Methodological developments in data fusion for lake water reflectance from satellite sensors

Zhaoyuan Zou

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

School of Mathematics and Statistics
College of Science and Engineering
University of Glasgow

January 2025

Abstract

Fusing satellite-sensed reflectance data from different sources is of interest to monitor lake water quality, and the satellite sensors have possibly different spatial, temporal and spectral supports. The nonparametric statistical downscaling (NSD) model is an existing state-of-the-art fusion model which can account for a change of spatial and temporal support between two remote sensors [Wilkie et al., 2019]. However, the NSD model is computationally demanding for large datasets and does not allow multivariate responses with an additional spectral dimension. Thus, the aim of this thesis is to improve the computational efficiency of the NSD model and then extend this model to provide an approach that is suitable for a multivariate response to enable the fuse of reflectance data with different spectral and temporal supports from two sensors. The NSD model assumes that the discrete data at each location within a lake from each data source are observations of smooth functions over time and that the coefficients of these smooth functions are modelled as spatially correlated via a covariance matrix. In this thesis, a novel approach proposes using a Gaussian predictive process to approximate the spatial varying coefficients in the NSD model, which requires the inversion of a matrix with smaller dimensions in the Gibbs sampling process and hence reduces the computational time for the parameter estimation. The predictive performance and computational efficiency of the proposed nonparametric statistical downscaling model with Gaussian predictive process (NSD-GPP) are compared to the NSD model through simulation and using satellite reflectance data from Lake Garda. It was found that the NSD-GPP model achieves a similar predictive performance as the NSD model using less computational time. To enable data fusion from the two sensors with a multivariate wavelength dimension, a novel method using the two-dimensional B-spline basis functions was developed where the basis functions were used to represent the reflectance over both time and wavelength at each location, and a different precision parameter was used for each wavelength. Lake Garda is used as an example of interest here, and methods are general for any lake of interest in principle. Overall, it is found that the proposed multivariate NSD-GPP model could be used to make predictions for the unobserved wavelengths and time points within the observed range. It may be beneficial to provide reflectance data at higher temporal and wavelength frequencies, and this model could in principle be extended to consider similar challenges in space.

Contents

List of Tables

List of Figures

Acknowledgements

Firstly, I would like to thank my supervisors, Dr Ruth O'Donnell, Prof Claire Miller, and Prof Duncan Lee, for their invaluable guidance and support throughout my PhD study. Their knowledge and experience guided me through the challenges and inspired me in my research. I also gratefully acknowledge the PhD scholarship received from the College of Science and Engineering. This PhD would not be possible without this scholarship.

Thank you to Xiaohan Liu and Stefan Smith from Plymouth Marine Laboratory for providing and advising on the reflectance data used in this thesis.

Thank you to everyone in the School of Mathematics and Statistics. To Dr Tereza Neocleous and Prof Janine Illian for their suggestions in the annual reviews. To Dr Craig Wilkie, for all his help in applying the nonparametric statistical downscaling model. To all colleagues, thank for their accompany and chats in the Math building.

Thank my family for their support and understanding through my PhD study. To my friends, thank you for supporting me through these years, for travels and parties.

Finally, thank you to Shuang for her love, trust, and encouragement. Her accompany helped me through the hard time of this journey. Life takes two, I will be with you.

Declaration

I certify that the thesis presented here for examination for a PhD degree of the University of Glasgow is solely my own work other than where I have clearly indicated that it is the work of others (in which case the extent of any work carried out jointly by me and any other person is clearly identified in it) and that the thesis has not been edited by a third party beyond what is permitted by the University's PGR Code of Practice.

The copyright of this thesis rests with the author. No quotation from it is permitted without full acknowledgement. I declare that the thesis does not include work forming part of a thesis presented successfully for another degree [unless explicitly identified and as noted below]

I declare that this thesis has been produced in accordance with the University of Glasgow's Code of Good Practice in Research.

I acknowledge that if any issues are raised regarding good research practice based on review of the thesis, the examination may be postponed pending the outcome of any investigation of the issues.

Part of the work in Chapter 5 has been presented as a poster in 38th International Workshop on Statistical Modelling (IWSM), Durham, UK, July 2024. The work presented in Chapter 5 has been published in the the Springer series "Contributions to Statistics" with the title "A computational efficient spatial-temporal fusion model for reflectance data".

Signature: Zhaoyuan Zou

Date: 27/09/2024

Chapter 1

Introduction

1.1 Background

Water is invaluable for people, businesses, and nature. To secure the sustainable availability of water, the European Union Water Framework Directive was issued in 2000 to protect all water bodies and prevent deterioration in water quality [Directive, 2000]. This Directive also pointed out the importance of monitoring and assessing the status of water bodies to fulfil the legislative requirement and to identify and establish action plans for the areas in need. Therefore, selected water bodies, including rivers, lakes, and coastal waters, are monitored by different agencies or research projects. Lakes are vital in ensuring the global biosphere is in good condition and conserving biological diversity because lakes support nutrient cycling and provide habitats for organisms [Fang et al., 2006]. Biodiversity is crucial to human well-being globally, because biodiversity contributes to nutrient cycling and food supply, and increases the stability of ecosystems to climate change [Reaka-Kudla et al., 1996]. Thus, it has developed into a political issue since the Convention on Biological Diversity (CBD) became effective in 1993 [Pauchard, 2017].

In order to achieve the long-term goal of preserving biodiversity and effectively managing the inland water ecosystem, research projects such as Diversity 2 [Wramner et al., 2015] and Globalakes [Tyler et al., 2016] collected water quality data such as chlorophyll-a, suspended matter concentration, and coloured dissolved organic matter (CDOM). Chlorophyll-a is produced by plants in absorbing energy from sunlight, and its concentration will increase in high-level nutrient situations [Spyrakos et al., 2018]. Thus, Chlorophyll-a is a commonly used indicator for eutrophication, which means the concentration of minerals and nutrients is abundant in the water body. Eutrophication will result in phytoplankton abundance and potentially toxic cyanobacteria blooms [O'Neil et al., 2012]. The resulting oxygen depletion and toxic water will harm the plant and animal life in the water, so it is important to monitor Chlorophyll-a levels. Suspended matter in water means the dissolved suspended particles which can be trapped by a filter. Suspended matter concentration measures the suspended matter in a unit volume from the sampled water

body, which directly affects the light penetration into the water body and thereafter the growth of hydrophyte [Zhang et al., 2010]. Thus, the water body with higher levels of suspended matter concentration will cause less dissolved oxygen to be released in water, indicating poorer water quality. Coloured dissolved organic matter (CDOM) is the coloured and optically measurable fraction of dissolved organic matter in the water, which consists of humic substances originated from sediments and plants, together with proteins and acids produced by plankton. CDOM has an effect on the water-leaving radiance level because of its absorption of the ultraviolet light, and this absorption property will protect phytoplankton from biological damage [Aulló-Maestro et al., 2017].

These water quality parameters can be estimated by in-situ or remote-sensing techniques. For in-situ techniques, researchers usually use tubes to collect water samples directly from the water surface and then analyse them in laboratories to get the water quality parameters. Samples can be analyzed by spectrophotometry, which measures the amount of the substance based on their differences in absorption and transmission of light [Morris, 2015]. Recently, automatic lake monitoring buoys have been introduced to monitor water quality [Tiberti et al., 2021]. These floating devices facilitate data collection up to very high time frequencies (e.g., one sample per minute), including air temperature, solar radiation and even underwater characteristics such as chlorophyll-a and potential of hydrogen (pH) [McBride et al., 2019]. In-situ techniques provide accurate measurements for water quality data, but these measurements are only available at limited predefined locations within the water bodies. The cost of in-situ measurements also makes continuous monitoring over long temporal periods prohibitive [Kiefer et al., 2015]. These limitations of in-situ sampling decrease the representativeness of in-situ monitoring due to the small number of data points in space and time [Kiefer et al., 2015]. Remote sensing is a method of obtaining the Earth's surface information by capturing and analysing the reflected electromagnetic energy from the interested targets [Joseph, 2005]. Remote-sensing satellites provide high temporal and spatial resolution data for the physical, chemical and biological processes [Dash and Ogutu, 2016]. The raw measurements from the remote sensors are reflectance data, which can be used to estimate water quality parameters by retrieval algorithms. Reflectance can be scatted by the atmosphere and bright objects such as clouds and ice, which will introduce uncertainty in measuring reflectance data [Liu et al., 2021]. Thus, the water quality data estimated from remote sensing have an advantage in finer spatial and temporal resolution, but a disadvantage in accuracy compared to in-situ data [Tyler et al., 2016].

There are numerous algorithms developed to calculate the water quality parameters from remotely-sensed reflectance data, such as the band ratio model which is based on the ratio of reflectance at two selected bands, and environmental scientists need to select the appropriate retrieval algorithms [Yang et al., 2022]. Neil et al. (2019) suggested selecting the chlorophyll-a retrieval

algorithms using the optical water types of the lakes, and they found that this method achieved a better estimation than using a single retrieval algorithm for the lakes with different optical water types [Neil et al., 2019]. Optical water types are identified by clustering the reflectance data over wavelength from a large number of water bodies into spectrally distinct groups [Spyrakos et al., 2018]. The optical water type is identified for the reflectance data at each pixel, and the retrieval algorithms are selected for the corresponding optical water types [Neil et al., 2019]. Since the optical water type can vary largely over space and time [Soomets et al., 2020], calculating water reflectance data over wavelength at the required time and location is essential to estimate the water quality parameters.

In this thesis reflectance data from three different types of remote sensors have been explored, which are the Medium Resolution Imaging Spectrometer (MERIS), Moderate Resolution Imaging Spectroradiometer (MODIS), and Ocean and Land Colour Instrument (OLCI). These sensors collect reflectance data at a set of spectral bands, which are regions of wavelengths with given centres and bandwidths. An example of spectral bands for these three sensors is shown in Figure 1.1 and the range of the bands are shown in Table 1.1. The bands within 400-1000 nm are selected because these bands are commonly used for water quality retrieval algorithms [Yang et al., 2022]. MERIS was launched on the satellite for the Earth observation mission, Envisat [Rast et al., 1999]. However, this mission ended in 2012 because the satellite lost its communication. The Sentinel-3 OLCI is the successor of MERIS, and it has additional spectral bands [Kravitz et al., 2020]. MODIS is also a satellite-based sensor launched by National Aeronautics and Space Administration (NASA) in 1999 [Salomonson et al., 2006]. The details of these three sensors are summarized in Table 1.2.

Table 1.2 and Figure 1.1 show that these sensors have different spatial, temporal, and spectral resolutions, and there is a trade-off between the resolutions of these dimensions. For example, MODIS has more frequent temporal observations and a longer total time period than MERIS and OLCI. However, the MERIS and OLCI reflectance data are assumed to be more suitable for lake water than MODIS data, because the MERIS and OLCI sensors have a finer spatial resolution and key spectral bands such as 708 nm to estimate water quality parameters [Palmer et al., 2015]. However, the MODIS sensor does not have the spectral band near 708 nm as shown in Table 1.1. These three sensors have different wavelength bands that cover different wavelength ranges. Additionally, none of these sensors could provide a complete and accurate reflectance dataset for any locations and time points from 1999 to the present.

Data fusion is of interest in remote sensing applications, which combines information from different data sources to provide a refined estimation of physical phenomena [Hall and Llinas,

MERIS		MODIS		OLCI	
band (nm)	range (nm)	band (nm)	range (nm)	band (nm)	range (nm)
				400	385-415
412	402-422	412	405-420	412	402-422
443	433-453	443	438-448	443	433-453
		469	459-479		
490	480-500	488	483-493	490	480-500
510	500-520			510	500-520
		531	526-536		
		551	546-556		
		555	545-565		
560	550-570			560	550-570
620	610-630			620	610-630
		645	620-670		
665	655-675	667	662-672	665	655-675
				674	666-681
681	674-689	678	673-683	681	674-689
708	698-718			708	698-718
		748	743-753		
754	746-761			754	746-761
761	759-764			761	759-764
				764	761-769
				767	765-770
779	764-794			779	764-794
865	845-885			865	845-885
		869	862-877		
885	875-895			885	875-895
900	890-910			900	890-910
		905	890-920		
		936	931-941		
		940	915-965	940	920-960
				1020	980-1060

Table 1.1: The bands and range for the MERIS, MODIS, and OLCI sensors. The bands with similar range are allocated to the same row for these three sensors.

Sensor	Spatial Resolution	Temporal resolution	Number of bands	Year Coverage
MERIS	300 m	3 days	15 (390-1040 nm)	2002-2012
MODIS	1 km	1-2 days	36 (405-14385 nm)	1999-Present
OLCI	300m	4 days	21 (400-1020nm)	2016-Present

Table 1.2: Details of MERIS, MODIS and OLCI.

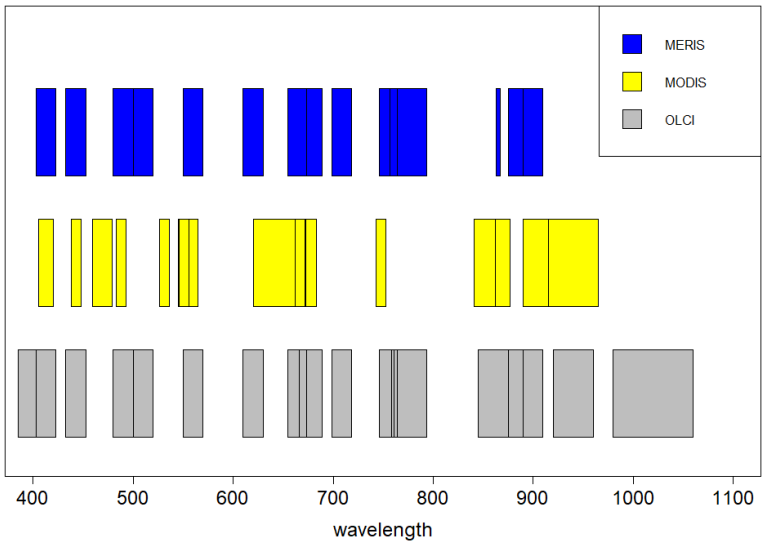


Figure 1.1: Spectral bands of the MERIS, MODIS and OLCI sensors within 400-1000 nm, each rectangle represents a spectral band.

2001]. Fusing the reflectance data from MODIS and MERIS, or MODIS and OLCI for their overlapping years will maximise the benefit of their advantages. This fused dataset can provide reflectance data over wavelength at the required time point and location, which can be used to determine the corresponding optical water type, and then assist environmental scientists in determining the retrieval algorithms and then estimating the water quality parameters. There are several fusion models which can be used to fuse the data from different sources, and they can be grouped into spatial-temporal fusion model and spatial-temporal-spectral fusion model. The spatial-temporal fusion models are designed for spatial-temporal data with univariate response at each location and time point. For example, Berrocal et al. (2010b) use a downscaling model to combine point-level and areal level ozone concentration data, which regresses the observed ozone data as a function of the numerical output, and the coefficients in this function could vary across time [Berrocal et al., 2010b]. Cressie and Johannesson (2008) propose a fixed rank kriging (FRK) approach, which uses a non-stationary covariance function defined by a small set of basis functions [Cressie and Johannesson, 2008]. Nguyen et al. (2014) build a spatial-temporal data fusion model upon this fixed rank kriging method, where the underlying true spatiotemporal processes are modelled at the areal level [Nguyen et al., 2014]. These models require the data to be available on the same temporal support, but the reflectance data used in this research have different temporal supports. Wilkie et al. (2019) present a nonparametric statistical downscaling (NSD) model, which can fuse the log(cholorophyll-a) data from the in-situ technique and satellite sensor [Wilkie et al., 2019]. This model uses smooth functions to represent data over time at each location, which reduces the dimensionality of the data by representing these functions as combinations of known basis functions. The spatially varying basis coefficients for

each data source are linked through a regression model, which allows data fusion from different sources. This model enables data fusion with different spatio-temporal supports and allows the prediction of in-situ data at an unobserved location or time.

The spatial-temporal-spectral fusion models are designed for data with a multivariate response at each time and location. Berrocal et al. propose a bivariate spatial-temporal fusion model to fuse ozone and particulate matter ($PM_{2.5}$) variables from monitoring sites and an air quality model called Community Mesoscale Air Quality (CMAQ) model [Berrocal et al., 2010a]. They assume the measured ozone data could be represented by a combination of the modelled ozone and $PM_{2.5}$ data, and the measured $PM_{2.5}$ data could be represented by a similar combination but with different sets of coefficients. Then, the correlations of these spatial-temporal varying coefficients are modelled by a coregionalization matrix. They also extend this model for data with a multivariate response for different pollutants. Another example of data fusion which has been considered for multi-spectral data is discussed in the work of Gevaert and García-Haro (2015). They propose an unmixing model to decompose the spatial-spectral data $\mathbf{X} \in \mathbb{R}^{L\times N}$ at each time point into an endmember matrix $\mathbf{E} \in \mathbb{R}^{L\times P}$ and abundance matrix $\mathbf{A} \in \mathbb{R}^{P\times N}$, where N is the number of locations, L is the number of wavelength bands, and P is the number of endmembers. Then, the relationship between the two data sources is built upon these endmember and abundance matrices. The temporal dependency is also modelled on these two matrices to separate the temporal variation in the spatial and spectral dimensions.

These fusion models have limitations of no spectral dimension or short-term temporal dependency, which are not suitable for the fusion of reflectance data in this research. The spatial-temporal models discussed above are designed for the data without the spectral dimension, while the multivariate spatial-temporal fusion models require the same set of variables for the two data sources. If the reflectance data at different wavelengths are treated as different variables, these models require the same set of wavelengths for the two sensors. However, the sensors used in this research have different spectral supports, which are shown in Table 1.2. The unmixing-based models only consider the temporal dependency for two consecutive time points, which makes it challenging to capture the long-term and yearly temporal patterns. However, this research aims to produce the fused reflectance data for a long-term temporal range. Therefore, a novel fusion model will be developed in this thesis to overcome these limitations.

1.2 Research aims

This research project is motivated by the fusion problem of lake water reflectance data from different remote sensors. Since there is no overlapping time for all these three sensors, the fusion of reflectance data in this thesis is considered for MODIS-MERIS and MODIS-OLCI

for their overlapping years. The reflectance data used in this research project are provided by Plymouth Marine Laboratory (PML). The reflectance data are from the European Space Agency Lakes Climate Change Initiative project, provided via the Centre for Environment Data Archival (CEDA) [Carrea et al., 2022]. This dataset contains lake water leaving reflectance and other climate variables collected by satellite sensors from 1992 to 2020 over 2000 inland lakes [Carrea et al., 2023]. In this dataset, the spatial resolutions are aggregated into 1km×1km for the MERIS and OLCI sensors, which provide the same spatial resolution as the MODIS sensor. Thus, the reflectance data used in this thesis have the same spatial support but different temporal and spectral supports for these three sensors. Considering the overlapping time between the sensors and the dataset provided by PML, the aim of this thesis is to develop a fusion model to fuse the reflectance data with different temporal and spectral supports from two sensors for a single lake.

In this thesis, a set of spatial-temporal-spectral fusion models are developed in Chapters 4, 5 and 6 to fuse the reflectance data with different temporal and spectral supports from two sensors. The idea is to start from a spatial-temporal fusion model and develop it to allow data fusion with an additional spectral dimension. The NSD model of Wilkie et al. (2019) is selected as a starting point for the developments and will be extended to solve the fusion problem in this research. Since the NSD model has not been applied to fuse reflectance data to the best of our knowledge, the aim of Chapter 4 is to assess how accurately the NSD model can predict the reflectance data for a single wavelength band at an unobserved location or time by fusing the data from two sensors. The predictive performance of the NSD model is compared to that of the FRK model of Cressie and Johannesson (2008), and the comparison shows that the NSD model has a slightly better prediction ability but requires a longer computational time than the FRK model. Thus, improving the computational efficiency of the NSD model will make it easier to apply to larger datasets.

Chapter 5 will introduce an approximation method for the NSD model, which can achieve similar predictive performance to the NSD model but requires less computational time. The parameters in the NSD model are estimated using Bayesian inference with MCMC simulation, which involves the inversion of a spatial covariance matrix at each stage of a MCMC algorithm with tens of thousands of iterations. This chapter improves the computational efficiency of the NSD model by endowing it with a Gaussian predictive process, which is called nonparametric statistical downscaling model with Gaussian predictive process (NSD-GPP) in this thesis. The predictive performance and the computational time of these two models are compared in a real-world reflectance dataset and simulated datasets.

Reflectance data are multivariate vectors over wavelengths for each location and time point, and the spectral and temporal supports are also typically different. Thus, a novel multivariate

NSD-GPP model is proposed in Chapter 6 to enable the multivariate response in this model and account for the different spectral and temporal supports for reflectance data. This multivariate NSD model uses two-dimensional B-splines basis functions to represent the reflectance data over time and wavelength at each location, and this representation will allow the different supports from two sensors. This multivariate NSD model extends the NSD model for multivariate responses and change of supports, which fulfils the fusion task in this research. This multivariate NSD-GPP model can be used to estimate the reflectance data at the required wavelength bands for any location and time point within the domain, which will provide a complete reflectance dataset to select the retrieval algorithms and then to estimate the water quality parameters.

The remainder of this thesis is structured as follows. Chapter 2 includes a review of the standard statistical theory that will be used and developed upon throughout the thesis. Chapter 3 provides the exploratory analysis to understand the properties and the spatial, temporal and spectral variations of the remote-sensing reflectance data from these three sensors. This exploratory analysis aims to identify the relationship between the reflectance data from these sensors and to decide which fusion techniques are suitable for this research. Based on the exploratory analysis, the NSD model is selected to fuse the reflectance data for one wavelength band in Chapter 4, and the predictive performance of the NSD model is assessed. Chapter 5 will introduce an approximation method for the NSD model, which can achieve similar predictive performance to the NSD model but requires less computational time using the Gaussian predictive process. Chapter 6 discusses a novel multivariate NSD-GPP model to fuse the reflectance data from the two sensors with different spectral and temporal supports. Chapter 7 summarises the work in this thesis and discusses the potential future research.

Chapter 2

Review of statistical methods

Chapter 2 presents a review of the statistical methods that will be used and developed upon in this thesis. The inference methods are first reviewed, and these methods are used for parameter estimation throughout this thesis. Section 2.1 introduces the likelihood-based inference, which includes maximum likelihood estimation and confidence interval. Section 2.2 introduces Bayesian inference, where the choice of prior distributions and the simulation methods for posterior distributions are discussed. Section 2.3 presents regression splines, where the different types of basis functions, estimation of the coefficients and the choice of smoothing parameters will be outlined. In this thesis, reflectance data will be represented as a combination of splines over time and wavelength to account for the different supports of the remote sensors. Therefore Section 2.5 discusses functional data analysis approaches, such as the distance between functional data and functional principal component analysis (FPCA), and these approaches will be used to analyse the reflectance data in Chapter 3. Section 2.6 presents the background of spatial geostatistics, because Gaussian geostatistical processes will be used to represent the spatially varying coefficients for the models developed in Chapters 5 and 6.

2.1 Likelihood-based inference

Suppose the variable Y with observations $\mathbf{y} = (y_1, y_2, \cdots, y_n)$ comes from a distribution with a known form and an unknown set of parameters $\boldsymbol{\theta} = (\theta_1, \theta_2, \cdots, \theta_p)$. The likelihood-based inference can be used to estimate these unknown parameters $\boldsymbol{\theta}$ [Mavrakakis and Penzer, 2021]. The likelihood function is defined as:

$$L(\boldsymbol{\theta}|\mathbf{y}) = f(\mathbf{y}|\boldsymbol{\theta}) = \prod_{i=1}^{n} f(y_i|\boldsymbol{\theta}),$$

where the latter equality holds if the data are independent. The likelihood function is the same as the joint density probability function, but it is used to measure how likely the parameter values of $\boldsymbol{\theta}$ are given the observed data $\mathbf{y}$. Maximum likelihood estimation (MLE) finds the point estimator

$\hat{\theta}$ that maximises the likelihood function or log-likelihood function $\ell(\boldsymbol{\theta}|\mathbf{y}) = \log[L(\boldsymbol{\theta}|\mathbf{y})]$. The MLE $\hat{\theta}$ can be calculated by solving the following equations:

$$\frac{\partial \ell(\boldsymbol{\theta}|\mathbf{y})}{\partial \theta_1} = 0, \quad \frac{\partial \ell(\boldsymbol{\theta}|\mathbf{y})}{\partial \theta_2} = 0, \quad \cdots, \quad \frac{\partial \ell(\boldsymbol{\theta}|\mathbf{y})}{\partial \theta_p} = 0.$$

If these equations cannot be solved analytically, then a numerical method such as Newton's method is applied [Mavrakakis and Penzer, 2021]. A confidence interval for $\boldsymbol{\theta}$ is found by calculating the Fisher's information [Mavrakakis and Penzer, 2021]. Take the parameter θ_1 as an example, Fisher's information is defined as:

$$I(\theta_1) = -\mathbb{E}\left[\frac{\partial^2 \ell(\boldsymbol{\theta}|\mathbf{y})}{\partial^2 \theta_1}\right].$$

If the sample size n is large, the MLE $\hat{\theta}_1$ approximately follows the distribution:

$$\hat{\theta}_1 \sim N\left(\theta_1, \frac{1}{I(\theta_1)}\right).$$

This property can be used to find the 95% confidence interval for θ_1, which is $\hat{\theta}_1 \pm z_{0.975}\frac{1}{\sqrt{I(\hat{\theta}_1)}}$, and $z_{0.975} = 1.96$ is the z score corresponding to the 95% confidence interval.

2.2 Bayesian inference

Bayesian inference is an alternative estimation approach to likelihood-based inference, which is used to fit the models in Chapters 4, 5 and 6. For a model defined by the parameters $\boldsymbol{\theta}$, the Bayesian inference uses the observed data $\mathbf{y}$ and additional beliefs before observing the data to make probability statements of the parameters $\boldsymbol{\theta}$ [Gelman et al., 2014]. Then, the parameters $\boldsymbol{\theta}$ are estimated by the posterior probability distribution, $f(\boldsymbol{\theta}|\mathbf{y})$. In Bayes' theorem, the posterior distribution is defined as:

$$f(\boldsymbol{\theta}|\mathbf{y}) = \frac{f(\mathbf{y}|\boldsymbol{\theta})f(\boldsymbol{\theta})}{f(\mathbf{y})}.$$

The likelihood $f(\mathbf{y}|\boldsymbol{\theta})$ represents the probability of seeing the data $\mathbf{y}$ conditional on the values of parameters $\boldsymbol{\theta}$, and prior $f(\boldsymbol{\theta})$ represents the prior knowledge about the parameters $f(\boldsymbol{\theta})$. The marginal likelihood $f(\mathbf{y}) = \int f(\mathbf{y}|\boldsymbol{\theta})f(\boldsymbol{\theta})d\boldsymbol{\theta}$, is a normalisation integral which is independent of the parameters $\boldsymbol{\theta}$, so the posterior distribution can be written as:

$$f(\boldsymbol{\theta}|\mathbf{y}) \propto f(\mathbf{y}|\boldsymbol{\theta})f(\boldsymbol{\theta}).$$

2.2.1 Prior distributions

The prior distribution $f(\boldsymbol{\theta})$ defines the knowledge of the parameters $\boldsymbol{\theta}$ before observing the data $\mathbf{y}$, and there are different methods to specify the prior distribution. An informative prior can be used if there is specific information about the parameters [Gelman et al., 2014]. If there is no specific or little information about the parameter, a weakly-informative prior can be used, such as the uniform prior that assumes the parameter takes equal probability for all values within the support. Since the uniform prior is a constant, the posterior is proportional to the likelihood. Another choice is the conjugate prior, which is specified so that the posterior distribution and prior distribution have the same functional form. An example of conjugate prior for the normal distribution is shown below [Gelman et al., 2014]. Suppose the observation y is from a normal distribution with $y \sim N(\mu, \frac{1}{p_y})$, where μ and p_y are unknown mean and precision parameters. If the precision parameter p_y is known and the conjugate prior of $\mu \sim N(\mu_0, \frac{1}{p_0})$ is used, the conditional posterior distribution of μ is:

$$f(\mu|\frac{1}{p_y}, y) \sim N\left((p_y + p_0)^{-1}(p_y y + p_0\mu_0),\ (p_y + p_0)^{-1}\right).$$

If the mean parameter μ known and the conjugate prior $y \sim \text{Gamma}(a_y, b_y)$ is used, the conditional posterior distribution of p_y is:

$$f(p_y|\mu, y) \sim \text{Gamma}\left(a_y + \frac{1}{2},\ b_y + \frac{1}{2}(y-\mu)^2\right).$$

These conjugate priors can be generalised to the multivariate observations $\mathbf{y} = (y_1, y_2, \cdots, y_n)$, which are used in the parameter estimation in Chapters 4, 5, and 6.

2.2.2 Markov Chain simulation

Markov Chain Monte Carlo (MCMC) simulation is a method to generate samples from the target distribution, and the Markov chain converges to its stationary distribution after a long run, which is the target posterior distribution [Gelman et al., 2014]. Two popular MCMC algorithms are the Metropolis Hastings algorithm and the Gibbs sampler are introduced in what follows.

Metropolis Hastings algorithm

Metropolis Hastings algorithm is used when the posterior distribution $f(\boldsymbol{\theta}|\mathbf{y})$ does not have a closed form. Suppose the parameter vector is written as $\boldsymbol{\theta} = \{\theta_1, \theta_2, ..., \theta_k\}$, the Metropolis Hastings algorithm algorithm starts from finding an initial parameter value $\boldsymbol{\theta^0}$ with $f(\boldsymbol{\theta}^0|\mathbf{y}) > 0$ that is chosen at random. Then, for $t = 1, 2, 3, ...T$

1. Draw a possible new value $\boldsymbol{\theta}^*$ from the proposal distribution $q(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{t-1})$.

2. Calculate the acceptance probability $r = \frac{f(\boldsymbol{\theta}^*|\mathbf{y})/q(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{t-1})}{f(\boldsymbol{\theta}^{t-1}|\mathbf{y})/q(\boldsymbol{\theta}^{t-1}|\boldsymbol{\theta}^*)}$.

3. Generate a sample u from Uniform distribution Uniform$(0,1)$.

4. If $u \leq \min\{r,1\}$, set $\boldsymbol{\theta}^t = \boldsymbol{\theta}^*$ as the new value. If $u > \min\{r,1\}$, set $\boldsymbol{\theta}^t = \boldsymbol{\theta}^{t-1}$ as the new value.

One choice of the proposal distribution is a normal distribution with mean $\boldsymbol{\theta}^{t-1}$ and a fixed variance $\sigma^2\mathbf{I}$. Using proposal distribution with a small variance σ^2 leads to the proposed new value close to the current value. Then, more of the proposed values are accepted, which results in a high acceptance rate [Gelman et al., 2014]. The acceptance rate means the fraction of proposed values that are accepted. The variance parameter in the proposal distribution is tuned to achieve the desired acceptance rate. If the target distribution is a d-dimensional multivariate normal distribution ($d > 5$), the optimal acceptance rate is 0.23 [Gelman et al., 2014].

Gibbs sampler

The Gibbs sampler uses the conditional distribution as the proposal distribution for each parameter, which makes the acceptance rate for this proposal 1 [Gelman et al., 2014]. Suppose the parameters vector are written as $\boldsymbol{\theta} = \{\theta_1, \theta_2, ..., \theta_k\}$. After generating an initial value $\boldsymbol{\theta}$, the samples are drawn by the following steps for $t = 1, 2, 3, ...T$:

- Draw a sample from the conditional distribution $f(\theta_1^t|\theta_2^{t-1}, \theta_2^{t-1}, ..., \theta_k^{t-1})$.
- Draw a sample from the conditional distribution $f(\theta_2^t|\theta_1^t, \theta_3^{t-1}, ..., \theta_k^{t-1})$.
- ...
- Draw a sample from the conditional distribution $f(\theta_k^t|\theta_1^t, \theta_2^t, ..., \theta_{k-1}^{t-1})$.

In the Metropolis Hastings algorithm and Gibbs sampler, the total number of iterations T is a large number, say 20000, and the initial iterations are discarded to reduce the influence of the starting values. If the conditional distributions have closed forms, the Gibbs sampler will be used. However, when the conditional distributions are not from the known family of distributions, the Metropolis Hastings algorithm will be used. For both algorithms, the convergence of the Markov chain is checked before any inference is made from the MCMC simulation. In this thesis, the Geweke's diagnostic [Geweke, 1992] is used to check the convergence of the Markov chain. This diagnostic is based on a test of the means of the first part (10%) and the last part (50%) of a Markov chain. If the mean of the first part and the last part are not statistically different, then it suggests the chain has converged.

2.3 Regression modelling

Regression modelling is used to find the relationship between variables of interest. This section discusses the correlation between random variables and the linear model.

2.3.1 Correlation

The covariance between two random variables Y and X is defined as:

$$\mathrm{Cov}(X,Y) = \mathbb{E}\left[(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))\right].$$

Then, the correlation between two random variables is defined as:

$$\mathrm{Corr}(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}}.$$

The Pearson correlation coefficient is used to measure the correlation between two sets of data samples, $\mathbf{x} = (x_1, x_2, \cdots, x_n)$ and $\mathbf{y} = (y_1, y_2, \cdots, y_n)$. The Pearson correlation coefficient is defined as:

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}},$$

where $\bar{x}$ and $\bar{y}$ are the sample means of $\mathbf{x}$ and $\mathbf{y}$. Pearson correlation coefficient has a range $[-1, 1]$. If the correlation between two variables is zero, there is no linear relationship between these two variables. If the correlation is close to -1 or 1, there is strong linear association between these two variables [Mavrakakis and Penzer, 2021].

2.3.2 Linear model

Regression modelling is used to find the relationship between a response variable $\mathbf{y} = (y_1, y_2, ..., y_n)^\top$ and a set of exploratory variables $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n)^\top$, where $\mathbf{x}_i = (1, x_{i1}, x_{i2}, ..., x_{ip})$ is the $p+1$-dimensional exploratory variable relating to observation i for $i = 1, 2, ..., n$. The linear model assumes the relationship between y_i and $\mathbf{x}_i$ is:

$$y_i = c_0 + \sum_{j=1}^{p} c_j x_{ij} + \varepsilon_i, \quad i = 1, 2, ..., n.$$

The error term $\boldsymbol{\epsilon} = (\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_n)$ is assumed to follow the independent and identical normal distribution $\varepsilon_i \sim \mathrm{N}(0, \sigma^2)$ for $i = 1, 2, ..., n$. Thus, the matrix form of these equations is:

$$\mathbf{y} = \mathbf{X}\mathbf{c} + \boldsymbol{\epsilon},$$

where

$$\mathbf{c} = \begin{pmatrix} c_0 \\ c_1 \\ \cdots \\ c_p \end{pmatrix}, \boldsymbol{\epsilon} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \cdots \\ \varepsilon_n \end{pmatrix}.$$

Then, using the likelihood-based inference introduced in Section 2.1, the maximum likelihood estimations are $\hat{\mathbf{c}} = (\mathbf{X}^\top \mathbf{X})^{-1}\mathbf{X}^\top \mathbf{y}$ and $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \mathbf{x}_i\hat{\mathbf{c}})^2$, where $\mathbf{x}_i$ is the $i-$th row of the design matrix $\mathbf{X}$. However, in practice $\hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \mathbf{x}_i\hat{\mathbf{c}})^2$ is often used.

2.4 Regression splines

The reflectance samples for the data in this thesis are only made at a limited number of points over time and wavelength, but the underlying process can be assumed to be continuous [Spyrakos et al., 2018]. Thus, the reflectance data are treated as observations from a smooth function f over time and wavelength. Firstly the methods used to represent data with a smooth curve will be described. Then, how to control the smoothness of the curve by using smoothing parameters and how to choose these parameters will be discussed.

For a response variable $\mathbf{y} = (y_1, y_2, ..., y_n)$ and a single explanatory variable $\mathbf{x} = (x_1, x_2, ..., x_n)$ such as time or wavelength, a general smooth relationship can be estimated using the following model:

$$y_i = f(x_i) + \varepsilon_i, \tag{2.1}$$

where $\varepsilon_i \sim N(0, \sigma_\varepsilon^2)$ are assumed to be independent random measurement errors. A pair of adjacent data values y_i and y_{i+1} are assumed to be similar if $(x_i,\ x_{i+1})$ are similar, so a smooth function will be needed, which means the function $f(x)$ has one or more continuous derivatives [Wood, 2017]. The shape of the function $f(x)$ is not pre-specified, so it should be data-driven. If additional information is available that $f(x)$ has a simple shape such as a straight line or high-order polynomial, then the problem reduces to linear or polynomial regression [Hastie et al., 2009]. However, high-order polynomials should be carefully used since they will suffer from numerical instability [Epperson, 1987]. An alternative approach is to capture the behavior of the data locally, where the fitted model f is only continuous and differentiable to some degree. Non-parametric methods to estimate $f(x)$ include kernel smoothing and regression splines or penalized regression splines [Ramsay and Silvermann, 2005]. This section will focus on regression splines, and each smooth function $f(x)$ is represented using basis functions to ensure the model is linear in its parameters. In this locally defined basis function method, the estimator

will be represented by a set of predefined basis functions:

$$f(x) = \sum_{k=1}^{K} c_k \phi_k(x), \tag{2.2}$$

where $\phi_k(x)$ is the k-th basis function and c_k is the coefficient to be estimated. By letting $\mathbf{c} = (c_1, c_2, ..., c_k)^\top$ and $\boldsymbol{\phi}(x) = (\phi_1(x), \phi_2(x), ..., \phi_K(x))^\top$, equation (2.2) can be expressed in matrix notation as:

$$f(x) = \boldsymbol{\phi}(x)^\top \mathbf{c}. \tag{2.3}$$

Thus, equation (2.1) can also be transformed into matrix form as

$$\mathbf{y} = \boldsymbol{\Phi}\mathbf{c} + \boldsymbol{\epsilon}, \tag{2.4}$$

where $\boldsymbol{\epsilon} = (\varepsilon_1, \varepsilon_2, ..., \varepsilon_n)^\top \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_n)$ and $\boldsymbol{\Phi} \in \mathbb{R}^{n \times K}$ is a matrix where $\boldsymbol{\Phi}_{ik} = \phi_k(x_i)$. There is a list of commonly used basis functions given by [Wood, 2017]. In what follows, polynomial splines, B-splines and Fourier bases will be discussed.

2.4.1 Piecewise polynomial spline

A commonly used spline is the piecewise polynomial, where the fitted function is a polynomial at sub-intervals and the values (and first few derivatives) are consistent at the boundary of the intervals so that the fitted function is smooth [Wood, 2017]. Cubic splines are a common choice because of their smoothness to the human eye and computational properties. To guarantee a unique solution of the parameters for cubic splines, additional requirements are made on the boundary of the spline [Wood, 2017]. For example, the natural cubic splines force the function to be linear beyond the boundary knots. Besides, when the data are periodic, the spline function is required to reach the same value at the start and end of the period. This spline is called a cyclic cubic spline.

2.4.2 B-spline

A B-spline is defined by de Boor [De Boor and De Boor, 1978] recursively as:

$$\phi_{j,1}(x) = \begin{cases} 1 & \text{for } t_j \leq x < t_{j+1} \\ 0 & \text{otherwise} \end{cases}, \tag{2.5}$$

$$\phi_{j,k}(x) = \frac{x - t_j}{t_{j+k-1} - t_j} \phi_{j,k-1}(x) + \frac{t_{j+k} - x}{t_{j+k} - t_{j+1}} \phi_{j+1,k-1}(x), \tag{2.6}$$

where $j = 1,2,...m$ is indexed over a set of knots $(t_1,t_2,...t_m)$, and $t_j < t_{j+1}$. B-splines with order $k = 4$ are continuous on their first and second derivatives, making the fitted function looks smooth to human eyes. B-splines also have compact support, where the B-splines with order k are positive over no more than k intervals (t_j,t_{j+1}). This property makes the inner product matrix $\mathbf{\Phi}^\top\mathbf{\Phi}$ zero outside a diagonal band, providing a great deal of advantage in computing the estimate of coefficients due to its sparsity.

In R, different packages use different boundary knot settings for B-splines. As shown in Figure 2.1 [Eilers et al., 2015], the left plot uses equally spaced knots while the right one uses multiple boundary knots. The *mgcv* package [Wood, 2017] uses equally spaced knots, which means the B-spline of order k needs additional $(k-1)$ knots beyond both ends of the boundary. However, this equally spaced knot setting will result in a loss of differentiability at each endpoint, so the inferences cannot be made outside the region of available data. The *fda* package [Ramsay et al., 2009] uses multiple boundary knots, while *mgcv* package [Wood, 2017] uses equally spaced knots, which means the B-spline of order k needs additional $(k-1)$ knots beyond both ends of the boundary. Additionally, the *mgcv* package also uses a basis $\phi_0(x) = 1$ to capture the intercept.

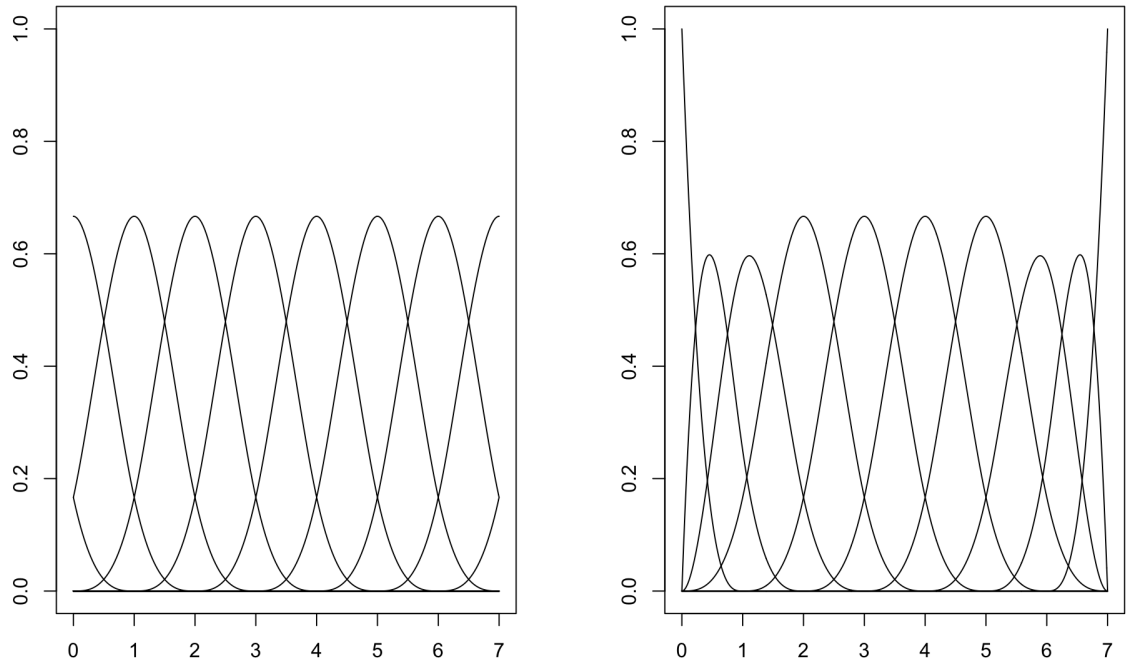


Figure 2.1: Examples of boundary of B-spline basis functions; left: equally spaced knots, right: multiple boundary knots.

2.4.3 Fourier basis

A Fourier basis may be preferred for periodic data, which is defined as below:

$$\begin{aligned}
&\phi_1(x) = 1 \\
&\phi_2(x) = \sin(\omega x) \\
&\phi_3(x) = \cos(\omega x) \\
&\phi_4(x) = \sin(2\omega x) \\
&\phi_5(x) = \cos(2\omega x) \\
&\vdots \\
&\phi_{2q}(x) = \sin(q\omega x) \\
&\phi_{2q+1}(x) = \cos(q\omega x),
\end{aligned}$$

where ω is the frequency of the basis function, and q is used to control the total number of basis functions with $m = 2q+1$. Thus, the fitted curve is

$$f(x) = c_1 + c_2\sin(\omega x) + c_3\cos(\omega x) + \cdots + c_{2q}\sin(q\omega x) + c_{2q+1}\cos(q\omega x). \tag{2.7}$$

The Fourier basis functions are considered for periodic data with period $2\pi/\omega$.

2.4.4 Parameter estimation

In this section, parameter estimation of the smoothing function will be introduced. For a chosen type of basis function, the smoothness of the fitted function is controlled by the number of basis functions, or equivalently the number of knots for the B-spline. One possible solution for choosing the degree of smoothness is to use backward selection through hypothesis testing. However the model with $k-1$ knots is not embedded within a model which has k evenly spaced knots, which makes this method problematic [Wood, 2017]. Alternatively, instead of just using the mean square error criteria, a penalty term is added to control the wiggleness of the fitted curve f. Thus, the objective criterion for the fitting of Equation 2.4 is:

$$\begin{aligned}
S(\boldsymbol{c}) &= ||\mathbf{y} - \mathbf{\Phi c}||^2 + \lambda \int_x [f''(x)]^2 dx \\
&= ||\mathbf{y} - \mathbf{\Phi c}||^2 + \lambda \mathbf{c}^\top \mathbf{D c},
\end{aligned} \tag{2.8}$$

where $\mathbf{D} = \int D^2\boldsymbol{\phi}(x)D^2\boldsymbol{\phi}(x)^\top dx$ and D^2 is the second order differential operator. The smoothing parameter λ works to control the smoothness of the function. The basis functions with relatively higher dimension than needed are selected and the exact value of the dimension should only slightly affect the fitted model [Wood, 2017]. To minimize this criterion, the derivative of (2.8) is taken with respect to the parameter vector $\mathbf{c}$ and set to zero, giving

$$-2\boldsymbol{\Phi}^\top\mathbf{y} + \boldsymbol{\Phi}^\top\boldsymbol{\Phi}\mathbf{c} + \lambda\mathbf{D}\mathbf{c} = 0. \tag{2.9}$$

Thus, the parameter vector $\mathbf{c}$ is estimated by:

$$\hat{\mathbf{c}} = (\boldsymbol{\Phi}^\top\boldsymbol{\Phi} + \lambda\mathbf{D})^{-1}\boldsymbol{\Phi}^\top\mathbf{y}. \tag{2.10}$$

This gives us the estimated fitted value:

$$\hat{\mathbf{y}} = \boldsymbol{\Phi}\hat{\mathbf{c}} = \boldsymbol{\Phi}(\boldsymbol{\Phi}^\top\boldsymbol{\Phi} + \lambda\mathbf{D})^{-1}\boldsymbol{\Phi}^\top\mathbf{y} = \mathbf{S}\mathbf{y}. \tag{2.11}$$

The matrix $\mathbf{S} = \boldsymbol{\Phi}(\boldsymbol{\Phi}^\top\boldsymbol{\Phi} + \lambda\mathbf{D})^{-1}\boldsymbol{\Phi}^\top$ is called the hat or projection matrix. We can also estimate the variance of the error σ_ε^2:

$$\hat{\sigma_\varepsilon^2} = \frac{\text{residual sum of squares}}{\text{degrees of freedom of error}} = \frac{||\mathbf{y} - \boldsymbol{\Phi}\hat{\mathbf{c}}||^2}{n - \text{trace}(2\mathbf{S} - \mathbf{S}\mathbf{S}^\top)}. \tag{2.12}$$

For $\lambda = 0$, the smoothness is controlled by the basis functions only as no penalty is applied. As $\lambda \to \infty$, the smoothness of the fitted function will increase as the penalty term becomes stronger and the function becomes a straight line. Other types of penalty are also available, such as the difference between the coefficients of adjacent B-splines [Eilers et al., 2015]. When first-order differences are used, the penalty term changes from $\int[f''(x)]^2dx$ to $\sum_{k=1}^{K-1}(c_{k+1} - c_k)^2$ [Eilers and Marx, 1996]. This penalty term will be used in Chapter 3 for smoothing reflectance data over time and wavelength. However, this penalty term is not used in chapters 4, 5 and 6 when applying the NSD model and the development of the NSD model. The smoothness of these models is controlled by the number of basis functions, and the optimal number of basis functions is selected by the cross-validation method, which will be discussed in Section 5.2.4.

2.4.5 Cross validation and Generalized cross validation

One possible approach to select the smoothing parameter is through assessing the predictive accuracy of the model. Leave-one-out cross validation is designed to achieve this goal, which works by leaving out each datum y_i in turn and estimating the average squared error from the model based on the remaining $n-1$ points [Wood, 2017]. Thus, the smoothing parameter λ is

selected to minimise the mean square error (MSE). That is

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}\{y_i - \hat{f}_{-i}(x_i)\}^2, \tag{2.13}$$

where $\hat{f}_{-i}(x_i)$ denotes the predicted value of the curve at x_i from the model constructed from the remaining data, excluding x_i. However, leave-one-out cross validation requires the model to be fitted n times for each potential smoothing parameter value, which is computationally expensive when the scale of n is large. To achieve computational efficiency, an approximation is applied to leave-one-out cross validation, which only requires the model to be fitted once with the full data for each potential smoothing parameter value. Craven and Wahba (1979) propose this approximation [Wahba, 1975], which is called generalized cross validation (GCV) and defined as:

$$\text{GCV} = \frac{n\,\text{RSS}}{(n - \text{trace}(\mathbf{S}))^2}, \tag{2.14}$$

where the residual sum of squares (RSS) is given by $||\mathbf{y} - \mathbf{\Phi}\hat{\mathbf{c}}||^2$. For each candidate smoothing parameter λ, GCV is calculated and the minimum value of GCV is used to indicate the optimal smoothing parameter.

2.5 Functional data analysis approaches

Section 2.4 introduces how splines can be used to estimate smooth functions. Functional data analysis is a statistical framework that treats the observations as realisations of smooth functions. After estimating the smooth functions from the observations, the functional data analysis approaches are carried out on the function level [Ramsay and Silvermann, 2005]. The reflectance data in this thesis can be converted into smooth functions over time or wavelength at each location of the lake. To investigate the variations in these functions, Functional principal component analysis will be used. Since the reflectance data are collected by multiple sensors, the similarity of the smooth functions from two sensors can be assessed by calculating the distance between two functional data. This section will discuss these two Functional data analysis approaches.

