** Perturbative Quantum Chromodynamics

**PSO** Particle Swarm Optimisation

**PWA** Partial Wave Analysis

**QCD** Quantum Chromodynamics

**RBF** Radial Basis Function

# Bibliography


[1] G. Carleo et al. “Machine learning and the physical sciences”. In: *Rev. Mod. Phys.* 91 (4 Dec. 2019), p. 045002. DOI: 10.1103/RevModPhys.91.045002. URL: https://link.aps.org/doi/10.1103/RevModPhys.91.045002.

[2] G. E. Karniadakis et al. “Physics-informed machine learning”. In: *Nature Reviews Physics* 3 (2021), pp. 422–440. DOI: 10.1038/s42254-021-00314-5. URL: https://doi.org/10.1038/s42254-021-00314-5.

[3] D. Guest et al. “Deep Learning and Its Application to LHC Physics”. In: *Annual Review of Nuclear and Particle Science* 68.1 (Oct. 2018), pp. 161–181. ISSN: 1545-4134. DOI: 10.1146/annurev-nucl-101917-021019. URL: http://dx.doi.org/10.1146/annurev-nucl-101917-021019.

[4] A. Radovic et al. “Machine learning at the energy and intensity frontiers of particle physics”. In: *Nature* 560 (2018), pp. 41–48. DOI: 10.1038/s41586-018-0361-2. URL: https://doi.org/10.1038/s41586-018-0361-2.

[5] A. Boehnlein et al. “Colloquium: Machine learning in nuclear physics”. In: *Rev. Mod. Phys.* 94 (3 Sept. 2022), p. 031003. DOI: 10.1103/RevModPhys.94.031003. URL: https://link.aps.org/doi/10.1103/RevModPhys.94.031003.

[6] C. E. Rasmussen and C. K. I. Williams. *Gaussian Processes for Machine Learning*. MIT Press, 2006. ISBN: 0-262-18253-X. URL: https://gaussianprocess.org/gpml/chapters/RW.pdf.

[7] M. Seeger. “Gaussian processes for machine learning”. In: *International Journal of Neural Systems* 14.2 (2004), pp. 69–106. DOI: 10.1142/S0129065704001899. URL: https://pubmed.ncbi.nlm.nih.gov/15112367/.

[8] A. P. Bartók et al. “Gaussian Approximation Potentials: The Accuracy of Quantum Mechanics, without the Electrons”. In: *Phys. Rev. Lett.* 104 (13 Apr. 2010), p. 136403. DOI: 10.1103/PhysRevLett.104.136403. URL: https://link.aps.org/doi/10.1103/PhysRevLett.104.136403.

[9] D. Higdon et al. “A Bayesian approach for parameter estimation and prediction using a computationally intensive model”. In: *Journal of Physics G: Nuclear and Particle Physics* 42.3 (Feb. 2015), p. 034009. DOI: 10.1088/0954-3899/42/3/034009. URL: https://doi.org/10.1088/0954-3899/42/3/034009.

[10] J. D. McDonnell et al. “Uncertainty Quantification for Nuclear Density Functional Theory and Information Content of New Measurements”. In: *Phys. Rev. Lett.* 114 (12 Mar. 2015), p. 122501. DOI: 10.1103/PhysRevLett.114.122501. URL: https://link.aps.org/doi/10.1103/PhysRevLett.114.122501.

[11] B. Acharya and S. Bacca. “Gaussian process error modeling for chiral effective-field-theory calculations of $np \leftrightarrow d\gamma$ at low energies”. In: *Physics Letters B* 827 (2022), p. 137011. ISSN: 0370-2693. DOI: 10.1016/j.physletb.2022.137011. URL: https://www.sciencedirect.com/science/article/pii/S0370269322001459.

[12] C. Drischler et al. “Quantifying uncertainties and correlations in the nuclear-matter equation of state”. In: *Physical Review C* 102.5 (Nov. 2020). ISSN: 2469-9993. DOI: 10.1103/physrevc.102.054315. URL: http://dx.doi.org/10.1103/PhysRevC.102.054315.