2.5.1 Distance between two functional data

In situations where multiple sensors providing data on the same process are available, multiple functions can be calculated. Thus, the difference between these functions can be defined. For any two functional data objects $f_i(x)$ and $f_j(x)$ formed by the same set of basis functions using Equation (2.3) where $\phi(x)$ is the vector of basis functions evaluated at x, the distance between

them is defined as:

$$d_{ij} = \sqrt{\int_x (f_i(x) - f_j(x))^2 dx} = \sqrt{\int_x (\mathbf{c}_i^\top \boldsymbol{\phi}(x) - \mathbf{c}_j^\top \boldsymbol{\phi}(x))^2 dx} = \sqrt{(\mathbf{c}_i - \mathbf{c}_j)^\top \mathbf{W} (\mathbf{c}_i - \mathbf{c}_j)}, \quad (2.15)$$

where the element of matrix $\mathbf{W} \in \mathbb{R}^{K \times K}$ is:

$$W_{ij} = \int_x \phi_i(x) \phi_j(x) dx, \quad (2.16)$$

and the matrix $\mathbf{W}$ will be evaluated by numerical integration [Febrero-Bande and De La Fuente, 2012]. This distance is the functional equivalent of the euclidean distance between two points. It will be computationally efficient for calculating the distance between multiple functional data with the same basis functions, since the matrix $\mathbf{W}$ is the same for every pair of functions.

2.5.2 Principal component analysis and Functional principal component analysis

Principal component analysis (PCA) is a method to analyse the multivariate data, which transforms original variables to a set of orthogonal ones, such that the new uncorrelated variables maximize the variance explained [Jolliffe, 2002]. However, it can not be applied directly to the discrete data with a large amount of missingness. Thus, the discrete data are converted to smooth functions, and the variation of the functions can be explored through Functional Principal Component Analysis (FPCA). The FPCA aims to find a set of orthonormal functions so that the lower dimensional representation uses the leading orthonormal functions to approximate the functional data as closely as possible [Ramsay and Silvermann, 2005]. In this section, the methods of PCA and FPCA will be discussed.

Principal component analysis (PCA)

For a data matrix $\mathbf{X} \in \mathbb{R}^{m \times n}$, PCA decomposes it to an orthogonal projection matrix $\mathbf{P}$ and a matrix $\mathbf{T}$ to represent the projection of $\mathbf{X}$ on to the new space, which is written as:

$$\mathbf{X} = \mathbf{T}\mathbf{P}^\top, \quad (2.17)$$

where the matrix $\mathbf{P} \in \mathbb{R}^{n \times r}$ is called the loading matrix and the matrix $\mathbf{T} \in \mathbb{R}^{m \times r}$ is referred to as the score matrix, and r is the rank of $\mathbf{X}$. The matrix $\mathbf{P}$ is calculated by an eigen-decomposition of the covariance matrix $\boldsymbol{\Sigma}$, such that $\boldsymbol{\Sigma} = \mathbf{P}\boldsymbol{\Lambda}\mathbf{P}^\top$, where each column of $\mathbf{P}$ represent an eigenvector and the diagonal entries of $\boldsymbol{\Lambda}$ are the corresponding eigenvalues in descending order. The sample covariance matrix $\boldsymbol{\Sigma}$ is calculated by $\boldsymbol{\Sigma} = \frac{\tilde{\mathbf{X}}^\top \tilde{\mathbf{X}}}{m-1}$, where $\tilde{\mathbf{X}}$ is the data matrix after mean subtraction and m is the number of observations of data. [Jolliffe, 2002]

For a spatial-temporal dataset, the variables of interest are observed at different locations and time points. In S-mode PCA [Richman, 1986], each column of the data matrix represents a location and each row represents time. The S-mode PCA will provide spatially indexed PCs, so that the PC loadings can be plotted spatially on a map, where the locations with similar loading values have the similar temporal patterns.

Functional Principal Component Analysis (FPCA)

After introducing PCA, the FPCA can be defined in a similar way. The basis function representation of functional data has been introduced in section 2.4, so that

$$f(x) = \sum_{j=1}^{K} c_j \phi_j(x).$$

Suppose there is a collection of functional data, such that

$$f_m(x) = \sum_{j=1}^{K} c_{mj} \phi_j(x), \quad m = 1, ..., M,$$

and this equation can be written into matrix form:

$$\mathbf{F}(x) = \mathbf{C}\boldsymbol{\phi}(x), \tag{2.18}$$

where $C_{m,j} = c_{mj}$. The FPCA aims to find K orthonormal function $\xi_j, j = 1, ..., K$, so that the expansion $\hat{f}_m(x) = \sum_{j=1}^{K} \alpha_{mj}\xi_j(x)$ will minimise the global measure of approximation $\sum_{m=1}^{M} ||f_m - \hat{f}_m||^2$, where $\alpha_{mj} = \int \hat{f}_m(x)\xi_j(x)dx$ and $||f_m - \hat{f}_m||^2 = \int [f_m(x) - \hat{f}_m(x)]^2 dx$ [Ramsay and Silvermann, 2005]. This is equivalent to solve the problem:

$$\int V(x_i, x)\xi(x)dx = \lambda\xi(x_i), \tag{2.19}$$

subject to $\int \xi(x)^2 dx = 1$ and $\int \xi_p(x)\xi_q(x)dx = 0$ for all $p \neq q$. The covariance function $V(x_i, x_k)$ for $i, k = 1, ..., n$ is defined as:

$$V(x_i, x_k) = \frac{1}{M} \sum_{m=1}^{M} f_m(x_i) f_m(x_k) = \frac{1}{n} \boldsymbol{\phi}(x_i)^\top \mathbf{C}^\top \mathbf{C} \boldsymbol{\phi}(x_k).$$

Suppose the orthonormal function $\xi(x)$ can also be expanded with the same basis function of the functional data, such that $\xi(x) = \sum_{j=1}^{K} d_j \phi_j(x) = \boldsymbol{\phi}(x)^\top \mathbf{d}$ [Ramsay and Silvermann, 2005].

Define the matrix $\mathbf{W} = \int \xi(x)\xi(x)^\top dx$, and the equation 2.19 can be written as:

$$\frac{1}{M}\phi(x_i)^\top \mathbf{C}^\top \mathbf{C}\mathbf{W}\mathbf{d} = \lambda\phi(x_i)^\top \mathbf{d}. \tag{2.20}$$

This equation holds for all x_i. Suppose $\mathbf{u} = \mathbf{W}^{1/2}\mathbf{d}$, the equation 2.20 becomes:

$$\frac{1}{M}\mathbf{W}^{1/2}\mathbf{C}^\top \mathbf{C}\mathbf{W}^{1/2}\mathbf{u} = \lambda\mathbf{u}. \tag{2.21}$$

Solving the eigen equation 2.21 provides K eigenvalues λ and eigenvector $\mathbf{u}$, and the eigenfunction can be computed by $\xi(x) = \phi(x)^\top \mathbf{d} = \phi(x)^\top \mathbf{W}^{-1/2}\mathbf{u}$. For each eigenfunction $\xi(x)$, the corresponding principal component score of the $f_m(x)$ is calculated by:

$$\alpha_m = \int \xi(x) f_m(x) dx, \quad m = 1, ..., M. \tag{2.22}$$

This PC score indicates the strength of the pattern $\xi(x)$ in the functional data $f_m(x)$.

2.6 Spatial Geostatistics

In this section, the spatial geostatistics methods are introduced because the reflectance data used in this thesis are treated as point-referenced data. In what follows, the definition of a geostatistical process, the spatial covariance functions and spatial prediction using kriging will be introduced.

2.6.1 Geostatistical process

A geostatistical process is defined as:

$$\{Y(\mathbf{s}) : \mathbf{s} \in A\},$$

where A is the study region [Diggle and Ribeiro, 2007], and A is assumed to be a subset of two-dimensional space $\mathbb{R}^2$ in this thesis. If the process $Y(\mathbf{s})$ are observed at n locations, the geostatistical data are defined as $\mathbf{Y} = \{Y(\mathbf{s}_1), Y(\mathbf{s}_2), \cdots, Y(\mathbf{s}_n)\}$. A key step of analysing geostatistical data is modelling the spatial correlation, which is represented by the covariance function. Suppose the process $Y(\mathbf{s})$ have a mean function of$\mu(\mathbf{s}) = \mathbb{E}[Y(\mathbf{s})]$, then the covariance function between two locations $\mathbf{s}$ and $\mathbf{s}'$ is defined as:

$$\mathrm{Cov}[Y(\mathbf{s}), Y(\mathbf{s}')] = \mathbb{E}[(Y(\mathbf{s}) - \mu(\mathbf{s}))(Y(\mathbf{s}') - \mu(\mathbf{s}'))].$$

Stationarity and isotropy

A geostatistical process $Y(\mathbf{s})$ is weakly stationary if $\mathbb{E}[Y(\mathbf{s})] = \mu(\mathbf{s}) = \mu$, and μ is a constant for all location $\mathbf{s}$. Additionally, the covariance $\text{Cov}[Y(\mathbf{s}), Y(\mathbf{s}+\mathbf{h})] = C(\mathbf{s}, \mathbf{s}+\mathbf{h}) = C(\mathbf{h})$ is a finite constant for any displacement vector $\mathbf{h}$ [Diggle and Ribeiro, 2007]. A further simplification for a weakly stationary geostatistical process is isotropy if the covariance $C(\mathbf{h}) = C(h)$, where $h = \|\mathbf{h}\|$ is the Euclidean distance between the two locations. For an isotropy process, the direction of $\mathbf{h}$ does not matter in the covariance function, and the covariance between two locations can be defined as a function of distance between them.

Gaussian process

Gaussian processes are widely used to model geostatistical data, because they incorporate different types of spatial correlation structures [Diggle and Ribeiro, 2007]. A geostatistical process $Y(\mathbf{s})$ is Gaussian process if for any collection of locations $\{\mathbf{s}_1, \mathbf{s}_2, \cdots, \mathbf{s}_n\}$, the joint distribution of $\mathbf{Y} = (Y(\mathbf{s}_1), Y(\mathbf{s}_2), \cdots, Y(\mathbf{s}_n))$ is multivariate Gaussian and is defined as:

$$\mathbf{Y} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma}),$$

which has a mean of $\boldsymbol{\mu}$ and a covariance of $\boldsymbol{\Sigma}$. A key property of the Gaussian process is the conditional multivariate Gaussian distribution. Suppose the vector $\mathbf{Y} = \begin{pmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \end{pmatrix}$ is jointly multivariate Gaussian which follows:

$$\mathbf{Y} = \begin{pmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \end{pmatrix} \sim N\left(\begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix} \right),$$

where $\boldsymbol{\Sigma}_{11}$ and $\boldsymbol{\Sigma}_{22}$ are the variance matrices for the vectors $\mathbf{Y}_1$ and $\mathbf{Y}_2$ respectively, and $\boldsymbol{\Sigma}_{12} = \boldsymbol{\Sigma}_{21}^{\top}$ is the covariance matrix between the vectors $\mathbf{Y}_1$ and $\mathbf{Y}_2$. Then, the conditional distribution of $\mathbf{Y}_1 | \mathbf{Y}_2$ is also multivariate Gaussian, which follows:

$$\mathbf{Y}_1 | \mathbf{Y}_2 \sim N(\boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{Y}_2 - \boldsymbol{\mu}_2),\ \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}). \tag{2.23}$$

This result is the basis of spatial prediction as outlined below. In this thesis, the weakly stationary and isotropic Gaussian process will be used to model geostatistical data.

Covariance functions

In geostatistical modelling, there are several models for the covariance function. One common choice is the weakly stationary and isotropic Matérn covariance function [Diggle and Ribeiro,

2007], which is defined as:

$$C(h) = \begin{cases} \sigma^2 + \tau^2 & \text{if } h = 0, \\ \sigma^2 \frac{(h/\phi)^\kappa}{2^{\kappa-1}\Gamma(\kappa)} K_\kappa(h/\phi) & \text{if } h > 0. \end{cases} \tag{2.24}$$

In this expression, $\Gamma(\kappa)$ is a Gamma function of κ, where κ is a smoothness parameter. For a fixed value of κ, $K_\kappa(\cdot)$ is a modified Bessel function of order κ. In addition, ϕ is the spatial decay parameter which controls how fast the covariance reaches zero, σ^2 is the partial sill which measures the amount of spatial variation, and τ^2 is the nugget which represents the amount of non-spatial variation. There are special cases of the smoothness parameter κ. If $\kappa = \frac{1}{2}$, this function is called an exponential covariance function, which can be written as:

$$C(h) = \begin{cases} \sigma^2 + \tau^2 & \text{if } h = 0, \\ \sigma^2 \exp(-h/\phi) & \text{if } h > 0. \end{cases} \tag{2.25}$$

If $\kappa \to \infty$, it is called a Gaussian covariance function, and is defined as:

$$C(h) = \begin{cases} \sigma^2 + \tau^2 & \text{if } h = 0, \\ \sigma^2 \exp(-(h/\phi)^2) & \text{if } h > 0. \end{cases} \tag{2.26}$$

Since the exponential covariance function is used in the NSD model [Wilkie et al., 2019] and the data fusion models proposed in Chapters 5 and 6 are developed based on the NSD model, the exponential covariance function is selected in this thesis. Other covariance functions could be explored. For the hierarchical models whose latent parameters follow a geostatistical process, the computational and storage challenges arise when the dimension of the covariance matrix is large [Jona Lasinio et al., 2013]. These latent parameters will be estimated in the Bayesian framework in this thesis, which requires inverting the matrix with a dimension of n in each iteration of the MCMC simulation. The details of the computational challenge associated with the covariance matrix will be discussed in Section 5.21.

Geostatistical model

Suppose a geostatistical process $\mathbf{Y}$ follows:

$$\mathbf{Y} \sim N(\mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Sigma}(\boldsymbol{\theta})), \tag{2.27}$$

where $\mathbf{X} \in \mathbb{R}^{n \times p}$ is the covariates matrix and $\boldsymbol{\beta}$ is a p-dimensional vector of the regression coefficients [Diggle and Ribeiro, 2007]. Suppose an exponential covariance function in Equation (2.25) is used, the covariance matrix can be written as $\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \sigma^2 \exp(-\mathbf{D}/\phi) + \tau^2 \mathbf{I}$, where $\mathbf{D}$ is the distance matrix between the data locations and $\mathbf{I}$ is the identity matrix. Then, the log-

likelihood function of Model 2.27 with observations $\mathbf{y}$ is:

$$\ell(\beta,\sigma^2,\tau^2,\phi) = -\frac{1}{2}\{n\log(2\pi)+\log|\sigma^2\exp(-\mathbf{D}/\phi)+\tau^2\mathbf{I}| \\ +(\mathbf{y}-\mathbf{X}\beta)^\top(\sigma^2\exp(-\mathbf{D}/\phi)+\tau^2\mathbf{I})^{-1}(\mathbf{y}-\mathbf{X}\beta)\}. \tag{2.28}$$

Suppose $\nu^2=\tau^2/\sigma^2$ and $\mathbf{V}=\exp(-\mathbf{D}/\phi)+\nu^2\mathbf{I}$, the log-likelihood function in Equation (2.28) is maximised at

$$\hat{\beta}(\mathbf{V}) = (\mathbf{X}^\top\mathbf{V}^{-1}\mathbf{X})\mathbf{X}^\top\mathbf{V}^{-1}\mathbf{y},$$
$$\hat{\sigma}^2(\beta,\mathbf{V}) = n^{-1}(\mathbf{y}-\mathbf{X}\beta)^\top\mathbf{V}^{-1}(\mathbf{y}-\mathbf{X}\beta).$$

In practice, the unbiased estimator is used for σ^2,

$$\hat{\sigma}^2(\beta,\mathbf{V}) = (n-p)^{-1}(\mathbf{y}-\mathbf{X}\beta)^\top\mathbf{V}^{-1}(\mathbf{y}-\mathbf{X}\beta).$$

2.6.2 Spatial correlation investigation

In this thesis, the spatial correlation of the residuals will be investigated to check if the model assumptions hold using the empirical semi-variogram. Suppose a weakly stationary and isotropic process $Y(\mathbf{s})$ is defined over a set of locations $\mathbf{s}=\{\mathbf{s}_1,\mathbf{s}_2,\cdots,\mathbf{s}_n\}$, the theoretical semi-variogram is defined as:

$$\gamma(h) = \frac{1}{2}\mathbb{E}[Y(\mathbf{s})-Y(\mathbf{s}+\mathbf{h})]^2.$$

For a dataset (or residuals) $\mathbf{y}=(y(\mathbf{s}_1),y(\mathbf{s}_2),\cdots,y(\mathbf{s}_n))$, the binned empirical semi-variogram is used to investigate the spatial correlation. Suppose the distance between all pairs of locations are grouped into K bins, $I_k=(h_{k-1},h_k]$ for $k=1,2,\cdots,K$ and $0=h_0<h_1<\cdots<h_k$. For the bin I_k, suppose there are n_k pairs of locations with distances within this bin. Then, the binned empirical semi-variogram can be defined as:

$$\hat{\gamma}(h_k^m) = \frac{1}{2n_k}\sum_{\|\mathbf{s}_i-\mathbf{s}_i\|\in I_k}(y(\mathbf{s}_i)-y(\mathbf{s}_j))^2,$$

where $h_k^m=(h_{k-1}+h_k)/2$ represents the centre of the bin I_k. To assess the spatial correlation, Monte Carlo envelopes for the empirical semi-variogram are computed [Diggle and Ribeiro, 2007]. For a large number of iterations, say 1000, the following two steps are repeated. First, a pseudo dataset is generated by permuting n data values randomly among the n locations. Then, the empirical semi-variogram is computed for each midpoint of the bins for the pseudo data set. After 1000 iterations, a 95% envelope for each distance h_k^m is computed as the $[2.5\%,97.5\%]$ percentiles of the empirical semi-variogram $\hat{\gamma}(h_k^m)$ among these 1000 iterations. If the empirical

semi-variogram falls outside the Monte Carlo envelopes particularly for small distance h, then there is significant spatial correlation in the data.

2.6.3 Kriging

Kriging is a method for spatial prediction of a process at a new location $\mathbf{s}_0$, and it will be used in Chapters 4 and 5 to predict the reflectance data at the unobserved locations. In this thesis, a weakly stationary and isotropic Gaussian process is considered. For a spatial process $\mathbf{Y} \sim N(\mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Sigma}(\boldsymbol{\theta}))$, Kriging aims to find the best linear unbiased predictor $\hat{Y}(\mathbf{s}_0)$ which minimises the mean squared prediction error (MSPE), $\mathbb{E}[(\hat{Y}(\mathbf{s}_0) - Y(\mathbf{s}_0))^2]$ [Diggle and Ribeiro, 2007]. The minimum of MSPE is achieved when $\hat{Y}(\mathbf{s}_0) = \mathbb{E}[Y(\mathbf{s}_0)|\mathbf{Y}]$. The joint distribution of $(Y(\mathbf{s_0}), \mathbf{Y})$ can be written as:

$$\begin{pmatrix} Y(\mathbf{s_0}) \\ \mathbf{Y} \end{pmatrix} \sim N\left(\begin{pmatrix} \mathbf{x}_0^\top \boldsymbol{\beta} \\ \mathbf{X}\boldsymbol{\beta} \end{pmatrix}, \begin{pmatrix} C(0,\boldsymbol{\theta}) & \mathbf{C}(\mathbf{s}_0,\boldsymbol{\theta})^\top \\ \mathbf{C}(\mathbf{s}_0,\boldsymbol{\theta}) & \boldsymbol{\Sigma}(\boldsymbol{\theta}) \end{pmatrix} \right),$$

where $\mathbf{x}_0$ is the covariates at location $\mathbf{s}_0$, $C(0,\boldsymbol{\theta})$ is the variance at location $\mathbf{s}_0$, and $\mathbf{C}(\mathbf{s}_0,\boldsymbol{\theta}) = (\mathrm{Cov}[Y(\mathbf{s}_0), Y(\mathbf{s}_1)], \mathrm{Cov}[Y(\mathbf{s}_0), Y(\mathbf{s}_2)], \cdots, \mathrm{Cov}[Y(\mathbf{s}_0), Y(\mathbf{s}_n)])$ is the covariance between the prediction location and the observed data locations. Using Equation 2.23, the conditional distribution of $Y(\mathbf{s}_0)|\mathbf{Y}$ follows:

$$Y(\mathbf{s}_0)|\mathbf{Y} \sim N(\mathbf{x}_0^\top \boldsymbol{\beta} + \mathbf{C}(\mathbf{s}_0,\boldsymbol{\theta})^\top \boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}),\ C(0,\boldsymbol{\theta}) - \mathbf{C}(\mathbf{s}_0,\boldsymbol{\theta})^\top \boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}\mathbf{C}(\mathbf{s}_0,\boldsymbol{\theta})).$$

Since the parameters μ_Y and $\boldsymbol{\theta}$ are unknown, the conditional distribution of $Y(\mathbf{s}_0)|\mathbf{Y}$ is:

$$Y(\mathbf{s}_0)|\mathbf{Y} \sim N(\mathbf{x}_0^\top \hat{\boldsymbol{\beta}} + \mathbf{C}(\mathbf{s}_0,\hat{\boldsymbol{\theta}})^\top \boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})^{-1}(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}),\ C(0,\hat{\boldsymbol{\theta}}) - \mathbf{C}(\mathbf{s}_0,\hat{\boldsymbol{\theta}})^\top \boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})^{-1}\mathbf{C}(\mathbf{s}_0,\hat{\boldsymbol{\theta}})),$$

where the parameters $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\theta}}$ are estimated by Bayesian inference in this thesis. The predictor $\hat{Y}(\mathbf{s}_0) = \mathbb{E}[Y(\mathbf{s}_0)|\mathbf{Y}] = \mathbf{x}_0^\top \hat{\boldsymbol{\beta}} + \mathbf{C}(\mathbf{s}_0,\hat{\boldsymbol{\theta}})^\top \boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})^{-1}(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})$ is called universal Kriging predictor.

Chapter 3

Exploratory analysis

This chapter presents an exploratory analysis of the remote sensing reflectance data from the MERIS, MODIS, and OLCI sensors using the statistical methods introduced in the previous chapter. The first aim of this exploratory analysis is to introduce the data and investigate the data over time, space and wavelength for each sensor. The second aim of this exploratory analysis is to identify the relationship between the reflectance data from these sensors. The data fusion models such as the work of [Berrocal et al., 2010b], and [Cressie and Johannesson, 2008] introduced in Chapter 1 build a linear regression model between the data from two different sources, assuming there is a linear relationship over time between them. Thus, the linear relationship between these sensors over time will be checked for each pixel. The relationship of reflectance between sensors over wavelength will also be investigated because the fusion model of reflectance over wavelength in Chapter 6 requires the closeness of the two sensors over wavelength. The third aim of the exploratory analysis is to understand the variations in the reflectance data. If there are common temporal patterns over space in the reflectance data for each sensor, this property could be used to represent the reflectance data in the data fusion model, such as in the nonparametric statistical downscaling (NSD) model [Wilkie et al., 2019] that I utilise in Chapters 4 and 5. The reflectance data from Lake Garda will be used to illustrate the properties of reflectance and check the performance of the proposed data fusion models in this research. Lake Garda is around the medium size of the lakes provided by Plymouth Marine Laboratory (PML) [Carrea et al., 2022], and its relatively small size makes it quicker to check the model performance in the following chapters.

This chapter is structured as follows. First, the reflectance data from Lake Garda will be introduced and illustrated by exploratory plots of the reflectance over space, time, and wavelength. Then, the relationships between the sensors will be investigated over time and wavelength for each location within Lake Garda. Since there is no overlapping time for all these three sensors as discussed in Chapter 1, the relationships will be evaluated for MODIS-MERIS and MODIS-OLCI for their overlapping years. The relationship over time will be assessed by scatter plots

and Peason's correlation coefficients at each location and each wavelength. Then, the relationship over wavelength will be explored using functional distance to ensure that data fusion over wavelength is appropriate for the Lake Garda dataset. Since reflectance data are observed at different wavelength bands from these sensors, the reflectance data are first converted to smooth functions over wavelength and the functional distance is calculated between these functions. After that, the common temporal patterns over space are investigated using functional principal component analysis (FPCA). Since this thesis will start from a spatial-temporal model as discussed in Chapter 1, the spatial-temporal patterns are the focus of this chapter and the variations are explored for each wavelength band separately. Finally, potentially suitable data fusion models for this thesis and the datasets used in the following chapters will be discussed.

3.1 Data description

This section describes the reflectance data from Lake Garda. Since this research aims to develop a general fusion method for lake reflectance data from two satellite sensors, the method is not sensitive to the choice of the lake, but Lake Garda is used in this thesis for illustration. Lake Garda is the largest lake in Italy with an area of 368 km^2, and it is vital for its tourism, agriculture and drinking purposes [Giardino et al., 2014]. The reflectance data at Lake Garda are observed by the MERIS, MODIS, and OLCI sensors. The MERIS reflectance data are available from 01/01/2009 to 30/03/2012. The temporal coverage of the MODIS and OLCI sensors are 01/01/2009-31/12/2019 and 01/05/2016-31/12/2019 respectively. The temporal coverage is different from the details in Table 1.2 because the data considered here are a subset of the remote-sensing data from these three sensors provided by PML. Additionally, the spatial resolutions for these two lakes are aggregated into 1km $\times$ 1km for the MERIS and OLCI sensors by PML, which gives the same spatial resolution as the MODIS sensor. The reflectance data from wavelength bands between 400 nm to 681 nm are explored in this section, because the majority of retrieval algorithms of water quality parameters described in Liu et al. (2021) and Yang et al. (2022) are based on the reflectance data within this range [Liu et al., 2021, Yang et al., 2022]. There are about 600 pixels for each sensor. For the days with observations, the mean of pixels with observations is about 300, which means half of the locations have observations on average. The missing rate are summarised in Figure 3.1 and Figure 3.2. For each location, the missing rate is calculated by the number of days with missing values divided by the total number of days. The missing rate is higher at the lake's boundary than at the inner locations for the MERIS and OLCI sensors. For the MODIS sensor, the northwest boundary has a higher missing rate than the rest of the locations, and this pattern could result from the different algorithms used by these sensors to distinguish land and water.

Reflectance data from Lake Garda for all pixels are plotted over time, space and wavelengths

to show the features of reflectance data. The temporal plots of reflectance data are shown in Figure 3.3 for the MERIS and MODIS sensors, and in Figure 3.4 for the MERIS and MODIS sensors for their overlapping time periods. These pairs of bands are selected because they have a similar coverage of wavelength as shown in Table 1.2. For each time point, the reflectance data are averaged over space to investigate the general temporal patterns over space. Figure 3.3 shows approximately one peak per year for the MERIS and MODIS sensors for the first three pairs of wavelength bands. The reflectance data for the other two pairs show relatively less temporal variations. Figure 3.4 also shows similar temporal patterns between the OLCI and MODIS sensors for these selected bands.

The reflectance data are averaged over time for each location to investigate the general spatial patterns for different pairs of wavelength bands in Figure 3.5 and Figure 3.6. These three pairs of bands are selected to illustrate the spatial patterns because they cover the wavelength range from 412 nm to 618 nm. The retrieval algorithms of water quality parameters rely on the reflectance data within this range [Yang et al., 2022]. These plots use different scales of colour for each pair of wavelength bands to show the spatial patterns because the averaged reflectance data could be affected by the extreme values. From the spatial plots, reflectance data are generally smooth over space for these sensors, and the spatial patterns are similar between MERIS-MODIS and OLCI-MODIS sensors for the selected pairs of bands. It also shows the difference in the average reflectance between the locations near the boundary of the lake and the inner part of the lake. The spatial plots for the other wavelength bands are also checked, showing generally similar spatial patterns between these sensors.

The boxplots of reflectance data at each wavelength band are shown in Figure 3.7 and Figure 3.8 for the MERIS-MODIS and OLCI-MODIS sensors for their overlapping date periods and all locations in Lake Garda. The outliers are not included in the boxplots because the aim is to show the variability of reflectance data at different wavelength bands. It is found that there is more variability for the wavelength bands within 400-600 nm than those within 600-700 nm for all sensors when comparing their interquartile ranges. This spectral pattern is similar in MERIS-MODIS and MODIS-OLCI sensors for their overlapping periods.

In summary, the plots of reflectance data suggest that the MERIS-MODIS and OLCI-MODIS sensors have similar temporal, spatial, and spectral patterns of reflectance data. In the following sections, these patterns and the relationship of reflectance data between these sensors will be further investigated.

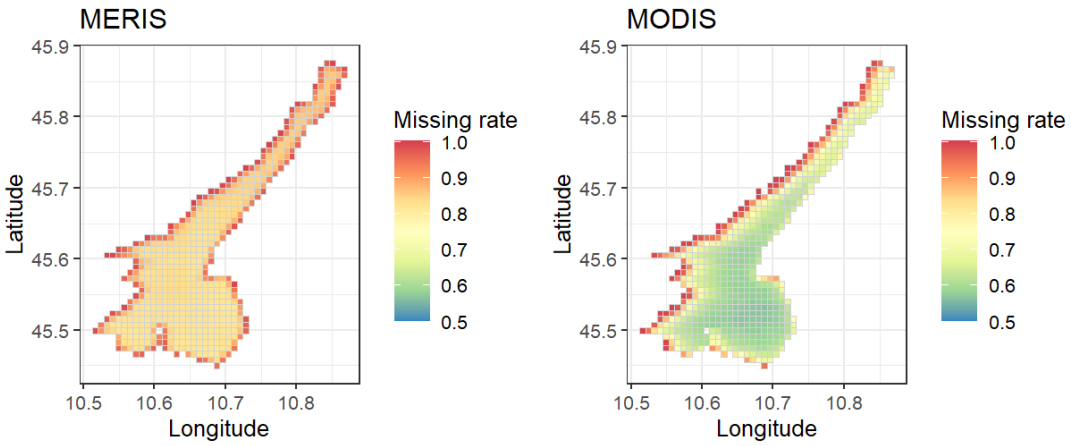


Figure 3.1: Missing rate of the MERIS and MODIS sensors from 01/01/2009 to 30/03/2012.

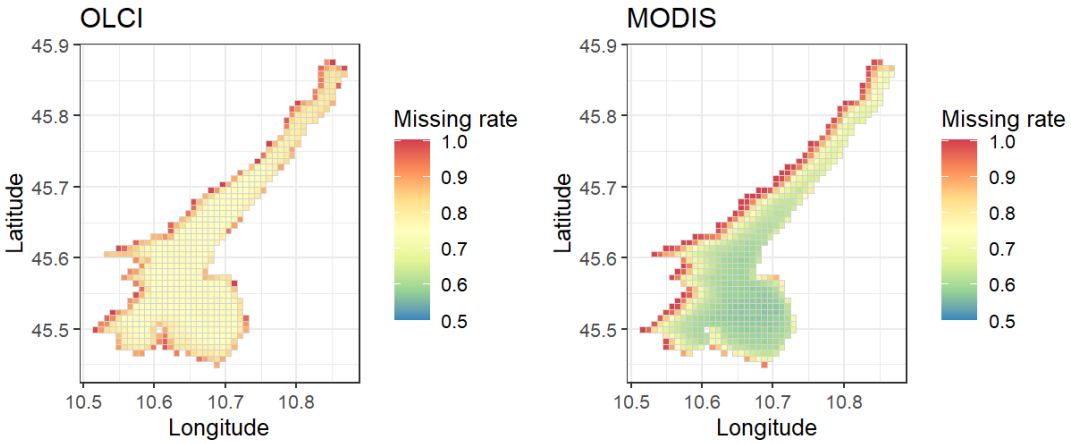


Figure 3.2: Missing rate of the OLCI and MODIS sensors from 01/01/2009 to 30/03/2012.

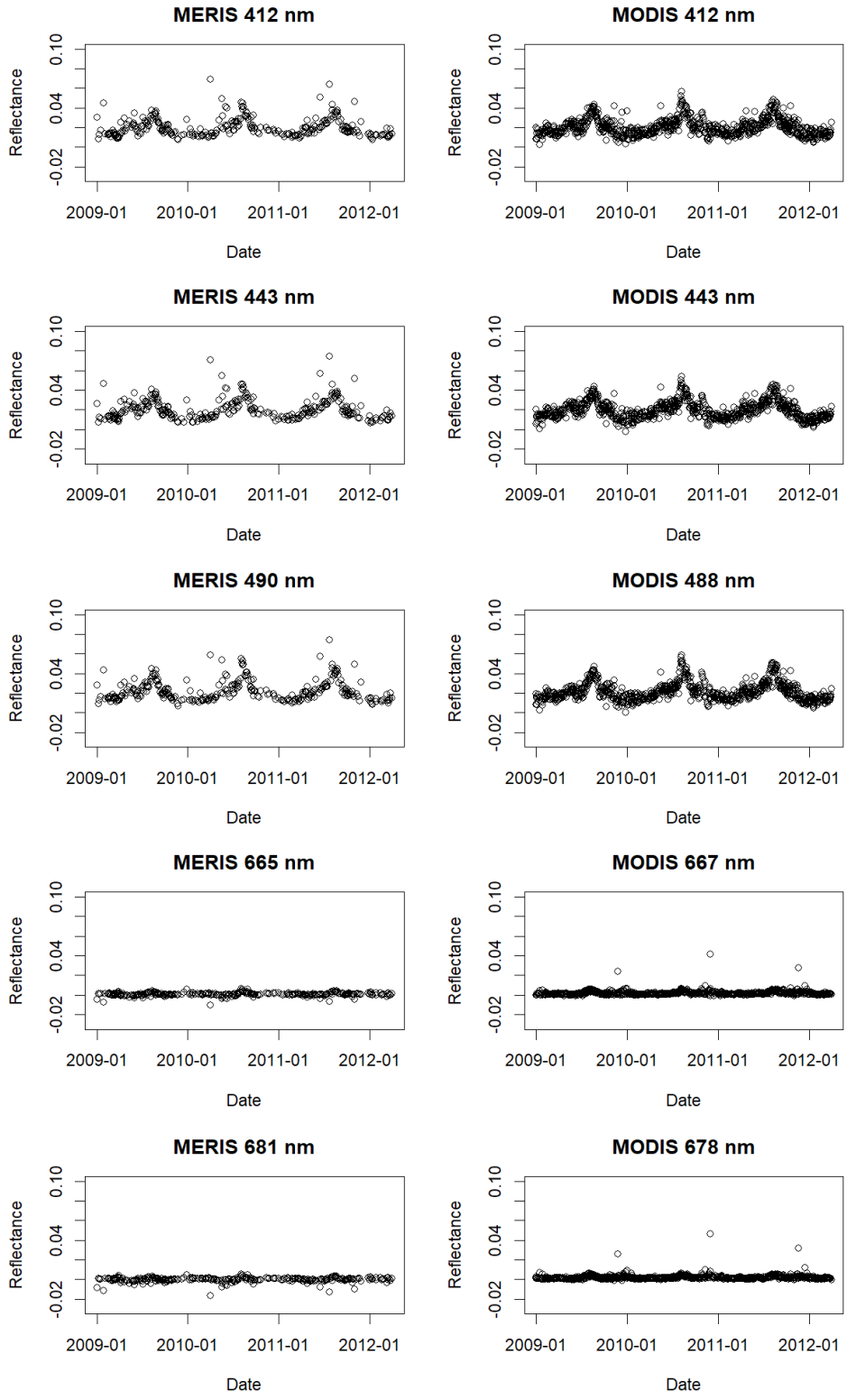


Figure 3.3: Averaged MERIS and MODIS reflectance data over space for each date and the selected pairs of wavelength bands for Lake Garda from 01/01/2009 to 30/03/2012.

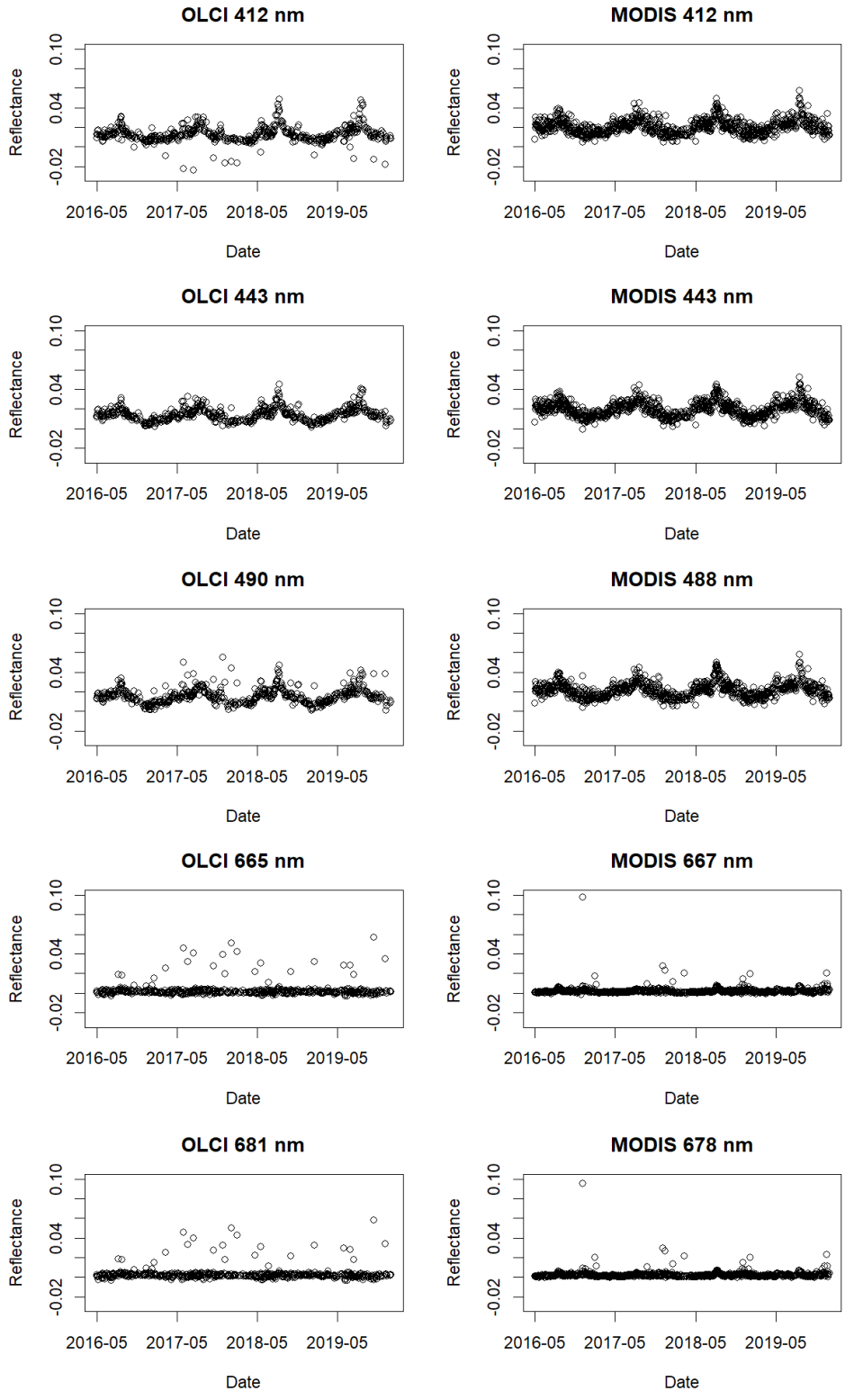


Figure 3.4: Averaged OLCI and MODIS reflectance data over space for each date and the selected pairs of wavelength bands for Lake Garda from 01/05/2016-31/12/2019.

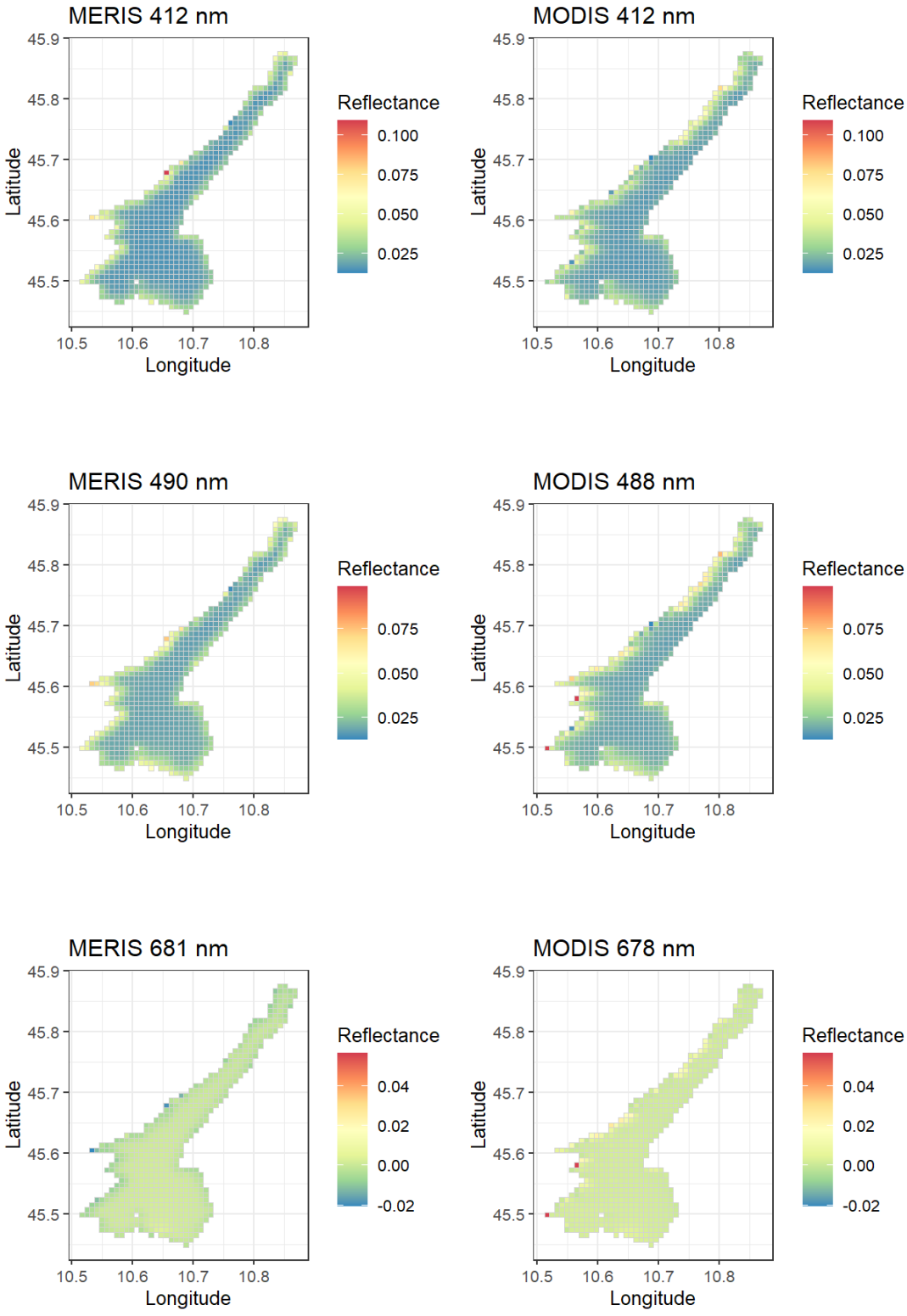


Figure 3.5: Averaged MERIS and MODIS reflectance data over time for each location in Lake Garda and the selected pairs of wavelength bands.

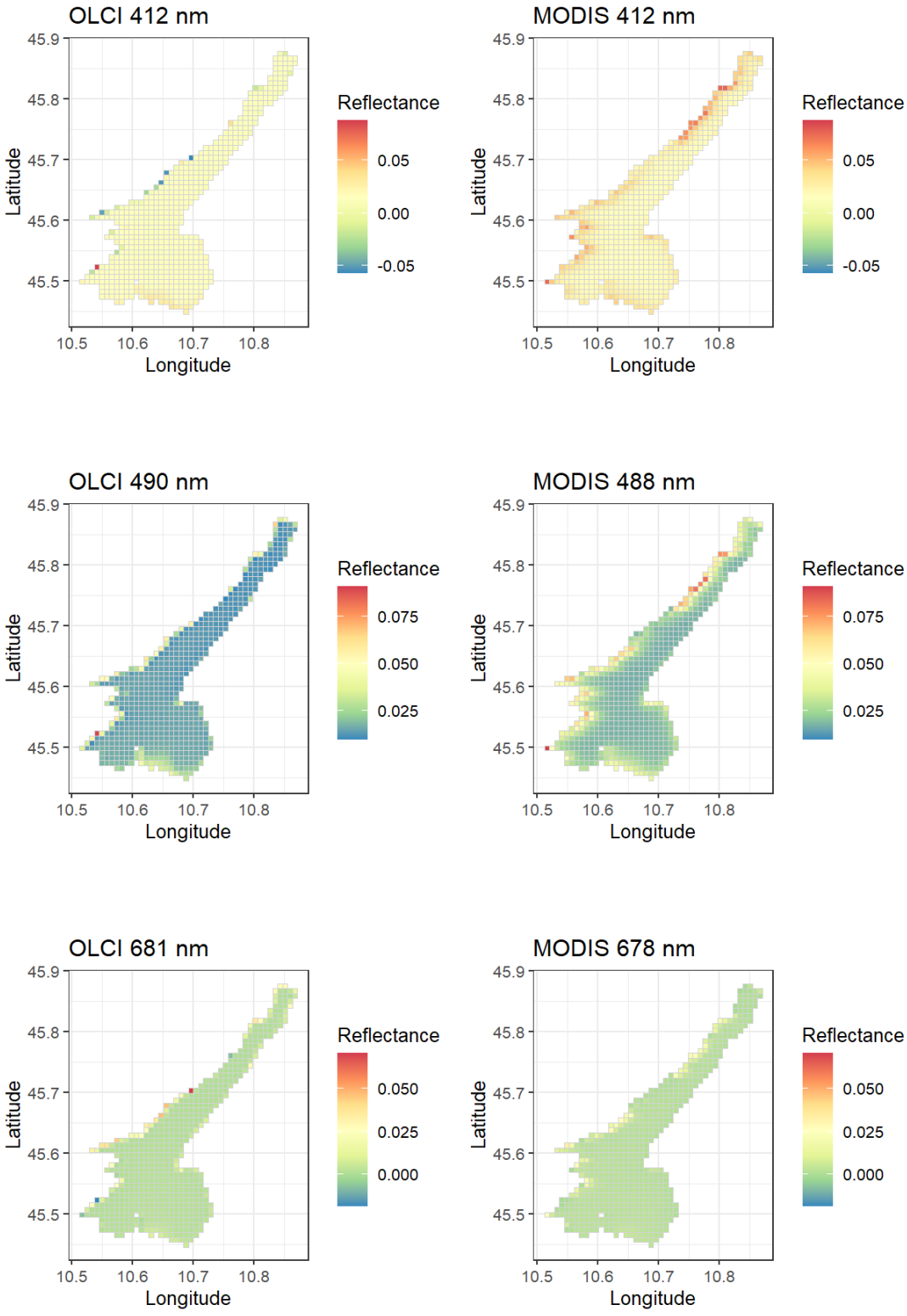


Figure 3.6: Averaged OLCI and MODIS reflectance data over time for each location in Lake Garda and the selected pairs of wavelength bands.

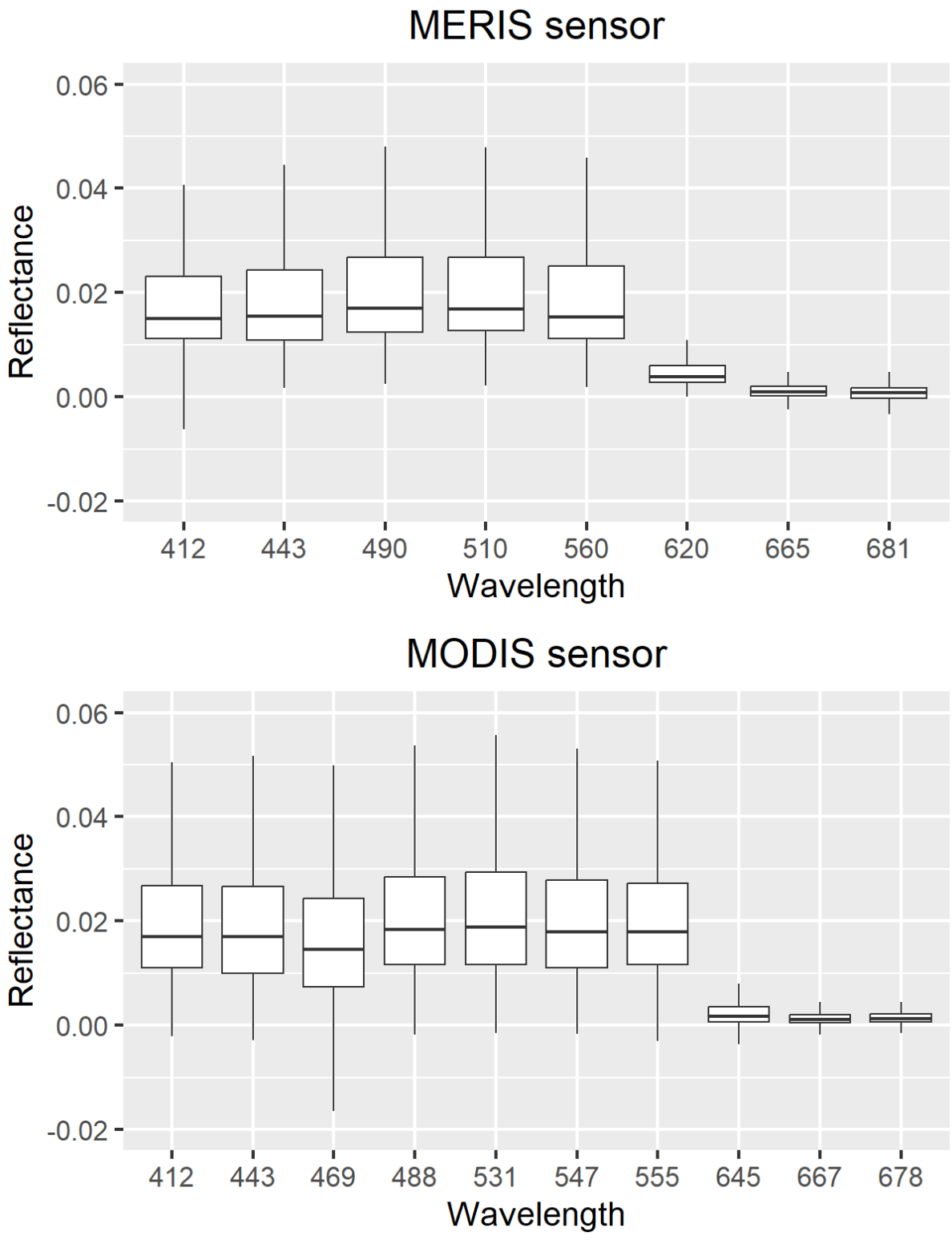


Figure 3.7: Boxplots of the MERIS and MODIS reflectance data at each wavelength band for all locations in Lake Garda from 01/05/2016-31/12/2019.

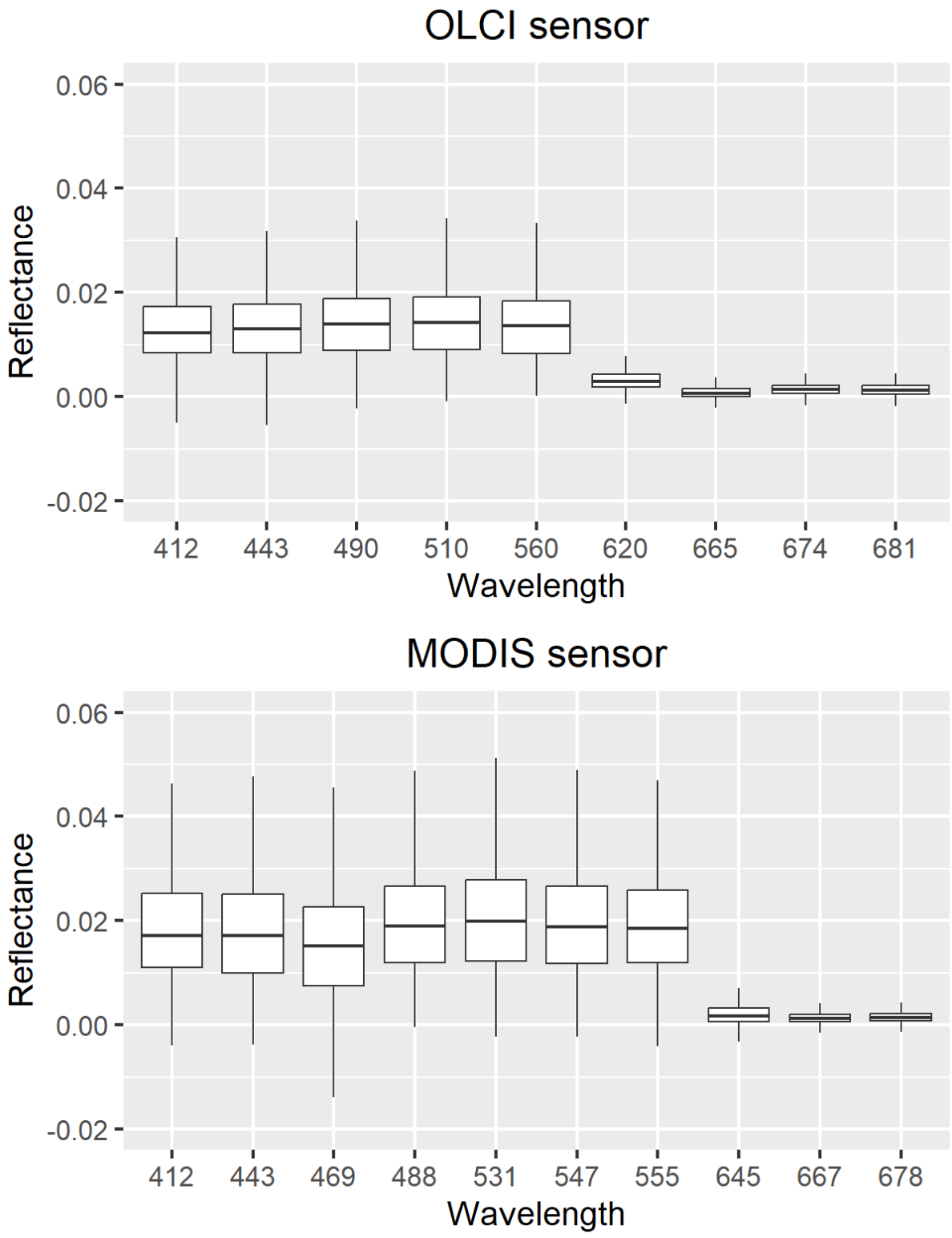


Figure 3.8: Boxplots of the OLCI and MODIS reflectance data at each wavelength band for all locations in Lake Garda from 01/01/2009 to 30/03/2012.

3.2 Assessing the relationship between the sensors over time

In this section, the relationship between the sensors is assessed using the scatter plots and Pearson correlation coefficients. The aim is to check if there are any linear relationships between the two series of reflectance data over time from the two sensors at the same location and similar wavelength bands, and the strength of the linear relationship at each pixel. Considering the coverage of wavelength bands in Table 1.2, five pairs of bands between each pair of sensors are selected, which are (MODIS, MERIS) = (412, 412), (443, 443), (488, 490), (667, 665), (678, 681) and (MODIS, OLCI) = (412, 412), (443, 443), (488, 490), (667, 665), (678, 681). These pairs of bands are assumed to be comparable since the reflectance data from these sensors are recorded at similar wavelengths for these pairs of bands. Then, the scatter plots of the MERIS reflectance data against the MODIS reflectance data and the OLCI reflectance data against the MODIS reflectance data are shown in Figure 3.10 and 3.11. These eight locations are randomly selected to illustrate the relationship between these sensors and the linear relationships can be assumed for these locations. The scatter plots for the other locations and the other pairs of wavelength bands are also checked for these sensors, and there is no strong evidence against the linear relationships between these sensors.

Then, the Pearson correlation coefficients introduced in Section 2.3 are calculated for each location within the lake to assess the strength of linear relationship. The correlation plots for Lake Garda between the MODIS and MERIS sensors are shown in Figure 3.12, and the correlation plots between the MODIS and OLCI sensors are shown in Figure 3.12. Correlation values are higher for the inner pixels than the edge pixels with correlations ranging from 0.5 to 1 for inner pixels and -1 to 0.5 for the edge pixels. The edge pixels and inner pixels for Lake Garda are defined in Figure 3.14, where the blue cells represent the edge pixels and the red cells represent the inner pixels. The edge pixel is defined as the boundary between the NA and non-NA values in the R package *raster*, which means the eight pixels surrounding the edge pixel have both NA and non-NA values [Hijmans et al., 2015]. This process is repeated twice, so that the remaining inner pixels have a relatively strong linear relationship between the sensors over time. The remotely sensed reflectance data at lake's edge are likely to be less accurate than the inner part of the lake because the pixel at the edge is a mixture of both land and lake water [Rodgers, 1990], which explains the lower correlations.

3.3 Assessing the relationship between sensors over wavelength

In this section the relationship between the sensors over wavelength is assessed by functional distance. Since this thesis aims to develop a fusion model to fuse the reflectance data with different temporal and spectral supports from two sensors, the relationship between sensors in the

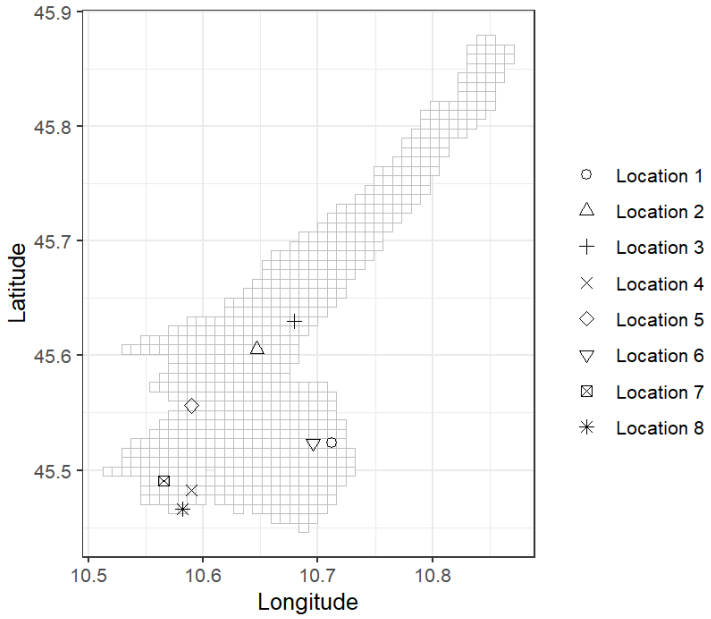


Figure 3.9: Scatter plot of the OLCI reflectance data against the MODIS reflectance data at four randomly picked locations for wavelength band pair (412, 412).

wavelength dimension is fundamental to apply this fusion model. The reflectance data between 412 nm and 681 nm are selected in this section, because these three sensors have a similar number of bands within this range, which is shown in Table 1.2. The functional distance between the MODIS-MERIS and the MODIS-OLCI sensors will be calculated for their overlapping years. This section will describe the details of converting the reflectance data to smooth function over wavelength and calculating functional distance.

Smoothing of reflectance data

The reflectance data in Lake Garda from the MERIS and MODIS sensors are selected as an example of smoothing reflectance data. Two major problems need to be addressed in the process of smoothing, which are the irregularly observed wavelengths from the two sensors and the smoothness of the functions. Since these smooth functions will be used to calculate functional distance in the following analysis, the same set of basis functions and smoothness are required. There are eight bands for the MERIS sensor and ten bands for the MODIS sensor within the 412 nm to 681 nm range. The MODIS reflectance data over wavelength are imputed using natural cubic interpolating splines, and the values are estimated at the same set of wavelengths as the MERIS data, which leads to eight observations over wavelength for both sensors. Then, ten B-splines basis functions with order four are used to smooth the reflectance data, because this is the maximum number of basis functions for eight observations within the range of wavelength. The smoothness of the functions is constrained by the smoothing parameter λ, which is selected by minimizing generalized cross-validation (GCV) with Equation (2.14). After calculating GCV for each of the smooth function from MERIS and MODIS, the smoothing parameter λ is chosen

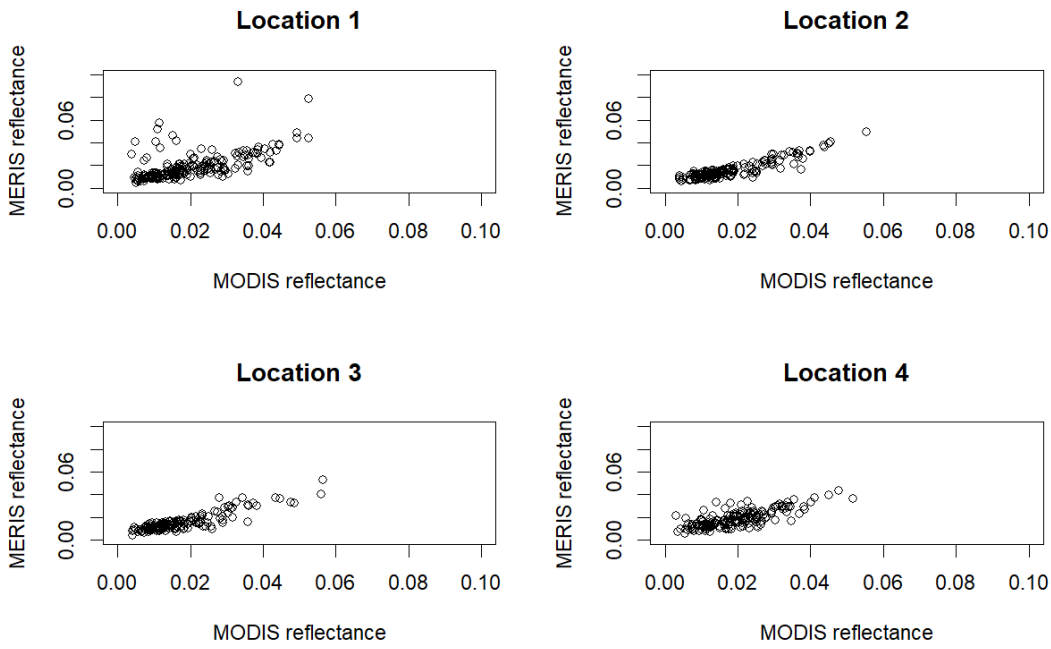


Figure 3.10: Scatter plot of the MERIS reflectance data against the MODIS reflectance data at four randomly picked locations for wavelength band pair (412, 412).

by minimising the mean of the GCV for all the fitted functions. Finally, the reflectance data from the MODIS and MERIS sensors can be converted to smooth functions with the selected number of basis functions and smoothing parameter through *fda* package [Ramsay et al., 2009].

Functional distance of reflectance data

After smoothing the reflectance data, the functional distance introduced in Section 2.5.1 can be calculated. The reflectance data observed by the MERIS and MODIS sensors near the centre of Lake Garda on 29/09/2009 are used to illustrate how to calculate the distance between two functional data objects. This date is selected because both the MERIS and MODIS sensors have observations at more than 80% of the pixels within Lake Garda. The functional data for the MODIS (blue line) and MERIS (black line) sensors are shown in Figure 3.15, and the functional distance is 0.018 for this location using Equation (2.15). The functional distance itself does not imply the relationship between the sensors over wavelength, but comparing the functional distance at all locations of the lake will indicate where the functional distance values are relative lower. A lower value of functional distance suggests the functions of reflectance over wavelength from the two sensors are closer, which represents a stronger relationship between the sensors over wavelength. Thus, the functional distance between MERIS and MODIS reflectance data are extended to all locations, which provides a functional distance map of Lake Garda in Figure 3.16. The high values of functional distance are located at the boundary of the lake, and this pattern could result from the different algorithms used by the two sensors to distinguish land

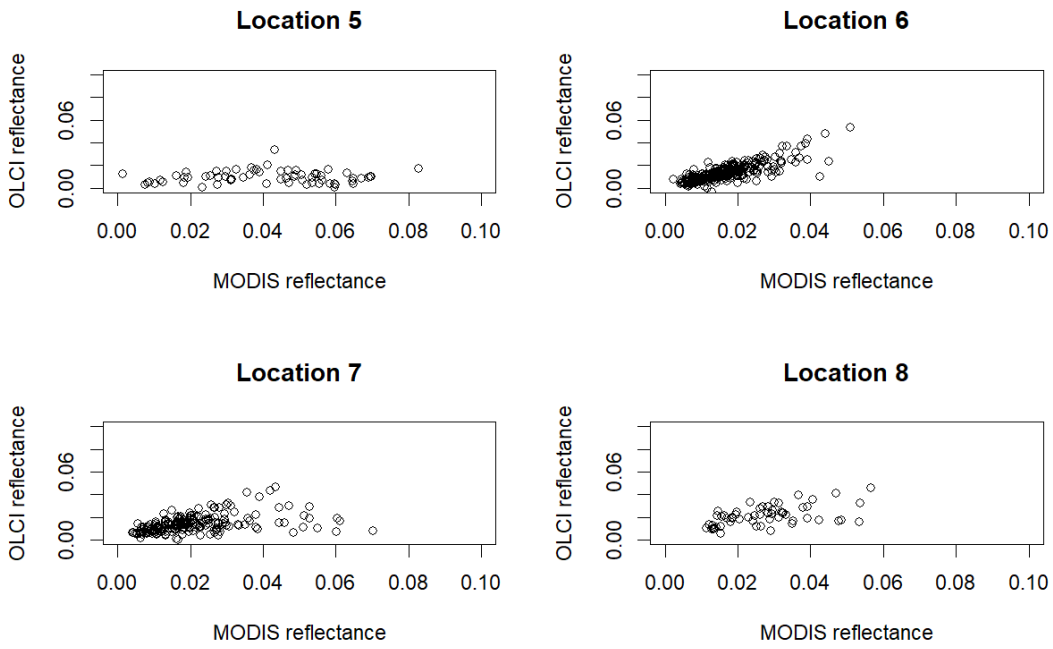


Figure 3.11: Scatter plot of the OLCI reflectance data against the MODIS reflectance data at four randomly picked locations for wavelength band pair (412, 412).

and lake water, which will affect the reflectance estimation near the boundary of Lake Garda. The values of functional distance are smooth over space, which suggests that nearby locations tend to have similar values to functional distance.

The functional distance is calculated between MODIS-MERIS and MODIS-OLCI sensors for their overlapping years to see how this spatial pattern changes over time. Since these sensors have different temporal supports as shown in Table 1.2 and they have different amount of missing values for their overlapping dates, it is hard to investigate the functional distance over the entire lake on a daily level. Thus, the reflectance data are aggregated by calculating the monthly average for each pixel so that there are more records over space at the monthly level. After calculating the monthly average reflectance, the functional distance for each month is calculated using the same method as above. The monthly functional distance map of Lake Garda in 2009 is shown in Figure 3.17. These plots show generally high values of functional distance near the edge of the lake and low values of functional distance for the inner part of the lake. The high values of functional distance are generally located near the edge of the lake since different algorithms used by the two sensors to distinguish land and lake water will affect the reflectance estimation near the boundary. Similar analyses are applied to the reflectance data in Lake Garda from the OLCI and MODIS sensors, which is shown in Figure 3.18. The functional distance maps between these two sensors also show higher values near the edge of the lake than the inner part of the lake. The functional distance maps in Figure 3.17 and Figure 3.18 are shown in this section for illustration, and functional distance maps for the rest of the months between these

sensors also show similar spatial patterns.

In conclusion, the function distance maps show high values near the edge of Lake Garda and low values for the inner part of the lake for MODIS-MERIS and MODIS-OLCI pairs of sensors for their overlapping years. This result suggests that the reflectance data over wavelength are more similar for the inner part of the lake. In Section 3.2, it is shown that the inner pixels have relatively higher Pearson's correlation coefficients. Considering the relationship between sensors over time and wavelength, it is more appropriate to apply the data fusion over time and wavelength for these locations compared to those near the edge of the lake. When modelling the lake water remote sensing data, the edge pixels could be removed to avoid the land masking problem. For example, the edge pixels are removed to analyse the remote sensing chlorophyll-a data at Lake Taruo [Gong et al., 2021a]. Thus, the edge pixels defined in Figure 3.14 will be removed in this thesis.

3.4 Identifying common temporal patterns in reflectance data

In this section, the temporal patterns in reflectance data are explored to check if there are common temporal patterns over space. This property will indicate how to represent the reflectance data in the data fusion models in the following chapters. Since reflectance data in this chapter have a large amount of missingness, the principal component analysis can not be applied to this dataset. Thus, reflectance data are first converted to smooth functions over time, and Functional data analysis (FPCA) is applied to these functions to explore the common temporal patterns. In the previous two sections, the results show that the locations at the edge of the lake have lower values of Pearson's correlation coefficient and larger values of functional distance compared to the locations at the inner part of the lake. Thus, the boundary with two pixels width will be removed for the analysis in this chapter.

Take the MERIS reflectance data for Lake Garda at 412 nm as an example, the reflectance data are converted to smooth functions over time using a similar method described above. A common smoothing parameter is used for all locations to ensure a fair analysis among these smooth functions. Thus, the number of basis functions is set to 171, which is the largest possible number of basis functions considering the number of observations over time for all locations within Lake Garda. Then, the smoothing parameter is selected by minimising the mean of GCV for these functions, which is 316 for this example. After that, the MERIS reflectance data are converted to smooth functions for all locations in Lake Garda, which is shown in Figure 3.19. Then, FPCA is applied to these smooth functions using *fda* package [Ramsay et al., 2009]. The variances explained by the leading four functional principle components (FPCs) are summarized in Table 3.1, which explains 70.5% of the total variance. The eigenfunctions of the first two FPCs are

plotted in Figure 3.20. On each plot, the solid line corresponds to the mean of all functional data curves, and the line shown by '+' and '-' corresponds to adding and subtracting a small amount of each FPC, since the FPCs represent the variation around the mean [Ramsay et al., 2009]. The first eigenfunction approximately represents a constant vertical shift of the mean function, and the second eigenfunction function represents the variation in December 2009-March 2010 and September 2010-March 2011. The corresponding FPC score maps are plotted in Figure 3.21. The similar values of PC score represents a similar temporal pattern among the corresponding locations. The spatial correlation of PC scores in Figure 3.21 suggests that the nearby locations tend to have a similar temporal pattern. A similar analysis is applied to the MODIS and OLCI reflectance for Lake Garda at 412 nm. The maps of the FPC scores corresponding to the first two FPCs are shown in Figure 3.22 for MODIS and Figure 3.23 for OLCI, and the FPC scores also show a strong spatial correlation. Similar analyses are also applied to the other wavelength bands for these sensors, and the FPC scores corresponding to the leading principal components also show strong spatial correlations. The FPC score represents the strength of the corresponding temporal patterns at each location, so the similar values of PC scores represent similar temporal patterns. Since the nearby pixels tend to have similar temporal patterns for the reflectance data of Lake Garda from these sensors, it is possible to build spatial models over these smooth functions over time to fuse reflectance data. Thus, the assumption of the common temporal patterns over space holds for the Lake Garda dataset, and the NSD model can be applied to this dataset in the following chapters.

	PC1	PC2	PC3	PC4
variance proportions	42.3%	14.8%	7.7%	5.7%

Table 3.1: The variance explained by first four functional principle components of the MERIS reflectance data for Lake Garda at 412 nm.

3.5 Conclusion

In this chapter, statistical methods are applied to analyse remote-sensing reflectance data for Lake Garda from the MERIS, MODIS, and OLCI sensors to identify the relationship between the reflectance data from the sensors and understand the variations in the reflectance data.

The scatter plots and Pearson's correlation coefficients are used to assess the linear relationship between the sensors over time for the selected pairs of wavelength bands. The results show that the linear relationship is stronger for the locations of the inner part of the lake than the edge of the lake. The assumption of linear relationship broadly holds for the Lake Garda datasets, which suggests the fixed rank kriging model can be used to fuse the reflectance data from these sensors.

The relationship between the sensors over wavelength is explored by functional distance, and the function distance has larger values for the edge pixels than the inner pixels of the lake, and the values of function distance are close to 0 for the inner pixels. Thus, the reflectance data can be fused over wavelength between these sensors. These two analyses suggest that the data fusion method seems reasonable for combining the reflectance data from the sensors, and the edge pixels will be removed to avoid the land masking problem near the edge of the lake in this thesis.

Functional principal component analysis is used to analyse the temporal variations in reflectance data. The results show that reflectance data at nearby locations tend to have a similar temporal pattern. Since the smooth functions of reflectance data over time have similar patterns at nearby locations, it is possible to build a spatial model over these functions. The NSD model introduced in Chapter 1 uses this property to represent spatial-temporal data from two different sources, so this model will be used to fuse reflectance data from the sensors at one wavelength band in the next chapter.

The exploratory analysis in this chapter suggests that it is reasonable to consider data fusion between the MERIS/MODIS and MODIS/OLCI sensors for Lake Garda in this thesis. As discussed in Chapter 1, this thesis will first consider the spatial-temporal fusion model of reflectance data for a single wavelength band in Chapters 4 and 5. Then, this model will be developed for multiple wavelength bands in Chapter 6. The datasets used in these chapters are decided as follows. In Chapters 4 and 5, the MERIS and MODIS reflectance data for wavelength band 412 will be used to investigate the predictive performance of the data fusion models. Since the current Chlorophyll-a retrieval algorithms, such as the Maximum band ratio algorithm and first order differential model algorithm, are developed based on the MERIS reflectance data, fusing the reflectance data from the MERIS and MODIS sensors will provide comparative studies with the existing research [Liu et al., 2021]. Additionally, including reflectance data at band 412 will make the estimation of Chlorophyll-a less sensitive to the uncertainty of the parameter estimation in the Maximum band ratio algorithm [O'Reilly and Werdell, 2019]. Thus, band 412 is selected as an example to investigate the data fusion models for a single wavelength band in Chapters 4 and 5. In Chapter 6, the reflectance data from the MODIS and OLCI sensors will be considered for the data fusion models with multiple wavelength bands. The first reason is that the MERIS sensor stopped its mission in 2012, and the OLCI sensor is the successor of the MERIS sensor [Kravitz et al., 2020]. Fusing the MODIS and OLCI sensors will be the focus of the future development of the data fusion techniques. The second reason is that the OLCI sensor has additional bands compared to the MERIS sensor, and including these bands will provide additional information to estimate the reflectance data over wavelength. Thus, the OLCI and MODIS reflectance data in Lake Garda are selected as an illustration to assess the performance of the data fusion model with multiple wavelength bands.

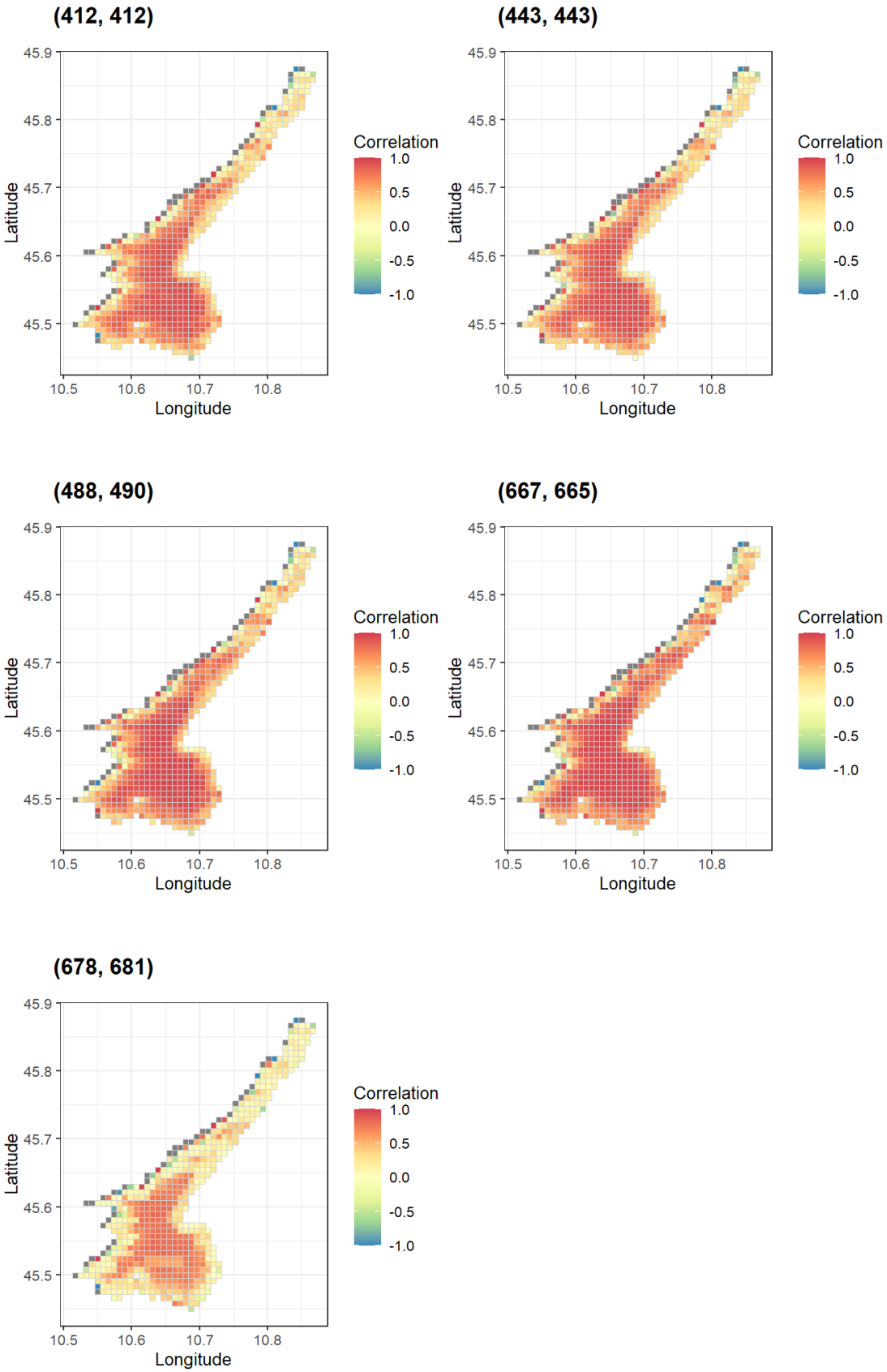


Figure 3.12: Correlation across time at all pixels in Lake Garda between the MODIS and MERIS sensors. The figure title represents the pair of wavelength bands (MODIS, MERIS).

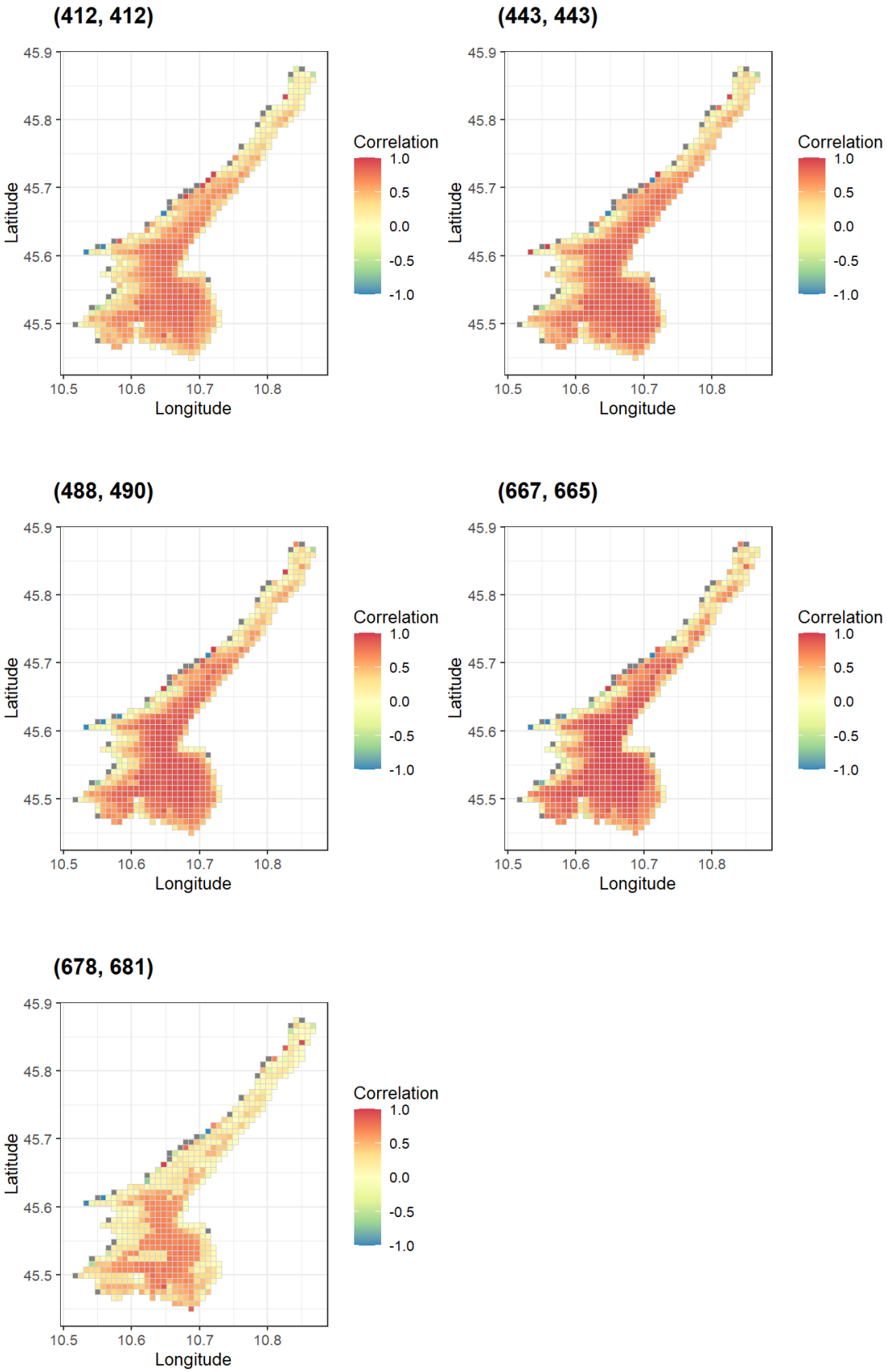


Figure 3.13: Correlation across time at all pixels in Lake Garda between the MODIS and OLCI sensors. The figure title represents the pair of wavelength bands (MODIS, OLCI).

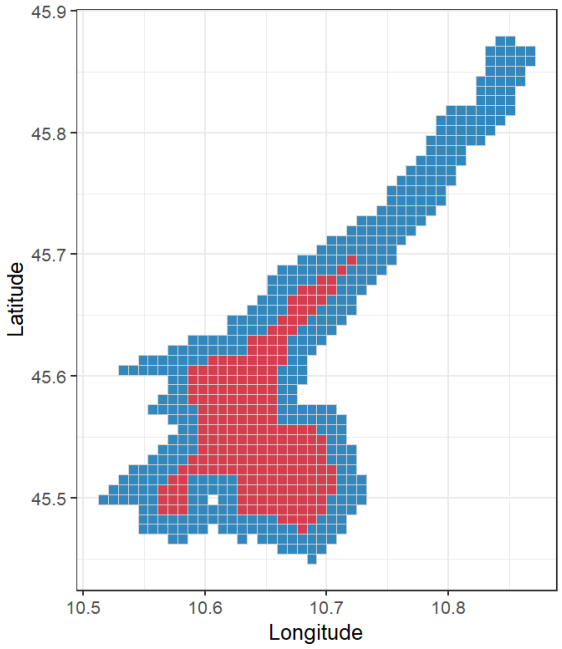


Figure 3.14: Plots of edge pixels and inner pixels for Lake Garda. Blue cells: edge pixels; red cells: inner pixels.

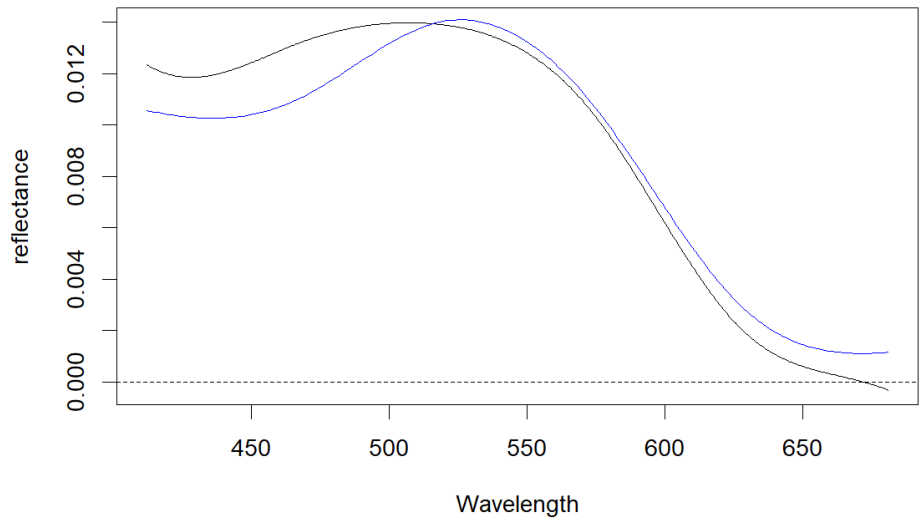


Figure 3.15: Plots of reflectance function near the centre of Lake Garda on 29/09/2009. Black lines: MERIS, blue lines: MODIS.

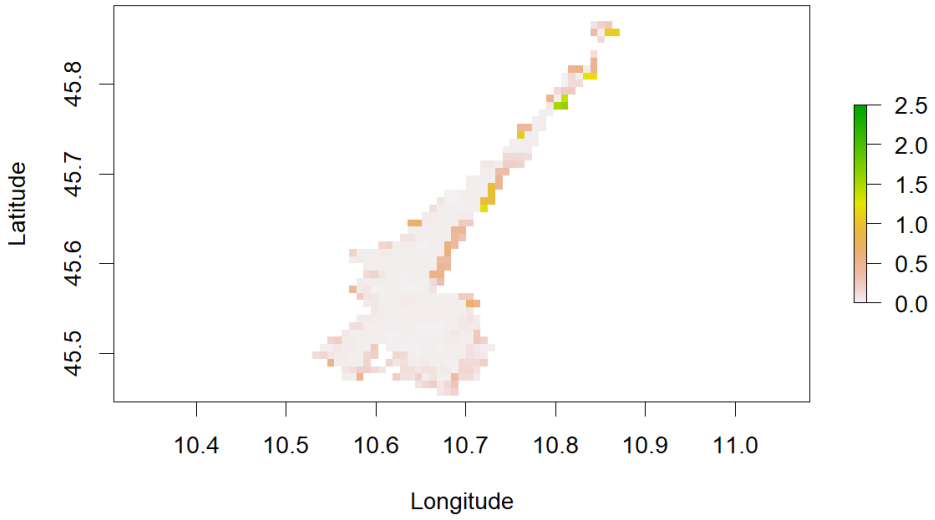


Figure 3.16: Map of functional distance between the MODIS and MERIS reflectance data of Lake Garda on 29/09/2009.

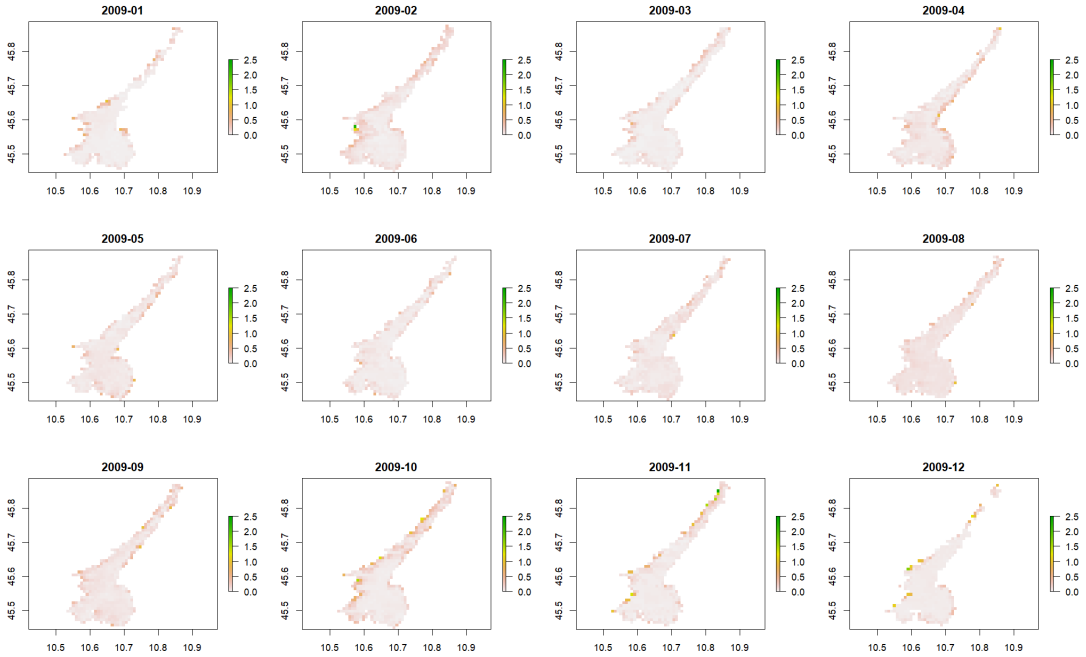


Figure 3.17: Functional distance maps between the MERIS and MODIS sensors for Lake Garda in 2009. The x-axis represents longitude and the y-axis represents latitude.

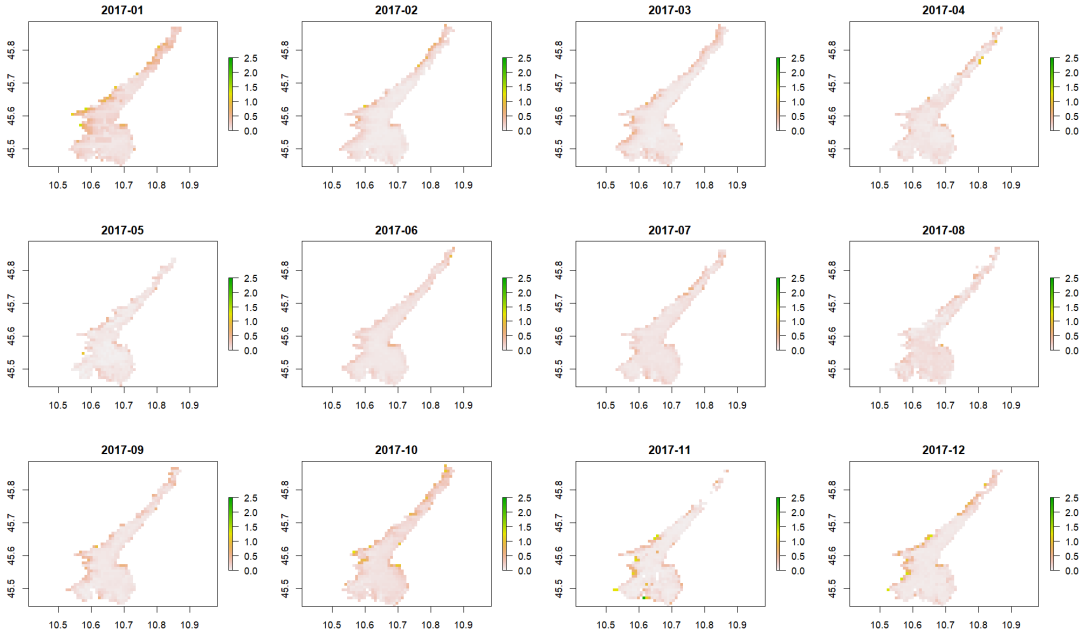


Figure 3.18: Functional distance maps between the OLCI and MODIS sensors for Lake Garda in 2017. The x-axis represents longitude and the y-axis represents latitude.

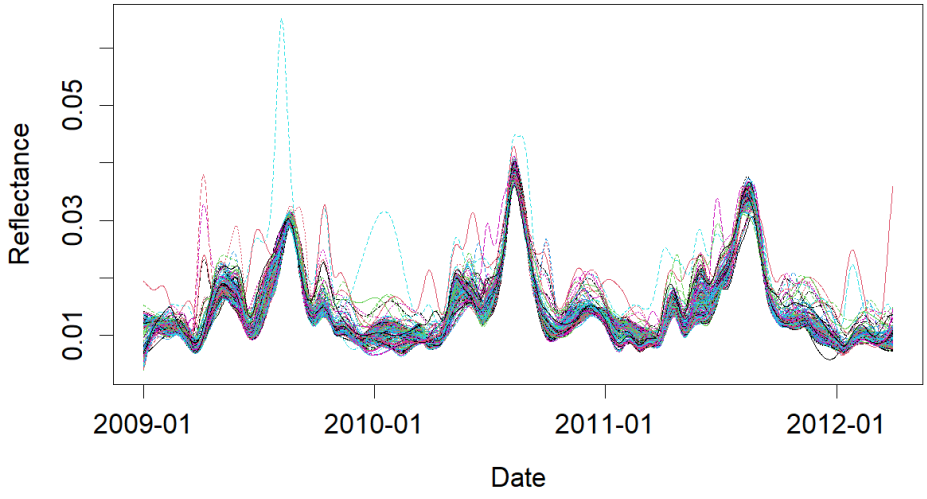


Figure 3.19: Smooth functions over time of the MERIS reflectance data for all locations in Lake Garda at 412 nm.

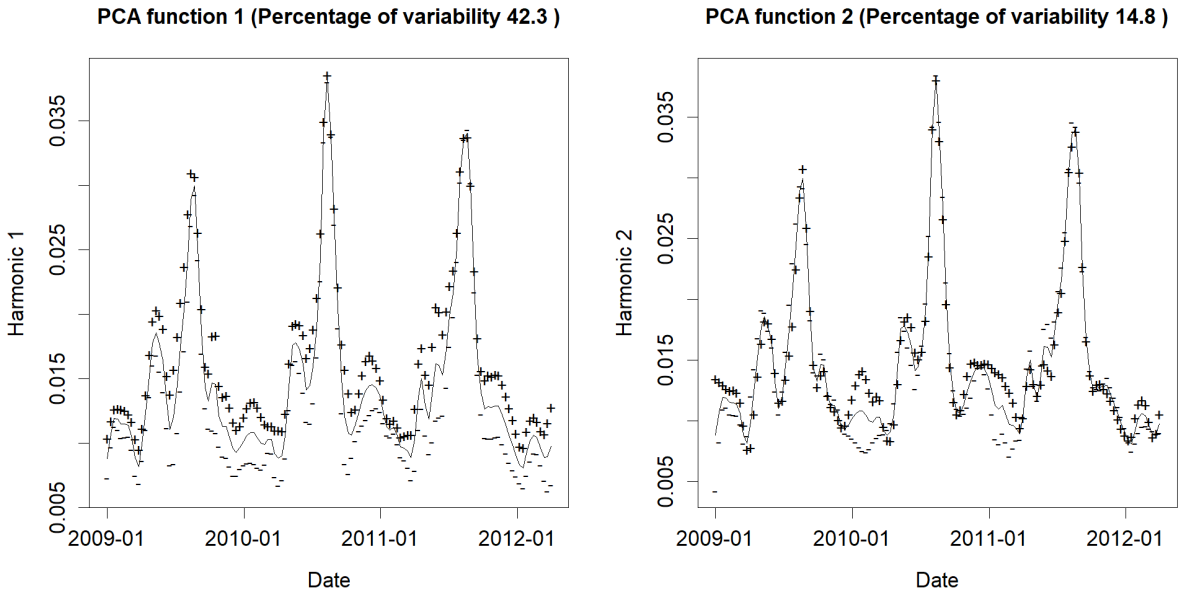


Figure 3.20: Eigenfucntions for the first two PC of the MERIS reflectance data for Lake Garda at 412 nm.

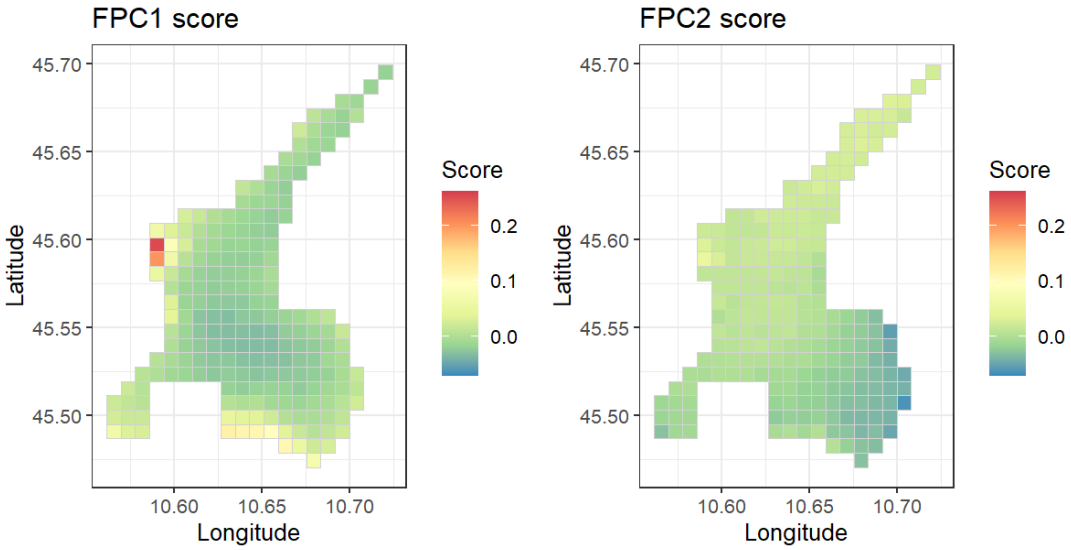


Figure 3.21: PC scores map for the first two PC of the MERIS reflectance data for Lake Garda at 412 nm.

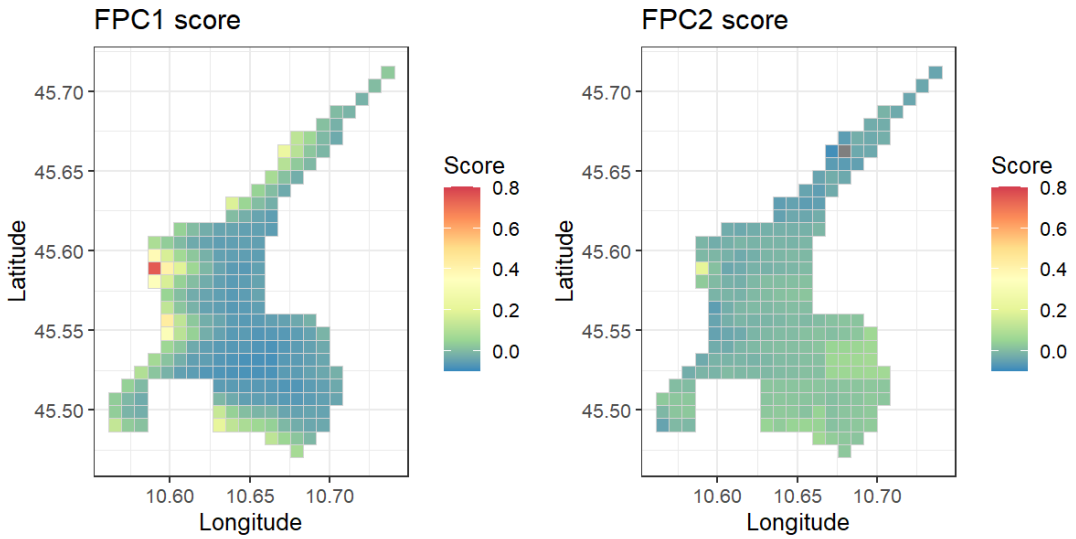


Figure 3.22: PC scores map for the first two PC of the MODIS reflectance data for Lake Garda at 412 nm. The percentage of variability is 73.8% for fPC1 and 4.4% for fPC2.

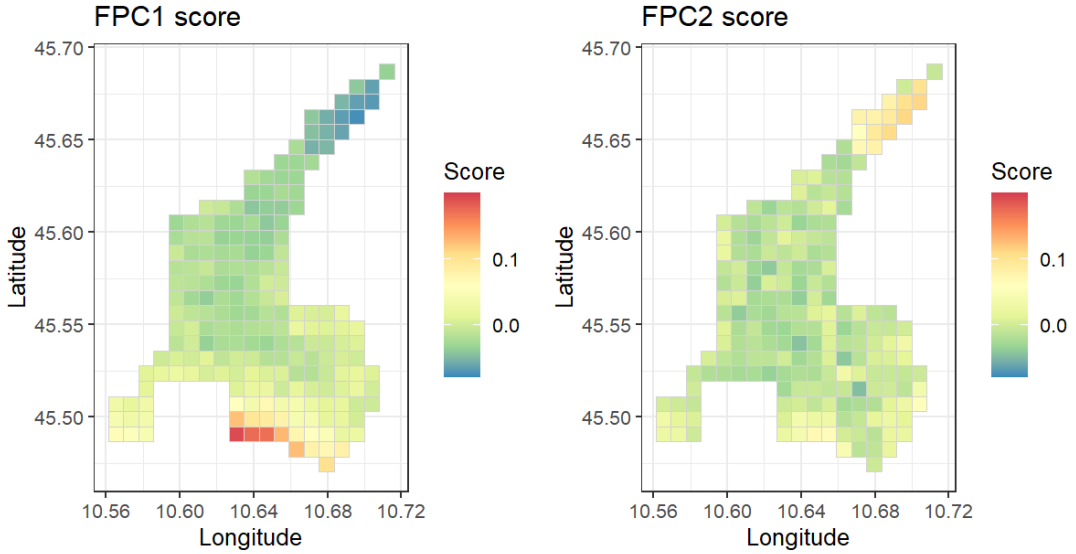


Figure 3.23: PC scores map for the first two PC of the OLCI reflectance data for Lake Garda at 412 nm. The percentage of variability is 27.3% for fPC1 and 14.7% for fPC2.