[13] J. Keller et al. “Nuclear Equation of State for Arbitrary Proton Fraction and Temperature Based on Chiral Effective Field Theory and a Gaussian Process Emulator”. In: *Phys. Rev. Lett.* 130 (7 Feb. 2023), p. 072701. DOI: 10.1103/PhysRevLett.130.072701. URL: https://link.aps.org/doi/10.1103/PhysRevLett.130.072701.

[14] M. M. Noack et al. “Gaussian processes for autonomous data acquisition at large-scale synchrotron and neutron facilities”. In: *Nature Reviews Physics* 3 (2021), pp. 685–697. DOI: 10.1038/s42254-021-00345-y. URL: https://doi.org/10.1038/s42254-021-00345-y.

[15] H. Iwamoto. “Generation of nuclear data using Gaussian process regression”. In: *Journal of Nuclear Science and Technology* 57.8 (2020), pp. 932–938. DOI: 10.1080/00223131.2020.1736202. eprint: https://doi.org/10.1080/00223131.2020.1736202. URL: https://doi.org/10.1080/00223131.2020.1736202.

[16] Y. Gal et al. *Baryons from Mesons: A Machine Learning Perspective*. 2020. arXiv: 2003.10445 [hep-ph]. URL: https://arxiv.org/abs/2003.10445.

[17] X. Wang et al. “Extrapolating exclusive charmonium photoproduction to the forward regime using Gaussian process”. In: *The European Physical Journal C* 86 (2026), p. 557. DOI: 10.1140/epjc/s10052-026-15799-6.

[18] B. J. Kröse and P. van der Smagt. *An Introduction to Neural Networks*. 8th ed. Amsterdam: University of Amsterdam, 1996. URL: https://www.infor.uva.es/~teodoro/neuro-intro.pdf.

[19] TseKiChun. *Neural Network Diagram*. Science Learning Hub, University of Waikato. CC BY-SA 4.0. 2023. URL: https://www.sciencelearn.org.nz/images/5156-neural-network-diagram.

[20] C. R. Harris et al. "Array programming with NumPy". In: *Nature* 585 (2020), pp. 357–362. DOI: 10.1038/s41586-020-2649-2.

[21] P. Virtanen et al. "SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python". In: *Nature Methods* 17 (2020), pp. 261–272. DOI: 10.1038/s41592-019-0686-2.

[22] R. F. Ferguson. *Gaussian-Process*. https://github.com/rferguson22/Gaussian-Process. GitHub repository. Accessed: June 2nd 2026. 2026.

[23] D. Griffiths. *Introduction to Elementary Particles*. 2nd. Wiley-VCH, 2008.

[24] M. Gell-Mann. *THE EIGHTFOLD WAY: A THEORY OF STRONG INTERACTION SYMMETRY*. Tech. rep. California Inst. of Tech., Pasadena. Synchrotron Lab., Mar. 1961. DOI: 10.2172/4008239. URL: https://www.osti.gov/biblio/4008239.

[25] M. Gell-Mann. "A SCHEMATIC MODEL OF BARYONS AND MESONS". In: *Phys. Letters* Vol: 8 (Feb. 1964). DOI: 10.1016/S0031-9163(64)92001-3. URL: https://www.osti.gov/biblio/4082875.

[26] G. Zweig. "An SU(3) model for strong interaction symmetry and its breaking. Version 1". In: (Jan. 1964). DOI: 10.17181/CERN-TH-401.

[27] Wikimedia Commons contributors. *Category: Hadron multiplets*. https://commons.wikimedia.org/wiki/Category:Hadron_multiplets. Accessed: April 8th 2026.

[28] K. G. Wilson. "Confinement of quarks". In: *Phys. Rev. D* 10 (8 Oct. 1974), pp. 2445–2459. DOI: 10.1103/PhysRevD.10.2445. URL: https://link.aps.org/doi/10.1103/PhysRevD.10.2445.