Chapter 4

Fuse lake reflectance data from different sensors by a downscaling model

4.1 Introduction

Remote sensing techniques have been widely used to monitor the inland water quality because of their better spatial coverage and temporal frequency compared to conventional monitoring approaches. The previous chapter shows that the reflectance data collected by Medium resolution imaging spectrometer (MERIS) and Moderate Resolution Imaging Spectroradiometer (MODIS) sensors are comparable near the centre of Lake Garda in the summer months. This leads us to consider fusing the reflectance data from different sensors.

The aim of this chapter is to apply an existing data fusion model to a lake water reflectance dataset and assess how accurately one can predict MERIS reflectance data at an unobserved location or time period by using the MODIS data at the same location. This assessment is based on the root mean squared error of the predictions, the coverage of 95% credible interval, the significance of the coefficients linking MODIS and MERIS and residual checking. This chapter will start with a literature review of data fusion methods and the details of a non-parametric downscaling model. Then, reflectance data at Lake Garda will be introduced and explored, which is the case study in this chapter. After that, the non-parametric downscaling model will be used to combine reflectance data from different sensors for Lake Garda. The accuracy of the downscaling model will be assessed by its root mean squared error (RMSE) and the coverage of 95% credible interval. Finally, the limitations and development of this model will be discussed.

4.2 Literature review of data fusion methods

In this section, the literature on data fusion methods will be summarised. Data fusion can be defined as a process of combining information from multiple data sources to achieve a refined

estimation of physical phenomena [Hall and Llinas, 2001]. Since the improving sensor technologies provide a great number of available data, data fusion methods become increasingly important in remote sensing. There are a few attempts to combine the remote sensing data from different sensors. Wynne et al. (2021) used linear regression to model the OLCI or MERIS cyanobacterial index data in Lake Erie (which are the products of the reflectance data by algorithms) as a constant times the MODIS cyanobacterial, which will provide a 20-year cyanobacterial bloom record [Wynne et al., 2021]. Zeng and Binding (2021) used the MODIS, MERIS and OLCI reflectance data as the input of a neural network model to produce algal bloom conditions from 2002 to 2020 [Zeng and Binding, 2021]. Although these papers focus on different lake water quality parameters, they are the products from the first-level reflectance data from these three sensors, and their data fusion methods generally model the MERIS/OLCI data as a function of MODIS data, and these methods aim to solve two difficulties, the potential large size of the dataset and different spatial or temporal supports of remote sensing instruments [Nguyen et al., 2012]. In what follows, the data fusion methods will be introduced to overcome these two problems.

The change of spatial support (the datasets have different spatial resolutions) problem could be solved by estimating the spatial data at one resolution based on the data at another resolution [Nguyen et al., 2012]. The following two data fusion methods focus on the change of support problem. Fuentes and Raftery (2005) propose a Bayesian melding model to combine the SO_2 concentration observations and the numerical outputs from a regional scale air quality model, which assumed the observations and the numerical outputs were driven by the same underlying process so that the joint distribution of them follows a Gaussian process [Fuentes and Raftery, 2005]. Berrocal et al. (2010) use a downscaling model to combine point-level and areal-level ozone concentration data, which regress the observed ozone data on the numerical output, and the coefficients in this function could vary across time [Berrocal et al., 2010b]. However, these two methods have computational cubic complexities with respect to their data size [Belyaev et al., 2014], which makes them not suitable for massive datasets.

Inference for spatial models for large data sets often requires MCMC methods, which involve matrix decompositions. Thus, dimension reduction methods are required as the computational time increases as cubic in the number of locations in the MCMC simulation [Banerjee et al., 2008]. Banerjee et al. (2008) use Gaussian predictive process models to approximate the predictions using a small set of locations to reduce the dimension of the data [Banerjee et al., 2008]. Cressie and Johannesson (2008) use a fixed rank kriging approach, where the non-stationary covariance function is constructed using a small set of basis functions [Cressie and Johannesson, 2008]. Nguyen et al. (2014) propose a spatial-temporal data fusion model based on this fixed rank kriging model, where the underlying true spatiotemporal processes are modelled at the

areal level [Nguyen et al., 2014]. In summary, the model proposed by Nguyen and Berrocal can fuse data from different spatial supports but requires the data to have the same temporal support, which could be difficult for remotely sensed reflectance data since the sensors have different recycle period. The data fusion method used to fuse the lake water reflectance data should tackle both the different supports and the potential large size of the dataset, since the remote-sensing data are daily with spatial coverage of the whole lake, and the observations from the sensors have different spatial and temporal resolutions.

Wilkie et al. (2019) use a nonparametric statistical downscaling model to fuse the log(chlorophyll-a) data from the in-situ source and satellite sensor [Wilkie et al., 2019]. This model solves the change of spatial and temporal supports problem, and allows the prediction of in-situ data at an unobserved location or time. This model reduces the dimensionality of the data by representing the reflectance data over time as combinations of known basis functions. The spatially varying basis coefficients for each data source are linked through a regression model, which allows data fusion from different sources. This model is fitted using MCMC, which will provide credible intervals for the estimated parameters and predictions from the model. This downscaling model could also be applied to the dataset with missingness in both space and time, and it could be processed with an existing R-package. Thus, this downscaling model is selected to fuse lake reflectance data, and the details of this model will be introduced in the following section.

4.3 Method

This section will introduce the nonparametric statistical downscaling (NSD) model proposed by Wilkie et al. [Wilkie et al., 2019] and the fixed rank kriging (FRK) model proposed by Cressie [Cressie and Johannesson, 2008].

4.3.1 NSD model

Model definition

Suppose $\mathbf{y}_i$ is the MERIS reflectance data vector at pixel i ($i = 1, 2, ..., n$) for times 1 to q_i and $\mathbf{y}_i = (y_{i1}, ..., y_{iq_i})^T$. The vector $\mathbf{y}_i$ is assumed to follow a multivariate normal distribution such that:

$$\mathbf{y}_i | \mathbf{c}_i, p_y \sim \mathrm{N}_{q_i}\left(\mathbf{\Phi}_i \mathbf{c}_i,\ \frac{1}{p_y}\mathbf{I}_{q_i}\right), \tag{4.1}$$

where p_y is the precision parameter, $\mathbf{I}_{q_i}$ is a $(q_i \times q_i)$ identity matrix, $\mathbf{\Phi}_i$ is the $(q_i \times m)$ matrix of m basis functions (Fourier basis function or B-splines basis function in this chapter) evaluated at times 1 to q_i, and $\mathbf{c}_i$ is the m dimensional vector of corresponding coefficients. Similarly the

distribution of MODIS reflectance vector $\mathbf{x}_i = (x_{i1}, ..., x_{ip_i})^T$ can be written as:

$$\mathbf{x}_i|\mathbf{d}_i, p_x \sim \mathrm{N}_{p_i}\left(\mathbf{\Psi}_i\mathbf{d}_i,\ \frac{1}{p_x}\mathbf{I}_{p_i}\right), \tag{4.2}$$

where p_x is the precision matrix, $\mathbf{I}_{p_i}$ is a $(p_i \times p_i)$ identity matrix, $\mathbf{\Psi}_i$ is the $(p_i \times m)$ matrix of m basis functions,and $\mathbf{d}_i$ is the m dimensional coefficients vector. The coefficients vector $\mathbf{c}_i$ and $\mathbf{d}_i$ are related by a regression model with spatially-varying coefficients such that:

$$\begin{aligned}
c_{ij}|\alpha_{ij}, \beta_{ij}, d_{ij}, p_c &\sim \mathrm{N}\left(\alpha_{ij} + \beta_{ij}d_{ij},\ \frac{1}{p_c}\right),\\
\boldsymbol{\alpha}_j|p_\alpha &\sim \mathrm{N}_n\left(\mathbf{0},\ \frac{1}{p_\alpha}\exp(-\phi_\alpha\mathbf{D})\right),\\
\boldsymbol{\beta}_j|p_\beta &\sim \mathrm{N}_n\left(\mathbf{1},\ \frac{1}{p_\beta}\exp(-\phi_\beta\mathbf{D})\right),
\end{aligned}$$

where $\mathbf{D}$ is the distance matrix between all pairs of the n locations of the MERIS data. The coordinates of the reflectance data are recorded in longitude and latitude, which means the unit of this distance matrix is degree. Since this downscaling model will only use the MODIS data at the locations where the MERIS data are also available, which means the MERIS and MODIS data are observed at the same location set. Thus, the distance matrix $\mathbf{D}_{\text{data}}$ is the same for the MERIS and the MODIS data. The prior distribution for these parameters above can be written as follows:

$$\begin{aligned}
p_y &\sim \mathrm{Ga}(a_y,\ b_y),\\
p_x &\sim \mathrm{Ga}(a_x,\ b_x),\\
\mathbf{d}_i &\sim \mathrm{N}_m(\boldsymbol{\mu}_d,\ \mathbf{\Sigma}_d),\\
p_\alpha &\sim \mathrm{Ga}(a_\alpha,\ b_\alpha),\\
p_\beta &\sim \mathrm{Ga}(a_\beta,\ b_\beta),\\
p_c &\sim \mathrm{Ga}(a_c,\ b_c).
\end{aligned}$$

In the following work, $\boldsymbol{\mu}_d$ is a length-m zero vector and $\mathbf{\Sigma}_d = 100 * \mathbf{I}_m$ where $\mathbf{I}_m$ is a $(m \times m)$ identity matrix to yield a weakly informative prior, since there is no additional information about the signs and the covariance structure of $\mathbf{d}_i$.The weakly informative prior is used for all these precision parameters. Take p_y as an example. The gamma prior for the precision p_y is equivalent to the inverse gamma prior for the variance σ_y^2, which means:

$$p_y \sim \mathrm{Gamma}(\mathrm{shape} = a_y, \mathrm{rate} = b_y) \Leftrightarrow \sigma_y^2 \sim \mathrm{inverse-Gamma}(\mathrm{shape} = a_y, \mathrm{scale} = b_y),$$

where a_y is the shape parameter and b_y is the scale parameter for the inverse-gamma distribution. Sahu (2006) suggests to use $\sigma_y^2 \sim \text{inverse} - \text{Gamma}(2,1)$, which has mean of 1 and a variance of infinity [Sahu et al., 2006].

Parameter estimation

The parameters of the NSD model are estimated by Gibbs sampling, and the full conditional posterior distributions are described as follows. Since the conjugate prior of p_α is the Gamma distribution, the posterior distribution of p_α can be written as follows:

$$f(p_\alpha|\cdot) \sim \text{Gamma}\left(a_\alpha + \frac{mn}{2},\ b_\alpha + \frac{1}{2}\text{tr}(\exp(-\phi_\alpha)^{-1}\boldsymbol{\alpha}^T\boldsymbol{\alpha})\right). \tag{4.3}$$

The conditional posteriors of p_β, p_y, p_x, p_c have similar forms as p_α, and they are shown as follows:

$$\begin{aligned}
f(p_\beta|\cdot) &\sim \text{Gamma}\left(a_\beta + \frac{mn}{2},\ b_\beta + \frac{1}{2}\text{tr}(\exp(-\phi_\beta)^{-1}(\boldsymbol{\beta}-\mathbf{1})^T(\boldsymbol{\beta}-\mathbf{1}))\right);\\
f(p_y|\cdot) &\sim \text{Gamma}\left(a_y + \sum_{i=1}^{n}\frac{q_i}{2},\ b_y + \frac{1}{2}\sum_{i=1}^{n}(\mathbf{y}_i - \boldsymbol{\Phi}_i\mathbf{c}_i)^T(\mathbf{y}_i - \boldsymbol{\Phi}_i\mathbf{c}_i)\right);\\
f(p_x|\cdot) &\sim \text{Gamma}\left(a_x + \sum_{i=1}^{n}\frac{p_i}{2},\ b_x + \frac{1}{2}\sum_{i=1}^{n}(\mathbf{x}_i - \boldsymbol{\Psi}_i\mathbf{d}_i)^T(\mathbf{x}_i - \boldsymbol{\Psi}_i\mathbf{d}_i)\right);\\
f(p_c|\cdot) &\sim \text{Gamma}\left(a_c + \frac{mn}{2},\ b_c + \frac{1}{2}\text{tr}(\mathbf{I}_n(\mathbf{c} - (\boldsymbol{\alpha} + \boldsymbol{\beta}\odot\mathbf{d}))^T\mathbf{I}_m(\mathbf{c} - (\boldsymbol{\alpha} + \boldsymbol{\beta}\odot\mathbf{d})))\right).
\end{aligned}$$

Since the conjugate prior of $\boldsymbol{\alpha}_j$ is a normal distribution, then the posterior distribution of $\boldsymbol{\alpha}_j$ is defined as:

$$\begin{aligned}
&f(\boldsymbol{\alpha}_j|\cdot) \sim \text{N}(\boldsymbol{\Sigma}_{\alpha_j}\mathbf{A}_{\alpha_j},\ \boldsymbol{\Sigma}_{\alpha_j}),\\
&\boldsymbol{\Sigma}_{\alpha_j} = (p_\alpha(\exp(-\phi_\alpha\mathbf{D}_{\text{data}}))^{-1} + p_c(\mathbf{I}_n)^{-1})^{-1},\\
&\mathbf{A}_{\alpha_j} = p_c(\mathbf{c}_j - \boldsymbol{\beta}_j\odot\mathbf{d}_j).
\end{aligned}$$

Then, the posterior of $\boldsymbol{\beta}_j$, $\boldsymbol{c}_j$ and $\boldsymbol{d}_j$ can be defined using the same method as follows:

$$\begin{aligned}
&f(\boldsymbol{\beta}_j|\cdot) \sim \text{N}(\boldsymbol{\Sigma}_{\beta_j}\mathbf{A}_{\beta_j},\ \boldsymbol{\Sigma}_{\beta_j}),\\
&\boldsymbol{\Sigma}_{\beta_j} = (p_\beta(\exp(-\phi_\beta\mathbf{D}_{\text{data}}))^{-1} + \mathbf{G}_j^\top\mathbf{G}_j)^{-1},\\
&\mathbf{A}_{\beta_j} = p_\beta\exp(-\phi_\beta\mathbf{D}_{\text{data}})^{-1}\mathbf{1} + p_c\mathbf{G}_j^\top(\mathbf{c}_j - \boldsymbol{\alpha}_j),
\end{aligned}$$

where $\mathbf{G}_j$ is a diagonal matrix with $\mathbf{d}_j$ as its diagonal.

$$
\begin{aligned}
&f(\mathbf{c}_i|\cdot) \sim \mathrm{N}(\boldsymbol{\Sigma}_{c_i}\mathbf{A}_{c_i},\ \boldsymbol{\Sigma}_{c_i}),\\
&\boldsymbol{\Sigma}_{c_i} = (p_y\boldsymbol{\Phi}_i^\top\boldsymbol{\Phi}_i + p_c(\mathbf{I}_m)^{-1})^{-1},\\
&\mathbf{A}_{c_j} = p_y\boldsymbol{\Phi}_i^\top\mathbf{y}_i + p_c(\boldsymbol{\alpha}_i + \boldsymbol{\beta}_i \odot \mathbf{d}_i);\\
&f(\mathbf{d}_i|\cdot) \sim \mathrm{N}(\boldsymbol{\Sigma}_{d_i}\mathbf{A}_{d_i},\ \boldsymbol{\Sigma}_{d_i}),\\
&\boldsymbol{\Sigma}_{d_i} = (\boldsymbol{\Sigma}_d^{-1} + p_x\boldsymbol{\Psi}_i^\top\boldsymbol{\Psi}_i + p_c\mathbf{F}_i^\top\mathbf{F}_i)^{-1},\\
&\mathbf{A}_{d_j} = \boldsymbol{\Sigma}_d^{-1}\boldsymbol{\mu}_d + p_x\boldsymbol{\Psi}_i^\top\mathbf{x}_i + p_c\mathbf{F}_i^\top(\mathbf{c}_i - \boldsymbol{\alpha}_i);
\end{aligned}
$$

where $\mathbf{F}_i$ is a diagonal matrix with $\boldsymbol{\beta}_i$ as its diagonal.

The spatial decay parameters ϕ_α and ϕ_β will be estimated by the cross-validation method, and the details of this process will be discussed in the Section 4.4.2.

To make predictions $\tilde{y_{ij}}$ at new times j ($j = 1,2,...,\tilde{q_i}$) and new locations i ($i = 1,2,...,\tilde{n}$), the posterior prediction distribution can be written as follows:

$$
\begin{aligned}
&\tilde{\mathbf{y}}_i|\tilde{\mathbf{c}}_i, \sigma_y^2 \sim N_{\tilde{q}_i}\left(\tilde{\boldsymbol{\Phi}}_i\tilde{\mathbf{c}}_i,\ \frac{1}{p_y}\mathbf{I}_{\tilde{q}_i}\right),\\
&\tilde{c}_{ij}|\tilde{\alpha}_{ij}, \tilde{\beta}_{ij}, \tilde{d}_{ij}, \sigma_c^2 \sim N\left(\tilde{\alpha}_{ij} + \tilde{\beta}_{ij}\tilde{d}_{ij},\ \frac{1}{p_c}\right),\\
&\tilde{\mathbf{d}}_i \sim N_m(\tilde{\boldsymbol{\Sigma}}_{d_i}\tilde{\mathbf{A}}_{d_i},\ \tilde{\Sigma}_{d_i}),\\
&\tilde{\boldsymbol{\Sigma}}_{d_i} = (\boldsymbol{\Sigma}_d^{-1} + p_x\tilde{\boldsymbol{\Psi}}_i^\top\tilde{\boldsymbol{\Psi}}_i)^{-1},\\
&\tilde{\mathbf{A}}_{d_i} = \boldsymbol{\Sigma}_d^{-1}\boldsymbol{\mu}_d + p_x\tilde{\boldsymbol{\Psi}}_i^\top\tilde{\mathbf{x}}_i.
\end{aligned}
$$

Suppose the matrix of distance between the MERIS training dataset and the MERIS prediction dataset can be written as:

$$
\mathbf{D} = \begin{pmatrix} \mathbf{D}_{\text{pred}} & \mathbf{D}_{\text{pred:data}} \\ \mathbf{D}_{\text{data:pred}} & \mathbf{D}_{\text{data}} \end{pmatrix},
$$

and the joint distribution of $\tilde{\boldsymbol{\alpha}}_j$ and $\boldsymbol{\alpha}_j$ follows a Gaussian process, then $\tilde{\boldsymbol{\alpha}}_j$ can be predicted by:

$$
\begin{aligned}
&\tilde{\boldsymbol{\alpha}}_j|\boldsymbol{\alpha}_j \sim \mathrm{N}\Bigg(\mathbf{0} + \exp(-\phi_\alpha\mathbf{D}_{\text{pred:data}})\exp(-\phi_\alpha\mathbf{D}_{\text{data}})^{-1}\boldsymbol{\alpha}_j,\\
&\frac{1}{p_\alpha}(\exp(-\phi_\alpha\mathbf{D}_{\text{pred}}) - \exp(-\phi_\alpha\mathbf{D}_{\text{pred:data}})\exp(-\phi_\alpha\mathbf{D}_{\text{data}})^{-1}\exp(-\phi_\alpha\mathbf{D}_{\text{data:pred}}))\Bigg).
\end{aligned}
$$

The conditional distribution of $\tilde{\boldsymbol{\beta}}_j|\boldsymbol{\beta}_j$ can be defined in the same way as above:

$$\tilde{\boldsymbol{\beta}}_j|\boldsymbol{\beta}_j \sim \mathrm{N}\Bigg(\mathbf{0}+\exp(-\phi_\beta \mathbf{D}_{\text{pred:data}})\exp(-\phi_\beta \mathbf{D}_{\text{data}})^{-1}\boldsymbol{\alpha}_j,$$
$$\frac{1}{p_\beta}(\exp(-\phi_\beta \mathbf{D}_{\text{pred}})-\exp(-\phi_\beta \mathbf{D}_{\text{pred:data}})\exp(-\phi_\beta \mathbf{D}_{\text{data}})^{-1}\exp(-\phi_\beta \mathbf{D}_{\text{data:pred}}))\Bigg).$$

Then the model could be fitted by Gibbs sampling, where each of the parameters is drawn from the full conditional distribution in turn, and this recycling will continue until all the parameters converge to the posterior distribution. The convergence of the Markov chain will be checked by the Geweke diagnostic introduced in Section 2.2.2.

4.3.2 FRK model

In the previous subsection, the NSD model is introduced. In order to compare the prediction ability of the NSD model with another spatio-temporal model, the Fixed rank kriging (FRK) model will be discussed. The FRK is based on a spatio-temporal random effect model, where the underlying process is represented by a set of pre-defined spatio-temporal basis functions. In the work of Zammit-Mangion et al. (2017), the FRK model has been compared to other models, such as the Lattice kriging model and Stochastic Partial Differential Equation-Gaussian Markov random fields (SPDE-GMRF) model, using both simulated and real data [Zammit-Mangion and Cressie, 2017]. The Lattice kriging model uses a set of basis functions to decompose the spatio-temporal process, but it uses a larger number of basis functions (>10000) than the FRK model [Nychka et al., 2016]. The SPDE-GMRF model also uses the spatio-temporal basis functions, but the resulting spatial random field approximates a Gaussian process with a Matern class covariance function [Lindgren and Rue, 2015]. The results show that the root mean-squared prediction error of the FRK model is similar to the other two models but the coverage of 90% credible interval of FRK are closer to 90% and larger than the other two models [Zammit-Mangion and Cressie, 2017]. Thus, the FRK model is selected in this chapter to compare to the prediction ability of the downscaling model. The FRK model and NSD model both use the basis functions to lower the dimensionality of the data. However, these two models have a couple of differences, which are summarised in Table 4.1. The major differences between these two models are their model type and the corresponding parameter estimation methods.

Difference	FRK model	NSD model
Method of inference	Frequentist inference	Baysian inference
Parameter estimation	Expectation-Maximization algorithm	Gibbs sampling
Basis functions	Tensor product of spatial and temporal basis	B-splines basis over time

Table 4.1: Differences of FRK and NSD model

Then, the FRK model can be defined as follows. Suppose $y(i,j)$ is MERIS reflectance data at

pixel $i(i = 1,2,...,n)$ for time $j(j = 1,2,...m)$, and $x(i,j)$ is MODIS reflectance data at pixel i for time j, the FRK model is:

$$y(i,j) = \alpha x(i,j) + \sum_{k=1}^{K} \theta_k(i,j)\eta_k + \xi(i,j), \tag{4.4}$$

where α is the regression coefficient of $x(i,j)$, $\theta_k(i,j)$ is the known spatial-temporal basis function evaluated at location i and time j. The $\theta_k(i,j)$ could be defined by the tensor product of a spatial basis function and a temporal basis function. An example of spatial basis functions is shown in Figure 4.1. In the left panel, the spatial basis functions with one resolution (spatial basis functions have the same radius) are used, which give 12 spatial basis functions in total to cover the area of Lake Garda. In the right panel, the spatial basis function with two resolution (spatial basis functions with two different radiuses) will give us 129 spatial basis functions. The temporal basis functions are shown in Figure 4.2, where there are 29 equally spaced exponential basis functions over the 39 months. The spatio-temporal basis functions are the tensor products of these two groups of functions. The choice of the number of basis functions in the FRK model will be discussed in Section 4.4.3. The $\xi(i,j)$ is assumed to be spatial and temporal uncorrelated and $\xi(i,j) \sim N(0,\sigma^2_\xi)$. The expectation maximization (EM) algorithm is used to estimate the parameters in the FRK model using the R package FRK [Zammit-Mangion and Cressie, 2017].

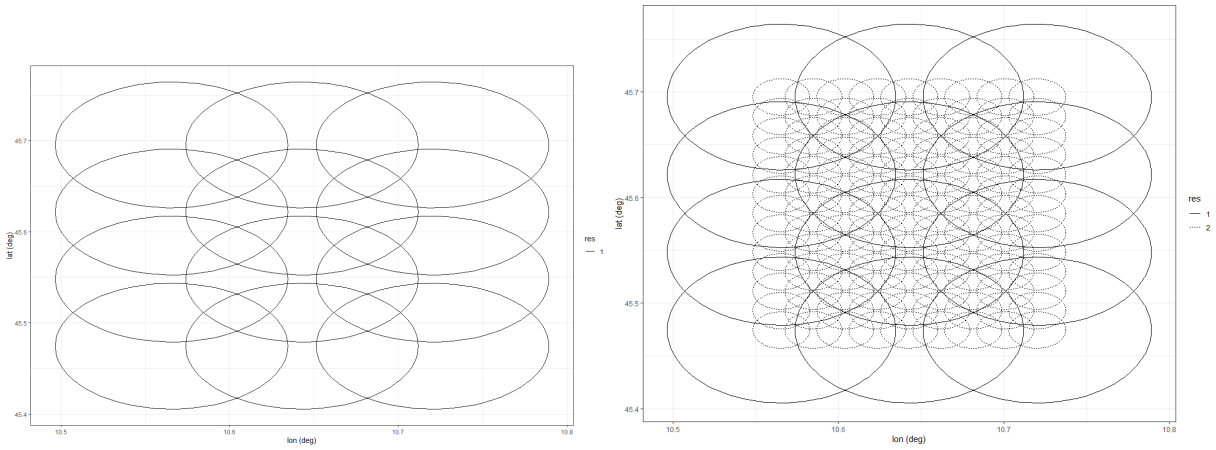


Figure 4.1: Spatial basis functions of FRK model with different resolution.

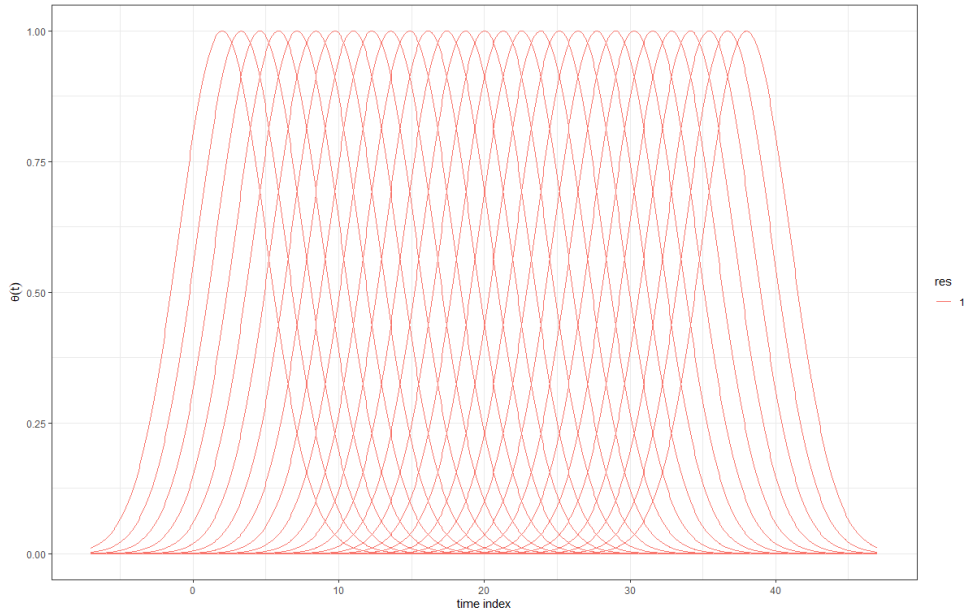


Figure 4.2: Temporal basis functions of FRK model.

4.3.3 Assess prediction ability of the model

After fitting each model, the prediction ability of each model is assessed by its root mean squared prediction error (RMSPE) and coverage of its 95% credible interval. The formulas for these two criteria are shown below:

$$\text{RMSPE} = \sqrt{\frac{\sum_{i=1}^{\tilde{n}} \sum_{j=1}^{\tilde{q}_i} (y_{ij} - \tilde{y_{ij}})^2}{\tilde{n} \times \tilde{q}_i}} \tag{4.5}$$

$$\text{coverage} = \frac{\sum_{i=1}^{\tilde{n}} \sum_{j=1}^{\tilde{q}_i} I(y_{ij} \in 95\%\text{credible interval})}{\tilde{n} \times \tilde{q}_i} \times 100\% \tag{4.6}$$

The model with a smaller value of RMSPE and a coverage closer to 95% will be preferred.

4.4 Results of NSD model on reflectance data set

In this section, the downscaling model described above will be used to fuse lake reflectance data from different sensors at Lake Garda. In what follows, the remotely-sensed reflectance data will be introduced and the initial exploratory analysis will be applied to Lake Garda data set. Then the downscaling model will be applied to this dataset and the model fitting will be assessed by the residuals and the coverage of 95% credible interval. Finally, the NSD model will be compared to another spatio-temporal model (FRK) based on their RMSPE of the prediction data set.

4.4.1 Lake Garda reflectance Data

Lake Garda is the largest lake in Italy with an area of 368.64 km^2, and it is important for its tourism, agriculture and drinking purposes [Giardino et al., 2014]. The reflectance data collected by the satellite sensors are useful in estimating the water quality parameters. The Lake Garda reflectance data observed by the MERIS, MODIS and OLCI sensors are introduced in Section 3.1. In this section, the reflectance data from the MERIS and MODIS sensors are selected as an illustration to compare the NSD model and the FRK model, because fusing the reflectance data from the MERIS and MODIS sensors will provide comparative studies with the existing research [Liu et al., 2021]. The overlapping period of MERIS and MODIS is 01/01/2009 - 31/03/2012 (1186 days). The wavelength 412 nm is selected, because the range of 400-500 nm is the blue light range and the reflectance data in this range to are used to calculate the chlorophyll-a concentration in the lake [Zeng and Binding, 2021]. The edge pixels are removed as discussed in Section 3.3, and there are 211 inner pixels left after removing the edge pixels. The reflectance data of the inner pixels at wavelength 412 nm are plotted over time in Figure 4.3 for MERIS and MODIS sensors, which show similar temporal patterns. For these remaining pixels, 21 pixels (approximately 10% of the inner pixels) are randomly selected as the test dataset and the remaining 190 pixels are treated as the training dataset to which the models will be fitted, and the locations of these prediction pixels are shown in Figure 4.4.

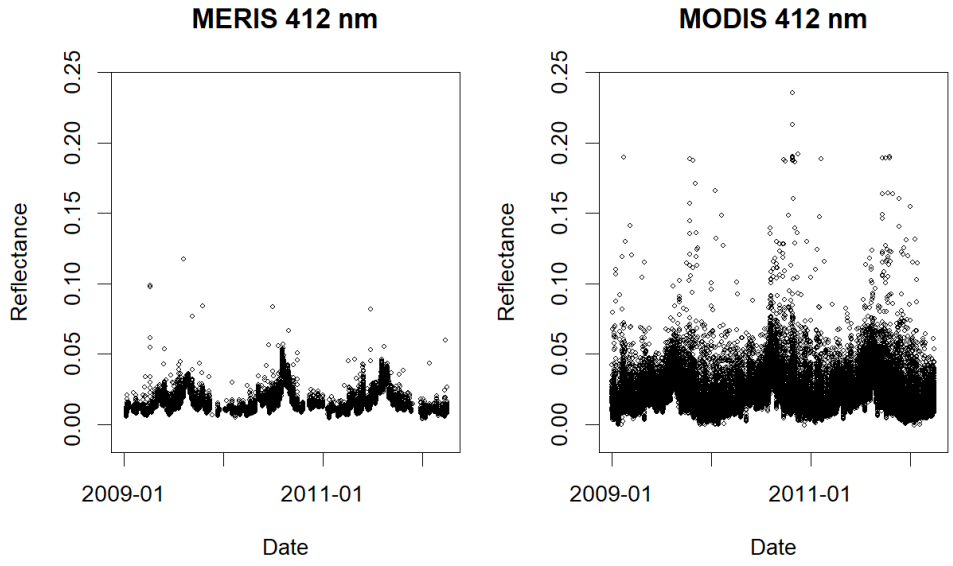


Figure 4.3: Reflectance data across the year for the inner pixels in Lake Garda.

4.4.2 Application of the NSD model on Lake Garda reflectance data

The previous section shows that the MERIS and MODIS reflectance data at Lake Garda are comparable after removing the border pixels. In this section, the NSD model is used to predict

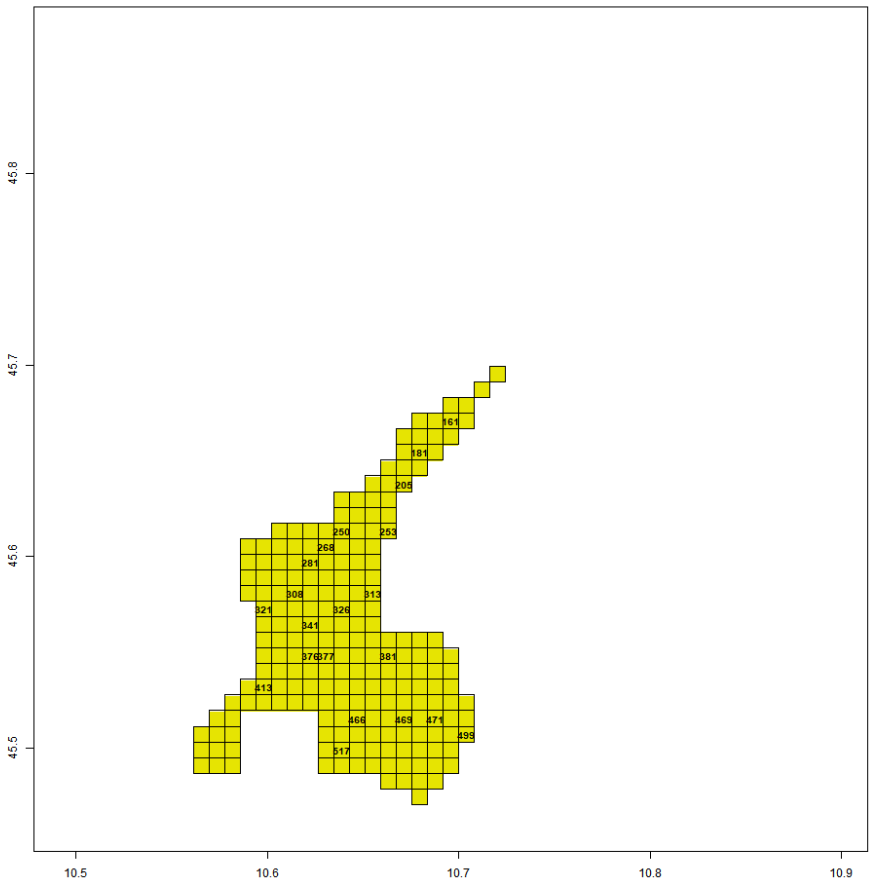


Figure 4.4: Pixels of Lake Garda after removing the 2-pixels-width border. The pixels with number labels are the prediction data set, and the rest of the pixels are the training data set.

the MERIS reflectance data based on the MODIS data. The out-of-sample prediction is used to assess the model performance on this dataset. The downscaling model is applied to the Lake Garda dataset using the MCMC simulation of 50,000 iterations with a burn-in of the first 5,000 iterations. The Markov chains are recorded in every 10 iteration for saving computer memory. It takes about 10 hours to run the downscaling model on this dataset.

Assess the convergence of Markov chains and residuals

To avoid redundancy, the convergence of the MCMC chains and a residual assessment of the NSD model with 29 B-splines function and spatial decay parameters equal to 0.001 are shown below (the selection of basis function and spatial decay parameters will be discussed in section 4.4.2). First, the trace plots of selected parameters are shown in Figure 4.5, which does not show strong evidence against the convergence of the Markov chains.

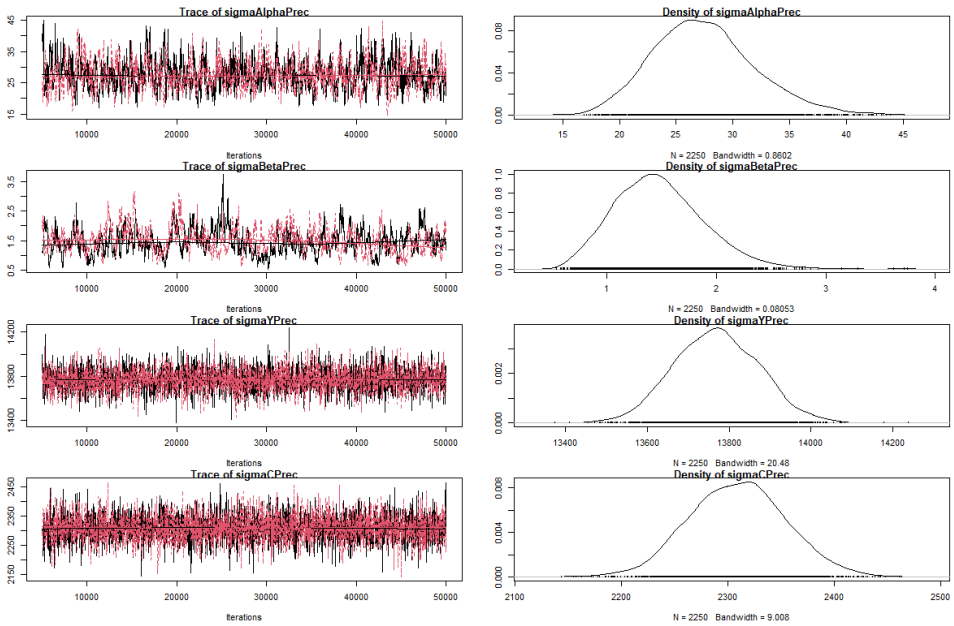


Figure 4.5: Trace and density plots for precision parameters p_α, p_β, p_y, p_c of the downscaling model.

Then the Geweke's diagnostic introduced in Section 2.2.2 is used to further assess the convergence of the Markov chains. The p-value of the Geweke diagnostic of p_α is 0.14, which suggests there is no evidence to reject the null hypothesis that the means of the first and the last part of the chain are equal, and it also suggests that the Markov chain converges. In what follows, the parameters in any downscaling model will be checked using the same diagnostics, although the p-value of the Geweke diagnostic will not be included to prevent repetition.

After checking the convergence of the Markov chains, the residuals of this model will be assessed to check the assumptions of normality and constant variance. The qq-plots of the residuals at 9 random locations are shown in Figure 4.6, and the points lies on a straight lines except for a couple of outliers, which shows the residuals may not quite be normally distributed at each locations. The outliers suggest that transformation could be considered to make the residuals closer to the normal distribution. The squared root transformation is applied to the reflectance data before the model fitting, and the resulting qq-plots are shown in Figure 4.7. There are still outliers outside the straight lines, which do not show strong improvements after transformation. Thus, the transformation of the reflectance data will not be used in this section.

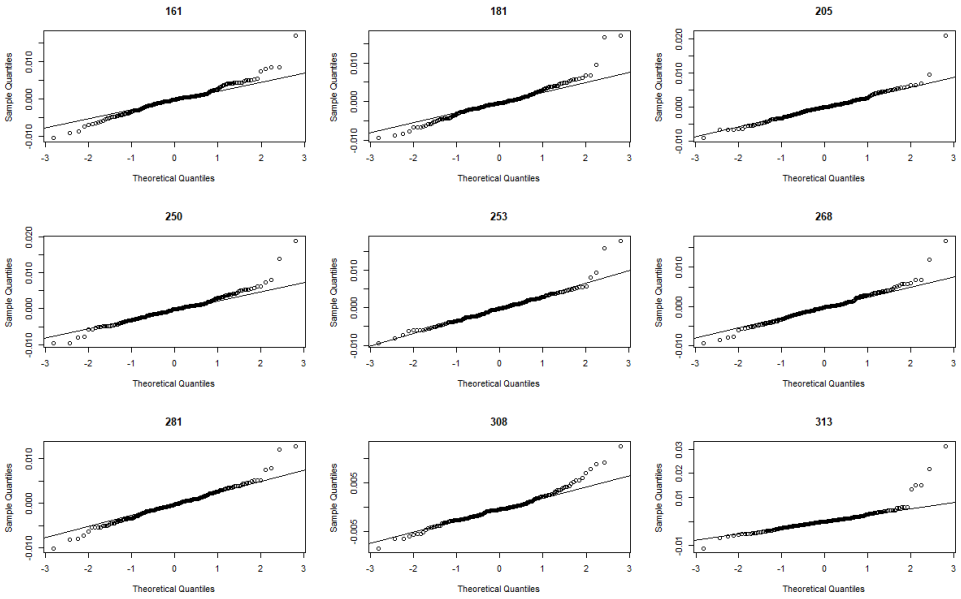


Figure 4.6: qq-plot of residuals for different locations.

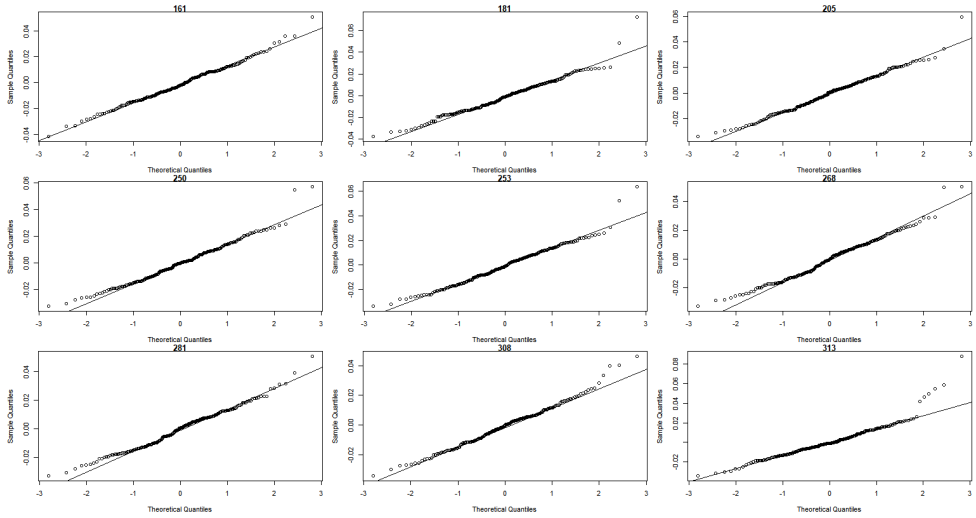


Figure 4.7: qq-plot of residuals for different locations with squared root transformation.

The residuals against fitted values plots for 9 randomly chosen locations in Figure 4.8 shows constant variance across the fitted values for each location. The residuals against time plots in Figure 4.9 suggest the variance of the residuals at each location are approximately the same, but there appear to be some small patterns across time as the absolute values of residual are high in summer months than winter months in each year. This pattern is not captured by the mean function in the downscaling model. However, these residual plots do not show strong evidence against the model assumptions.

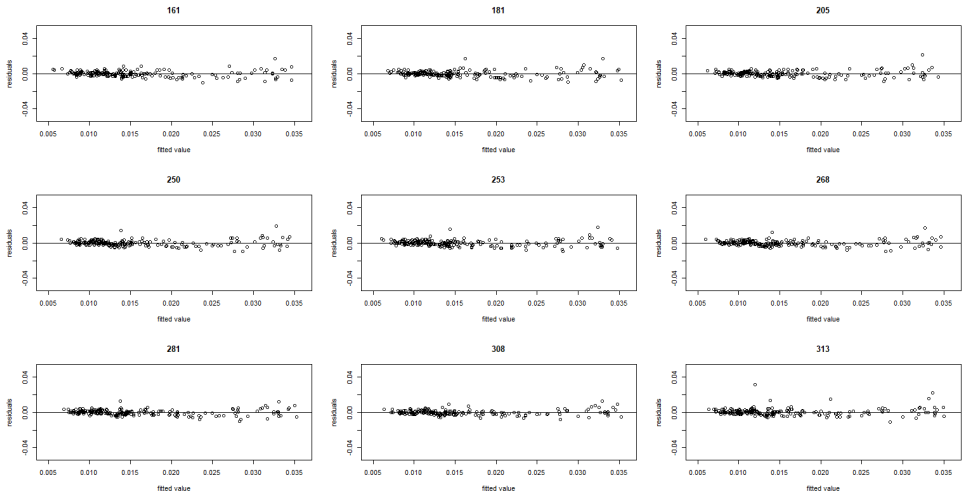


Figure 4.8: Residuals against fitted values for different locations.

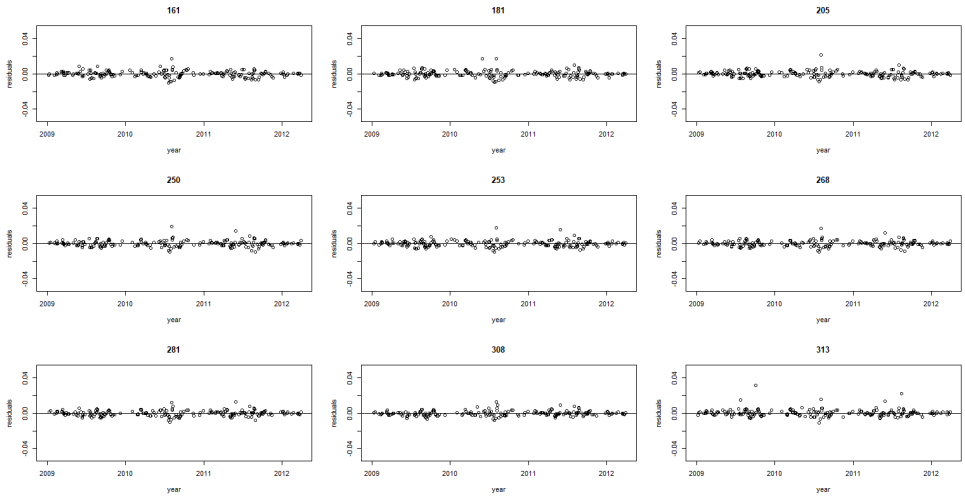


Figure 4.9: Residuals against year for different locations.

Select basis functions and spatial decay parameters

The type and number of the basis functions and the spatial decay parameters in the NSD model should be chosen before the model fitting. By the suggestion in Wilkie (2019), the number of basis functions could be estimated by the formula:

$$\text{Basis dimension} = (\text{total months} \times \frac{2 \times \text{r}}{12}) + 3, \tag{4.7}$$

where *r* is the expected number of peaks in each year [Wilkie et al., 2019]. If we assume there are $\{2,3,4,5\}$ peaks in the MERIS refelctance data in each year, it suggests using $\{16,23,29,36\}$ B-splines basis functions.

The number of the basis functions and the spatial decay parameters could be selected by cross

validation methods, which compare the RMSPE or other statistics of the resulting predictions. However, the computation time of the downscaling model on the Lake Garda dataset is about 10 hours with 29 B-splines basis functions, and this computation time will increase with the number of basis functions. Thus, the leave-one-out cross-validation method is not used to chose these hyperparameters in the NSD model. In this section, the different combinations of the basis function and the spatial decay parameters are used to fit the NSD model for the training data set (90% of the locations) of Lake Garda, and then out-of-sample predictions (10% of the locations) from the model are used to compute the RMSPE and coverage of the 95% credible interval, which are summarised in the Table 4.2. From this table, the coverage of the credible interval will decrease when the spatial decay parameters increase, and this suggests using smaller spatial decay parameters. The RMSPE changes only at four decimal places and the coverage changes at two decimal places at the bottom-right two by two matrix in this table, showing that the NSD model is relatively robust to the choices of these hyper-parameters. In what follows, the NSD model with 29 B-splines basis functions and spatial decay parameters equal to 0.001 will be used as it achieves the minimum RMSPE.

RMSPE/Coverage: Basis functions \ Spatial decay	0.1	0.01	0.001
16	0.0067/1	0.0059/0.99	0.0059/0.99
23	0.0054/1	0.0043/1	0.0043/0.99
29	0.0049/1	0.0036/1	0.0035/0.99
36	0.0054/1	0.0037/0.99	0.0036/0.99

Table 4.2: Result of RMSPE and for different choices of basis functions and spatial decay parameters.

Credible intervals of the coefficients

The credible intervals of the coefficients $\{\beta_{ij}\}$ and $\{\tilde{\beta}_{ij}\}$ will be checked to find if the NSD model could identify the relationship between the MERIS and the MODIS reflectance data. The credible intervals of 4 randomly picked coefficients are summarised in Table 4.3. This table shows that the 90% and 95% credible intervals of $\beta_{1,5}$ and $\tilde{\beta}_{1,15}$ all contain 0. However, the 95% credible interval of $\alpha_{1,5}$ and $\tilde{\alpha}_{1,15}$ do not contains 0. The credible intervals are also computed for other coefficients and the results show that all of the 90% and 95% credible intervals of $\beta_{i,j}$ and $\tilde{\beta}_{i,j}$ all contains 0. From Formula 1.4 and 1.5, both of $\beta_{i,j}$ and $\alpha_{i,j}$ vary over space, which will result in identifiable problems of $\beta_{i,j}$ and α_{ij}. For the predictive purpose, the identification of β_{ij} and α_{ij} will not be a big problem.

Coefficients \ Quantile(%)	2.5	5	50	95	97.5
$\alpha_{1,5}$	0.002	0.004	0.019	0.033	0.036
$\beta_{1,5}$	-0.31	-0.22	0.26	0.72	0.82
$\tilde{\beta}_{1,15}$	-0.41	-0.32	0.12	0.55	0.63
$\tilde{\alpha}_{1,15}$	0.010	0.016	0.047	0.079	0.085

Table 4.3: quantile of Markov chains for each coefficients.

Prediction of MERIS data

There are 21 randomly selected prediction pixels in Lake Garda, and their locations are shown in Figure 4.4. Based on the order of their labels, the first pixel is location 161 and the prediction of MERIS reflectance data at this location is shown in Figure 4.10. The prediction lines (black solid line) across the time capture the main pattern of the MERIS data (grey dots), and the credible intervals contain most of the MERIS data points. Predictions at four other locations are shown in Figure 4.11, where the prediction lines still capture the primary pattern across time and the credible intervals cover the most of the MERIS data points.