[29] M. Thomson. *Modern Particle Physics*. Cambridge University Press, 2013.

[30] H. Fritzsch et al. "Advantages of the color octet gluon picture". In: *Physics Letters B* 47.4 (1973), pp. 365–368. ISSN: 0370-2693. DOI: https://doi.org/10.1016/0370-2693(73)90625-4. URL: https://www.sciencedirect.com/science/article/pii/0370269373906254.

[31] D. J. Gross and F. Wilczek. “Ultraviolet Behavior of Non-Abelian Gauge Theories”. In: *Physical Review Letters* 30 (1973), pp. 1343–1346. DOI: 10.1103/PhysRevLett.30.1343.

[32] H. D. Politzer. “Reliable Perturbative Results for Strong Interactions?” In: *Physical Review Letters* 30 (1973), pp. 1346–1349. DOI: 10.1103/PhysRevLett.30.1346.

[33] M. Breidenbach et al. “Observed Behavior of Highly Inelastic Electron-Proton Scattering”. In: *Phys. Rev. Lett.* 23 (16 Oct. 1969), pp. 935–939. DOI: 10.1103/PhysRevLett.23.935. URL: https://link.aps.org/doi/10.1103/PhysRevLett.23.935.

[34] S. Bethke. “Experimental tests of asymptotic freedom”. In: *Progress in Particle and Nuclear Physics* 58.2 (2007), pp. 351–386. ISSN: 0146-6410. DOI: https://doi.org/10.1016/j.ppnp.2006.06.001. URL: https://www.sciencedirect.com/science/article/pii/S0146641006000615.

[35] A. Thiel et al. “Light Baryon Spectroscopy”. In: *Progress in Particle and Nuclear Physics* 125 (July 2022), p. 103949. ISSN: 0146-6410. DOI: 10.1016/j.ppnp.2022.103949. URL: http://dx.doi.org/10.1016/j.ppnp.2022.103949.

[36] R. G. Edwards et al. “Excited state baryon spectroscopy from lattice QCD”. In: *Phys. Rev. D* 84 (7 Oct. 2011), p. 074508. DOI: 10.1103/PhysRevD.84.074508. URL: https://link.aps.org/doi/10.1103/PhysRevD.84.074508.

[37] C. Gattringer and C. B. Lang. *Quantum Chromodynamics on the Lattice*. Vol. 788. Springer, 2010. DOI: 10.1007/978-3-642-01850-3.

[38] S. Navas et al. “Review of Particle Physics”. In: *Phys. Rev. D* 110 (3 Aug. 2024), p. 030001. DOI: 10.1103/PhysRevD.110.030001. URL: https://link.aps.org/doi/10.1103/PhysRevD.110.030001.

[39] T. Hyodo et al. “Origin of resonances in the chiral unitary approach”. In: *Phys. Rev. C* 78 (2008), p. 025203. DOI: 10.1103/PhysRevC.78.025203. eprint: arXiv:0803.2550[nucl-th].

[40] F. Pittler et al. “Universal parameters of the Λ(1380), the Λ(1405), and their isospin partners from a combined analysis of lattice QCD and experimental results”. In: *Phys. Rev. D* 112.7 (2025), p. 074037. DOI: 10.1103/ls4c-6f2y. arXiv: 2507.14283 [hep-ph].

[41] M. Mai et al. “Inclusion of KΛ electroproduction data in a coupled channel analysis”. In: *European Physical Journal A* 59.12 (2023), p. 286. DOI: 10.1140/epja/s10050-023-01188-0.

[42] J. Bulava et al. “Two-Pole Nature of the $\Lambda(1405)$ Resonance from Lattice QCD”. In: *Phys. Rev. Lett.* 132 (5 Jan. 2024), p. 051901. DOI: 10.1103/PhysRevLett.132.051901. URL: https://link.aps.org/doi/10.1103/PhysRevLett.132.051901.

[43] J. Bulava et al. “Lattice QCD study of $\pi\Sigma - \overline{K}N$ scattering and the $\Lambda(1405)$ resonance”. In: *Phys. Rev. D* 109 (1 Jan. 2024), p. 014511. DOI: 10.1103/PhysRevD.109.014511. URL: https://link.aps.org/doi/10.1103/PhysRevD.109.014511.

[44] J. M. M. Hall et al. “Lattice QCD evidence that the $\Lambda(1405)$ resonance is an antikaon–nucleon molecule”. In: *Phys. Rev. Lett.* 114.13 (2015), p. 132002. DOI: 10.1103/PhysRevLett.114.132002. eprint: arXiv:1411.3402[hep-lat].

[45] R. Koniuk and N. Isgur. “Where Have All the Resonances Gone? An Analysis of Baryon Couplings in a Quark Model with Chromodynamics”. In: *Phys. Rev. Lett.* 44 (13 Mar. 1980), pp. 845–848. DOI: 10.1103/PhysRevLett.44.845. URL: https://link.aps.org/doi/10.1103/PhysRevLett.44.845.

[46] Y.-R. Liu et al. “Pentaquark and Tetraquark States”. In: *Progress in Particle and Nuclear Physics* 107 (2019), pp. 237–320. ISSN: 0146-6410. DOI: 10.1016/j.ppnp.2019.04.003.

[47] P. Bicudo. “Tetraquarks and pentaquarks in lattice QCD with light and heavy quarks”. In: *Physics Reports* 1039 (2023), pp. 1–49. ISSN: 0370-1573. DOI: https://doi.org/10.1016/j.physrep.2023.10.001. URL: https://www.sciencedirect.com/science/article/pii/S0370157323003137.

[48] K. J. Peters. “A Primer on partial wave analysis”. In: *Int. J. Mod. Phys. A* 21 (2006). Ed. by F. Melchiorri and Y. Rephaeli, pp. 5618–5624. DOI: 10.1142/S0217751X06034811. arXiv: hep-ph/0412069.

[49] L. Clark. “Hadron spectroscopy of pseudoscalar and vector meson photoproduction using linearly polarised photons at CLAS”. PhD thesis. University of Glasgow, 2022. DOI: 10.5525/gla.thesis.83236. URL: https://theses.gla.ac.uk/83236/.

[50] W.-T. Chiang and F. Tabakin. “Completeness rules for spin observables in pseudoscalar meson photoproduction”. In: *Physical Review C* 55.4 (1997), pp. 2054–2066. DOI: 10.1103/PhysRevC.55.2054.

[51] S. Fegan. “Polarisation Observables from Strangeness Photoproduction on a Frozen Spin Target with CLAS at Jefferson Lab”. PhD Thesis. University of Glasgow, 2011. URL: https://theses.gla.ac.uk/3158/.

[52] L. Lyons. *Statistics for Nuclear and Particle Physicists*. Cambridge University Press, 1986.

[53] C. A. Paterson. "Polarization Observables in Strangeness Photoproduction with CLAS at Jefferson Lab". PhD Thesis. University of Glasgow, 2008. DOI: 10.5525/gla.thesis.74042. URL: https://theses.gla.ac.uk/74042/.

[54] D. G. Ireland et al. "Photoproduction reactions and non-strange baryon spectroscopy". In: *Progress in Particle and Nuclear Physics* 111 (2020), p. 103752. ISSN: 0146-6410. DOI: https://doi.org/10.1016/j.ppnp.2019.103752. URL: https://www.sciencedirect.com/science/article/pii/S0146641019300870.

[55] A. M. Sandorfi et al. "Determining pseudoscalar meson photo-production amplitudes from complete experiments". In: *J. Phys. G: Nucl. Part. Phys.* 38 (2011), p. 053001. DOI: 10.1088/0954-3899/38/5/053001.

[56] B. Krusche and S. Schadmand. "Study of Non-Strange Baryon Resonances with Meson Photoproduction". In: *Progress in Particle and Nuclear Physics* 51 (2003), pp. 399–485. DOI: 10.1016/S0146-6410(03)90005-2.