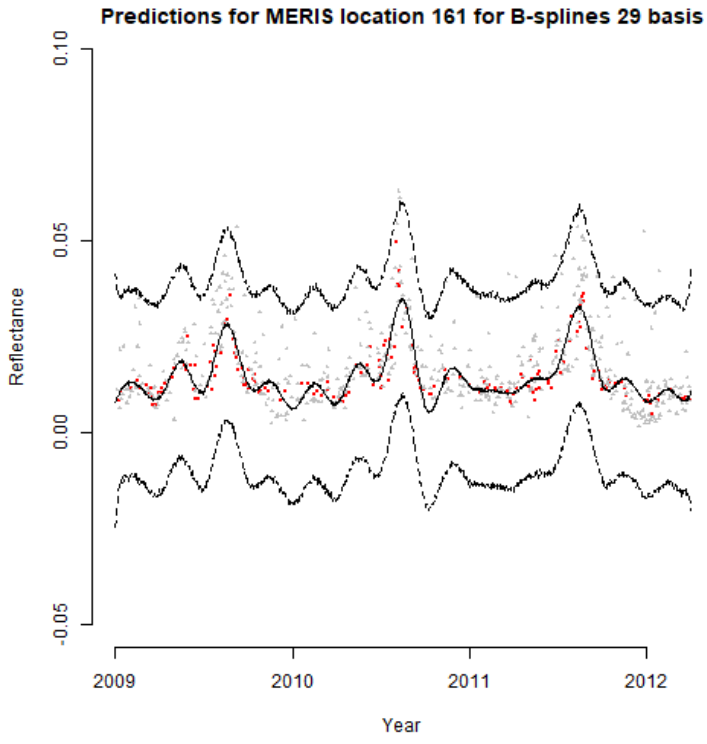


Figure 4.10: Predictions for MERIS reflectance at location 161 with 29 B-splines functions and spatial decay parameters equals to 0.001. Red points: MERIS data, grey points:MODIS data, black solid line in the middle: predictions of MERIS from downscaling model, black solid line in the top and bottom: 95% credible interval.

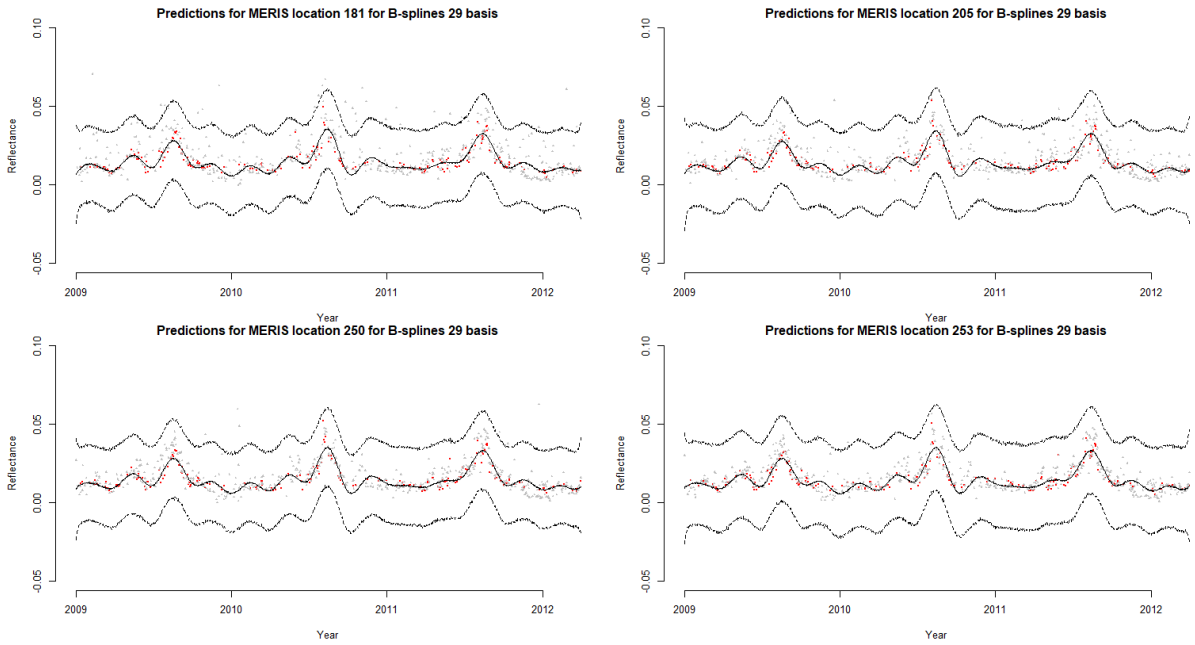


Figure 4.11: Predictions for MERIS reflectance at location 181, 205, 250 with 29 B-splines functions and spatial decay parameters equals to 0.001. Red points: MERIS data, grey points:MODIS data, black solid line in the middle: predictions of MERIS from downscaling model, black solid line in the top and bottom: 95 percent credible interval.

4.4.3 Comparison of Downscaling model and Fixed rank kriging model

In the previous subsection, the NSD model is applied to the Lake Garda reflectance dataset. In order to assess the prediction ability of the NSD model, the FRK model is used to make a comparison. In what follows, the data pre-processing, model fitting and the results from these two models will be discussed.

Data pre-processing

Since the R package that fits the FRK model (FRK package) requires the training dataset to have no missingness in space and time, gap filling must be applied to the reflectance data before model comparison. However, there are 50% and 82% missingness in the MERIS and MODIS training datasets respectively at the daily scale, which will make the gap filling problematic with this large amount of missingness. Thus, the daily reflectance data from MERIS and MODIS are first averaged over months. These monthly average reflectance data only have 2% missingness in the MERIS training dataset and no missingness in the MODIS dataset. Then, the missingness in the monthly averaged data are filled by interpolating the MERIS data by a natural cubic spline at each location. An example of gap filling for the monthly-averaged data at location 129 (the first location of the training dataset based on its label) is shown in Figure 4.12, where the gap filling value (black circle) is evaluated by the interpolation cubic line at January 2010.

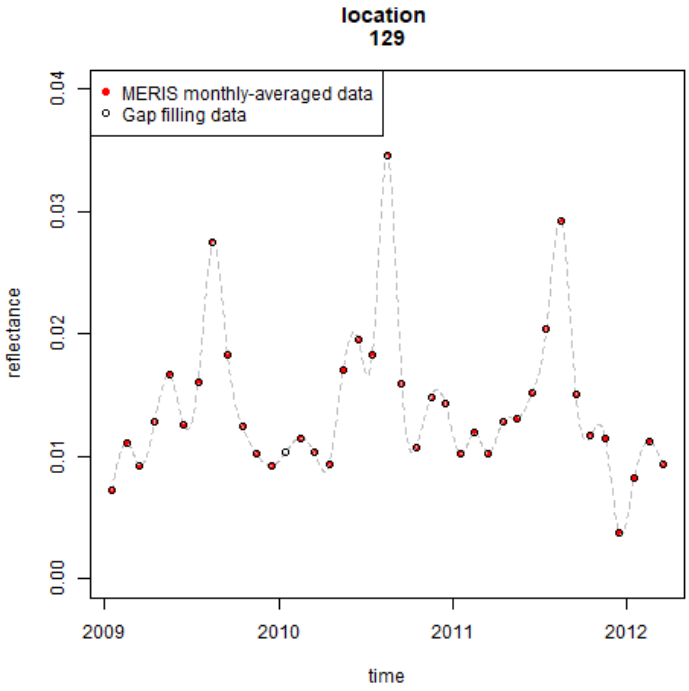


Figure 4.12: Gap filling of monthly-averaged MERIS reflectance data at location 129.

Model fitting and residual checking

Then, the NSD model and the FRK model are applied to these monthly datasets after gap filling. The model fitting process of the NSD model is similar to section 4.4.2. The FRK model is fitted by the expectation-maximization algorithm, where the tolerance between successive iterations is a difference of 0.001 in the log-likelihood. Based on this threshold, the expectation-maximization algorithm stops after 2 iterations. After fitting the model, the residuals are checked using the following steps. The qq-plots of the NSD and the FRK model residuals are shown in Figure 4.14 and 4.13, where the points are approximately straight lines except for a couple of outliers, which shows the normality assumption holds for both models.

The residuals against time plots in Figure 4.15 and 4.16 show that the variance of residuals for each location are approximately the same, but there are still some seasonal patterns(for example location 281 in Figure 4.16). The temporal correlations of the residuals are further checked by the autocorrelation function (ACF) in Figure 4.17 and Figure 4.18. The ACF show that there is no statistically significant temporal autocorrelation at these locations after lag 1 since the autocorrelation almost lies within the 95% confidence interval. Thus, the assumption of temporal uncorrelated residuals holds for both the NSD and the FRK model.

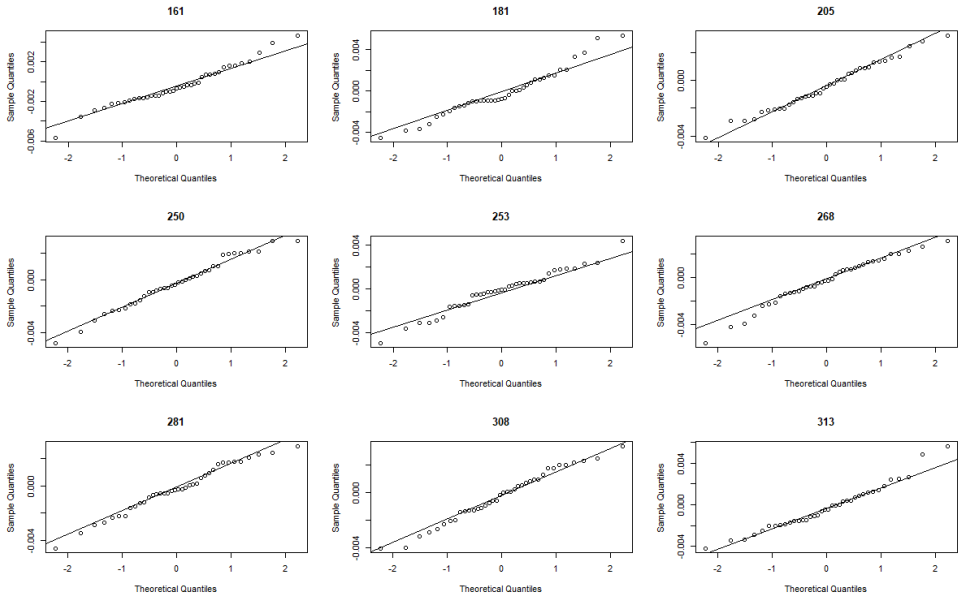


Figure 4.13: qq-plot of the residuals for different locations from the NSD model.

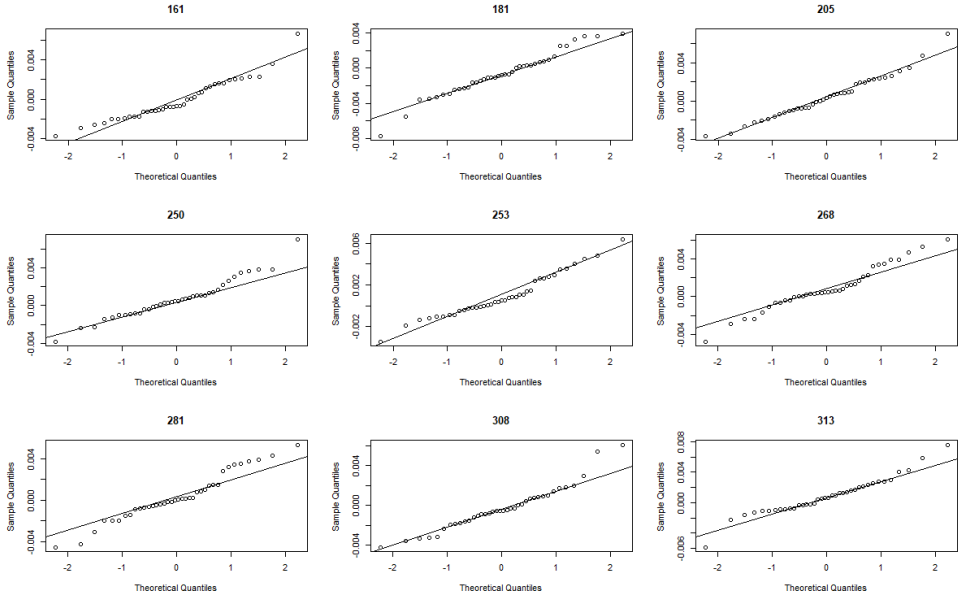


Figure 4.14: qq-plot of the residuals for different locations from the FRK model.

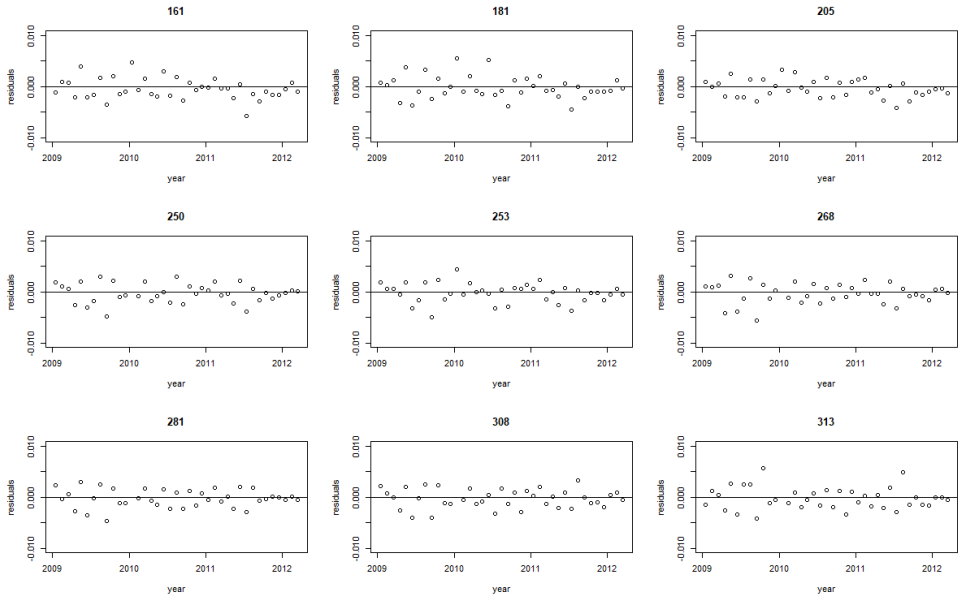


Figure 4.15: Residuals against time for different locations of NSD model.

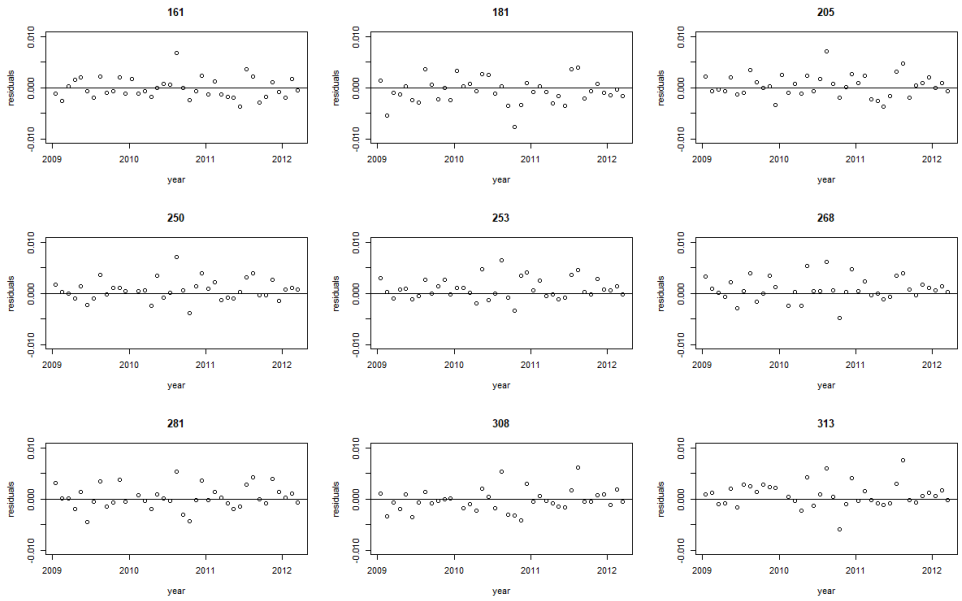


Figure 4.16: Residuals against time for different locations of FRK model.

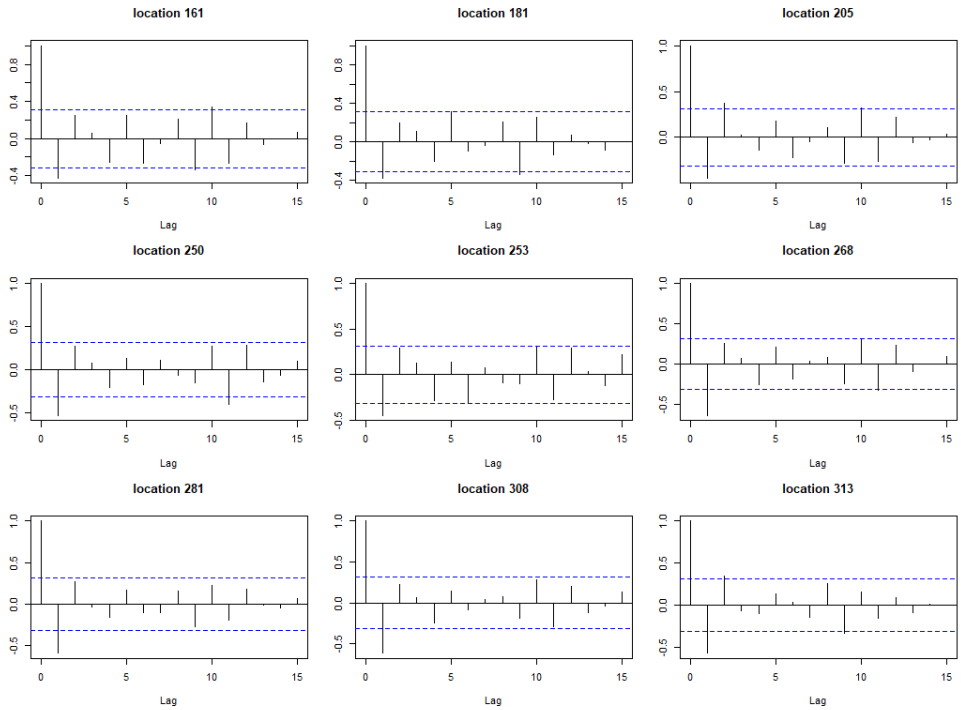


Figure 4.17: Autocorrelation function of residuals of the NSD for each prediction location

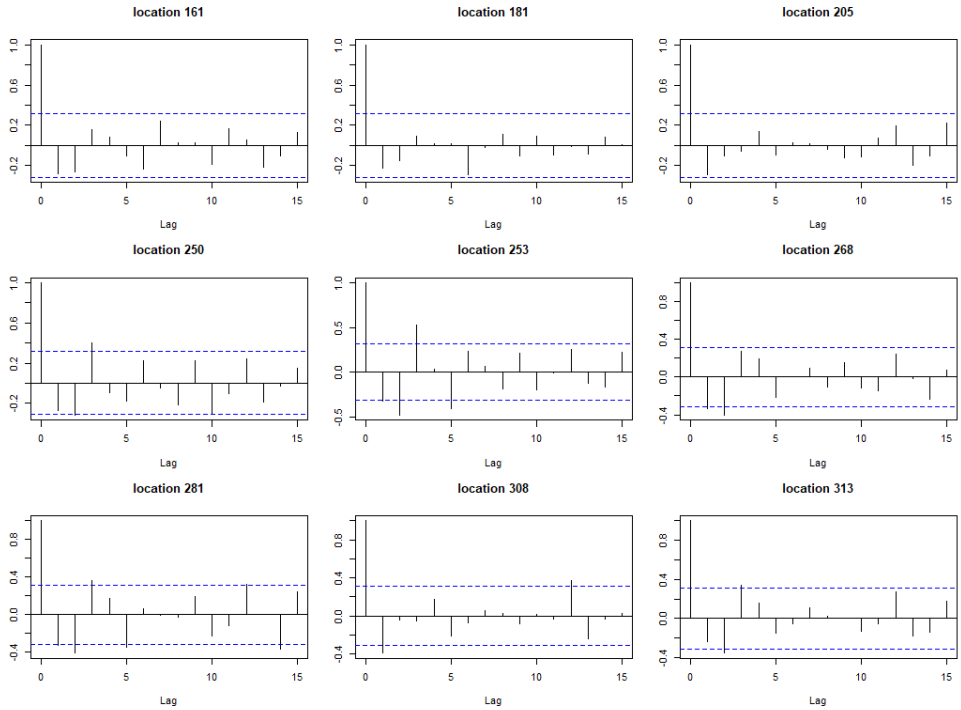
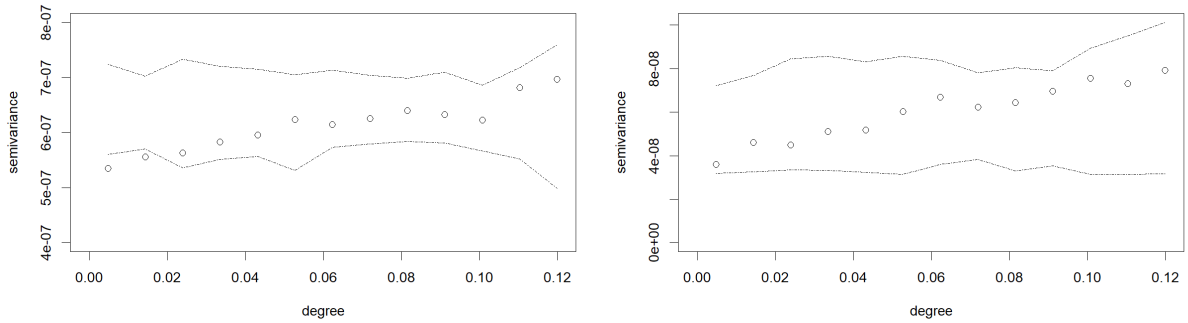


Figure 4.18: Autocorrelation function of residuals of the FRK for each prediction location

In the FRK and NSD models, the residuals are assumed to be spatially uncorrelated for each time period. Since the NSD model makes predictions at the point level, the spatial autocorrelation of the residuals could be assessed by the variograms with Monte Carlo envelopes. However, the FRK model makes predictions at an areal level, the variogram is used again for consistency

and the centroids are used to represent the location of each grid square. Since the maximum of distance between all pairs of locations is about 0.25 degrees, the empirical semi-vairogram is calculated over half of the maximum of distance, which is 0.125 degrees. The variogram plots for a randomly chosen month (Jan 2009) in Figure 4.19 show no statistically significant spatial autocorrelation since the empirical semi-variogram almost lies within the Monte Carlo envelopes. The variogram plots for the rest of the months are also checked and no significant spatial autocorrelation is found in the residuals. Thus, the assumption of spatially uncorrelated residuals holds for both the NSD and the FRK models. In conclusion, the analysis of the residuals from the NSD and the FRK models shows no significant evidence against the model assumptions.



(a) NSD model (b) FRK model

Figure 4.19: Variogram plots of the residuals from the NSD (left panel) and the FRK model (right panel) in January 2009.

Results of FRK and NSD model

After checking the model fitting, the RMSPE of the predictions are summarised in the Table 4.4. The number of temporal basis functions in the FRK model is selected to be the same as the number of B-splines basis functions in the NSD model, so that the temporal trends are modelled by the same number of temporal basis functions in both models. The number of spatial basis functions in the FRK model is selected to cover Lake Garda using the bi-square spatial function with 1 resolution (12 basis functions) and 2 resolutions (129 basis functions). From the table, the RMSPE of the NSD model with 29 B-spline basis functions and spatial decay parameter equal to 0.001 is less than the RMSPE of the FRK model with 12 or 129 spatial basis functions, which shows that the NSD model and the FRK model have similar predictive performance for the Lake Garda reflectance data. The predictions at location 161 in Lake Garda are shown in Figure 4.20, where the NSD model captures the peak in the years 2010 and 2011 better than the FRK model.

Model	RMSPE
NSD, 23 B basis, spatial decay=0.001	0.0031
NSD, 29 B basis, spatial decay=0.001	0.0020
FRK, 12 spatial basis, 29 temporal basis	0.0025
FRK, 129 spatial basis, 29 temporal basis	0.0021

Table 4.4: RMSPE for different choices of basis functions and spatial decay parameters.

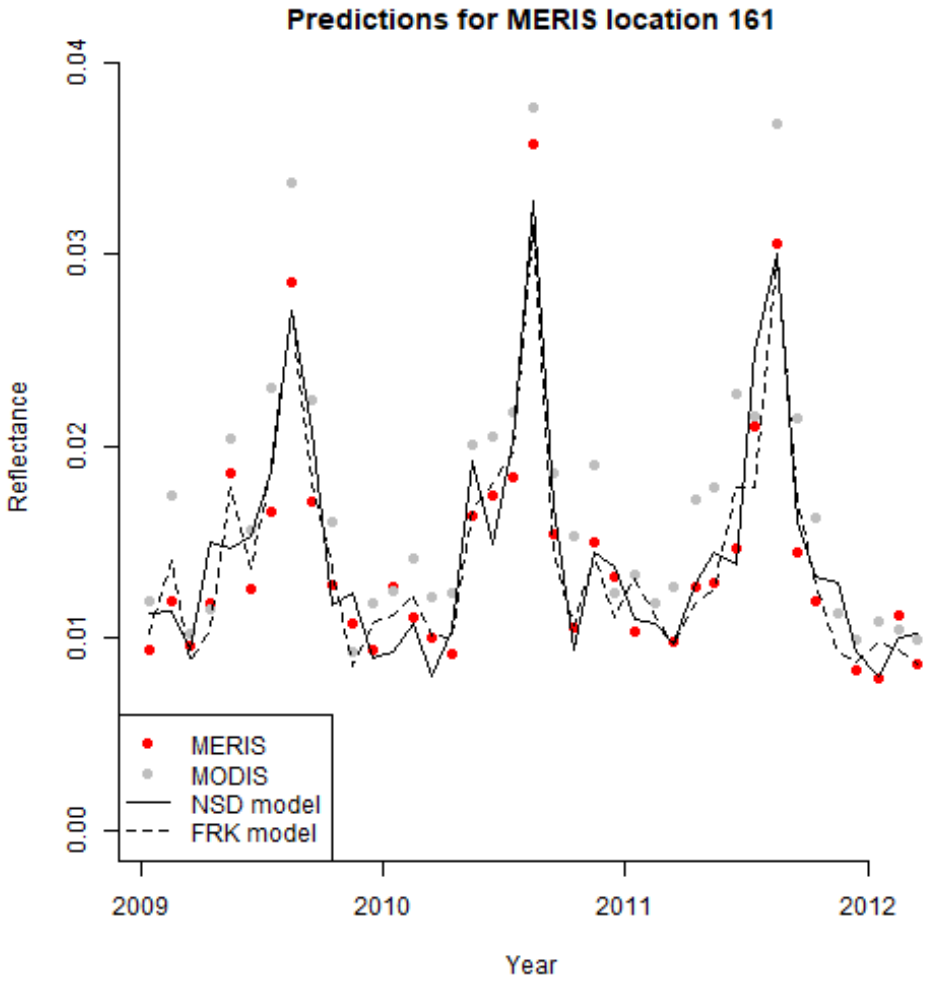


Figure 4.20: Predictions for monthly averaged MERIS reflectance data at location 161 from the FRK model (12 spatial basis, 29 temporal basis) and the NSD model (29 B-splines basis, spatial decay=0.001).

To check the prediction over space of these two models, August of 2010 is selected since the peak in the reflectance data appears in August each year, and the predictions over the lake at this month are shown in Figure 4.21. From this plot, both the NSD model and the FRK model capture the spatial pattern of the MERIS data in August 2010 quite well. However, the NSD model predictions are almost constant over space, while the FRK model predictions vary over space and captures the high value on the left side. The former is caused by the small value of the spatial decay parameters in the NSD model, which results in the high spatial correlation of

the coefficients β_{ij} and α_{ij} in the NSD model. The FRK model is able to predict the high value of the MERIS data on the left side of Lake Garda, since the MODIS data at this location is high and the FRK model uses the MODIS data as a covariate. The scatter plots of predicted against observed values for four randomly picked months (June 2011 to September 2011) are shown in Figure 4.22and Figure 4.23. In these plots, the points generally lie on the $y = x$ line, which suggests the model predictions match the observations well for these two models.

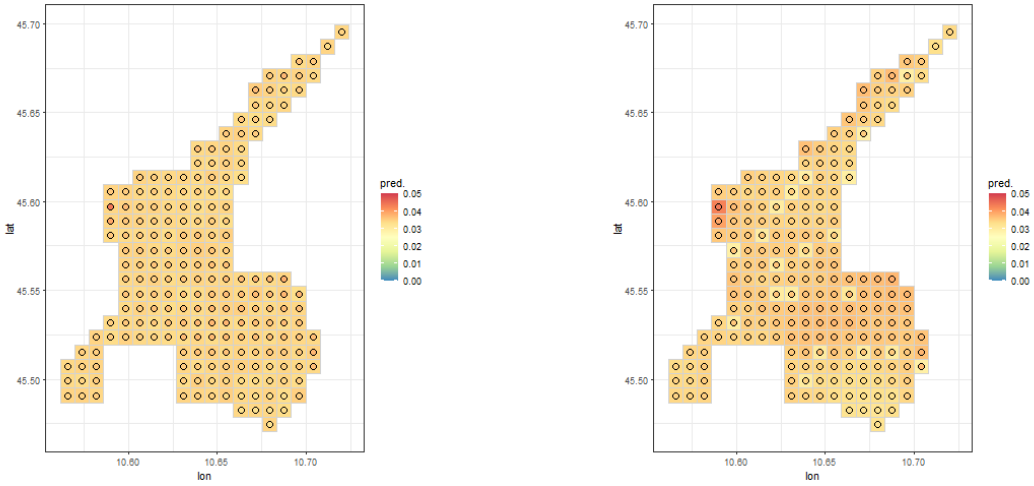


Figure 4.21: Spatial prediction of Lake Garda reflectance data in August 2010 from the NSD model (left panel) and the FRK model (right panel). The circled dots are the observed MERIS reflectance data.

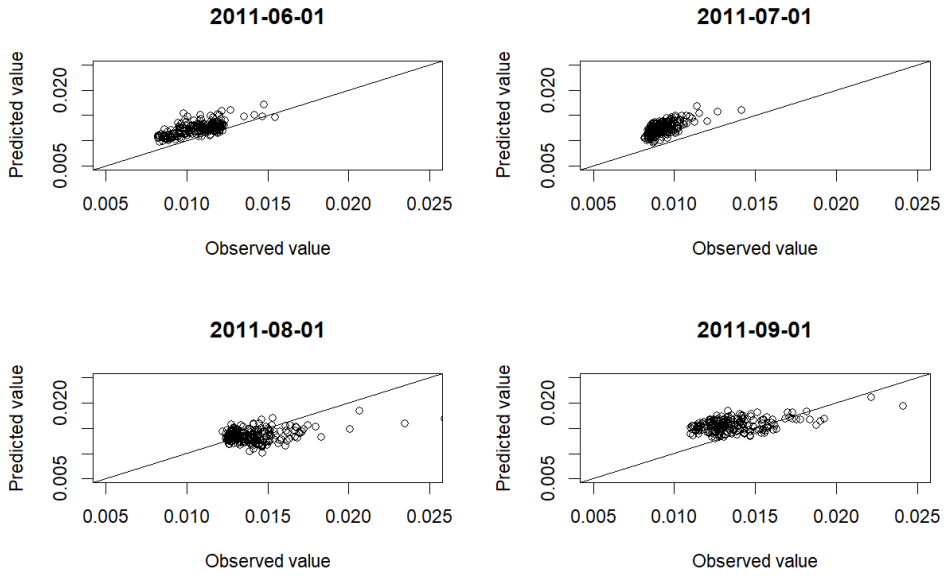


Figure 4.22: Predicted values using the NSD model against the observed values for all pixels from June 2011 to September 2011.

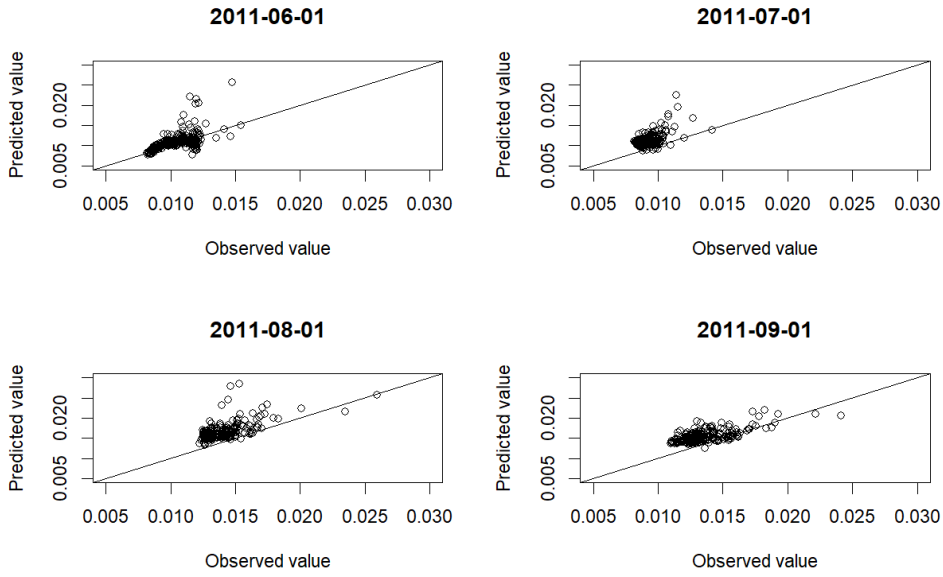


Figure 4.23: Predicted values using the NSD model against the observed values for all pixels from June 2011 to September 2011.

The comparison between these two models shows that the prediction ability of the NSD model is slightly better than the FRK model. However, the limitation of the NSD model is its computational time, which takes about 10 hours while the FRK model with 12 spatial basis functions could be fitted within 5 minutes. Another limitation of the NSD model is the prediction over space, since it is almost constant over space and does not capture the local high value of the reflectance data. This could be improved by selecting different spatial decay parameters, but it requires a long computational time. On balance, the NSD model is used to predict the MERIS reflectance data because it can handle the missingness in both space and time, while the FRK model could only be applied to the dataset without missingness. This advantage makes the NSD model more suitable for the remote-sensing data with a large amount of missingness in both space and time.

4.5 Conclusion

In this chapter, we use Lake Garda reflectance data from MERIS and MODIS sensors to assess the possibility of making predictions of MERIS data at unobserved locations across the time dimension by an existing downscaling model. For this dataset, the NSD model could be used to predict the MERIS reflectance data by fusing the MERIS and MODIS data. This chapter is motivated by the different temporal coverage of the two remote sensors, and the fusion of reflectance data from the different sensors will provide a series of reflectance maps across the lake within the time range of these sensors.

The comparison between the NSD and the FRK model shows that the prediction ability of the these two models are similar. However, the limitation of the NSD model is its computational time. There are about 25000 parameters estimated in each iteration of the MCMC for the NSD model, while there are only about 350 parameters estimated in each iteration of the EM algorithm for the FRK model. Although the NSD model has its advantage in handling the missing in both space and time, this computational time makes it hard to apply the NSD to larger lakes with thousands of different pixels. Thus, the future study will focus on decreasing the computational time of the NSD model. One possible approach is to use Integrated Nested Laplace Approximations (INLA) to approximate the posterior marginal distribution in the NSD model rather than using the Gibbs sampling [Lindgren and Rue, 2015]. Another approach is to use the spatial basis functions as the FRK model and time-varying coefficients, which will decrease the number of parameters in the NSD model and the computational time.

In this chapter, the FRK model is selected as a competitor to assess the model performance of the NSD model. It is possible to use a tensor product in a generalised additive model using the mgcv R package, which allows the missing values in the temporal dimension [Wood, 2017]. Thus, future work could compare the model performance of the generalised additive model and the NSD model on the original data without averaging over month.

In conclusion, the NSD model could be used to fuse the reflectance data from different sensors and then provide a longer series of reflectance maps to monitor the lake water quality, and improving the computational time of the NSD model will make it easier to be applied to the larger lakes. In the next chapter, improving the computational efficiency of the NSD model will be discussed.

Chapter 5

Improving the computational efficiency of the NSD model

5.1 Introduction

In the previous chapter, the non-parametric statistical downscaling (NSD) model is used to make predictions of the MERIS reflectance data using the MODIS data at Lake Garda. The NSD model can be used to make predictions at unobserved locations or time points, and provide smaller root mean square predictive error than the Fixed ranking kriging (FRK) model that was also explored. The NSD model could handle the spatial-temporal change of support problem, but this function has not been implemented in the FRK R-package. However, the NSD model is more computationally expensive than the FRK model. The computational time of the NSD model makes it prohibitive to apply to large datasets, such as Lake Erie with approximately 20000 locations.

The aim of this chapter is to propose an approximation for the NSD model, which can achieve similar predictive performance to the NSD model but requires less computational time. This chapter will start by describing the approximation method, which is the Gaussian predictive process (GPP). The model proposed in this chapter is therefore called the NSD-GPP model. The NSD-GPP model will be used to predict the MERIS reflectance data for Lake Garda, and the results will be compared to the NSD model. After that, a simulation study will be carried out to compare the performances of the NSD-GPP model and the NSD model for a range of datasets with different spatial patterns. Finally, the limitations and the development of the NSD-GPP model will be discussed.

5.2 Methods

In this section, the NSD model will be reviewed, and its computational efficiency will be discussed. Then, the approximation method, the Gaussian predictive process, will be introduced. Finally, a new NSD model which fuses the original model with the Gaussian predictive process will be proposed.

5.2.1 Spatial process in the NSD model

In this study, the NSD model is used to predict the MERIS reflectance data $\mathbf{y}_i$ using the MODIS data $\mathbf{x}_i$ for location $i = 1, 2, ..., n$. The NSD model is defined as follows. Suppose $\mathbf{y}_i = (y_{i1}, ..., y_{iq_i})$ is the vector of the MERIS reflectance data from time 1 to q_i, and $\mathbf{x}_i = (x_{i1}, ..., x_{ip_i})$ represents the MODIS reflectance data from time 1 to p_i. The NSD model assumes the vectors $\mathbf{y}_i$ and $\mathbf{x}_i$ are the observations of smooth functions over time :

$$\begin{aligned} \mathbf{y}_i|\mathbf{c}_i, \frac{1}{p_y} &\sim \mathrm{N}_{q_i}\left(\mathbf{\Phi}_i\mathbf{c}_i, \frac{1}{p_y}\mathbf{I}_{q_i} \right), \\ \mathbf{x}_i|\mathbf{d}_i, \frac{1}{p_x} &\sim \mathrm{N}_{p_i}\left(\mathbf{\Psi}_i\mathbf{d}_i, \frac{1}{p_x}\mathbf{I}_{p_i} \right), \end{aligned} \tag{5.1}$$

where $\mathbf{\Phi}_i, \mathbf{\Psi}_i$ are the temporal basis function matrices. In the NSD model, the same set of basis functions is used for $\mathbf{y}_i$ and $\mathbf{x}_i$, which means the number of columns of $\mathbf{\Phi}_i$ and $\mathbf{\Psi}_i$ equals the number of basis functions m. The number of rows of these two matrices are q_i and p_i respectively, which represents the basis functions evaluated at times q_i and p_i. The vectors c_i and d_i with a dimension of m are the corresponding coefficient vectors. p_y and p_x are the precision parameters.

Then, the NSD model assumes the coefficients c_{ij} and d_{ij} are related by a linear regression model with a slope parameter β_{ij} and an intercept parameter α_{ij} for $j = 1, 2, ..., m$. The parameters $\boldsymbol{\alpha}_j$ and $\boldsymbol{\beta}_j$ are assumed to vary over space and follow a Gaussian process with a fixed mean vector and the covariance function as follows:

$$\begin{aligned} c_{ij} &\sim \mathrm{N}\left(d_{ij}\beta_{ij} + \alpha_{ij}, \frac{1}{p_c} \right), \\ \boldsymbol{\alpha}_j &\sim \mathrm{N}\left(\mathbf{0}, \frac{1}{p_\alpha}\exp(-\phi_\alpha \mathbf{D}) \right), \\ \boldsymbol{\beta}_j &\sim \mathrm{N}\left(\mathbf{1}, \frac{1}{p_\beta}\exp(-\phi_\beta \mathbf{D}) \right), \\ \mathbf{d}_i &\sim \mathrm{N}_m(\boldsymbol{\mu}_d, \boldsymbol{\Sigma}_d), \end{aligned} \tag{5.2}$$

where p_α, p_β are the precision parameters and ϕ_α, ϕ_β are the spatial decay parameters, and the matrix $\mathbf{D}$ is the distance matrix between the n locations. For the vector $\mathbf{d}_i$, the weakly informative prior is used with $\boldsymbol{\mu}_d = \mathbf{0}$ and $\boldsymbol{\Sigma}_d = 100 * \mathbf{I}_m$.

The NSD model assumes that the MERIS coefficients c_{ij} and the MODIS coefficients d_{ij} are related by a spatial regression model. In this regression model, α_{ij} are the spatial varying intercepts and β_{ij} are the spatial varying slopes. The vector $\boldsymbol{\alpha}_j = (\alpha_{1j}, \alpha_{1j}, ..., \alpha_{nj})$ and $\boldsymbol{\beta}_j = (\beta_{1j}, \beta_{1j}, ..., \beta_{nj})$ are assumed to follow a Gaussian process with dimension of n, which has an exponential correlation function. The Gamma prior is used for the precision p_α with $p_\alpha \sim \text{Ga}(a_\alpha, b_\alpha)$, where a_α is the shape parameter and b_α is the rate parameter. Then, using the formula for conjugate prior [Gelman et al., 2014], the posterior distribution of $\boldsymbol{\alpha}_j$ is defined as:

$$
\begin{aligned}
&f(\boldsymbol{\alpha}_j|\cdot) \sim \text{N}(\ \boldsymbol{\Sigma}_{\alpha_j}\mathbf{A}_{\alpha_j},\ \boldsymbol{\Sigma}_{\alpha_j}\ ),\\
&\boldsymbol{\Sigma}_{\alpha_j} = (p_\alpha(\exp(-\phi_\alpha\mathbf{D}))^{-1} + p_c(\mathbf{I}_n)^{-1})^{-1},\\
&\mathbf{A}_{\alpha_j} = p_c(\mathbf{c}_j - \boldsymbol{\beta}_j \odot \mathbf{d}_j),
\end{aligned}
$$

where $\odot$ is the element-wise multiplication. The parameters in the NSD model are estimated by Bayesian inference with Markov chain Monte Carlo (MCMC). In this formula, computing $\boldsymbol{\Sigma}_{\alpha_j}$ requires inverting of a $n \times n$ matrix in each iteration of the Gibbs sampling. Since the prior and the likelihood of $\boldsymbol{\beta}_j$ have a similar form as $\boldsymbol{\alpha}_j$, sampling from the posterior distribution of the parameters $\boldsymbol{\beta}_j$ will also need this $n \times n$ matrix inversion. The computational time of this inversion will increase as $O(n^3)$ in the number of locations n [Banerjee et al., 2008]. Since the estimation of the other parameters of the NSD model does not require inverting a matrix of dimension $n \times n$, using a lower dimensional estimation of $\boldsymbol{\alpha}_j$ and $\boldsymbol{\beta}_j$ will substantially reduce the computational time of the NSD model. This suggests using an approximation for the original spatial process of $\boldsymbol{\alpha}_j$ and $\boldsymbol{\beta}_j$.

5.2.2 Gaussian predictive process

Gaussian predictive process (GPP) [Banerjee et al., 2008] is one approximation method for representing the spatial process of $\boldsymbol{\alpha}_j$ and $\boldsymbol{\beta}_j$. A GPP model projects the original spatial process to a lower dimensional subspace, and treats the lower-dimensional process as an approximation, which reduces the computational burden of estimating the original process [Banerjee et al., 2008]. A modified GPP for improving the estimation of the spatial error is introduced in [Guhaniyogi et al., 2011], but the additional metropolis steps in the parameter estimation will also require more computational time. Since the aim of this study focus on the predictive performance and the computational efficiency of GPP, this modified version is not included in this study.

Take $\boldsymbol{\alpha}_j$ as an example, suppose the spatial process $\boldsymbol{\alpha}_j$ is defined at locations $S = \{s_1, s_2, ..., s_n\}$, a lower-dimensional spatial process $\boldsymbol{\alpha}_j^*$ is defined over a set of knots $S^* = \{s_1^*, s_2^*, ..., s_r^*\}$. The number of knots r is much less than the number of observed data n. Then, the Gaussian predictive process [Banerjee et al., 2008] is defined as:

$$
\begin{aligned}
&\boldsymbol{\alpha}_j = \mathbf{Z}\boldsymbol{\alpha}_j^*, \\
&\boldsymbol{\alpha}_j^* \sim \mathrm{N}_r(\, \mathbf{0},\, \frac{1}{p_\alpha}\exp(-\phi_\alpha \mathbf{D}^*)\,), \\
&\mathbf{Z} = \exp(-\phi_\alpha \mathbf{D}^T)(\exp(-\phi_\alpha \mathbf{D}^*))^{-1},
\end{aligned}
$$

where:

- $\mathbf{D}^*$ is the $r \times r$ distance matrix between the selected knots $S^* = \{s_1^*, s_2^*, ..., s_r^*\}$.
- $\mathbf{D}^T$ is the $n \times r$ distance matrix between the observed locations $S = \{s_1, s_2, ..., s_n\}$ and the selected knots $S^* = \{s_1^*, s_2^*, ..., s_r^*\}$.

Then, the posterior distribution of $\boldsymbol{\alpha}_j^*$ can be derived as:

$$
\begin{aligned}
&f(\boldsymbol{\alpha}_j^* | \cdot) \propto \mathrm{N}(\, \boldsymbol{\Sigma}_{\alpha_j^*}\mathbf{A}_{\alpha_j^*},\, \boldsymbol{\Sigma}_{\alpha_j^*}\,), \\
&\boldsymbol{\Sigma}_{\alpha_j^*} = (p_\alpha(\exp(-\phi_\alpha \mathbf{D}^*))^{-1} + p_c \mathbf{Z}^T\mathbf{Z})^{-1}, \\
&\mathbf{A}_{\alpha_j^*} = p_c \mathbf{Z}^T(\mathbf{c}_j - \boldsymbol{\beta}_j \odot \mathbf{d}_j).
\end{aligned}
$$

In the NSD model, the estimation of spatial decay parameters ϕ_α and ϕ_β has an identifiability problem when estimating the spatial precision parameters p_α and p_β at the same time [Wilkie et al., 2019]. Thus, the grid search over a range of possible values of the spatial decay parameter is used for the NSD model [Sahu et al., 2006]. Since the value of ϕ_α is chosen before the model is fitted, the matrix $\mathbf{Z}$ is fixed after selecting the knots S^*. Thus, the estimation of $\boldsymbol{\alpha}_j$ is approximated by estimating the lower-dimensional process $\boldsymbol{\alpha}_j^*$ in each iteration. In the formula of the posterior distribution of $\boldsymbol{\alpha}_j^*$, the dimension of the variance matrix $\boldsymbol{\Sigma}_{\alpha_j^*}$ is a $r \times r$. Thus, the inversion of the $n \times n$ matrix also becomes the inversion of a $r \times r$ matrix with $r \ll n$. Similarly, the Gaussian predictive process for the coefficients $\boldsymbol{\beta}_j$ in the NSD model is defined as:

$$
\begin{aligned}
&\boldsymbol{\beta}_j = \mathbf{V}\boldsymbol{\beta}_j^*, \\
&\boldsymbol{\beta}_j^* \sim \mathrm{N}\left(\, \mathbf{1},\, \frac{1}{p_\beta}\exp(-\phi_\beta \mathbf{D}^*)\, \right), \\
&\mathbf{V} = \exp(-\phi_\beta \mathbf{D}^T)(\exp(-\phi_\beta \mathbf{D}^*))^{-1}.
\end{aligned}
$$

This is an approximation for $\boldsymbol{\beta}_j$, and it also requires an inversion of a $r \times r$ matrix in the parameter estimation. In conclusion, the Gaussian predictive process uses a lower-dimensional approximation and the estimation of the process involves the inversion of a lower-dimensional matrix, which then reduces the computational time in each iteration of the Gibbs sampling.

5.2.3 NSD-GPP model

In the above section, the computational efficiency of the Gaussian predictive process compared to the original spatial process is described. In this section, the NSD model will be combined with the Gaussian process. The NSD-GPP model has the similar form as the NSD model in Equation 5.1 and 5.2, except that the $\boldsymbol{\alpha}_j$ and $\boldsymbol{\beta}_j$ are approximated by $\mathbf{Z}\boldsymbol{\alpha}_j^*$ and $\mathbf{V}\boldsymbol{\beta}_j^*$. The NSD-GPP model is defined as:

$$
\begin{aligned}
&\mathbf{y}_i|\mathbf{c}_i, p_y \sim \mathrm{N}_{q_i}\left(\ \boldsymbol{\Phi}_i\mathbf{c}_i,\ \frac{1}{p_y}\mathbf{I}_{q_i}\ \right),\\
&\mathbf{x}_i|\mathbf{d}_i, p_x \sim N_{p_i}\left(\ \boldsymbol{\Psi}_i\mathbf{d}_i,\ \frac{1}{p_x}\mathbf{I}_{p_i}\ \right),\\
&c_{ij} \sim \mathrm{N}\left(\ d_{ij}\mathbf{V}_i\boldsymbol{\beta}_j^* + \mathbf{Z}_i\boldsymbol{\alpha}_j^*,\ \frac{1}{p_c}\ \right),\\
&\mathbf{d}_i \sim \mathrm{N}_m(\ \boldsymbol{\mu}_d,\ \boldsymbol{\Sigma}_d\ ),\\
&\boldsymbol{\alpha}_j^* \sim \mathrm{N}\left(\ \mathbf{0},\ \frac{1}{p_\alpha}\exp(-\phi_\alpha\mathbf{D}^*)\ \right),\\
&\boldsymbol{\beta}_j^* \sim \mathrm{N}\left(\ \mathbf{1},\ \frac{1}{p_\beta}\exp(-\phi_\beta\mathbf{D}^*)\ \right),\\
&\mathbf{Z} = \exp(-\phi_\alpha\mathbf{D}^T)(\exp(-\phi_\alpha\mathbf{D}^*))^{-1},\\
&\mathbf{V} = \exp(-\phi_\beta\mathbf{D}^T)(\exp(-\phi_\beta\mathbf{D}^*))^{-1},
\end{aligned}
$$

where:

- $\mathbf{c}_j = (c_{1j}, c_{2j}, ...c_{nj})^T$ and $\mathbf{d}_j = (d_{1j}, d_{2j}, ...d_{nj})^T$ are the coefficients for the j^{th} basis function.
- $\mathbf{D}^*$ is the $r \times r$ distance matrix between the selected knots $S^* = \{s_1^*, s_2^*, ..., s_r^*\}$.
- $\mathbf{D}^T$ is the $n \times r$ distance matrix between the observed locations $S = \{s_1, s_2, ..., s_n\}$ and the selected knots $S^* = \{s_1^*, s_2^*, ..., s_r^*\}$.