[57] I. Barker et al. "Complete experiments in pseudoscalar photoproduction". In: *Nuclear Physics B* 95.2 (1975), pp. 347–356. ISSN: 0550-3213. DOI: https://doi.org/10.1016/0550-3213(75)90049-8. URL: https://www.sciencedirect.com/science/article/pii/0550321375900498.

[58] F. Sauli. "From bubble chambers to electronic systems: 25 years of evolution in particle detectors at CERN (1979–2004)". In: *Physics Reports* 403–404 (2004), pp. 471–504. DOI: 10.1016/j.physrep.2004.08.023.

[59] V. Crede and W. Roberts. "Progress towards understanding baryon resonances". In: *Reports on Progress in Physics* 76.7 (June 2013), p. 076301. DOI: 10.1088/0034-4885/76/7/076301. URL: https://doi.org/10.1088/0034-4885/76/7/076301.

[60] C. Grupen and B. Shwartz. *Particle Detectors*. 2nd ed. Cambridge Monographs on Particle Physics, Nuclear Physics and Cosmology. Cambridge University Press, 2008.

[61] R. L. Workman et al. "Unified Chew-Mandelstam SAID analysis of pion photoproduction data". In: *Phys. Rev. C* 86 (2012), p. 015202. DOI: 10.1103/PhysRevC.86.015202.

[62] D. Drechsel et al. "Unitary isobar model –MAID2007". In: *European Physical Journal A* 34 (2007), pp. 69–97. DOI: 10.1140/epja/i2007-10490-6.

[63] R. A. Arndt et al. "Extended partial-wave analysis of NN scattering data". In: *Phys. Rev. C* 74 (2006), p. 045205. DOI: 10.1103/PhysRevC.74.045205.

[64] A. Gelman et al. *Bayesian Data Analysis*. 3rd. Boca Raton, FL: Chapman and Hall/CRC, 2013. ISBN: 978-1-4398-4095-5.

[65] V. Burkert et al. "The CLAS12 Spectrometer at Jefferson Laboratory". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 959 (2020), p. 163419. ISSN: 0168-9002. DOI: https://doi.org/10.1016/j.nima.2020.163419. URL: https://www.sciencedirect.com/science/article/pii/S0168900220300243.

[66] A. V. Anisovich et al. "Photoproduction of pions and properties of baryon resonances from a Bonn–Gatchina partial-wave analysis". In: *The European Physical Journal A* 44.2 (2010), pp. 203–220. URL: https://doi.org/10.1140/epja/i2010-10950-x.

[67] D. Rönchen et al. "Coupled-channel dynamics in the reactions $\pi N \to \pi N, \eta N, K\Lambda, K\Sigma$". In: *European Physical Journal A* 49 (2013), p. 44. DOI: 10.1140/epja/i2013-13044-5.

[68] N. Suzuki et al. "Extraction of resonance poles from a coupled-channel model". In: *Physical Review Letters* 104 (2010), p. 042302. DOI: 10.1103/PhysRevLett.104.042302.

[69] G. Casella and R. L. Berger. *Statistical Inference*. 2nd ed. Pacific Grove, CA: Duxbury, 2002.

[70] G. Cowan. *Statistical Data Analysis*. Oxford, UK: Oxford University Press, 1998. ISBN: 9780198501565.

[71] S. M. Ross. *A First Course in Probability*. 9th ed. Boston: Pearson, 2014.

[72] D. S. Sivia and J. Skilling. *Data Analysis: A Bayesian Tutorial*. 2nd ed. Oxford: Oxford University Press, 2006. ISBN: 9780198568322.

[73] C. M. Bishop. *Pattern Recognition and Machine Learning*. Springer, 2006.

[74] R. M. Neal. *Probabilistic Inference Using Markov Chain Monte Carlo Methods*. Technical Report CRG-TR-93-1. University of Toronto, 1993.

[75] P. R. Bevington and D. K. Robinson. *Data Reduction and Error Analysis for the Physical Sciences*. 3rd ed. McGraw-Hill, 2003. ISBN: 9780072472271.

[76] D. J. C. MacKay. *Information Theory, Inference, and Learning Algorithms*. Cambridge University Press, 2003. ISBN: 9780521642989.

[77] A. Kendall and Y. Gal. *What Uncertainties Do We Need in Bayesian Deep Learning for Computer Vision?* 2017. arXiv: 1703.04977 [cs.CV]. URL: https://arxiv.org/abs/1703.04977.

[78] A. P. Dawid. “Present Position and Potential Developments: Some Personal Views: Statistical Theory: The Prequential Approach”. In: *Journal of the Royal Statistical Society. Series A (General)* 147.2 (1984), pp. 278–292.

[79] National Institute of Standards and Technology. *Normal Distribution*. NIST/SEMATECH e-Handbook of Statistical Methods. 2026. URL: https://www.itl.nist.gov/div898/handbook/eda/section3/eda3661.htm.

[80] S. M. Ross. *Introduction to Probability and Statistics for Engineers and Scientists*. 5th. Boston, MA: Academic Press, 2014.

[81] B. Matern. *Spatial Variation*. Vol. 36. Lecture Notes in Statistics. Berlin, Germany: Springer-Verlag, 1986. ISBN: 978-3-642-61544-3.

[82] L. Clark et al. “Photoproduction of the $\Sigma^+$ hyperon using linearly polarized photons with CLAS”. In: *Phys. Rev. C* 111 (2 2025), p. 025204. DOI: 10.1103/PhysRevC.111.025204. URL: https://link.aps.org/doi/10.1103/PhysRevC.111.025204.

[83] M. D. McKay et al. “A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code”. In: *Technometrics* 21.2 (1979), pp. 239–245. ISSN: 00401706. URL: http://www.jstor.org/stable/1268522.

[84] P. Saves et al. “SMT 2.0: A Surrogate Modeling Toolbox with a focus on hierarchical and mixed variables Gaussian processes”. In: *Advances in Engineering Software* 188 (2024), p. 103571. ISSN: 0965-9978. DOI: https://doi.org/10.1016/j.advengsoft.2023.103571. URL: https://www.sciencedirect.com/science/article/pii/S096599782300162X.

[85] D. Foreman-Mackey et al. “emcee: The MCMC Hammer”. In: *Publications of the Astronomical Society of the Pacific* 125.925 (Mar. 2013), pp. 306–312. DOI: 10.1086/670067. URL: https://doi.org/10.1086/670067.

[86] F. Pedregosa et al. “Scikit-learn: Machine Learning in Python”. In: *Journal of Machine Learning Research* 12 (Jan. 2012).

[87] P. J. Rousseeuw. “Silhouettes: A graphical aid to the interpretation and validation of cluster analysis”. In: *Journal of Computational and Applied Mathematics* 20 (1987), pp. 53–65. ISSN: 0377-0427. DOI: https://doi.org/10.1016/0377-0427(87)90125-7. URL: https://www.sciencedirect.com/science/article/pii/0377042787901257.

[88] J. Kennedy and R. Eberhart. “Particle swarm optimization”. In: *Proceedings of ICNN’95 - International Conference on Neural Networks*. Vol. 4. 1995, 1942–1948 vol.4. DOI: 10.1109/ICNN.1995.488968.

[89] R. Eberhart and Y. Shi. “Particle swarm optimization: developments, applications and resources”. In: *Proceedings of the 2001 Congress on Evolutionary Computation (IEEE Cat. No.01TH8546)*. Vol. 1. 2001, 81–86 vol. 1. DOI: 10.1109/CEC.2001.934374.

[90] M. Arjovsky et al. “Wasserstein Generative Adversarial Networks”. In: *Proceedings of the 34th International Conference on Machine Learning*. Ed. by D. Precup and Y. W. Teh. Vol. 70. Proceedings of Machine Learning Research. PMLR, Aug. 2017, pp. 214–223. URL: https://proceedings.mlr.press/v70/arjovsky17a.html.