The prior of the precision parameters follow the Gamma distribution with known shape and rate parameter, where

$$
\begin{aligned}
p_y &\sim \text{Ga}(a_y, b_y),\\
p_x &\sim \text{Ga}(a_x, b_x),\\
p_\alpha &\sim \text{Ga}(a_\alpha, b_\alpha),\\
p_\beta &\sim \text{Ga}(a_\beta, b_\beta),\\
p_c &\sim \text{Ga}(a_c, b_c).
\end{aligned}
$$

The weakly informative prior is used for all these precision parameters using Gamma$(2,1)$, which is discussed in Section 4.3.1.

5.2.4 Parameter estimation

The parameters of NSD-GPP model are estimated using Bayesian inference with MCMC. The full conditional posterior distributions for all the parameters are proper distributions except for the spatial decay parameters ϕ_α and ϕ_β which do not have closed forms for their posterior distributions. The spatial decay parameters will be selected by the cross-validation method. Since the conjugate prior of p_α is the Gamma distribution, the posterior distribution of p_α can be written as follows:

$$
f(p_\alpha|\cdot) \sim \text{Gamma}\left(a_\alpha + \frac{mr}{2},\ b_\alpha + \frac{1}{2}\text{tr}[\exp(-\phi_\alpha \mathbf{D}^*)^{-1}\boldsymbol{\alpha}^{*T}\boldsymbol{\alpha}^*] \right). \tag{5.3}
$$

The conditional posterior of p_β, p_y, p_x, p_c have a similar form as the p_α, and they are shown as follows:

$$
\begin{aligned}
&f(p_\beta|\cdot) \sim \text{Gamma}\left(a_\beta + \frac{m}{2},\ b_\beta + \frac{1}{2}(\boldsymbol{\beta}^* - \mathbf{1})^T(\boldsymbol{\beta}^* - \mathbf{1}) \right);\\
&f(p_y|\cdot) \sim \text{Gamma}\left(a_y + \sum_{i=1}^{n}\frac{q_i}{2},\ b_y + \frac{1}{2}\sum_{i=1}^{n}(\mathbf{y}_i - \boldsymbol{\Phi}_i\mathbf{c}_i)^T(\mathbf{y}_i - \boldsymbol{\Phi}_i\mathbf{c}_i) \right);\\
&f(p_x|\cdot) \sim \text{Inv}-\text{Gamma}\left(a_x + \sum_{i=1}^{n}\frac{p_i}{2},\ b_x + \frac{1}{2}\sum_{i=1}^{n}(\mathbf{x}_i - \boldsymbol{\Psi}_i\mathbf{d}_i)^T(\mathbf{x}_i - \boldsymbol{\Psi}_i\mathbf{d}_i) \right);\\
&f(p_c|\cdot) \sim \text{Gamma}\left(a_c + \frac{mn}{2},\ b_c + \frac{1}{2}\text{tr}(\mathbf{I}_n(\mathbf{c} - (\mathbf{Z}\boldsymbol{\alpha}^* + (\mathbf{V}\boldsymbol{\beta}^*)\odot\mathbf{d}))^T\mathbf{I}_m(\mathbf{c} - (\mathbf{Z}\boldsymbol{\alpha}^* + (\mathbf{V}\boldsymbol{\beta}^*)\odot\mathbf{d}))) \right).
\end{aligned}
$$

Since the conjugate prior of $\boldsymbol{\alpha}_j^*$ is the normal distribution, then the posterior distribution of $\boldsymbol{\alpha}_j$ is defined as:

$$
\begin{aligned}
&f(\boldsymbol{\alpha}_j^*|\cdot) \sim \mathrm{N}(\, \boldsymbol{\Sigma}_{\alpha_j^*}\mathbf{A}_{\alpha_j^*},\, \boldsymbol{\Sigma}_{\alpha_j^*}\,),\\
&\boldsymbol{\Sigma}_{\alpha_j^*} = (p_\alpha(\exp(-\phi_\alpha \mathbf{D}^*))^{-1} + p_c\mathbf{Z}^T\mathbf{Z})^{-1},\\
&\mathbf{A}_{\alpha_j^*} = p_c\mathbf{Z}^T(\mathbf{c}_j - (\mathbf{V}\boldsymbol{\beta}_j^*)\odot\mathbf{d}_j).
\end{aligned}
$$

Then, the posterior of $\boldsymbol{\beta}_j^*$, $\boldsymbol{c}_j$ and $\boldsymbol{d}_j$ can be defined with the same method as follows:

$$
\begin{aligned}
&f(\boldsymbol{\beta}_j^*|\cdot) \sim \mathrm{N}(\, \boldsymbol{\Sigma}_{\beta_j^*}\mathbf{A}_{\beta_j^*},\, \boldsymbol{\Sigma}_{\beta_j^*}\,),\\
&\boldsymbol{\Sigma}_{\beta_j^*} = (p_\beta(\exp(-\phi_\beta \mathbf{D}^*))^{-1} + p_c\mathbf{V}^T\mathbf{G}_j^T\mathbf{G}_j\mathbf{V})^{-1},\\
&\mathbf{A}_{\beta_j^*} = p_\beta\exp(-\phi_\beta\mathbf{D}_{\text{data}})^{-1}\mathbf{1} + p_c\mathbf{V}^T\mathbf{G}_j^T(\mathbf{c}_j - \mathbf{Z}\boldsymbol{\alpha}_j^*),
\end{aligned}
$$

where $\mathbf{G}_j$ is a diagonal matrix with $\mathbf{d}_j$ as its diagonal.

$$
\begin{aligned}
&f(\mathbf{c}_i|\cdot) \sim \mathrm{N}(\, \boldsymbol{\Sigma}_{c_i}\mathbf{A}_{c_i},\, \boldsymbol{\Sigma}_{c_i}\,),\\
&\boldsymbol{\Sigma}_{c_i} = (p_y\boldsymbol{\Phi}_i^T\boldsymbol{\Phi}_i + p_c\mathbf{I}_m)^{-1},\\
&\mathbf{A}_{c_i} = p_y\boldsymbol{\Phi}_i^T\mathbf{y}_i + p_c(\mathbf{Z}_i\boldsymbol{\alpha}^* + (\mathbf{V}_i\boldsymbol{\beta}^*)\odot\mathbf{d}_i).
\end{aligned}
$$

$$
\begin{aligned}
&f(\mathbf{d}_i|\cdot) \sim \mathrm{N}(\, \boldsymbol{\Sigma}_{d_i}\mathbf{A}_{d_i},\, \boldsymbol{\Sigma}_{d_i}\,),\\
&\boldsymbol{\Sigma}_{d_i} = (\boldsymbol{\Sigma}_d^{-1} + p_x\boldsymbol{\Psi}_i^T\boldsymbol{\Psi}_i + p_c\mathbf{F}_i^T\mathbf{F}_i)^{-1},\\
&\mathbf{A}_{d_j} = \boldsymbol{\Sigma}_d^{-1}\boldsymbol{\mu}_d + p_x\boldsymbol{\Psi}_i^T\mathbf{x}_i + p_c\mathbf{F}_i^T(\mathbf{c}_i - \mathbf{Z}_i\boldsymbol{\alpha}^*),
\end{aligned}
$$

where $\mathbf{F}_i$ is a diagonal matrix with $\mathbf{V}_i\boldsymbol{\beta}$ as its diagonal.

To make predictions $\tilde{y_{ij}}$ using x_{ij} at new times j ($j = 1,2,...,\tilde{q_i}$) and new locations i ($i = 1,2,...,\tilde{n}$), the following equation is used:

$$
\begin{aligned}
&\tilde{\mathbf{y}}_i|\tilde{\mathbf{c}}_i, \sigma_y^2 \sim \mathrm{N}_{\tilde{q}_i}\left(\tilde{\boldsymbol{\Phi}}_i\tilde{\mathbf{c}}_i,\, \frac{1}{p_y}\mathbf{I}_{\tilde{q}_i}\right),\\
&\tilde{c}_{ij} \sim \mathrm{N}\left(\, \tilde{\alpha}_{ij} + \tilde{\beta}_{ij}\tilde{d}_{ij},\, \frac{1}{p_c}\,\right),\\
&\tilde{\mathbf{d}}_i \sim \mathrm{N}_m\left(\, \tilde{\boldsymbol{\Sigma}}_{d_i}\tilde{\mathbf{A}}_{d_i},\, \tilde{\Sigma}_{d_i}\,\right),\\
&\tilde{\boldsymbol{\Sigma}}_{d_i} = (\boldsymbol{\Sigma}_d^{-1} + p_x\tilde{\boldsymbol{\Psi}}_i^T\tilde{\boldsymbol{\Psi}}_i)^{-1},\\
&\mathbf{A}_{d_i} = \boldsymbol{\Sigma}_d^{-1}\boldsymbol{\mu}_d + p_x\tilde{\boldsymbol{\Psi}}_i^T\tilde{\mathbf{x}}_i,
\end{aligned}
$$

where the current values of the Markov chain for other parameters such as p_c, p_y are used to get the prediction $\tilde{\mathbf{y}}_i$.

The coefficients $\tilde{\alpha}_{ij}$ and $\tilde{\beta}_{ij}$ for the new locations i and for each basis function j are estimated conditional on the coefficients $\boldsymbol{\alpha}_j^*$ and $\boldsymbol{\beta}_j^*$ of the observed locations. Since $\boldsymbol{\alpha}_j^*$ and $\boldsymbol{\beta}_j^*$ are Gaussian process, the property of the multivariate normal distribution is used [Gelman et al., 2014] and the conditional distribution could be written as:

$$\begin{aligned}
&\tilde{\boldsymbol{\alpha}}_j|\boldsymbol{\alpha}_j^* \sim \mathrm{N}\Bigg(\mathbf{0} + \exp(-\phi_\alpha \mathbf{D}_{12})\exp(-\phi_\alpha \mathbf{D}^*)^{-1}(\boldsymbol{\alpha}_j^* - \mathbf{0}), \\
&\frac{1}{p_\alpha}(\exp(-\phi_\alpha \tilde{\mathbf{D}}) - \exp(-\phi_\alpha \mathbf{D}_{12})\exp(-\phi_\alpha \mathbf{D}^*)^{-1}\exp(-\phi_\alpha \mathbf{D}_{21})) \Bigg); \\
&\tilde{\boldsymbol{\beta}}_j|\boldsymbol{\beta}_j^* \sim \mathrm{N}\Bigg(\mathbf{1} + \exp(-\phi_\beta \mathbf{D}_{12})\exp(-\phi_\beta \mathbf{D}^*)^{-1}(\boldsymbol{\beta}_j^* - \mathbf{1}), \\
&\frac{1}{p_\beta}(\exp(-\phi_\beta \tilde{\mathbf{D}}) - \exp(-\phi_\beta \mathbf{D}_{12})\exp(-\phi_\beta \mathbf{D}^*)^{-1}\exp(-\phi_\beta \mathbf{D}_{21})) \Bigg);
\end{aligned}$$

where:

- $\tilde{\mathbf{D}}$ is the $\tilde{n} \times \tilde{n}$ distance matrix between the prediction locations.

- $\mathbf{D}_{12}$ is the $\tilde{n} \times r$ distance matrix between the prediction locations and the knots of the Gaussian predictive process.

- $\mathbf{D}_{21}$ is the $r \times \tilde{n}$ distance matrix between the knots of the Gaussian predictive process and the prediction locations.

Choosing the number of temporal basis functions and the spatial decay parameters

One common way of selecting the number of temporal basis functions and the spatial decay parameters is using the cross-validation. In this chapter, two types of cross-validation are used, which are spatial cross-validation and temporal cross-validation.

- **spatial cross-validation** In spatial cross-validation, the locations of the reflectance data are randomly divided into k groups with an equal number of locations. Then, in each iteration, one group is chosen as the out-of-sample prediction and the other $k-1$ groups are treated as the training groups. The predictions are made for the selected locations in the prediction group for all months.

- **temporal cross-validation** In temporal cross-validation, the months of the reflectance data are randomly divided into k groups with an equal number of months. Then, in each iteration, one group is chosen as the out-of-sample prediction and the other $k-1$ groups are treated as the training groups. The predictions are made for the selected months in the prediction group for all locations.

The NSD-GPP model is fitted to the data of the training groups and the predictions are compared with the true values by the root mean square prediction error (RMSPE). The averages of the k iterations of the cross-validation are compared between different combinations of the number of temporal basis functions and the spatial decay parameters. The combination corresponding to the minimum of the averaged RMSPE is chosen. These two types of cross-validation methods are applied to the Lake Garda dataset, and the results and limitations will be discussed in section 5.3.

Choosing the knots for Gaussian predictive process

From the paper of Banerjee et al (2008), the number of knots is the primary factor of the parameter estimation and the predictive performance of the GPP model, and the equally spaced knots are preferred when lacking additional background information of the target spatial process [Banerjee et al., 2008]. Thus, the equally spaced knots will be used in this chapter. The effect of different number of knots on the prediction performance will be investigated by a simulation study in section 1.4.

5.3 Model comparison on Lake Garda dataset

In this section, the NSD-GPP and the NSD model will be applied to the Lake Garda dataset. The aim is to compare the prediction performance and the computational time of these two models. First, the results of the spatial ten-fold cross-validation are presented. Then, the limitation of the spatial cross-validation for the Lake Garda dataset is discussed and the temporal cross-validation for the Lake Garda is introduced. Finally, a sensitivity analysis for the precision parameters is used to check if the RMSPE is stable with different prior choices.

5.3.1 Spatial cross-validation

In the parameter estimation of the NSD-GPP model, the spatial decay parameter and the number of basis functions are treated as hyperparameters and are fixed before model fitting. Thus, these two parameters are selected by cross-validation, and the choices are made based on the out-of-sample prediction performance for different combinations of these two parameters. In this section, the Lake Garda reflectance dataset is used as an example and how to select the spatial decay parameter and the number of basis functions based on the spatial cross-validation will be

discussed.

The Lake Garda reflectance data were observed by MERIS and MODIS remote sensors at 211 locations from January 2009 to March 2012 (39 months in total). This dataset is provided by the Plymouth Marine Laboratory [Carrea et al., 2022], and the reflectance data are converted to the same spatial resolution for both of the MERIS and MODIS sensor. The NSD-GPP is used to predict the MERIS data based on the MODIS data at unobserved locations. First, the spatial locations are divided into training (190 locations) and prediction (21 locations) data sets, and an example of the locations for the training and the prediction datasets is shown in Figure 5.1. The parameters in the NSD-GPP model are estimated based on the training data set, and predictions for all time points within the time span are made for all prediction locations. The reflectance data are plotted across time for both training and prediction locations in Figure 5.2.

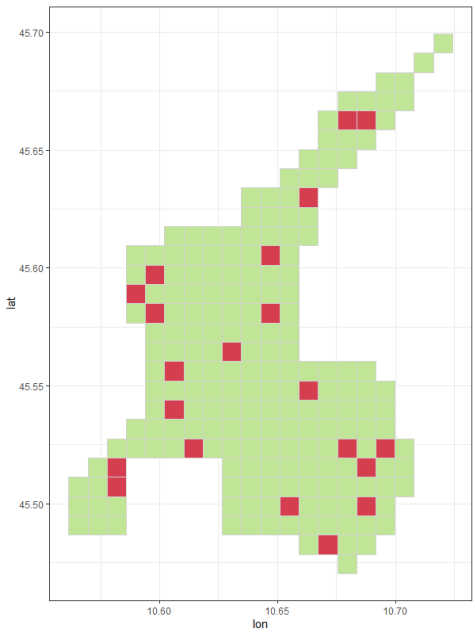


Figure 5.1: Map of Lake Garda. Red cell: prediction locations, green cells: training locations.

In Figure 5.2, there are approximately two or three peaks of the reflectance data in each year. In the paper of Wilkie et al. (2019), the number of basis functions is estimated as:

$$\text{Basis dimension} = (\text{total months} \times \frac{2 \times \text{r}}{12}) + 3, \tag{5.4}$$

where r is the expected number of peaks in each year [Wilkie et al., 2019]. Based on this equation, the assumptions with $\{1,2,3\}$ peaks per year will be explored, which leads to $\{10,16,23\}$ B-splines basis functions. Additionally, the 18 and 21 B-splines basis functions are added to explore this range further, since there are approximately two or three peaks per year from Figure 5.2.

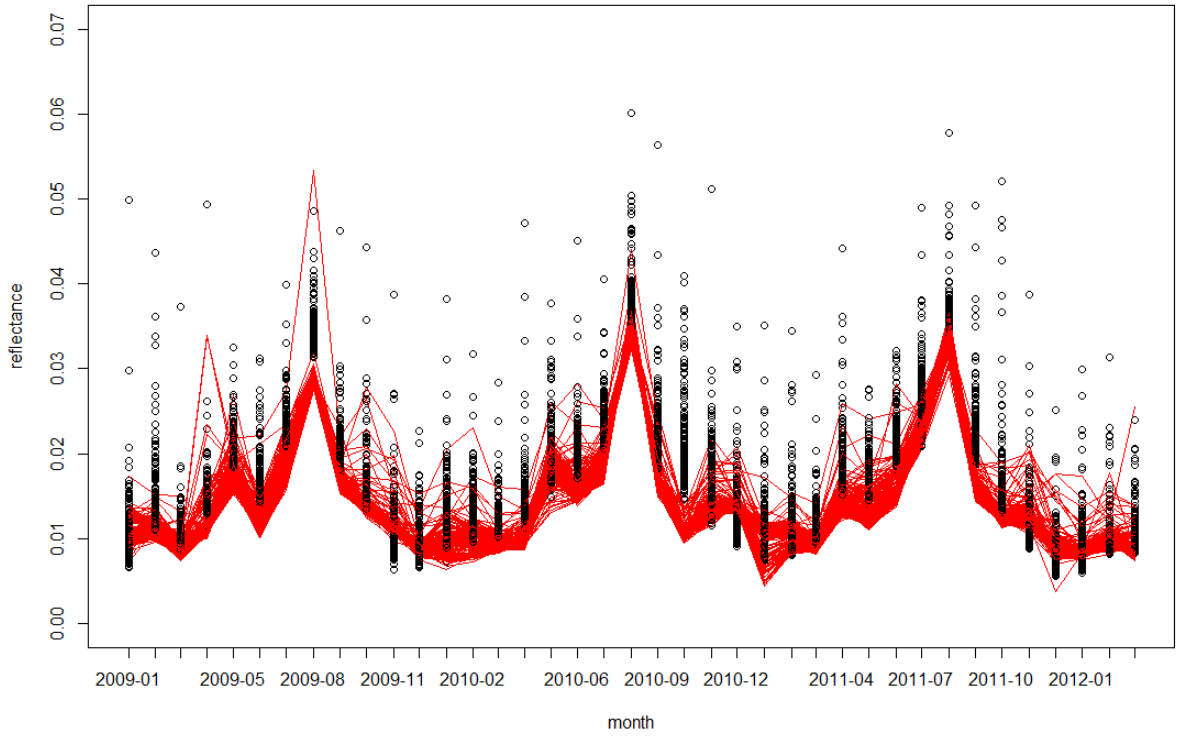


Figure 5.2: Reflectance data for all locations at Lake Garda from January 2009 to March 2012. Red lines: MERIS sensor, black points: MODIS sensor.

The spatial decay parameter controls the spatial correlation of the coefficients α_j for each basis function. The maximum distance between all pairs of all locations in Lake Garda is 0.256 degrees. If this distance is used as the effective range (semi-variogram reaches 95 percent of the sill at this distance) of α_j, the spatial decay is approximately equals to 10 and this value represent a strong spatial correlation for Lake Garda. The effect of the decay parameters is explored within the range of $\{0.1, 1, 10, 100\}$.

The prior for the precision parameters $p_y, p_x, p_\alpha, p_\beta, p_c$ is selected based on the background knowledge of the water reflectance data. In this study, the water reflectance data is a percentage value and the value is within the range from 0 to 10% [Huete, 2004]. The study of remote sensing data of lakes in South-Eastern Estonia suggests that the water reflectance data at wavelength 400 nm are within 0 to 1.5% [Kutser et al., 2016]. Another study of remote sensing reflectance data for Lake Dobczyce in southern Poland also shows the reflectance data at wavelength 400 are within 0 to 2% [Bielski and Toś, 2022]. Thus, the prior for variance parameter σ_y^2 of the water reflectance data is assumed to have a mean of 0.01, which is the same scale as the general lake water reflectance data at wavelength 400 nm and infinite variance, which leads to $\sigma_y^2 \sim \text{inverse} - \text{Ga}(2, 0.01)$ and $p_y \sim \text{Ga}(2, 0.01)$. The same prior is used for all these precision parameters, which means $a_y = a_x = a_\alpha = a_\beta = a_c = 2$ and $b_y = b_x = b_\alpha = b_\beta = b_c = 0.01$. The

sensitivity analysis for the prior choice of the precision parameters will be discussed in Section 5.3.3.

The equally spaced knots are used for the NSD-GPP model. The effect of the number of knots in the NSD-GPP model is explored in this study. For the NSD-GPP model with 25 knots, 49 knots and 100 knots are used for the Lake Garda dataset, which correspond to approximately 10%, 25% and 50% of the total locations in Lake Garda.

Then the NSD-GPP model and the NSD model are fitted with different combinations of spatial decay parameters and the number of basis functions. The parameters are estimated by Gibbs sampling with 20000 iterations with every ten iterations recorded, and the first 2000 iterations are treated as the burn-in period. The convergence of the Markov chains are checked by Geweke's diagnostic, and the results do not show evidence against the convergence. The assumptions of the model are checked by the residual plot and the QQ-plot in Figure 5.3, and the plots show that the residuals are approximately normally distributed and have a constant variance. The autocorrelation plot in Figure 5.4 for a randomly selected location shows no strong temporal correlation after lag 1. The variogram plot in Figure 5.4 for a randomly selected month suggests no significant spatial correlation in the residuals. The autocorrelation plots for the other locations and the variogram plots for the other months have similar results and do not show strong evidence against the assumption of independent residuals.

The root mean square prediction error (RMSPE), bias, the coverage and the width of the 95 percent credible interval of the predictions are summarised in the Table 5.1. Additionally, the percentage difference, the spatial decay parameter ϕ for the minimum RMSPE, and the computational time are also included. Since the number of spatial knots in the NSD-GPP model is selected before the model fitting, the percentage difference is used to compare the predictive performance of the NSD-GPP model with different numbers of knots. The percentage difference is calculated by the difference of the RMSPE between the NSD-GPP model and the NSD model divided by the RMSPE of the NSD model. Since the spatial decay parameters are selected by minimising the RMSPE from 10-fold cross-validation and there are little differences of the RMSPE between the results from these set of spatial decay parameters, only the results corresponding to the minimum of the RMSPE for each model are shown in Table 5.1. The computational time represents the running time of the NSD-GPP model and the NSD model for one iteration of the ten-fold cross-validation, and this value is the average over the ten iterations for robustness.

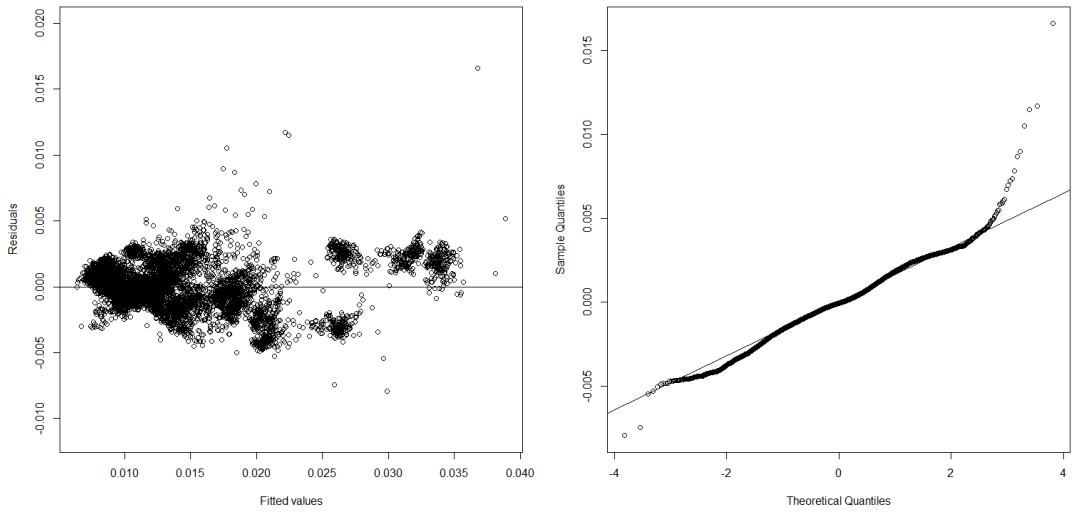


Figure 5.3: Left panel: Residuals versus fitted values for the NSD-GPP model with 25 knots, Right panel: QQ-plot for the residuals of the NSD-GPP model with 25 knots.

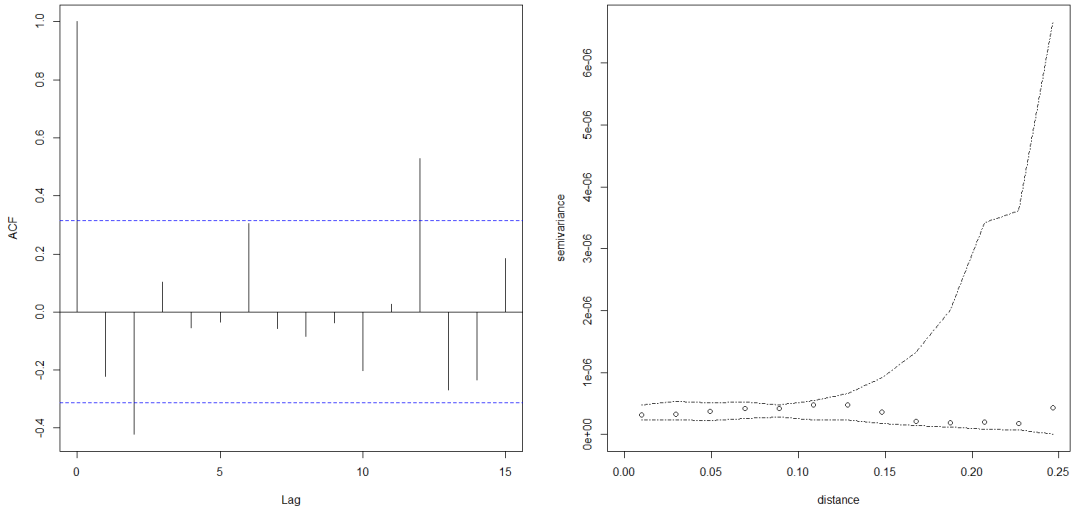


Figure 5.4: Left panel: ACF for the residuals at a random location of the NSD-GPP model with 25 knots ,Right panel: Variogram with Monte Carlo envelope at a random month for the residuals of the NSD-GPP model with 25 knots.

basis functions	model	ϕ	rmspe	difference (%)	coverage (%)	bias	width	time (hour)
10	NSD-GPP 25 knots	10	0.0051	7.22	97.74	-0.001	0.035	0.35
	NSD-GPP 49 knots	10	0.0050	5.90	97.55	0.000	0.032	0.52
	NSD-GPP 100 knots	10	0.0050	5.64	97.25	0.000	0.031	0.98
	NSD	10	0.0047	0.00	97.39	0.000	0.028	1.12
16	NSD-GPP 25 knots	10	0.0047	13.03	99.14	-0.001	0.033	0.49
	NSD-GPP 49 knots	10	0.0046	12.46	98.61	-0.001	0.031	0.74
	NSD-GPP 100 knots	10	0.0047	13.12	98.37	-0.001	0.030	1.46
	NSD	10	0.0041	0.00	97.98	0.000	0.026	1.52
18	NSD-GPP 25 knots	10	0.0044	17.21	99.63	-0.001	0.032	0.54
	NSD-GPP 49 knots	10	0.0043	15.07	99.37	-0.001	0.030	0.82
	NSD-GPP 100 knots	1	0.0043	15.01	98.54	-0.001	0.027	1.56
	NSD	10	0.0038	0.00	98.56	0.000	0.024	1.67
21	NSD-GPP 25 knots	1	0.0042	20.29	99.63	-0.001	0.027	0.60
	NSD-GPP 49 knots	1	0.0042	17.88	99.69	-0.001	0.027	0.93
	NSD-GPP 100 knots	10	0.0042	18.58	99.77	-0.001	0.029	1.84
	NSD	10	0.0035	0.00	99.95	0.000	0.024	1.95
23	NSD-GPP 25 knots	10	0.0038	26.88	99.86	-0.001	0.030	0.64
	NSD-GPP 49 knots	10	0.0037	24.46	99.75	-0.001	0.028	1.06
	NSD-GPP 100 knots	10	0.0036	22.65	99.69	-0.001	0.027	2.09
	NSD	10	0.0030	0.00	99.94	0.000	0.022	2.23

Table 5.1: Spatial cross-validation results of the NSD-GPP model and the NSD model with different number of basis functions for the Lake Garda dataset. ϕ: the spatial decay parameters corresponding to the minimum of the RMSPE, difference: the percentage difference between the RMSPE of the NSD model and the NSD-GPP model, coverage: the coverage of the 95% credible interval, width: the average width of the 95% credible interval.

From this table, most of the NSD and NSD-GPP models with different numbers of temporal basis functions achieve their minimum RMSPE with $\phi = 10$ with only three exceptions. The spatial decay parameter $\phi = 10$ represents a strong spatial correlation of the parameters α_j and β_j in each model for the Lake Garda dataset. The percentage difference is calculated by the difference of the RMSPE between the NSD-GPP model and the NSD model, then divided by the RMSPE of the NSD model. A smaller value of the percentage difference presents a better approximation for the NSD-GPP model. The RMSPE decreases when the number of spatial knots increases in the NSD-GPP model. However, there are no substantial differences in the RMSPE between the NSD-GPP model with different number of knots. Take the model with 10 temporal basis function in Table 5.1 as an example, the percentage difference is similar between the NSD-GPP model with different number of spatial knots. This situation also happens in the models with 16, 18, 21, 23 temporal basis functions. In each month, the reflectance data are smooth and do not have much spatial variation. Thus, a small number of spatial knots in the NSD-GPP model will be enough to capture the spatial patterns for the Lake Garda dataset, and increasing the number of knots in the NSD-GPP model will not result in a large improvement

in the RMSPE for spatial cross-validation. The percentage difference also increases when the number of basis functions increases. Since the Gaussian predictive process is used for the coefficients of every basis function, there are more approximations for a larger number of basis functions. It can be seen from Table 1.1 that there is very little difference among the bias of the models. The coverage and the width of the 95% credible interval generally decrease when the number of knots in the NSD-GPP model and the values for the NSD model are smaller than the NSD-GPP model. Since the NSD-GPP model has less parameters than the NSD model, and a larger number of samples gives narrower interval. The ratio of the NSD-GPP model and the NSD model is similar for different number of basis functions. The computational time of NSD-GPP model with 25 knots is approximately 30% of the NSD model, and this ratio is about 50% for the NSD-GPP model with 49 knots. The NSD-GPP model with 100 knots requires a similar computational time as the NSD model, which suggests that the knots for the NSD-GPP model should be less than 100 for time saving. The reduction in computational time for the NSD-GPP compared to the NSD is likely to become even more pronounced if the NSDGPP model code is further optimised in C++. At the moment, the NSD-GPP has been written in R, while the matrix inversions are computed using C++ for the NSD model.

Another finding is that the RMSPE will decrease when the number of basis functions increases. This is because the temporal trend is similar for both training and test datasets based on the reflectance plot in Figure 5.2. Using a large number of basis functions to smooth the time series in each location will track the data in the training data set. Since the temporal trends are similar to those in the test set, the RMSPE in the prediction data set will be also small when using a large number of basis functions. The predictions of the MERIS reflectance data at a randomly selected location are plotted in Figure 5.5. As shown in this, the prediction line will also track the true values with a large number of basis functions. The minimum RMSPE is achieved with 29 B-splines basis functions. However, using a larger number of basis functions will have over-fitting problem, and this would not be the best choice for the dataset with different temporal patterns over different locations. Thus, an exploration of the temporal cross-validation with different numbers of basis functions will be shown in the next section.

5.3.2 Temporal cross-validation

In the above section, the results and the limitations of the spatial cross-validation of the NSD-GPP model and the NSD model for Lake Garda dataset are discussed. In this section, the effects of the spatial decay parameter and number of basis functions of the NSD-GPP model in temporal cross-validation are explored.

The temporal cross-validation method is also used to select the best combination of the number of basis functions and the spatial decay parameters which minimises the RMSPE. In the tempo-

ral cross-validation, the MERIS reflectance data of the 39 months are divided into ten groups, and each group has about four months of the reflectance data for all locations. Then, in each iteration, one group is selected as the prediction dataset , and the remaining nine groups are treated as the training dataset. Then, the NSD-GPP model and the NSD model are fitted to the training dataset, and the out-of-sample predictions are made for the prediction months for all locations.

The model fitting process and the checking for the convergence of the Markov chains and the assumptions of the NSD-GPP model and the NSD model are the same as the previous section of the spatial cross-validation. The results show no strong evidence against the convergence and the model assumptions.

The results of the prediction performance for temporal cross-validation are summarised in Table 5.2. The minimum of the RMSPE of the temporal cross-validation of the NSD-GPP model and the NSD model is achieved with $\phi = 10$ with one exception, which suggests the parameters α_j and β_j have a strong spatial correlation for the Lake Garda dataset. The results corresponding to the minimum of the RMSPE are shown in this table, but the RMSPE are similar with different spatial decay parameters. The RMSPE of NSD-GPP model and the NSD model increases when the number of basis functions increases up to 21, and then decreases for 23 B-splines basis functions. An example of the temporal predictions from the NSD-GPP model with 25 knots is shown in Figure 5.6, where the NSD model with 10 basis function could capture the temporal trend of the reflectance data, and the predictions are close to the true values.

The predictions from the model with more basis functions are also close to the true values. Considering both the RMSPE and the coverage of the 95% credible interval in Table 5.2, the 10 B-splines basis function is selected for temporal cross-validation for both of the NSD-GPP model and the NSD model for the Lake Garda dataset. The RMSPE decreases when the number of the knots of the NSD-GPP model increases, and the improvement in the percentage difference is larger for a larger number of basis functions. The coverage and width of the 95% credible interval are similar for the NSD model and the NSD-GPP model with different number of the knots. The computational time of the temporal cross-validation is longer than the spatial cross-validation in the previous section. Since in the temporal cross-validation, the model fitting values are computed for all locations for each month in every iteration of the Gibbs sampling, and the spatial cross-validation only computes the prediction for about 10% of the locations for each month. However, the ratio between the NSD-GPP model and the NSD model is similar in these two types of cross validation. The computational of temporal cross-validation for the NSD-GPP model with 25 knots is about 33% of the NSD model, and this ratio is about 50% for the NSD-GPP model with 49 knots, and the ratio is about 1 for the NSD-GPP model with 100

knots. From this table, the NSD-GPP model with 25 knots could achieve a similar RMSPE for temporal cross-validation with 10 B-splines basis functions, and it only require about $\frac{1}{3}$ of the computational time for the NSD model.

basis functions	model	ϕ	rmspe	difference (%)	coverage (%)	bias	width	time (hour)
10	NSD-GPP 25 knots	10	0.0051	1.28	92.89	0.000	0.024	0.60
	NSD-GPP 49 knots	10	0.0051	1.06	92.71	0.000	0.023	0.88
	NSD-GPP 100 knots	10	0.0051	1.21	92.91	0.000	0.023	1.78
	NSD	10	0.0051	0.00	94.04	0.000	0.024	2.01
16	NSD-GPP 25 knots	10	0.0063	3.74	88.82	0.000	0.023	0.84
	NSD-GPP 49 knots	10	0.0064	4.21	88.18	0.000	0.023	1.62
	NSD-GPP 100 knots	1	0.0064	4.20	88.10	0.000	0.023	2.43
	NSD	10	0.0061	0.00	90.08	0.001	0.024	2.51
18	NSD-GPP 25 knots	10	0.0069	8.10	86.53	0.000	0.022	0.89
	NSD-GPP 49 knots	10	0.0068	6.81	85.53	0.000	0.022	1.36
	NSD-GPP 100 knots	10	0.0068	6.00	86.73	0.000	0.022	2.55
	NSD	10	0.0064	0.00	89.76	0.001	0.023	2.74
21	NSD-GPP 25 knots	10	0.0074	15.48	86.18	-0.001	0.023	1.02
	NSD-GPP 49 knots	10	0.0073	13.24	83.90	-0.001	0.023	1.58
	NSD-GPP 100 knots	10	0.0070	8.68	85.62	0.000	0.022	3.12
	NSD	10	0.0064	0.00	89.74	0.000	0.024	3.31
23	NSD-GPP 25 knots	10	0.0066	18.88	87.69	0.000	0.022	1.1
	NSD-GPP 49 knots	10	0.0063	12.99	85.30	0.000	0.020	1.82
	NSD-GPP 100 knots	10	0.0062	11.34	84.65	0.000	0.021	3.59
	NSD	10	0.0056	0.00	90.35	0.000	0.022	3.83

Table 5.2: Temporal cross-validation results of the NSD-GPP model and the NSD model with different number of basis functions for the Lake Garda dataset. ϕ: the spatial decay parameters corresponding to the minimum of the RMSPE, difference: the percentage difference between the RMSPE of the NSD model and the NSD-GPP model, coverage: the coverage of the 95% credible interval, width: the average width of the 95% credible interval.

5.3.3 Prior sensitivity analysis

In section 5.2.3, the prior for the precision parameters $p_y, p_x, p_c, p_\alpha, p_\beta$ is set to be Gamma$(2, 0.01)$ based on the background of the lake water reflectance data from the study of other lakes. This prior represents that the corresponding variance parameters have a mean of 0.01 and infinite variance. This prior is used for both the NSD model and the NSD-GPP model. A sensitivity analysis is discussed in this section to check if the averaged root mean squared prediction error is sensitive to the choice of this prior.

The NSD-GPP model and the NSD model with 10 basis functions are selected as an example for sensitivity analysis. From the above sections, the minimum RMSPE of the NSD-GPP model

and the NSD model with 10 basis functions is achieved with $\phi = 10$. Thus, the spatial decay parameters are fixed to 10, and the models are fitted with different prior choices. Since there is no additional knowledge of the precision parameters of the NSD model and the NSD-GPP model, a weakly informative prior is used with an infinite variance for the prior of the corresponding variance parameters $\sigma_y^2, \sigma_x^2, \sigma_c^2, \sigma_\alpha^2, \sigma_\beta^2$. This means the shape parameters are set to be 2 in the inverse-Gamma distribution of prior for the variance parameters [Gelman et al., 2014]. Then, the scale parameters of the inverse-Gamma distribution are set to be $\{0.001, 0.01, 0.1\}$, which leads to the mean of the prior of the variance parameters equal to $0.001, 0.01, 0.1$ respectively. This is equivalent to using the Gamma prior for the precision parameters with a shape parameter equal to 2 and a rate parameter equal to $0.001, 0.01, 0.1$.

The results for the spatial cross-validation and the temporal cross-validation with 10 B-splines basis functions and $\phi = 10$ are summarised in Table 5.3 and 5.4. From these two tables, the RMSPE and the bias in both spatial cross-validation and temporal cross-validation have small difference with different prior choices for both of the NSD-GPP model and the NSD model. The coverage and the width of the 95% credible interval will increase when using prior Gamma$(2, 0.1)$, since this prior assume the variance parameters have a larger prior mean. Since the aim of this study is to minimise the RMSPE among different combinations of the basis functions and the spatial decay parameters, the results from these two tables suggest that the RMSPE is stable within the sensitivity analysis. Thus, the prior Gamma$(2, 0.01)$ is used for the precision parameters for both the NSD-GPP model and the NSD model for the reflectance datasets in this chapter.

prior	model	rmspe	coverage (%)	bias	width
Gamma	NSD-GPP 25 knots	0.0051	96.49	0.000	0.029
(2, 0.001)	NSD-GPP 49 knots	0.0051	96.44	0.000	0.029
	NSD-GPP 100 knots	0.0051	96.40	0.000	0.029
	NSD	0.0047	96.35	0.000	0.025
Gamma	NSD-GPP 25 knots	0.0052	96.49	0.000	0.029
(2, 0.01)	NSD-GPP 49 knots	0.0052	96.44	0.000	0.029
	NSD-GPP 100 knots	0.0052	96.40	0.000	0.029
	NSD	0.0047	96.35	0.000	0.025
Gamma	NSD-GPP 25 knots	0.0054	99.98	-0.001	0.051
(2, 0.1)	NSD-GPP 49 knots	0.0054	99.99	-0.001	0.051
	NSD-GPP 100 knots	0.0054	99.99	-0.001	0.050
	NSD	0.0048	99.98	0.000	0.046

Table 5.3: Summary table for the spatial cross-validation of the NSD-GPP model and the NSD model with 10 B-splines basis functions and $\phi = 10$ with different prior choices of the precision parameter.

prior	model	rmspe	coverage (%)	bias	width
Gamma (2, 0.001)	NSD-GPP 25 knots	0.0051	92.07	0.000	0.022
	NSD-GPP 49 knots	0.0050	92.15	0.000	0.022
	NSD-GPP 100 knots	0.0050	91.99	0.000	0.022
	NSD	0.0050	90.60	0.000	0.020
Gamma (2, 0.01)	NSD-GPP 25 knots	0.0051	92.89	0.000	0.024
	NSD-GPP 49 knots	0.0051	92.71	0.000	0.023
	NSD-GPP 100 knots	0.0051	92.91	0.000	0.023
	NSD	0.0051	94.04	0.000	0.024
Gamma (2, 0.1)	NSD-GPP 25 knots	0.0051	96.60	0.000	0.034
	NSD-GPP 49 knots	0.0051	96.47	0.000	0.034
	NSD-GPP 100 knots	0.0051	96.46	0.000	0.034
	NSD	0.0051	96.63	0.000	0.033

Table 5.4: Summary table for the temporal cross-validation of the NSD-GPP model and the NSD model with 10 B-splines basis functions and $\phi = 10$ with different prior choices of the precision parameter.

5.4 Simulation study

In the previous section, the NSD-GPP model is applied to the Lake Garda dataset, and results show that the NSD-GPP model could achieve a similar RMSPE as the NSD model with only 30% of the computational time. In order to apply the NSD-GPP model to the other lake water reflectance datasets with different spatial patterns compared to Lake Garda, the simulation study will be conducted in this section to see if these findings generalise.

The simulation study is used to compare the predictive performance of the NSD-GPP model and the NSD model. Additionally, the effect of the number of the knots of the NSD-GPP model on the spatial prediction will be investigated for the datasets with different spatial patterns. Since this simulation study changes the spatial patterns in different scenarios, the spatial cross-validation which leaves out the spatial locations in each iteration, is used to assess the predictive performance of these two models. The results of the spatial cross-validation will be summarised by root mean square prediction error (RMSPE), bias and coverage of 95 % credible interval.

In this section, the model used to simulate the datasets will be firstly introduced. Then, the nine different scenarios of the simulated datasets will be illustrated. Finally, the model fitting on the simulated datasets will be discussed, and the conclusions are drawn from the summary tables for all these nine scenarios.

5.4.1 Data simulation

The data are simulated to mimic the remote-sensing reflectance data at Lake Garda in this study. The formula of this simulated model is defined as:

$$
\begin{aligned}
\mathbf{y}_i &\sim \mathrm{N}_{q_i}\left(\alpha_i + \beta_i \mathbf{x}_i, \frac{1}{p_y}\mathbf{I}_{q_i}\right), \\
\mathbf{x}_i &\sim \mathrm{N}_{p_i}\left(\boldsymbol{\Psi}_i \mathbf{d}_i, \frac{1}{p_x}\mathbf{I}_{p_i}\right), \\
\mathbf{d}_j &\sim \mathrm{N}_n\left(\mathbf{0}, \frac{1}{p_d}\exp(-\phi_d \mathbf{D})\right), \\
\boldsymbol{\alpha} &\sim \mathrm{N}_n\left(\mathbf{0}, \frac{1}{p_\alpha}\exp(-\phi_\alpha \mathbf{D})\right), \\
\boldsymbol{\beta} &\sim \mathrm{N}_n\left(\mathbf{1}, \frac{1}{p_\beta}\exp(-\phi_\beta \mathbf{D})\right), \\
p_y &= p_x = 100.
\end{aligned}
\tag{5.5}
$$

In this formula, $\mathbf{y}_i$ and $\mathbf{x}_i$ are the MERIS and the MODIS reflectance data at location i. $\boldsymbol{\Psi}_i$ are the matrices of the Fourier basis functions with dimension three, which aims to generate $\mathbf{x}_i$ with one peak per year [Wilkie et al., 2019]. The Fourier basis functions could generate the data with regular temporal peaks in each year. $\mathbf{d}_i$ are the corresponding coefficients. α_{ij} and β_{ij} are the spatial varying intercepts and slopes for the regression model between the coefficients c_{ij} and d_{ij}. ϕ_d, ϕ_α, ϕ_β are the spatial decay parameters and $\mathbf{D}$ is the distance matrix between the n locations. p_y, p_x, p_α, p_β, p_d are the precision parameters. The precision p_y, p_x are the precision of the observed data $\mathbf{y}_i$ and $\mathbf{x}_i$. Since the reflectance data have the scale proportional to 0.01, $p_y = p_x = 100$ are used to simulate the data with relative large observed error. This model will generate the MODIS reflectance data $\mathbf{x}$ with spatially correlated coefficients $\mathbf{d}_j$ so that the simulated $\mathbf{x}$ are also spatially correlated.

The data are simulated within a square domain with 225 locations to mimic the number of locations in Lake Garda, which equals 211. The square shape of the simulated dataset aims to compare the NSD-GPP model and the NSD model with a more general spatial shape, so that the results are not strongly influenced by one particular shape of the dataset. The distance between two adjacent locations is 0.01 because the unit of the distance between reflectance data is the degree of longitude and latitude, and the distance between two adjacent locations in lake Garda is approximately equal to 0.01. The number of months for the simulated datasets is chosen to be 39, which is the same as the Lake Garda dataset.

5.4.2 Scenarios

The scenarios of the simulated datasets are considered based on two factors, which are the spatial variation and the spatial correlation of the reflectance data. In the Lake Garda dataset, the results show that the data have a strong spatial correlation, but do not have strong spatial variation over the lake. Thus, in order to generalise the NSD-GPP model to other reflectance datasets for spatial cross-validation, these two factors will be varied for different scenarios.

- **Spatial variation**. This is controlled by the precision parameters p_d, p_α and p_β. A small value of the precision parameter represents a large spatial variation for the coefficients $\mathbf{d}_j$, $\boldsymbol{\alpha}_j$, and $\boldsymbol{\beta}_j$. Since p_y, p_x are fixed to 100 for the simulated datasets, The spatial variance of the simulated dataset is controlled by p_d, p_α and p_β to be 10%, 50% or 100% of the p_y, p_x to represent different level of spatial variation. Thus, $p_\alpha = p_\beta = p_d = 10$ is used for a large spatial variation, $p_\alpha = p_\beta = p_d = 50$ is used for a moderate spatial variation, and $p_\alpha = p_\beta = p_d = 100$ is used for a small spatial variation.

- **Spatial correlation**. This is controlled by the spatial decay parameters ϕ_α, ϕ_β, ϕ_d. Since the coefficient $\mathbf{d}_j$, $\boldsymbol{\alpha}_j$, and $\boldsymbol{\beta}_j$ are generated from an exponential covariance function, the effective range for an isotropic exponential covariance matrix is defined as the minimum distance where the two observations have a correlation greater than 0.05 [Irvine et al., 2007]. The effective range ξ is calculated by:

  $$\xi = -(1/\phi)\log(0.05)$$

  Thus, the spatial decay parameter could be estimated based on the effective range by:

  $$\phi = -(1/\xi)\log(0.05)$$

  The minimum distance between all pairs of the locations in the simulated dataset is 0.01 degrees. If the correlation between two observations is less than 0.05 when their distance is 0.01 in the simulation data, then any two observations do not have a significant correlation even for the nearest pair. This means the spatial correlation is weak for the dataset, and this gives $\phi_\alpha = \phi_\beta = \phi_d = 300$. The maximum distance between all pairs of the locations in the simulated dataset is 0.20 degrees. If the furthest two observations still have a correlation of 0.05, the spatial correlation of the dataset is strong. In this case, $\phi_\alpha = \phi_\beta = \phi_d = 15$ represents a strong spatial correlation. Similarly, the medium distance between all pairs of the locations in the simulated dataset is 0.076 degree, which suggests $\phi_\alpha = \phi_\beta = \phi_d = 40$ for a moderate spatial correlation.

Three situations are considered for each of these two factors, which give us nine different combinations (summarised in Table 5.5). For each scenario, 100 datasets with different random

seed are generated and the output statistics are averaged over these 100 datasets. An example of the simulated y_i and x_i are shown in Figure5.7 and Figure 5.8. Figure5.7 is used to show the observed data with a fixed precision of $p_\alpha = p_\beta = p_d = 10$ but different values of spatial decay parameters ϕ_α, ϕ_β and ϕ_d. From the first column of Figure5.7, the observed data are less spatially smooth with a larger number of spatial decay. Figure5.8 shows observed data with different levels of spatial variation. Since the spatial decay is fixed in Figure5.8, the spatial smoothness is the same among these three spatial maps. The spatial variation decreases when the precision increases based on the range of the reflectance data in each panel.

Parameter / Scenario	$\phi_d, \phi_\alpha, \phi_\beta$	p_d, p_α, p_β
1	15	10
2	15	50
3	15	100
4	40	10
5	40	50
6	40	100
7	300	10
8	300	50
9	300	100

Table 5.5: Different scenarios of the simulated data set. ϕ_d, ϕ_α, ϕ_β: spatial decay parameters. p_d, p_α, p_β: precision parameters.

5.4.3 Model fitting

From the temporal plot above, there is approximately one peak per year for the reflectance dataset, and the study of the Lake Garda dataset suggests that ten B-splines basis functions would be enough for one peak per year. Thus, ten B-splines basis functions are used for all the NSD-GPP model and the NSD model in this simulation study.

Then, the ten-fold cross-validation is carried out with the NSD model and the NSD-GPP model to make the out-of-sample spatial predictions. The root mean square prediction error (RMSPE), coverage of the 95% credible interval, bias, width of the 95% credible interval and computational time will be used for model comparison. The NSD-GPP model with 25, 49 and 100 knots are considered in this study which are approximately 10%, 25% and 50% of the total observed data locations in the cross-validation process. The simulated locations and the knots for these NSD-GPP models are shown in Figure 5.9. The process of simulating the datasets and summarising the results are shown in the following algorithm. For each scenario, 100 datasets with different random seed are generated. Then, for each dataset, the ten-fold cross-validation is used to select the optimal ϕ_α and ϕ_β for each model with the minimum of the RMSPE and then

record the corresponding coverage, bias and the width of 95% credible interval. This process is repeated for 100 datasets, and the criteria are averaged to summarise the results in a table. There will be nine tables for these nine different scenarios.

Algorithm 1 Simulation study

```
for Scenario= 1 to 9 do
   for s=1 to 100 do
      set random seed s, generate one dataset.
      for     Model in {NSD, NSD-GPP 25 knots, NSD-GPP 49 knots,
NSD-GPP 100 knots } do
         for φ in {0.1, 1, 10, 100, 1000} do
            run the 10-fold cross validation
         end for
         record the results corresponding to the minimum of the
RMSPE among different input of φ
      end for
   end for
   Average the RMSPE, cover, bias, width of credible interval
over 100 datasets for each model.  Summarise the results in a
table.
end for
```

Model fitting check

Both NSD model and the NSD-GPP model are fitted using Gibbs sampling with 20000 iterations with the burnin of the first 2000 iterations and the Markov chains are recorded in every 10 iteration for saving computer memory. The convergence of the Markov chain is checked by the Geweke's diagnostic, which does not show evidence against the convergence of the chain. The assumptions of the residuals for the NSD-GPP model are checked by the residual plots as follows. An example of the NSD-GPP model with 5 knots and $\phi_\alpha = \phi_\beta = 0.1$ is shown below, and the simulated data are generated with the parameters of the scenarios 1 in the Table 5.5. In Figure 5.10, the residuals versus fitted values plot shows no evidence against the assumptions of zero mean and constant variance, and the QQ-plot suggests the residuals are approximately normally distributed. The spatial-temporal correlation of the residuals is checked by the ACF plot at each location and the variogram plot for each month. The ACF at a random location and the variogram plot at a random month in Figure 5.11 are used for illustration. Since the ACF are within the 95% confidence interval after lag 1 and the semivariogram are within the Monte Carlo envelopes, there are no significant temporal correlation at location 1 and no significant spatial correlation for month 1. The ACF and variogram plots for the rest of the locations and months are also checked, the exceptions only occur at 2 month, which are about 5% of the total months. This could be a result of the small size of the simulated dataset. Thus, there are no strong evidence against the assumption that the residuals are spatially and temporally uncorrelated.

5.4.4 Results

After check the model fitting of the NSD model and the NSD-GPP model, the prediction performance of these models are compared by the summary statistics such as the RMSPE, bias, coverage of the 95% credible interval, width of the 95% credible interval. Take one dataset in scenario 1 as an example, the predictions at a randomly picked location from the four models are shown in Figure 5.12. The predictions (black lines) capture the trend over time for all these four models and the major differences between these four plots appear at the starting and ending period of the time axis. There are no strong differences in the prediction plots between the NSD model and the NSD-GPP model with different numbers of knots.

model	NSD-GPP 25 knots	NSD-GPP 49 knots	NSD-GPP 100 knots	NSD
computational time (hour)	0.36	0.55	1.02	1.15

Table 5.7: Computational time of the NSD-GPP model and the NSD-GPP model for the spatial cross-validation of the simulated dataset.

The results of the average over 100 datasets in each scenario are summarised in Table 5.6, and the computational time for each model is shown in Table 5.7. For each dataset, only the result corresponding to the minimum of the RMSPE with different spatial decay parameters is recorded. Since the spatial decay parameters among these 100 datasets is not the focus of this study and they do not vary much across the different models corresponding to the minimum of the RMSPE, the spatial decay parameters are not included in this table. The bias for the NSD-GPP model and the NSD model are almost the same up to three decimal place. The coverage and the width of the 95% credible interval have the similar pattern as the spatial cross-validation for Lake Garda dataset, where the coverage and the width decrease with a larger number of knots. This result suggest that NSD-GPP model with a larger number of knots will achieve a better predictive performance for the datasets with different spatial patterns.

The first finding is that the NSD-GPP model with 25 knots could achieve a similar RMSPE but require about 30% of the time of the NSD model for the simulation datasets with different scenarios. From Table 5.6, the NSD-GPP model approximates the NSD relatively well in the RMSPE of the spatial cross-validation for these nine scenarios since the maximum percentage difference is 13.084%. The computational time of the NSD model and the NSD-GPP model with different number of knots in one single run of the cross-validation process in R with personal computer is show in Table 5.7. From this table, the computational time of the NSD-GPP model with 25 knots is about one third of the NSD model, the time of NSD-GPP model with 49 knots is about half of the NSD model, and the time of the NSD-GPP model is almost the same as the NSD model. This is because the R-code NSD-GPP model has not been optimised with C++ for the matrix inversion as the NSD model. However, it still requires less computational

time and achieves the percentage difference of RMSPE less than 13% for all scenarios.

The second finding from the tables is that, when the precision parameter fixed, the improvement with larger number of knots for NSD-GPP model becomes smaller in the dataset with weaker spatial correlation. Take scenarios 1, 2, 3 as an example, the precision for parameter these three scenarios are 10, but the spatial decay parameters are 15, 40, and 300 respectively. Thus, the spatial correlation becomes weaker from scenarios 1 to 3. The percentage difference becomes smaller with a larger number of knots in scenarios 1, 2, 3. However, the improvement becomes smaller from scenarios 1 to 3. In scenario 1, the percentage difference of NSD-GPP model with 25 knots is 10.8%, and the percentage difference of NSD-GPP model with 100 knots is 3.5%. The improvement of percentage difference between 25 knots and 100 knots is 7.3%, and this improvement is 2.3% for scenario 2 and 2.5% for scenario 3. Scenario 3 is an extreme case when the correlation between $\mathbf{d}_j$, $\boldsymbol{\alpha}_j$, and $\boldsymbol{\beta}_j$ in the simulation model with Equation 5.5 is less than or equal to 0.05 beyond the minimum distance of all pairs of the locations in the simulated dataset. This situation could also been observed in the percentage difference for scenarios 4, 5, 6, but this situation is not obvious when comparing scenarios 7, 8, 9. Again, the scenarios 7, 8, 9 are also extreme cases when the nugget and the partial sill of the covariance are the same, which suggests the spatial variation is small. In general, increasing the number of knots in the NSD-GPP model will improve the prediction performance compared to the NSD model for the datasets with relatively stronger spatial correlation of the reflectance data.

The third finding is that the percentage difference becomes smaller for datasets with larger precision parameters. Take scenarios 1, 4, 7 as an example, the percentage difference becomes smaller from scenarios 1 to 7 for any type of the model. The possible reason is that, when the precision parameter becomes larger, the nugget and the partial sill of the covariance becomes closer. The partial sill is ten times of the nugget for scenario 1, the partial sill ten 10 times of the nugget for scenario 4, and the partial sill is the same as the nugget for scenario 7. Thus, when the partial sill becomes smaller, the spatial variation becomes smaller compared to the random error in the data. This makes it harder for the NSD-GPP model to distinguish the spatial and non-spatial variation in the dataset with the larger precision.

In conclusion, the NSD-GPP model could achieve the similar predictive performance as the NSD model but require less computational time. Increasing the number of knots will improve the predictive performance of the NSD-GPP model when the data have relatively large spatial variation and strong spatial correlation. From these scenarios, NSD-GPP model with 25 knots and 49 knots approximate the NSD model well, and the NSD-GPP model with 100 knots require almost the same time as the NSD model but does not improve the RMSPE much. Considering the computational time and the improvement of the RMSPE, the NSD-GPP model with a num-

ber of knots less to one-fourth of the total number of locations will be considered.

5.4.5 Simulation study for dataset with 625 locations

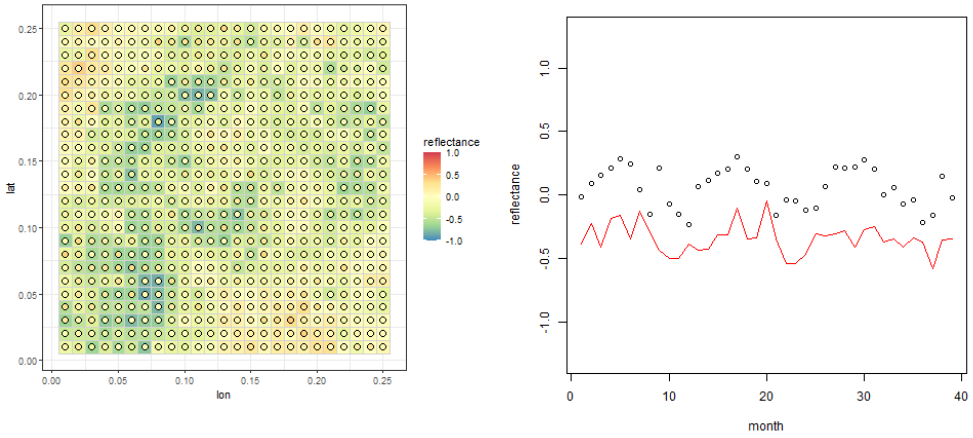
In the simulation study above, the NSD-GPP model and the NSD model are compared on the datasets with 225 locations in different scenarios. Since some real lake water reflectance datasets have a large number of locations, a simulation study with 625 locations will be investigated in this section to compare the computational time and the predictive performance of these two models. The aim of this study is to investigate the computational time of the NSD-GPP model with different numbers of knots compared to the NSD model.

Since the computational time will increase with the number of the locations of the dataset for both the NSD model and NSD-GPP model, the dataset with 625 locations are simulated with 20 different random seeds, and the results are averaged over these 20 datasets. Additionally, only one scenario is investigated in this small study, which uses $\phi_\alpha = \phi_\beta = \phi_d = 9$ and $p_\alpha = p_\beta = p_d = 10$ in the Equation 1.5. This scenario represents the simulated data with relatively strong spatial correlation and large spatial variation.

The prior choice of the precision parameter p_a is Gamma$(2, 0.01)$, which is the same as that used in the previous study for the Lake Garda dataset and the simulation of the datasets with 225 locations. The B-splines basis functions with dimension of 10 is used in this study for the assumption of one peak in the reflectance data per year from the temporal plot in Figure 5.13. The prior choice of input spatial decay parameter is $\{0.1, 1, 10, 100\}$, which is the same as that used in above study, and the optimal spatial decay parameter is selected by the spatial cross-validation method. The number of knots in the NSD-GPP model is selected as an approximation of 10%, 25% and 50% of the total number of the locations. Thus, the knots choices will be $\{64, 144, 289\}$ for a datasets with 625 locations.

The model fitting is checked by the residual plots and the convergence of the Markov chains is checked by Geweke's diagnostic. The plots of the predictions at a randomly selected location are shown in Figure 5.14, and the results of the predictive performance are summarised in Table 5.8. There is a systematic difference between the simulated MERIS and MODIS data in Figure 5.14 because Equation 5.5 assumes a linear relationship between these two sensors with a fixed α and β for each location. The RMSPE and the coverage of 95% credible interval decrease as the number of knots in the NSD-GPP model increases. The percentage difference of RMSPE between the NSD-GPP model with 64 knots and the NSD is about 11% and the computational time of the NSD-GPP model is about 17% of the NSD model. This suggests that using 10% of the total number of locations as the knots of the NSD-GPP model could achieve a similar RMSPE in the spatial cross-validation for the reflectance data. Another finding is that, when

the ratio between the number of knots and the total locations is fixed, the computational time ratio between the NSD-GPP model and the NSD model will decrease. In the simulation study of 225 locations, the time ratio of the computational time is about 30%, and this ratio decreases to 17% for the dataset with 625 locations. This simulation study suggests that using 10% of the locations as knots in the NSD-GPP model would result in a significant reduction of computational time for the larger datasets. This would be useful for larger lakes with thousands or ten thousands locations.



(a) spatial plot in the first month (b) temporal plot at location 1

Figure 5.13: Temporal and spatial plots for simulated data with 3 Fourier basis functions, $\phi_\alpha = \phi_\beta = \phi_d = 9$, $p_\alpha = p_\beta = p_d = 10$. The red lines are the MERIS data and the black dots are the MODIS data for panel. The color in the square represents the MERIS data and the color in the circle represents the MODIS data.

model	rmspe	difference (%)	coverage (%)	bias	width	time (hour)
NSD-GPP 64 knots	0.161	11.034	97.186	0.000	0.715	4.18
NSD-GP 144 knots	0.155	6.897	96.730	0.000	0.664	9.38
NSD-GP 289 knots	0.152	4.828	96.093	0.000	0.633	18.69
NSD	0.145	0.000	93.973	0.000	0.548	23.88

Table 5.8: Spatial cross-validation results of the NSD-GPP model and the NSD model for the simulation datasets with 625 locations. difference: the percentage difference between the RMSPE of the NSD model and the NSD-GPP model, coverage: the coverage of the 95% credible interval, width: the average width of the 95% credible interval.

5.5 Conclusion

In this chapter, an approximation method for the NSD model with a Gaussian predictive process is developed, which represents the spatial process in the NSD model with a lower-dimensional

process times a transformation matrix. This NSD-GPP model uses a smaller number of knots compared to number of observer locations, which could reduce the computational time in the Gibbs sampling process in the parameter estimations for the NSD model.

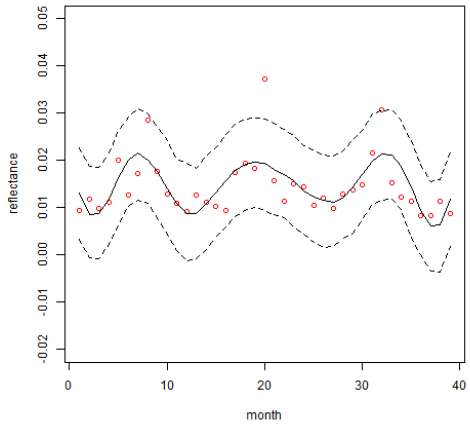
In the application of the NSD-GPP model to the Lake Garda dataset, the predictive performance and the computational time are compared to the NSD model. For both the spatial cross-validation and the temporal cros-validation, the results from the 10 fold cross validation suggest the NSD-GPP model with 25 knots and 10 B-splines basis function could achieve a similar RMSPE as the NSD model but use about 30% of the computational time. Since the temporal patterns are similar among different locations, the temporal cross-validation with ten-fold cross-validation would be a better method to select the number of basis functions for both the NSD model and the NSD-GPP model.

To extend the use of the NSD-GPP model for the datasets with different spatial patterns, a simulation study with different scenarios of the spatial correlation and spatial variation is presented. The results suggest that increasing the number of knots in the NSD-GPP model will decrease the RMSP, and this effect is more obvious for the dataset with strong spatial correlation and large spatial variation. The simulation study of the 625 locations also suggests that use 10% of the number of total locations as knots in the NSD-GPP model could a achieve a similar RMSPE for the dataset with strong spatial correlation and large spatial variation. Additionally, when the ratio of the knots in the NSD-GPP model and the number of total observed locations is fixed, the time ratio between the NSD-GPP model and the NSD model will decrease as the number of total locations increases. This would benefit the application to the lakes with a large number of locations. This NSD-GPP model could also be applied to other spatial-temporal dataset. However, the number of basis functions, the spatial decay parameters will rely on the initial exploratory analysis. The prior choice of the precision parameter requires the background knowledge, and this choice could also be checked by a sensitivity analysis.

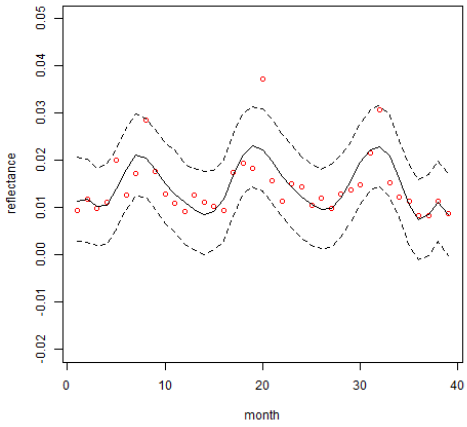
In this chapter, the training and test groups of the spatial cross-validation are randomly sampled over the observed locations. However, this validation method will result in poor prediction performance because of the spatial dependence between the training and test groups [Ploton et al., 2020]. Robert et al.(2017) suggested a blocking method, where the data are split into training groups and test groups are distinct blocks to increase the independence between these groups [Roberts et al., 2017]. This blocking method will increase the prediction error estimates and possibly improve the model's predictive performance at new locations.

One limitation of the NSD-GPP model is that the number of spatial knots is selected before the model fitting, which is estimated by the cross-validation method. In the application, this re-

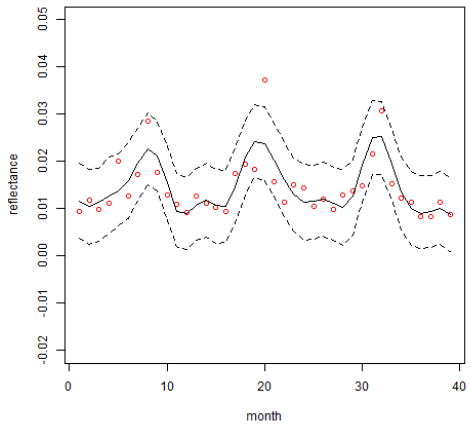
quires the users to fit the NSD-GPP models with different number of the spatial knots and then select the number based on both the predictive performance and the computational time. Based on the simulation study in this chapter, it is suggested to start from the knots with 10% and 25% of total locations, considering both the prediction performance and the computational efficiency. In this chapter, the fusion of reflectance data at a single wavelength band is explored. In the next chapter, the NSD-GPP model will be extended to handle the fusion of reflectance data over wavelength.



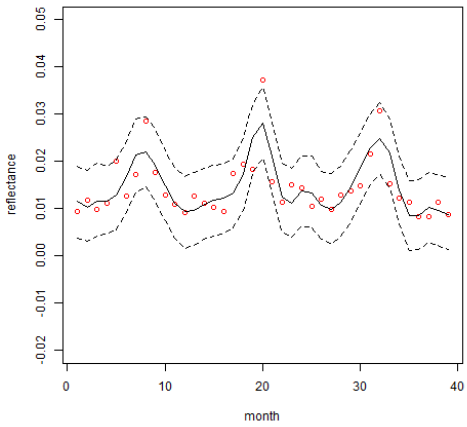
(a) 10 B-splines basis



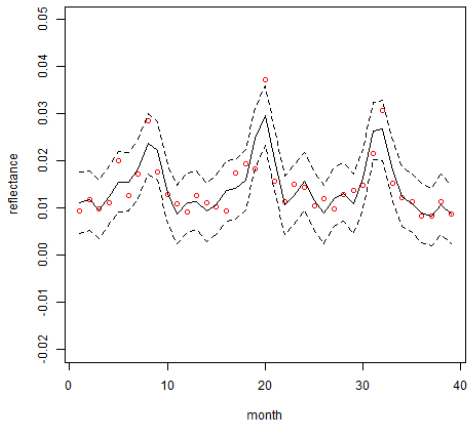
(b) 16 B-splines basis



(c) 18 B-splines basis

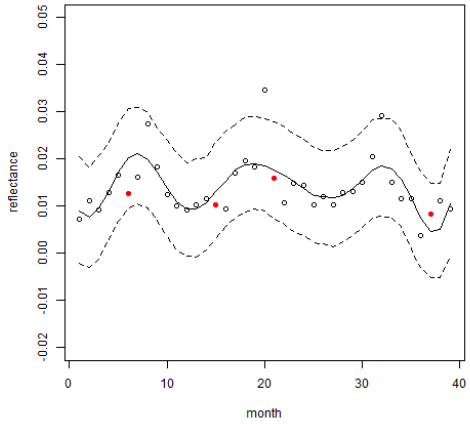


(d) 21 B-splines basis

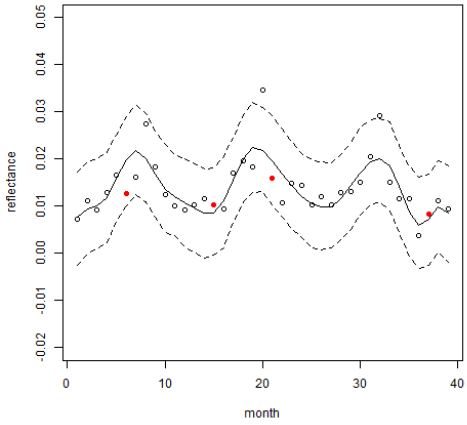


(e) 23 B-splines basis

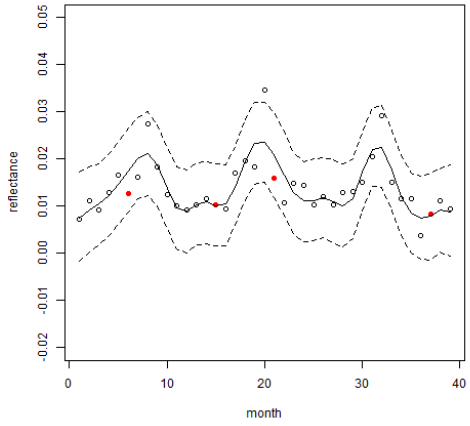
Figure 5.5: Spatial predictions from the NSD-GPP model with 25 knots at a randomly selected location with 10, 16, 18, 21, 23 B-splines basis function spatial decay equal to 10. Red circles: true values of the MERIS reflectance, Black solid lines: predictions from the NSD-GPP model, Black dashed lines: 95 percent credible interval of the predictions.



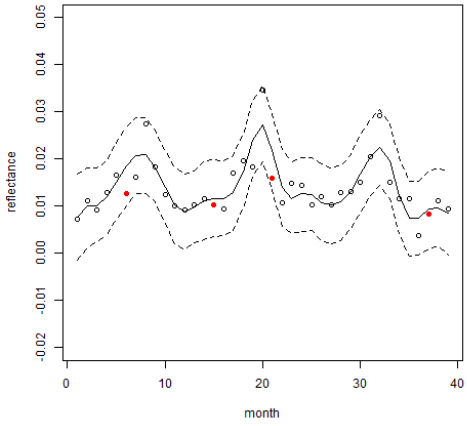
(a) 10 B-splines basis



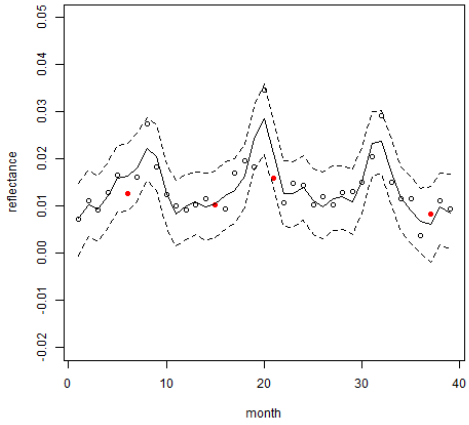
(b) 16 B-splines basis



(c) 18 B-splines basis

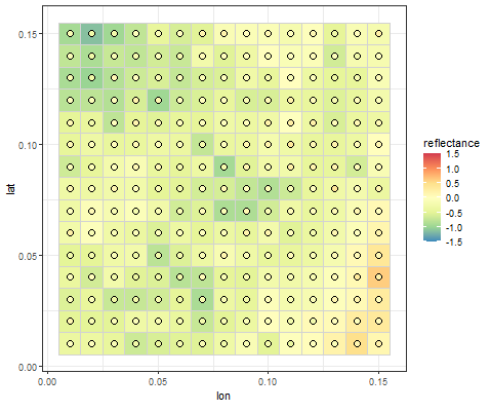


(d) 21 B-splines basis

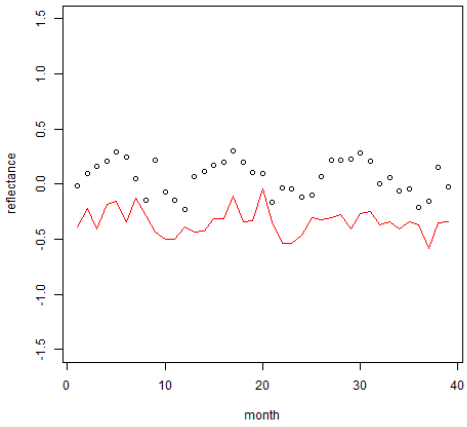


(e) 23 B-splines basis

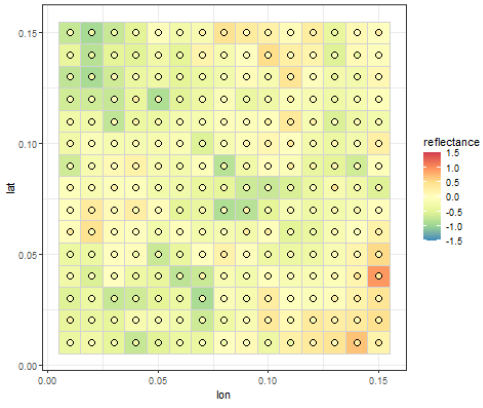
Figure 5.6: Temporal predictions from the NSD-GPP model with 25 knots at a randomly selected location with 10, 16, 18, 21, 23 B-splines basis function and spatial decay equal to 10. Black circles: observed MERIS reflectance data, Red points: true values of the unobserved MERIS reflectance, Black solid lines: predictions from the NSD-GPP model, Black dashed lines: 95 percent credible interval of the predictions.



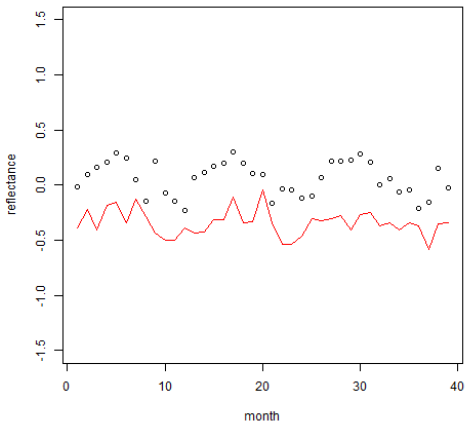
(a) spatial plot in the first month, $\phi = 15$



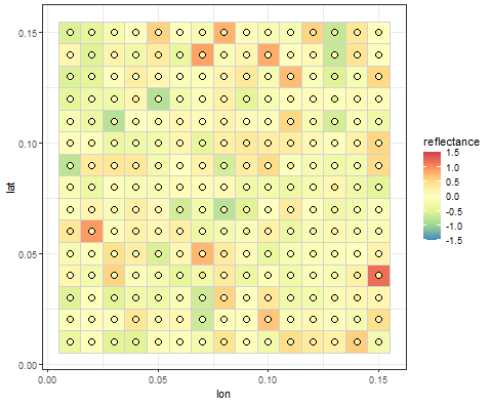
(b) temporal plot at location 1, $\phi = 15$



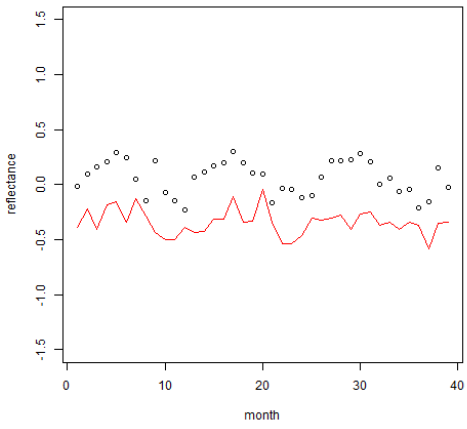
(c) spatial plot in the first month, $\phi = 40$



(d) temporal plot at location 1, $\phi = 40$

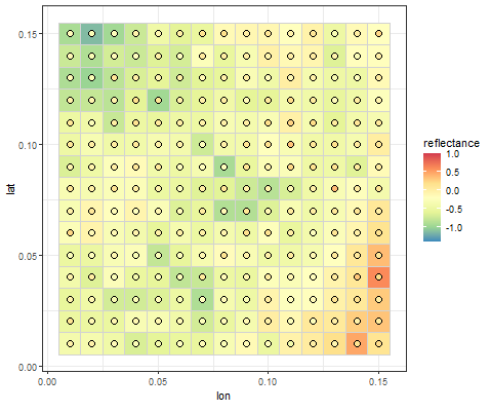


(e) spatial plot in the first month, $\phi = 300$

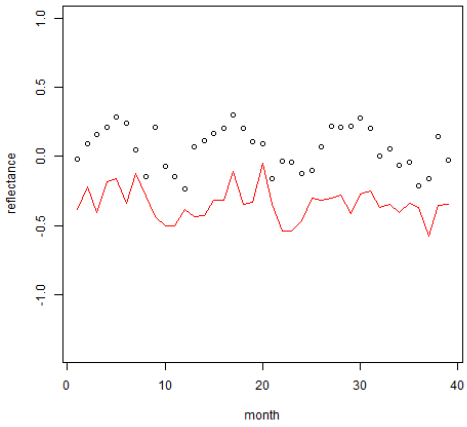


(f) temporal plot at location 1, $\phi = 300$

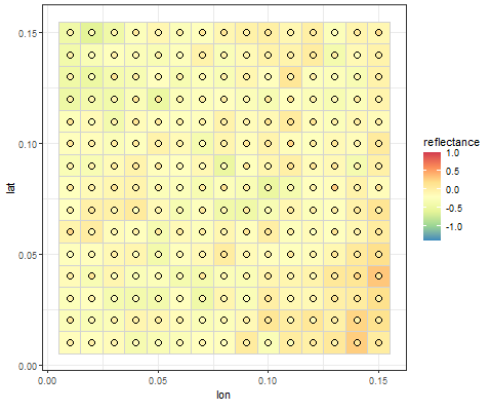
Figure 5.7: Temporal and spatial plots for simulated data with three Fourier basis functions, $p_\alpha = p_\beta = p_d = 10$. The red lines are the MERIS data and the black dots are the MODIS data for panel. The color in the square represents the MERIS data and the color in the circle represents the MODIS data.



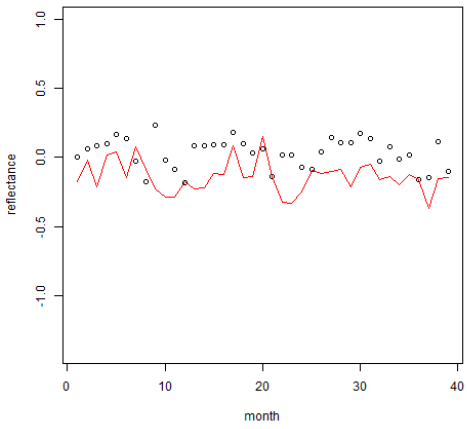
(a) spatial plot in the first month, $p = 10$



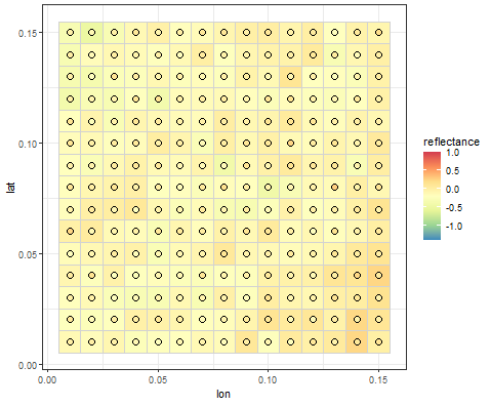
(b) temporal plot at location 1, $p = 10$



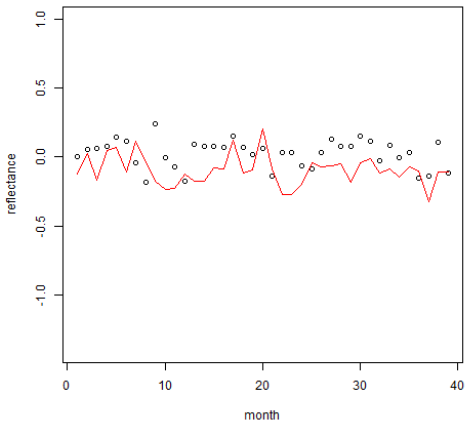
(c) spatial plot in the first month, $p = 50$



(d) temporal plot at location 1, $p = 50$



(e) spatial plot in the first month, $p = 100$



(f) temporal plot at location 1, $p = 100$

Figure 5.8: Temporal and spatial plots for simulated data with 3 Fourier basis functions, $\phi_\alpha = \phi_\beta = \phi_d = 15$. The red lines are the MERIS data and the black dots are the MODIS data for panel. The color in the square represents the MERIS data and the color in the circle represents the MODIS data.

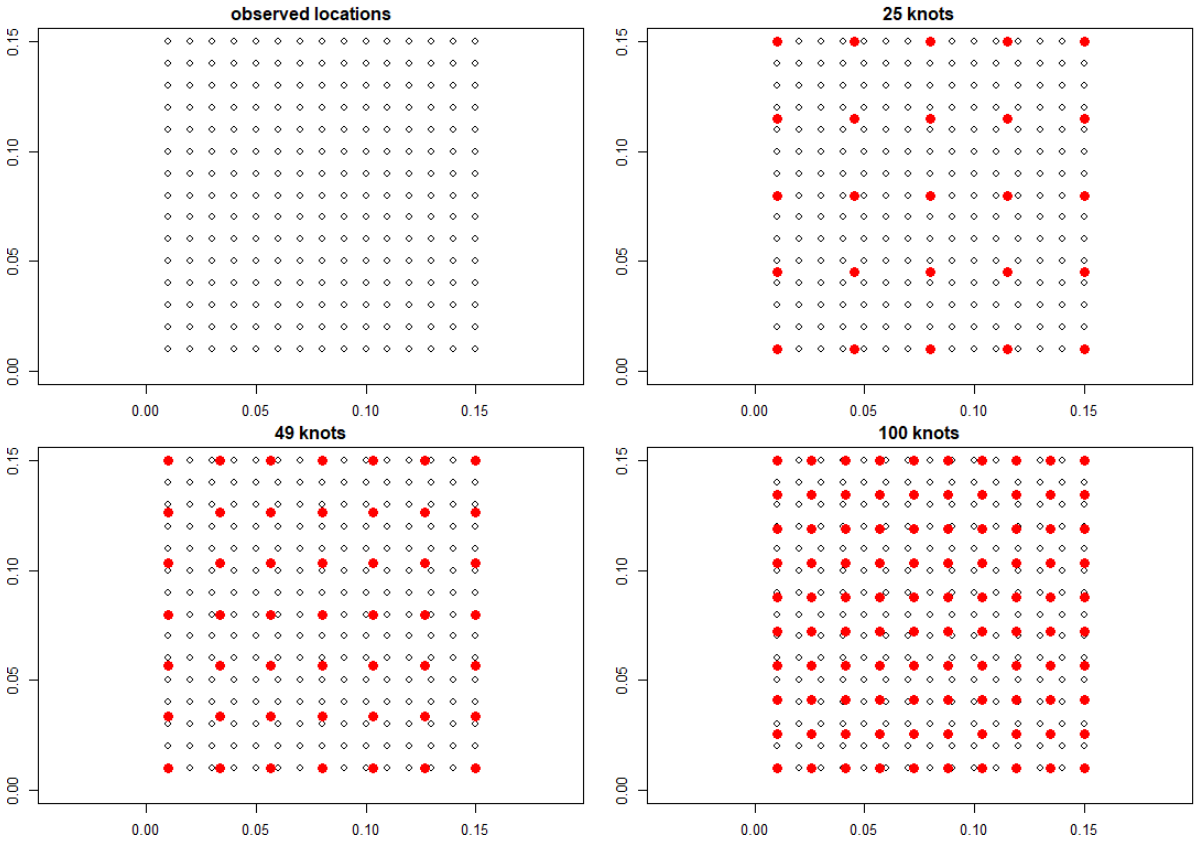
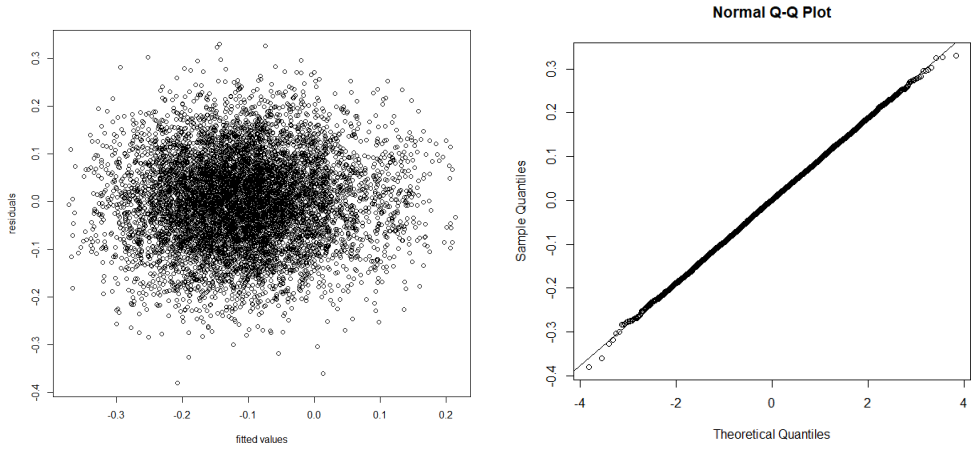
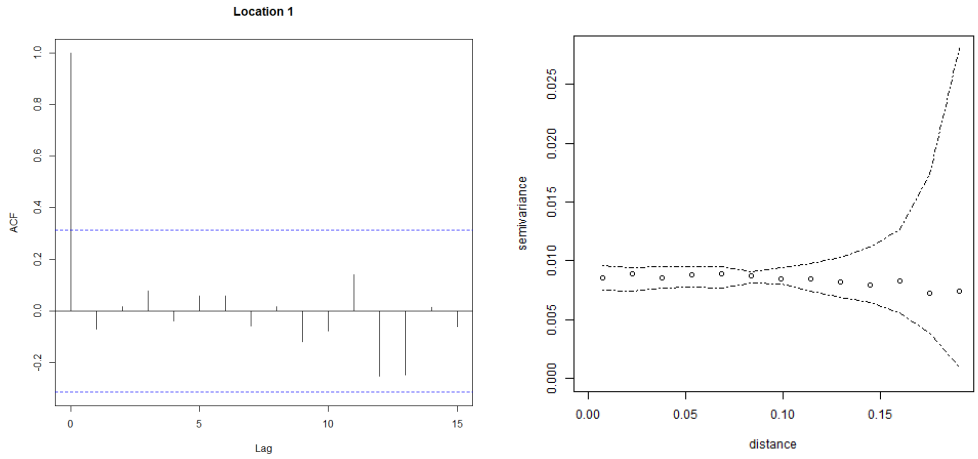


Figure 5.9: Locations of the simulated data (black circle) and the knots (red solid circle) for NSD-GPP model. Top left panel: locations of the 225 simulated data. Top right panel: locations of the 25 knots. Bottom left panel: locations of the 49 knots. Bottom right panel: locations of the 100 knots.



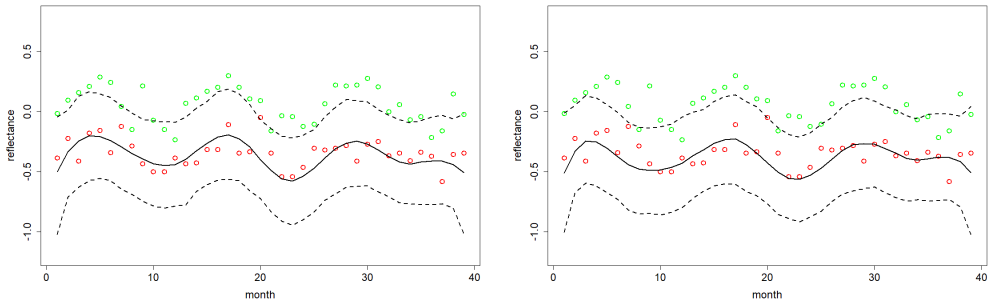
(a) (b)

Figure 5.10: (a):Residuals versus fitted values for the NSD-GPP model with 25 knots, (b): QQ-plot for the residuals of the NSD-GPP model with 25 knots.

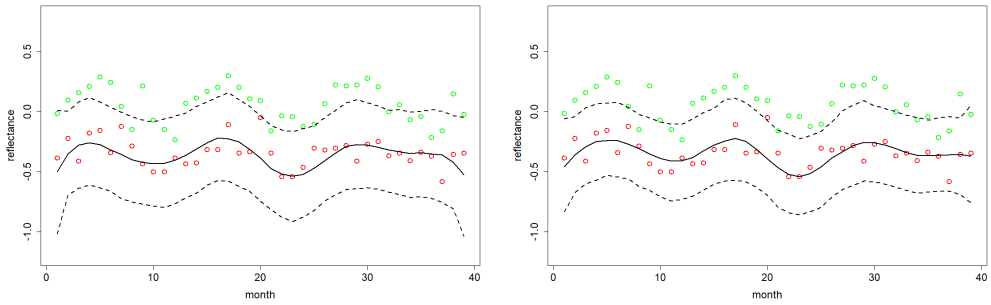


(a) (b)

Figure 5.11: (a):ACF for the residuals at a random location of the NSD-GPP model with 25 knots , (b):Variogram with Monte Carlo envelope at a random month for the residuals of the NSD-GPP model with 25 knots.



(a) NSD-GPP model with 25 knots (b) NSD-GPP model with 49 knots

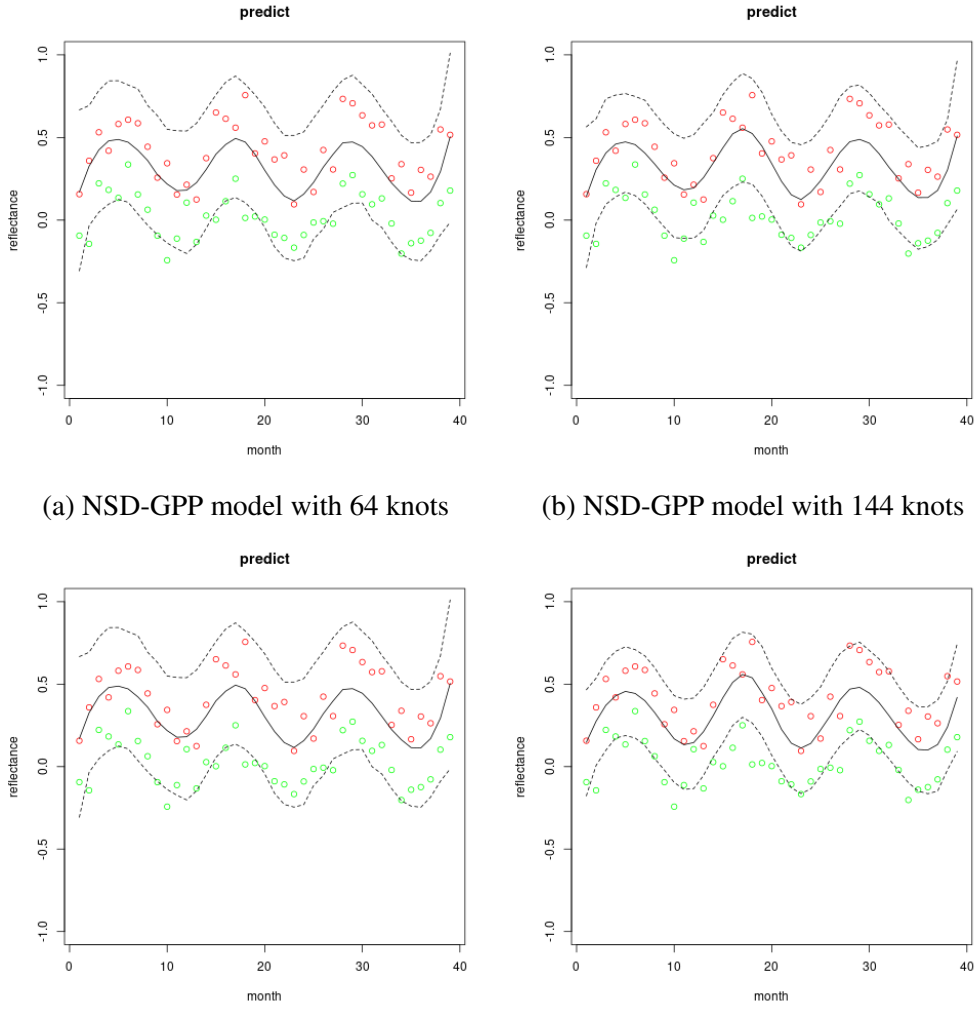


(c) NSD-GPP model with 100 knots (d) NSD model

Figure 5.12: Spatial prediction of the simulated MERIS reflectance data set (scenario 1) at a randomly selected location with four different models. Red circles: true values of the MERIS reflectance, Green circle: observed value of MODIS reflectance, Black solid lines: predictions from the model, Black dashed lines: 95 percent credible interval of the predictions.

scenario	model	rmspe	difference (%)	coverage (%)	bias	width
1	NSD-GPP 25 knots	0.181	10.926	97.071	-0.001	0.811
	NSD-GPP 49 knots	0.173	5.899	96.631	0.000	0.747
	NSD-GPP 100 knots	0.169	3.507	96.101	0.000	0.711
	NSD	0.163	0.000	93.346	0.000	0.606
2	NSD-GPP 25 knots	0.130	10.718	96.962	-0.001	0.570
	NSD-GPP 49 knots	0.129	9.234	96.558	-0.001	0.550
	NSD-GPP 100 knots	0.128	8.429	96.712	0.000	0.554
	NSD	0.118	0.000	94.578	0.000	0.456
3	NSD-GPP 25 knots	0.125	13.084	97.065	-0.001	0.551
	NSD-GPP 49 knots	0.123	11.311	97.756	0.001	0.577
	NSD-GPP 100 knots	0.122	10.531	97.644	0.000	0.559
	NSD	0.110	0.000	94.769	0.000	0.429
4	NSD-GPP 25 knots	0.244	10.791	95.377	0.000	1.000
	NSD-GPP 49 knots	0.231	5.001	95.184	0.000	0.935
	NSD-GPP 100 knots	0.224	1.686	96.094	-0.001	0.974
	NSD	0.220	0.000	91.754	-0.001	0.776
5	NSD-GPP 25 knots	0.147	8.042	96.505	-0.001	0.631
	NSD-GPP 49 knots	0.144	5.589	96.230	0.000	0.605
	NSD-GPP 100 knots	0.142	4.287	96.750	0.000	0.621
	NSD	0.137	0.000	93.921	0.000	0.515
6	NSD-GPP 25 knots	0.132	9.371	97.167	0.000	0.589
	NSD-GPP 49 knots	0.131	7.925	97.325	0.000	0.590
	NSD-GPP 100 knots	0.129	6.678	97.547	0.000	0.586
	NSD	0.121	0.000	94.419	0.000	0.466
7	NSD-GPP 25 knots	0.339	0.170	88.709	-0.001	1.104
	NSD-GPP 49 knots	0.338	0.022	89.203	0.000	1.115
	NSD-GPP 100 knots	0.338	0.014	88.816	-0.001	1.100
	NSD	0.338	0.000	88.510	-0.001	1.089
8	NSD-GPP 25 knots	0.178	0.050	92.196	0.000	0.637
	NSD-GPP 49 knots	0.178	0.033	92.205	0.000	0.637
	NSD-GPP 100 knots	0.178	0.014	92.177	-0.001	0.636
	NSD	0.178	0.000	92.153	0.000	0.636
9	NSD-GPP 25 knots	0.147	0.273	93.520	0.000	0.546
	NSD-GPP 49 knots	0.147	0.326	93.475	0.000	0.545
	NSD-GPP 100 knots	0.147	0.316	93.469	0.000	0.545
	NSD	0.146	0.000	93.341	0.000	0.540

Table 5.6: Spatial cross-validation results of the NSD-GPP model and the NSD model for the simulation datasets with different scenarios. difference: the percentage difference between the RMSPE of the NSD model and the NSD-GPP model, coverage: the coverage of the 95% credible interval, width: the average width of the 95% credible interval.



(a) NSD-GPP model with 64 knots

(b) NSD-GPP model with 144 knots

(c) NSD-GPP model with 289 knots

(d) NSD model

Figure 5.14: Spatial prediction of the simulated MERIS reflectance data set (scenario 1) at a randomly selected location with 4 different models. Red circles: true values of the MERIS reflectance, Green circle: observed value of MODIS reflectance, Black solid lines: predictions from the model, Black dashed lines: 95 percent credible interval of the predictions.

Chapter 6

Multivariate NSD-GPP model for the fusion of multi-wavelength reflectance data

6.1 Introduction

Water leaving reflectance is the ratio between upwelling and downwelling radiance of a water surface and is influenced by substances within the water body [Odermatt and Gege, 2022]. This property could be used to retrieve various water quality parameters such as Chlorophyll–a, coloured dissolved Organic Matter (CDOM) and total suspended matter (TSM), which are important parameters to monitor the lake water quality [Yang et al., 2022]. Currently, retrieval algorithms of water quality parameters rely on multispectral reflectance data, mostly with 3-10 bands [Yang et al., 2022]. The selection of wavelength bands for retrieval algorithms depends on the retrieved parameter and the water type of the water body from which the reflectance data are collected [Carrea et al., 2022]. Recently, reflectance data have been collected by remote sensors such as MODIS and OLCI, but these sensors may not have the observations at the required wavelength bands for retrieval algorithms. Thus, estimating the reflectance data for the required bands will benefit the retrieval algorithms for different water quality parameters and different water types.

The remote sensors MODIS and OLCI have different spatial resolutions. Take Lake Garda as an example, OLCI provides the reflectance data with a spatial resolution of 0.09 km^2, while MODIS provides observations with a spatial resolution of 1 km^2. The reflectance data provided by the Plymouth Marine Laboratory [Carrea et al., 2022] have been aggregated to 1 km^2 resolution so that the OLCI data have the same spatial resolution as the MODIS data. Thus, this chapter will not focus on the spatial support.

The remote sensors MODIS and OLCI also have different temporal resolutions and different wavelength bands. There are 479 days when the OLCI reflectance data are available from

01/05/2016 to 31/12/2019 (1340 days), and there are 868 days with observations for MODIS within this time range. Both OLCI and MODIS sensors have 10 wavelength bands within 400 nm and 700 nm, but the bands from the two sensors have different ranges and centres. The details of the bands for OLCI and MODIS will be introduced in Section 6.3.