[91] O. Thas. *Comparing Distributions*. Springer Series in Statistics. New York: Springer, 2010. DOI: 10.1007/978-3-642-04898-2.

[92] M. de Berg et al. *Computational Geometry: Algorithms and Applications*. 3rd. Springer, 2008. DOI: 10.1007/978-3-540-77974-2.

[93] G. S. Abrams and B. Sechi-Zorn. “Charge-Exchange Scattering of Low-Energy $K^-$ Mesons in Hydrogen”. In: *Phys. Rev.* 139 (2B July 1965), B454–B457. DOI: 10.1103/PhysRev.139.B454. URL: https://link.aps.org/doi/10.1103/PhysRev.139.B454.

[94] W. Alles et al. “On $K^-$-hydrogen interactions”. In: *Il Nuovo Cimento (1955–1965)* 11 (1959), pp. 771–775. DOI: 10.1007/BF02732540. URL: https://doi.org/10.1007/BF02732540.

[95] R. O. Bangerter et al. “Reactions $K^- p \to \Sigma^- \pi^+$ and $K^- p \to \Sigma^+ \pi^-$ in the momentum range from 220 to 470 MeV/c”. In: *Phys. Rev. D* 23 (7 Apr. 1981), pp. 1484–1499. DOI: 10.1103/PhysRevD.23.1484. URL: https://link.aps.org/doi/10.1103/PhysRevD.23.1484.

[96] J. Ciborowski et al. “Kaon scattering and charged Sigma hyperon production in $K^-$p interactions below 300 MeV/c”. In: *Journal of Physics G: Nuclear Physics* 8.1 (Jan. 1982), p. 13. DOI: 10.1088/0305-4616/8/1/005. URL: https://dx.doi.org/10.1088/0305-4616/8/1/005.

[97] M. Csejthey-Barth et al. “The interactions of low-energy $K^-$ mesons with protons”. In: *Physics Letters* 16.1 (1965), pp. 89–91. DOI: 10.1016/0031-9163(65)90415-4.

[98] D. H. Davis et al. “Elastic Scattering of Low-Energy $K^-$ Mesons on Protons”. In: *Phys. Rev. Lett.* 6 (3 Feb. 1961), pp. 132–134. DOI: 10.1103/PhysRevLett.6.132. URL: https://link.aps.org/doi/10.1103/PhysRevLett.6.132.

[99] D. Evans et al. “Charge-exchange scattering in $K^-$p interactions below 300 MeV/c”. In: *Journal of Physics G: Nuclear Physics* 9.8 (Aug. 1983), p. 885. DOI: 10.1088/0305-4616/9/8/011.

[100] R. G. Glasser et al. "$K^-$ Interactions with Free Protons in Nuclear Emulsion". In: *Il Nuovo Cimento* 9 (1958), pp. 1085–1090. DOI: 10.1007/BF02777960. URL: https://doi.org/10.1007/BF02777960.

[101] W. E. Humphrey and R. R. Ross. "Low-Energy Interactions of $K^-$ Mesons in Hydrogen". In: *Phys. Rev.* 127 (4 Aug. 1962), pp. 1305–1323. DOI: 10.1103/PhysRev.127.1305. URL: https://link.aps.org/doi/10.1103/PhysRev.127.1305.

[102] J. K. Kim. *Multichannel Phase-Shift Analysis of $\bar{K}N$ Interaction*. Report No. NEVIS-149. Unpublished. Columbia University, Nevis Laboratories, 1966.

[103] J. K. Kim. "Multichannel Phase-Shift Analysis of $\overline{K}N$ Interaction in the Region 0 to 550 MeV/c". In: *Phys. Rev. Lett.* 19 (18 Oct. 1967), pp. 1074–1078. DOI: 10.1103/PhysRevLett.19.1074. URL: https://link.aps.org/doi/10.1103/PhysRevLett.19.1074.

[104] W. Kittel et al. "The K- proton charge exchange interactions at low energies and scattering lengths determination". In: *Physics Letters* 21.3 (1966), pp. 349–351. ISSN: 0031-9163. DOI: https://doi.org/10.1016/0031-9163(66)90845-6. URL: https://www.sciencedirect.com/science/article/pii/0031916366908456.

[105] T. S. Mast et al. "Reactions $K^-p \to \Sigma^0\pi^0$ and $K^-p \to \Lambda\pi^0$ in the momentum range 240 to 450 MeV/c". In: *Phys. Rev. D* 11 (11 June 1975), pp. 3078–3091. DOI: 10.1103/PhysRevD.11.3078. URL: https://link.aps.org/doi/10.1103/PhysRevD.11.3078.

[106] T. S. Mast et al. "Elastic, charge-exchange, and total $K^-p$ cross sections in the momentum range 220 to 470 MeV/c". In: *Phys. Rev. D* 14 (1 July 1976), pp. 13–27. DOI: 10.1103/PhysRevD.14.13. URL: https://link.aps.org/doi/10.1103/PhysRevD.14.13.

[107] M. Sakitt et al. "Low-Energy $K^-$ Meson Interactions in Hydrogen". In: *Physical Review* 139.B3 (1965), pp. 719–728. DOI: 10.1103/PhysRev.139.B719.

[108] G. Thomas and P. Jain. "Low-energy $K^-$ meson elastic scattering from hydrogen". In: *Nuclear Physics B* 13.2 (1969), pp. 385–400. ISSN: 0550-3213. DOI: https://doi.org/10.1016/0550-3213(69)90243-0. URL: https://www.sciencedirect.com/science/article/pii/0550321369902430.

[109] A. D. Martin. "Kaon–nucleon parameters". In: *Nuclear Physics B* 179 (1981), pp. 33–48. DOI: 10.1016/0550-3213(81)90247-9.

[110] D. Nowak et al. “Charged Σ hyperon production by $K^-$ meson interactions at rest”. In: *Nuclear Physics B* 139 (1978), pp. 61–71. DOI: 10.1016/0550-3213(78)90179-7.

[111] J. S. Vesterstrøm and J. Riget. “A Diversity-Guided Particle Swarm Optimizer – the ARPSO”. In: *EVALife Technical Report* 2002-02 (2002).

[112] H. Edelsbrunner. “Triangulations and Meshes in Computational Geometry”. In: *Acta Numerica* 9 (2000), pp. 133–213. DOI: 10.1017/S0962492900001331.

[113] H. Dembinski and P. O. et al. “scikit-hep/iminuit”. In: (Dec. 2020). DOI: 10.5281/zenodo.3949207. URL: https://doi.org/10.5281/zenodo.3949207.

[114] G. J. McLachlan and D. Peel. *Finite Mixture Models*. Wiley Series in Probability and Statistics. John Wiley & Sons, 2000. ISBN: 9780471006268. DOI: 10.1002/0471721182.

[115] A. Belitsky et al. “Theory of deeply virtual Compton scattering on the nucleon”. In: *Nuclear Physics B* 629.1 (2002), pp. 323–392. ISSN: 0550-3213. DOI: https://doi.org/10.1016/S0550-3213(02)00144-X. URL: https://www.sciencedirect.com/science/article/pii/S055032130200144X.

[116] A. Gelman and D. B. Rubin. “Inference from Iterative Simulation Using Multiple Sequences”. In: *Statistical Science* 7.4 (1992), pp. 457–472. DOI: 10.1214/ss/1177011136. URL: https://doi.org/10.1214/ss/1177011136.

[117] A. Vehtari et al. “Rank-Normalization, Folding, and Localization: An Improved $\widehat{R}$ for Assessing Convergence of MCMC (with Discussion)”. In: *Bayesian Analysis* 16.2 (2021), pp. 667–718. DOI: 10.1214/20-BA1221. URL: https://doi.org/10.1214/20-BA1221.

[118] Python Software Foundation. *time — Time access and conversions*. https://docs.python.org/3/library/time.html. Accessed: May 13th 2026. 2026.