Additionally, OLCI reflectance data are assumed to be more accurate for lakes than MODIS data, which is discussed in Section 1.1. However, OLCI has fewer observations than MODIS in the temporal dimension. Thus, this chapter aims to estimate the smooth function of water leaving reflectance over wavelengths between 400 nm to 700 nm with a spatial resolution of 1km $\times$ 1km in Lake Garda for each day from 01/05/2016 to 31/12/2019 by fusing both sensors to make maximum benefit of both their advantages.

In the previous chapter, a nonparametric downscaling model with Gaussian predictive process (NSD-GPP model) was developed to fuse the reflectance data from two sensors at one wavelength band for Lake Garda. In this chapter, since OLCI and MODIS sensors have different wavelength bands, the proposed fusion method is extended to account for the different spectral supports of these two sensors and predict at multiple wavelengths. Additionally, the smooth function of reflectance over wavelength could be used to identify the optical water type of the lake and estimate the reflectance data for the required bands for the retrieval algorithm, which is discussed in Section 1.1. Thus, an extension of the NSD-GPP model will be developed to enable data fusion from two sensors with a multivariate wavelength dimension.

This chapter will start with a literature review of spatial-temporal-spectral data fusion for remote-sensing data. Then, the Lake Garda reflectance dataset will be introduced to describe the motivation of the model extension. After that, an extension of the NSD-GPP model with a spectral dimension will be introduced. Then, the predictive performance of this extended NSD-GPP model will be compared to the NSD-GPP model introduced in the last chapter. After that, the extended NSD-GPP model will be applied to fuse the reflectance data from the MODIS and the OLCI sensors for Lake Garda and make predictions over both time and wavelengths. Finally, the limitations of the proposed data fusion model will be discussed.

6.2 Literature review

The fusion task in this chapter could be viewed as multivariate spatial-temporal fusion, and the reflectance data at different wavelengths are treated as different variates. Berrocal et al. (2010) propose a bivariate spatial-temporal fusion model to fuse ozone and particulate matter ($PM_{2.5}$) variables from monitoring sites and an air quality model called Community Mesoscale Air Quality (CMAQ) model [Berrocal et al., 2010a]. They assume the measured ozone data could be rep-

resented by a combination of the modelled ozone and $PM_{2.5}$ data, and the measured $PM_{2.5}$ data could be represented by a similar combination but with different sets of coefficients. Then, the correlations of these spatial-temporal varying coefficients are modelled by a coregionalization matrix. They also extend this model for data with a multivariate response for different pollutants.

Rundel et al. (2015) develop Berrocal's model to fuse five major $PM_{2.5}$ sources from the in-situ monitoring sites and the numerical outputs from the air quality model, which is the same as in Berrocal et al. [Rundel et al., 2015]. This model aims to explore the contribution of each source of $PM_{2.5}$ across time and space. Rundel et al. (2015) also suggest that estimating the coregionalization matrix for these five sources would be difficult since this matrix is only defined on the latent level of the model [Rundel et al., 2015]. Thus, they only consider the coregionalization matrix for the two species of the $PM_{2.5}$ data with the highest correlation and then use the spatial-temporal models for the other three species separately.

Gong et al. (2021) propose a model to fuse the in-situ and modelled data for 12 pollutants [Gong et al., 2021b]. In the first stage, the slope parameters relating the in-situ and modelled data are estimated by assuming these slopes are independent over different locations. Then, a multivariate spatial-temporal model is built for the residuals when the slope parameters are fixed using the estimated values from the first stage. They also use the Integrated Nested Laplace Approximation (INLA) instead of the MCMC method in Berrocal's model to estimate the parameters in the model. These modifications are used to improve the computational efficiency of Berrocal's model.

There are more examples of multivariate spatial-temporal fusion models based on Berrocal's model, such as the models proposed by [Gilani et al., 2019], [Boaz et al., 2019] and [Huang et al., 2018]. However, these multivariate spatial-temporal fusion models require the same set of variables for the two data sources. If the reflectance data at different wavelengths are treated as different variables, these models require the same set of wavelengths for the two sensors. However, the OLCI and the MODIS sensors have different spectral supports. Thus, these multivariate spatial-temporal fusion models do not appear to be suitable for this chapter's fusion problem due to the misalignment of the spectral supports.

Another example of data fusion, which has considered multi-spectral data, is discussed in [Gevaert and García-Haro, 2015]. They propose an unmixing model to decompose the spatial-spectral data $\mathbf{X} \in \mathbb{R}^{L\times N}$ at each time point into an endmember matrix $\mathbf{E} \in \mathbb{R}^{L\times P}$ and abundance matrix $\mathbf{A} \in \mathbb{R}^{P\times N}$, where N is the number of locations, L is the number of wavelength bands, and P is the number of endmembers. Then, the relationship between the two data sources is built upon these endmember and abundance matrices. The temporal dependency is also modelled on these

two matrices to separate the temporal variation in the spatial and spectral dimensions. Peng et al. (2021) develop a geographically weighted unmixing model to improve the estimation of the endmember matrix and abundance matrix [Peng et al., 2021].

There are more attempts of the unmixing-based spatial-temporal-spectral fusion model proposed by [Zhou et al., 2022], [Shen et al., 2016] and [Xie et al., 2016]. However, these unmixing-based models only consider the temporal dependency for two consecutive time points, which makes it difficult to capture the long-term and yearly temporal patterns. Since this chapter aims to produce the fused reflectance data for a three-year time range, the unmixing-based models do not appear to be suitable for this application. In summary, the methods described in this section are not suitable for the Lake Garda dataset because of the misalignment of the spectral supports and a long-term temporal range.

6.3 Lake Garda reflectance data

In this section, the Lake Garda reflectance dataset is introduced to show the spatial, temporal, and spectral patterns of the reflectance data and the correlation in the reflectance data from the two sensors. For this dataset, reflectance data are observed by the MODIS and OLCI sensors at 203 pixels from 01/05/2016 to 31/12/2019. MODIS and OLCI reflectance data are observed at different wavelength bands, and the ranges of wavelength bands are shown in Table 6.1, where the band represents the centre of the corresponding wavelength range. The wavelength bands within 400-700 nm are selected for the MODIS and OLCI sensors since retrieval algorithms of chlorophyll-a are based on the reflectance data within this range [Liu et al., 2021], and Chlorophyll-a concentration is an important parameter to monitor the lake water quality. Additionally, the atmospheric correction algorithm performs better within 400-700 nm than 700-800 nm for this dataset [Liu et al., 2021]. For the Lake Garda dataset provided by Plymouth Marine Laboratory, atmospheric correction has been applied to the raw reflectance data to remove the effect of bright objects, such as clouds and snow, and the over-correction of atmospheric effect sometimes produces negative reflectance data.

MODIS		OLCI	
band (nm)	range (nm)	band (nm)	range (nm)
		400	393-408
412	405-420	412	408-418
443	438-448	443	438-448
469	459-479		
488	483-493	490	485-495
		510	505-515
531	526-536		
551	546-556		
555	545-565		
		560	555-565
		620	615-625
645	620-670		
667	662-672	665	660-670
		674	670-678
678	673-683	681	678-685

Table 6.1: The bands and range for the MODIS and OLCI sensors. The five pairs of MODIS and OLCI bands with a similar range are highlighted with the same colour. The blank cells represent the bands with different ranges for these two sensors.

MODIS and OLCI reflectance data are plotted for each dimension to illustrate their patterns. The spectral plots in Figure 6.1 and 6.2 show that the reflectance data at a pixel near the northeast of the lake have one peak across the wavelength dimension, which occurs at different wavelengths, suggesting that spectral patterns may change for different time points. The spatial patterns in reflectance data on 16/07/2016 at different wavelengths are shown in Figure 6.3 and 6.4 for OLCI and MODIS. There is more spatial variations for reflectance data within the 400-600 nm range than those within the 600-700 nm range because the reflectance data are almost identical over the lake for wavelength 600-700 nm. These plots suggest that the spatial variation is likely to be different at different wavelengths. Then, a random location is selected to show the temporal patterns of reflectance data for each wavelength band. The temporal plots in Figure 6.5 and 6.6 for MODIS and OLCI show one peak of reflectance data per year at the first six bands, and the peaks occur at a similar time for these bands. Additionally, these temporal plots show more temporal variations for the reflectance data within 400-600 nm than those within 600-700 nm.

In order to explore the relationship between the MODIS and OLCI reflectance data, the correlation of the reflectance data across the temporal dimension from two sensors are calculated for each pixel and each pair of wavelength bands with a similar range. Five pairs of bands are selected to illustrate the correlation of the reflectance data from the two sensors, which are (MODIS, OLCI) = (412,412), (443,443), (488,490), (667,665), (678,681). These pair of bands are assumed to be comparable since the reflectance data from these two sensors are recorded at

similar wavelengths for these pairs of bands. The correlation plots are shown in Figure 3.13. The correlation plots show a strong correlation across time for each pixel and each wavelength band between these two sensors. Thus, the data fusion method seems reasonable for combining the reflectance data from both sensors and then providing the most informative dataset within the time range. In this application, the univariate NSD-GPP model introduced in the last chapter can be applied to fuse the reflectance data for these single pairs of wavelength bands and then make predictions of OLCI reflectance data at unobserved time points. The aim of this chapter is to estimate the OLCI reflectance data at any wavelength within 400 nm to 700 nm. However, the univariate NSD-GPP model could not make spectral predictions at unobserved wavelength bands. Additionally, the univariate NSD-GPP model is not suitable to fuse the reflectance data from the sensors with different spectral supports. Table 6.1 shows the misalignment of the spectral supports of MODIS and OLCI, and the univariate NSD-GPP model could be applied to the five pairs of bands (highlighted with the same colour in the table) with a similar range. For the rest of the MODIS and OLCI bands with different ranges, the univariate NSD-GPP model is not suitable for fusing the reflectance data for these two sensors. Thus, an extension of the univariate NSD-GPP model will be developed in the next section to facilitate this change of spectral support problem for the two sensors and then make predictions for reflectance data over both wavelengths and time.

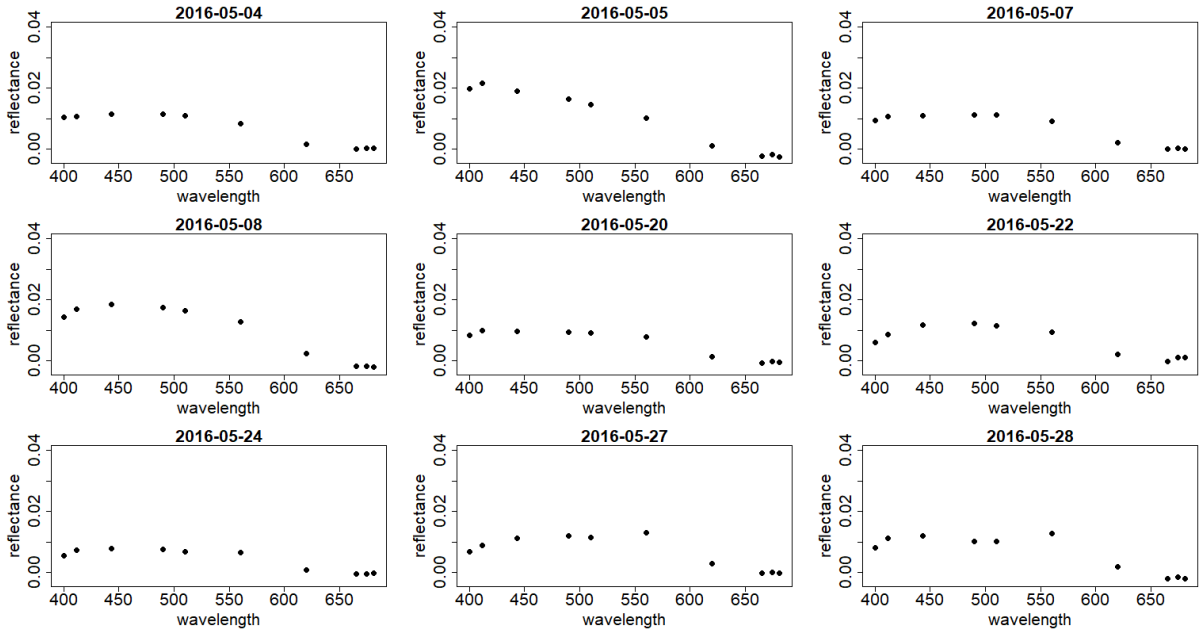


Figure 6.1: OLCI reflectance data over wavelengths for multiple days at a randomly selected location (near the northeast of the lake).

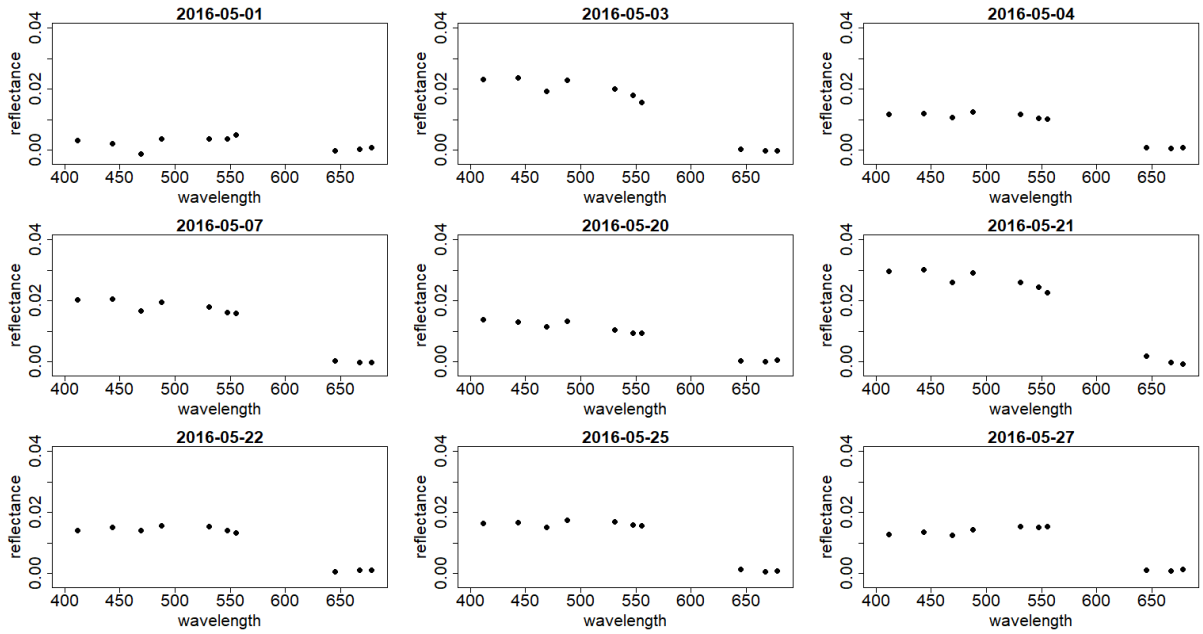


Figure 6.2: MODIS reflectance data over wavelengths for multiple days at a randomly selected location (near the northeast of the lake).

6.4 Method

In this section, a novel multivariate NSD-GPP model will be proposed. This model is motivated by the Lake Garda reflectance data set introduced in the previous section. The multivariate NSD-GPP model is developed to fuse the reflectance data from two sensors with the same spatial resolution, but different temporal resolutions and different wavelength bands. These assumptions of spatial, temporal and spectral resolutions are based on the following application for the Lake Garda dataset. Suppose Sensor 2 has more observations in the temporal dimension than Sensor 1, but the reflectance data from Sensor 1 are more accurate than the data from Sensor 2. The multivariate NSD-GPP model aims to solve the change of spectral support of Sensor 1 and Sensor 2, and then make predictions of reflectance data for Sensor 1 over both time and wavelength by fusing the data from both sensors. In what follows, the notation of the reflectance data from the two sensors will be defined. Then, the multivariate NSD-GPP model will be defined, and the parameter estimation and hyper-parameter selection for this model will be introduced.

6.4.1 Basis function representation

In the NSD-GPP model in the previous chapter, the time series at each location is represented by a series of known one-dimensional basis functions over time. In this extended NSD-GPP model, the spectral-temporal data at a single location are represented by a series of known two-dimensional basis functions over time and wavelength. These two-dimensional functions could

be the tensor product of splines basis functions produced by the Kronecker product of two sets of univariate B-splines basis functions. Since the reflectance data for different wavelengths and time points are treated as observations from a smooth two-dimensional function, this extended NSD model can handle different spectral and temporal supports of Sensor 1 and Sensor 2 using the same set of two-dimensional basis functions. For each spatial location i, the reflectance data from Sensor 1 is assumed to be a function over time t and wavelength λ. Then, the function of reflectance y can be represented by:

$$y_i(t,\lambda) = \sum_{j=1}^{m} \Phi_j(t,\lambda) c_{ji} + \varepsilon_i(t,\lambda),$$

where $\Phi_j(t,\lambda)$ is the two-dimensional basis function and $\varepsilon_i(t,\lambda)$ are i.i.d random errors. The basis function $\Phi_j(t,\lambda)$ can be constructed using the tensor product of two sets of one-dimensional basis functions. Suppose $\mathbf{\Phi}_1(t)$ and $\mathbf{\Phi}_2(\lambda)$ are two sets of fourth-order B-splines basis functions over time and wavelength respectively:

$$\mathbf{\Phi}_1(t) = \{\phi_{11}(t), \phi_{12}(t), ..., \phi_{1m_1}(t)\},$$
$$\mathbf{\Phi}_2(\lambda) = \{\phi_{21}(\lambda), \phi_{22}(\lambda), ..., \phi_{2m_2}(\lambda)\},$$

where m_1 and m_2 are the number of basis functions for $\mathbf{\Phi}_1(t)$ and $\mathbf{\Phi}_2(\lambda)$. The fourth-order B-splines basis function means that each function in $\mathbf{\Phi}_1(t)$ and $\mathbf{\Phi}_2(\lambda)$ is a cubic spline, which is smooth to the human eye. The number of basis functions m_1 and m_2 will be selected by a cross-validation method, and the details will be discussed in Section 6.4.2. Then, the two dimension basis functions $\mathbf{\Phi}(t,\lambda)$ can be constructed by:

$$\mathbf{\Phi}(t,\lambda) = \mathbf{\Phi}_1(t) \otimes \mathbf{\Phi}_2(\lambda) = \{\phi_{11}(t)\phi_{21}(\lambda), \phi_{11}(t)\phi_{22}(\lambda), ..., \phi_{1m_1}(t)\phi_{2m_2}(\lambda)\},$$

where $\mathbf{\Phi}(t,\lambda)$ contains m two-dimensional basis functions and $m = m_1 \times m_2$. Suppose the reflectance data from Sensor 1 at location i are observed at time points $t = t_1, t_2, t_3, ..t_{q_i}$ and wavelengths $\lambda = \lambda_{y_1}, \lambda_{y_2}, ...\lambda_{y_{w_1}}$. The reflectance data $\mathbf{y}_i$ can be represented by a $w_1 \times q_i$ matrix, but it will be more convenient to convert it into a $w_1 q_i \times 1$ vector for the further steps. $\mathbf{y}_i = (\mathbf{y}_i(\lambda_{y_1}), \mathbf{y}_i(\lambda_{y_2}), ..., \mathbf{y}_i(\lambda_{y_{w_1}})^T$, where $\mathbf{y}_i(\lambda_{y_1})$ is the vector of reflectance data for location i and wavelength λ_{y_1}. Then, the basis function matrix $\mathbf{\Phi}_i \in \mathbb{R}^{w_1 q_i \times m}$ represents the basis functions $\mathbf{\Phi}(t,\lambda)$ evaluated at these $w_1 q_i$ combinations of time and wavelength at location i. Similarly, suppose that the reflectance data from Sensor 2 at location i are observed at time points $t = t_1, t_2, t_3, ..t_{p_i}$ and wavelengths $\lambda = \lambda_{x_1}, \lambda_{x_1}, ...\lambda_{x_{w_2}}$. Then, $\mathbf{x}_i = (\mathbf{x}_i(\lambda_{x_1}), \mathbf{x}_i(\lambda_{x_2}), ..., \mathbf{y}_i(\lambda_{x_{w_2}})^T$ is a $w_2 p_i \times 1$ vector and $\mathbf{\Psi}_i$ is the basis function matrix for $\mathbf{\Psi}(t,\lambda)$ evaluated at these $w_2 p_i$ combinations of time and wavelength at location i for Sensor 2.

6.4.2 Multivariate NSD-GPP model

Now, the multivariate NSD-GPP model can be defined as follows:

$$
\begin{aligned}
&\mathbf{y}_i \sim \mathrm{N}_{w_1 q_i}\left(\, \mathbf{\Phi}_i \mathbf{c}_i,\, \mathbf{P}_{y_i}^{-1} \,\right), \\
&\mathbf{x}_i \sim \mathrm{N}_{w_2 p_i}\left(\, \mathbf{\Psi}_i \mathbf{d}_i,\, \mathbf{P}_{x_i}^{-1} \,\right), \\
&c_{ij} \sim \mathrm{N}\left(d_{ij}\mathbf{V}_i \boldsymbol{\beta}_j^* + \mathbf{Z}_i \boldsymbol{\alpha}_j^*,\, \frac{1}{p_c} \right), \\
&\mathbf{d}_i \sim \mathrm{N}_m(\, \boldsymbol{\mu}_d,\, \mathbf{\Sigma}_d \,), \\
&\boldsymbol{\alpha}_j^* \sim \mathrm{N}\left(\mathbf{0},\, \frac{1}{p_\alpha}\exp(-\phi_\alpha \mathbf{D}^*) \right), \\
&\boldsymbol{\beta}_j^* \sim \mathrm{N}\left(\mathbf{1},\, \frac{1}{p_\beta}\exp(-\phi_\beta \mathbf{D}^*) \right), \\
&\mathbf{Z} = \exp(-\phi_\alpha \mathbf{D}^T)(\exp(-\phi_\alpha \mathbf{D}^*))^{-1}, \\
&\mathbf{V} = \exp(-\phi_\beta \mathbf{D}^T)(\exp(-\phi_\beta \mathbf{D}^*))^{-1},
\end{aligned}
\tag{6.1}
$$

where:

- $\mathbf{c}_j = (c_{1j}, c_{2j}, ...c_{nj})^T$ and $\mathbf{d}_j = (d_{1j}, d_{2j}, ...d_{nj})^T$ are the coefficients for the j^{th} basis function for Sensor 1 and Sensor 2 respectively.
- $\mathbf{D}^*$ is the $r \times r$ distance matrix between the selected knots $S^* = \{s_1^*, s_2^*, ..., s_r^*\}$ for the Gaussian predictive process.
- $\mathbf{D}^T$ is the $n \times r$ distance matrix between the observed locations $S = \{s_1, s_2, ..., s_n\}$ and the selected knots $S^* = \{s_1^*, s_2^*, ..., s_r^*\}$.

In the first line of Model 6.1, the Sensor 1 reflectance data $\mathbf{y}_i$ are assumed to follow a multivariate normal distribution with the mean equal to the coefficients $\mathbf{c}_i$ times the basis functions $\mathbf{\Phi}_i$ and the variance equal to $\mathbf{P}_{y_i}^{-1}$. Similarly, the Sensor 2 reflectance data $\mathbf{x}_i$ are assumed to follow a multivariate normal distribution with the mean equal to the coefficients $\mathbf{d}_i$ times the basis functions $\mathbf{\Psi}_i$ and the variance equal to $\mathbf{P}_{x_i}^{-1}$. The plots in section 6.3 show that reflectance data have different temporal variations for the bands within the 400-600 nm range and those in the 600-700 nm range. Thus, it is assumed that the precision parameters are not constant over wavelength bands. In Model 6.1, the precision parameters are defined for each wavelength band, and the precision matrices $\mathbf{P}_{y_i}$ and $\mathbf{P}_{x_i}$ are block diagonal matrices, which can be represented as:

$$
\mathbf{P}_{y_i} = \begin{bmatrix}
p_{y_1}\mathbf{I}_{q_i} & \mathbf{0} & \cdots & \mathbf{0} \\
\mathbf{0} & p_{y_2}\mathbf{I}_{q_i} & \cdots & \mathbf{0} \\
\vdots & \vdots & \ddots & \vdots \\
\mathbf{0} & \mathbf{0} & \cdots & p_{y_{w_1}}\mathbf{I}_{q_i}
\end{bmatrix},
$$

$$\mathbf{P}_{x_i} = \begin{bmatrix} p_{x_1}\mathbf{I}_{p_i} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & p_{x_2}\mathbf{I}_{p_i} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & p_{x_{w_2}}\mathbf{I}_{p_i} \end{bmatrix}.$$

The precision parameters $p_{y_1}, p_{y_2}, ..., p_{y_{w_1}}$ are defined for Sensor 1 reflectance data at w_1 different wavelengths $\lambda_{y_1}, \lambda_{y_2}, ..., \lambda_{y_{w1}}$. Similarly, the precision parameters $p_{x_1}, p_{x_2}, ..., p_{x_{w2}}$ are defined for Sensor 2 reflectance data at ten different wavelengths $\lambda_{x_1}, \lambda_{x_2}, ..., \lambda_{x_{w2}}$.

The priors of the precision parameters follow the Gamma distribution with known shape and rate parameters, where

$$\begin{aligned} p_{y_k} &\sim \text{Ga}(a_{y_k},\, b_{y_k}),\ k = 1,2,3,...,w_1, \\ p_{x_k} &\sim \text{Ga}(a_{x_k},\, b_{x_k}),\ k = 1,2,3,...,w_2, \\ p_\alpha &\sim \text{Ga}(a_\alpha,\, b_\alpha), \\ p_\beta &\sim \text{Ga}(a_\beta,\, b_\beta), \\ p_c &\sim \text{Ga}(a_c,\, b_c). \end{aligned}$$

Then, the coefficients c_{ij} for Sensor 1 are related to d_{ij} for Sensor 2 by a linear model. The slope $\mathbf{V}_i\boldsymbol{\beta}_j^*$ and the intercept $\mathbf{Z}_i\boldsymbol{\alpha}_j^*$ are defined from the corresponding Gaussian predictive process $\mathbf{V}\boldsymbol{\beta}_j^*$ and $\mathbf{Z}\boldsymbol{\alpha}_j^*$. The coefficients $\mathbf{d}_i$ are assumed to follow a multivariate normal distribution with mean $\boldsymbol{\mu}_d$ and $\boldsymbol{\Sigma}_d$. In this chapter, $\boldsymbol{\mu}_d$ is a length-m zero vector and $\boldsymbol{\Sigma}_d = 100 \times \mathbf{I}_m$ where to yield a weakly informative prior, since there is no additional information about the signs and the covariance structure of d_i. The spatial process $\boldsymbol{\alpha}_j^*$ and $\boldsymbol{\beta}_j^*$ are defined in the fifth and sixth lines of Model 6.1, and the transformation matrices $\mathbf{Z}$ and $\mathbf{V}$ are defined in the seventh and eighth lines. These transformation matrices are discussed in Section 5.2.2.

Parameter estimation

The parameters of Model 6.1 are estimated using Bayesian inference with MCMC simulation. The full conditional posterior distributions for all the parameters are proper distributions except for the spatial decay parameters ϕ_α and ϕ_β, which do not have closed forms for their posterior distributions. The spatial decay parameters ϕ_α and ϕ_β will be selected by the cross-validation method, which will be introduced in the following section. Details of these full conditional distributions are given below. Since the conjugate prior of p_α is the Gamma distribution, the posterior distribution of p_α can be written as follows:

$$f(p_\alpha|\cdot) \sim \text{Gamma}\left(a_\alpha + \frac{mn}{2},\, b_\alpha + \frac{1}{2}\text{tr}[\exp(-\phi_\alpha \mathbf{D}^*)^{-1}\boldsymbol{\alpha}^{*T}\boldsymbol{\alpha}^*] \right),$$

where $\boldsymbol{\alpha}^* = (\boldsymbol{\alpha}_1^*, ..., \boldsymbol{\alpha}_2^*, \boldsymbol{\alpha}_m^*)$. The conditional posterior of p_β, p_c have a similar form as p_α, and they are shown as follows:

$$f(p_\beta|\cdot) \sim \text{Gamma}\left(a_\beta + \frac{mn}{2},\ b_\beta + \frac{1}{2}\text{tr}[\exp(-\phi_\beta(\boldsymbol{\beta}^* - \mathbf{1})^T(\boldsymbol{\beta}^* - \mathbf{1})] \right),$$
$$f(p_c|\cdot) \sim \text{Gamma}\left(a_c + \frac{mn}{2},\ b_c + \frac{1}{2}\text{tr}(\mathbf{I}_n(\mathbf{c} - (\mathbf{Z}\boldsymbol{\alpha}^* + (\mathbf{V}\boldsymbol{\beta}^*) \odot \mathbf{d}))^T \mathbf{I}_m(\mathbf{c} - (\mathbf{Z}\boldsymbol{\alpha}^* + (\mathbf{V}\boldsymbol{\beta}^*) \odot \mathbf{d}))) \right),$$

where $\boldsymbol{\beta}^* = (\boldsymbol{\beta}_1^*, ..., \boldsymbol{\beta}_2^*, \boldsymbol{\beta}_m^*)$.

The precision parameters $p_{y_1}, ..., p_{y_{w_1}}$ are estimated separately. Take p_{y_1} as an example:

$$f(p_{y_1}|\cdot) \sim \text{Gamma}\left(a_{y_1} + \sum_{i=1}^{n} \frac{q_i}{2},\ b_{y_1} + \frac{1}{2}\sum_{i=1}^{n} (\mathbf{y}_i(\lambda_{y_1}) - \boldsymbol{\Phi}_i(\lambda_{y_1})\mathbf{c}_i)^T (\mathbf{y}_i(\lambda_{y_1}) - \boldsymbol{\Phi}_i(\lambda_{y_1})\mathbf{c}_i) \right),$$

where $\mathbf{y}_i(\lambda_{y_1})$ is the vector of the reflectance data at location i and wavelength λ_{y_1}, and $\boldsymbol{\Phi}_i(\lambda_{y_1})$ is the basis function matrix at location i and wavelength λ_{y_1}. The conditional posterior distributions for $p_{y_2}, .., p_{y_{w_1}}$ can be calculated by changing the corresponding reflectance data vector and basis function matrix.

Similarly, the precision parameters $p_{x_1}, ..., p_{x_{w_2}}$ are estimated separately. Take p_{x_1} as an example:

$$f(p_{x_1}|\cdot) \sim \text{Gamma}\left(a_{x_1} + \sum_{i=1}^{n} \frac{p_i}{2},\ b_{x_1} + \frac{1}{2}\sum_{i=1}^{n} (\mathbf{x}_i(\lambda_{x_1}) - \boldsymbol{\Psi}_i(\lambda_{x_1})\mathbf{d}_i)^T (\mathbf{x}_i(\lambda_{x_1}) - \boldsymbol{\Psi}_i(\lambda_{x_1})\mathbf{d}_i) \right),$$

where $\mathbf{x}_i(\lambda_{x_1})$ is the vector of the reflectance data at location i and wavelength λ_{x_1}, and $\boldsymbol{\Psi}_i(\lambda_{x_1})$ is the basis function matrix at location i and wavelength λ_{x_1}.

Since the conjugate prior of $\boldsymbol{\alpha}_j^*$ is the normal distribution, then the posterior distribution of $\boldsymbol{\alpha}_j$ is defined as:

$$f(\boldsymbol{\alpha}_j^*|\cdot) \sim \text{N}(\ \boldsymbol{\Sigma}_{\alpha_j^*}\mathbf{A}_{\alpha_j^*},\ \boldsymbol{\Sigma}_{\alpha_j^*}\ ),$$
$$\boldsymbol{\Sigma}_{\alpha_j^*} = (p_\alpha(\exp(-\phi_\alpha\mathbf{D}^*))^{-1} + p_c\mathbf{Z}^T\mathbf{Z})^{-1},$$
$$\mathbf{A}_{\alpha_j^*} = p_c\mathbf{Z}^T(\mathbf{c}_j - (\mathbf{V}\boldsymbol{\beta}_j^*) \odot \mathbf{d}_j).$$

Then, the posterior of $\boldsymbol{\beta}_j^*$, $\boldsymbol{c}_j$ and $\boldsymbol{d}_j$ can be defined with the same method as follows:

$$
\begin{aligned}
&f(\boldsymbol{\beta}_j^*|\cdot) \sim \mathrm{N}(\ \boldsymbol{\Sigma}_{\beta_j^*}\mathbf{A}_{\beta_j^*},\ \boldsymbol{\Sigma}_{\beta_j^*}\ ),\\
&\boldsymbol{\Sigma}_{\beta_j^*} = (p_\beta(\exp(-\phi_\beta\mathbf{D}^*))^{-1} + p_c\mathbf{V}^T\mathbf{G}_j^T\mathbf{G}_j\mathbf{V})^{-1},\\
&\mathbf{A}_{\beta_j^*} = p_\beta + p_c\mathbf{V}^T\mathbf{G}_j^T(\mathbf{c}_j - \mathbf{Z}\boldsymbol{\alpha}_j^*),
\end{aligned}
$$

where $\mathbf{G}_j$ is a diagonal matrix with $\mathbf{d}_j$ as its diagonal.

$$
\begin{aligned}
&f(\mathbf{c}_i|\cdot) \sim \mathrm{N}(\ \boldsymbol{\Sigma}_{c_i}\mathbf{A}_{c_i},\ \boldsymbol{\Sigma}_{c_i}\ ),\\
&\boldsymbol{\Sigma}_{c_i} = (\boldsymbol{\Phi}_i^T\mathbf{P}_{y_i}\boldsymbol{\Phi}_i + p_c\mathbf{I}_m)^{-1},\\
&\mathbf{A}_{c_i} = \boldsymbol{\Phi}_i^T\mathbf{P}_{y_i}\mathbf{y}_i + p_c(\mathbf{Z}_i\boldsymbol{\alpha}^* + (\mathbf{V}_i\boldsymbol{\beta}^*)\odot\mathbf{d}_i).
\end{aligned}
$$

$$
\begin{aligned}
&f(\mathbf{d}_i|\cdot) \sim \mathrm{N}(\ \boldsymbol{\Sigma}_{d_i}\mathbf{A}_{d_i},\ \boldsymbol{\Sigma}_{d_i}\ ),\\
&\boldsymbol{\Sigma}_{d_i} = (\boldsymbol{\Sigma}_d^{-1} + \boldsymbol{\Psi}_i^T\mathbf{P}_{x_i}\boldsymbol{\Psi}_i + p_c\mathbf{F}_i^T\mathbf{F}_i)^{-1},\\
&\mathbf{A}_{d_j} = \boldsymbol{\Sigma}_d^{-1}\boldsymbol{\mu}_d + \boldsymbol{\Psi}_i^T\mathbf{P}_{x_i}\mathbf{x}_i + p_c\mathbf{F}_i^T(\mathbf{c}_i - \mathbf{Z}_i\boldsymbol{\alpha}^*),
\end{aligned}
$$

where $\mathbf{F}_i$ is a diagonal matrix with $\mathbf{V}_i\boldsymbol{\beta}$ as its diagonal.

Hyper-parameters selection

In the multivariate NSD-GPP model, the hyper-parameters will be selected by the cross-validation method. There are three types of hyper-parameters, which are the number of basis functions, the spatial decay parameters ϕ_α and ϕ_β, and the number of knots in the Gaussian predictive process.

Since the temporal prediction of Sensor 1 reflectance data over wavelengths is the aim of this model, the hyper-parameters will be selected by temporal cross-validation for this sensor. The temporal cross-validation is designed as follows:

- Find out the days when Sensor 1 has observations, then randomly divide these days into ten groups.
- In each of the ten iterations, select one group as the validation group and treat the other nine groups of days as the training group.
- For each combination of the hyper-parameters, fit the model and compute the out-of-sample temporal prediction of Sensor 1 reflectance data for the days in the validation group.
- After ten iterations, calculate the root mean square prediction error (RMSPE) and select the combination of hyper-parameters which minimises the RMSPE for the multivariate

NSD-GPP model.

The candidate values for each hyper-parameter are based on the following rules, which will be calculated for the applied dataset. For the study in the last chapter, it is suggested to use 12%, 25 % and 50% of the total locations as the number of knots, considering both the predictive performance compared to the NSD model and the computational efficiency. The temporal basis functions are selected based on the number of peaks in the reflectance data in each year, and the formula used to estimate the number of temporal basis functions in the last chapter is:

$$\text{Number of basis functions} = \left(\text{total months} \times \frac{2 \times r}{12}\right) + 3,$$

where r is the expected number of peaks in each year. Since this chapter will explore the daily reflectance data, the above formula is adjusted to the unit of day for this dataset:

$$\text{Number of basis functions} = \left(\text{total days} \times \frac{2 \times r}{365}\right) + 3.$$

From the formula of the multivariate NSD-GPP model, the coefficients $\boldsymbol{\alpha}_j^*$ and $\boldsymbol{\beta}_j^*$ are defined as:

$$\boldsymbol{\alpha}_j^* \sim \mathrm{N}\left(\mathbf{0},\ \frac{1}{p_\alpha}\exp(-\phi_\alpha \mathbf{D}^*)\right),$$
$$\boldsymbol{\beta}_j^* \sim \mathrm{N}\left(\mathbf{1},\ \frac{1}{p_\beta}\exp(-\phi_\beta \mathbf{D}^*)\right),$$

where $\mathbf{D}^*$ is the $r \times r$ distance matrix between the selected knots $S^* = \{s_1^*, s_2^*, ..., s_r^*\}$ for the Gaussian predictive process.

Take $\boldsymbol{\alpha}_j^*$ as an example, the covariance function of $\boldsymbol{\alpha}_j^*$ can be defined by the exponential correlation function of distance h between two locations:

$$C(h) = \frac{1}{p_\alpha}\exp(-\phi_\alpha h).$$

The correlation $\rho(h)$ between two locations is calculated by:

$$\rho(h) = \frac{C(h)}{C(0)} = \frac{\frac{1}{p_\alpha}\exp(-\phi_\alpha h)}{\frac{1}{p_\alpha}} = \exp(-\phi_\alpha h).$$

Suppose $\rho(h_1) = \{0.1, 0.5, 0.9\}$, which represents weak, moderate and strong correlation between two adjacent pixels with distance h_1, the corresponding set of candidate spatial decay

parameters can be calculated by:

$$\phi_\alpha = -\frac{1}{h_1}\log(\rho(h_1)).$$

The process $\boldsymbol{\alpha}_j^*$ is defined over the knots of the Gaussian predictive process, and the distance matrix $\mathbf{D}^*$ will change for different choices of the number of knots. The Gaussian predictive process $\mathbf{Z}\boldsymbol{\alpha}_j^*$ is the approximation of the spatial process defined over the pixel level, so h_1 is selected to be the distance between two adjacent pixels and this value will not change as the knots change. Since the coefficients $\boldsymbol{\beta}_j^*$ have a similar form as $\boldsymbol{\alpha}_j^*$, the spatial decay parameter ϕ_β can be explored with a similar idea. In this chapter, the spatial decay parameters ϕ_α and ϕ_β are set to be the same value for simplicity, and it is possible to explore the different values of ϕ_α and ϕ_β.

Spectral prediction

After fitting the multivariate NSD-GPP model, the Sensor 1 reflectance data at unobserved wavelengths (within the range of observed wavelengths) could be predicted as follows. Suppose the aim is to predict the reflectance $\mathbf{y}_i(\lambda^*)$ for location i at wavelength λ^* and time points $t = t_1, t_2, t_3, ..t_{q_i}$, where $\mathbf{y}_i(\lambda^*)$ is a q_i-dimensional vector. Then, the corresponding basis function matrix for $\mathbf{y}_i(\lambda^*)$ is $\boldsymbol{\Phi}_i(\lambda^*) \in \mathbb{R}^{q_i \times m}$, which represents the basis functions $\boldsymbol{\Phi}(t, \lambda)$ evaluated at wavelength λ^* and time points $t = t_1, t_2, t_3, ..t_{q_i}$. The reflectance vector $\mathbf{y}_i(\lambda^*)$ is obtained by drawing the samples from its posterior predictive distribution, which is defined as follows:

$$\mathbf{y}_i(\lambda^*) \sim \mathrm{N}_{q_i}\left(\boldsymbol{\Phi}_i(\lambda^*)\mathbf{c}_i,\ \frac{1}{p_y(\lambda^*)}\mathbf{I}_{q_i}\right),$$

where $\mathbf{c}_i$ is the coefficients for location i and $p_y(\lambda^*)$ is the precision parameter for wavelength λ^*, because the multivariate NSD-GPP model assumes a different precision parameter for each wavelength. The coefficients $\mathbf{c}_i$ are estimated from its posterior distribution, and the details are shown in Section 6.4.2. The precision parameter $p_y(\lambda^*)$ is estimated from the linear interpolation of the precision parameters $p_{y_1}, p_{y_2}, ..., p_{y_{w_1}}$ at the observed wavelengths, because λ^* is within the range of the observed wavelengths and the precision parameters are assumed to change linearly between every two observed wavelengths for simplicity. The precision parameters $p_{y_1}, p_{y_2}, ..., p_{y_{w_1}}$ are estimated from their posterior distributions, and the details are described in Section 6.4.2.

6.5 Application to Lake Garda dataset

In this section, the multivariate NSD-GPP model will be applied to the Lake Garda dataset to evaluate its temporal predictive performance and then make predictions of OLCI reflectance

data over time and wavelength. The temporal prediction of the multivariate NSD-GPP model is evaluated because the MODIS sensor and the OLCI sensor have different temporal resolutions. The MODIS sensor has more observations in the temporal dimension than the OLCI sensor. However, the reflectance data from the OLCI sensors are assumed to be more accurate than the MODIS sensor. Thus, the temporal prediction of the reflectance data is explored by fusing the observations from both sensors. Additionally, the spatial resolution of the Lake Garda dataset in this chapter is the same for the MODIS and the OLCI sensor, so spatial prediction is not the focus of this application. Since the multivariate NSD-GPP model is an extension of the univariate NSD-GPP model, the predictive performance of the multivariate NSD-GPP model will be compared to the univariate NSD-GPP model. The temporal predictions of these two models will be compared for the selected wavelength bands. Since reflectance data are assumed to be smooth over wavelengths, comparing the predictions over multiple wavelength bands should demonstrate the temporal predictive performance of the multivariate NSD-GPP model over wavelengths. This section is organised as follows. First, the temporal prediction performance of the multivariate NSD-GPP model and the univariate NSD-GPP model will be compared. Then, the multivariate NSD-GPP model will be used to predict OLCI reflectance data over wavelengths at the unobserved time points for Lake Garda.

6.5.1 Comparing the predictive performance

In this section, the multivariate NSD-GPP model and the univariate NSD-GPP model are applied to the Lake Garda reflectance dataset to compare the temporal predictive performance of these two models at the selected wavelength bands. This chapter aims to predict OLCI reflectance data over wavelengths at the unobserved time points. If the multivariate model could achieve similar temporal predictive performance as the univariate model at the selected wavelength bands, then the multivariate model could be used to predict OLCI reflectance data over wavelengths for the unobserved time, because the reflectance data are smooth over wavelengths. Additionally, the multivariate NSD-GPP model can facilitate the change of spectral support problem, which does not require the OLCI and MODIS data observed at the same wavelength bands. In what follows, the selection of wavelength bands will be introduced. Then, the process of temporal prediction will be described. After that, the choices of prior and the hyper-parameters will be discussed. Finally, the results of the temporal prediction from these two models will be summarised.

Selection of wavelength bands

The univariate NSD-GPP model requires the OLCI and MODIS reflectance data observed at the same wavelength. In this application, the wavelength bands from MODIS and OLCI are selected with a similar wavelength range. It is assumed that the reflectance data are smooth within

the range of each band so that the univariate NSD-GPP model can be applied separately to the reflectance for the pairs of bands with similar ranges from the two sensors. Thus, five pairs of bands are selected from Table 6.1, which are $(\text{MODIS, OLCI}) = \{(412,412),(443,443),(488,490),(667,665),(678,681)\}$.

Process of temporal prediction

The temporal prediction performance of the multivariate NSD-GPP model and the univariate NSD-GPP model will be compared based on their bias, root mean square prediction error (RMSPE) and coverage of 95% credible interval. The temporal prediction process is designed as follows:

- Find out the days when OLCI sensor has observations, randomly divide the days into five groups. Use one group as the out-of-sample prediction dataset and the rest four groups of days as the training dataset.
- For the training dataset, use temporal cross-validation to select the hyper-parameters of the multivariate NSD-GPP model and the univariate NSD-GPP model which minimise the RMSPE.
- Use the hyper-parameters selected by cross-validation to predict the OLCI reflectance data for the days in the prediction dataset by the multivariate NSD-GPP model and the univariate NSD-GPP model applied to the entire training dataset.
- Repeat the above steps five times and calculate bias, RMSPE, coverage for each model.

The above processes are repeated five times to check the consistency of the results. It is expected that the results will be stable among the different iterations. Thus, five iterations were selected for computational efficiency. In each iteration, the hyper-parameters are selected by the ten-fold cross-validation method introduced in section 6.4.2.

Choices of prior and hyper-parameters

The priors for the precision parameters $p_{y_1},...,p_{y_{w_1}},p_{x_1},...p_{x_{w_2}},p_\alpha,p_\beta,p_c$ in the multivariate NSD-GPP model are assumed to follow the Gamma distribution $\text{Gamma}(\text{shape} = 2, \text{rate} = 0.01)$, which is used in the last chapter for the application of the univariate NSD-GPP model to the Lake Garda reflectance data at 412nm from the MERIS and the MODIS sensor. The reflectance data are investigated for the same lake at multiple wavelength bands in this chapter, and it is assumed that the priors of the precision parameters are the same for each wavelength band. Thus, the same choices of prior are used for consistency.

The hyper-parameters for the multivariate NSD-GPP model and the univariate NSD-GPP model

are the number of basis functions, spatial decay parameters and number of knots for the Gaussian predictive process. The candidate values of hyper-parameters explored in this section are based on the Lake Garda dataset, and the formulas to find these values are discussed in Section 6.4.2. The number of spectral basis functions is fixed to be four for this dataset. The B-splines basis functions with order four represent the cubic functions and they look smooth to human eyes. There are only five observations over spectral dimension, and four B-spline basis functions are used for spectral dimension because this is the minimum number of basis functions with order four. Then, the number of temporal basis functions will be selected by the cross-validation method for both multivariate and univariate NSD-GPP models. The candidate values of the number of temporal basis functions are $\{10, 18, 25\}$ which represent $\{1, 2, 3\}$ peaks of reflectance data per year. Since the distance between two adjacent pixels is 0.008 degrees for Lake Garda, the spatial decay parameters will be explored for $\phi_\alpha = \phi_\beta = \{13, 87, 288\}$, which represents the correlation $\rho(h_1) = \{0.9, 0.5, 0.1\}$. In summary, there are 27 combinations of these hyper-parameters, and the optimal combination is selected by the temporal cross-validation described in Section 6.4.2.

Convergence of the Markov chain and Model fitting

In this section, the convergence of the Markov chain and the model fitting will be checked when applying the multivariate NSD-GPP model to the reflectance data. Take one iteration of the temporal prediction described in Section 6.5.1 as an example, the hyper-parameters of Model 6.1 are selected by the 10-fold cross-validation. The combination of hyper-parameters which minimises the RMSPE is {number of temporal basis functions=25, spatial decay parameters=288, number of knots for Gaussian predictive process=49}. Using this combination of hyper-parameters which minimise the RMSPE, the model is fitted and the residuals are calculated. There are 5000 iterations in the MCMC simulation, and the first 500 iterations are treated as the burn-in period. Then, the Markov chain is thinned by every ten iterations, yielding 450 samples. The convergence of the Markov chain is checked by Geweke's diagnostic. Take the parameter c_{11} in Model 6.1 as an example, the p-value of Geweke's diagnostic is 0.47, which suggests there is no evidence against the hypothesis that the means of the first 10% and the last 50% of the samples from the Markov chain are equal. Geweke's diagnostic is applied to a number of selected parameters in Model 6.1, and the corresponding p-values are bigger than 0.1. Thus, there is no evidence of the divergence of the Markov chains for these parameters.

Then, the assumptions of constant variance and normality of the residuals of Model 6.1 are checked by residual plots. Since the non-constant precision parameters for each wavelength band are used in Model 6.1, the variance of the residuals is different for each wavelength band. Then, the residuals for each band are standardised to check the assumptions of constant variance and normality. Take the residuals for OLCI reflectance data at location i and wavelength band

λ_{y_1} as an example, the standardised residuals $R_i^{\text{sd}}(\lambda_{y_1})$ are calculated by:

$$\begin{aligned}
&\mathbf{y}_i(\lambda_{y_1}) \sim \mathrm{N}_{q_i}\left(\mathbf{\Phi}_i(\lambda_{y_1})\hat{\mathbf{c}}_i, \frac{1}{\hat{p_{y_1}}}\mathbf{I}_{q_i}\right),\\
&\sqrt{\hat{p_{y_1}}}\mathbf{y}_i(\lambda_{y_1}) \sim \mathrm{N}_{q_i}(\sqrt{\hat{p_{y_1}}}\mathbf{\Phi}_i(\lambda_{y_1})\hat{\mathbf{c}}_i, \mathbf{I}_{q_i}),\\
&R_i^{\text{sd}}(\lambda_{y_1}) = \sqrt{\hat{p_{y_1}}}(\mathbf{y}_i(\lambda_{y_1}) - \mathbf{\Phi}_i(\lambda_{y_1})\hat{\mathbf{c}}_i),
\end{aligned}$$

where $\mathbf{y}_i(\lambda_{y_1})$ and $\mathbf{\Phi}_i(\lambda_{y_1})$ are the reflectance data vector and basis function matrix for OLCI reflectance data at location i and wavelength band λ_{y_1}. The parameters $\hat{\mathbf{c}}_i$ and $\hat{p_{y_1}}$ are the posterior median of the samples from the MCMC simulation. After computing the standardised residuals for each location and wavelength band, the standardised residuals are plotted in Figure 6.7. The standardised residuals vs fitted values plot shows heteroscedasticity, where the variance of the standardised residuals increases as the fitted values increase. The heavy tails in the top-right part of the Q-Q plot suggest that there are much more data located at extremes of the distribution compared to the normal distribution. Thus, the constant variance and normality assumption for Model 6.1 do not hold. Thus, the standardisation of the reflectance data is applied to improve the model fitting. It is assumed that fulfilling the assumptions of the multivariate NSD-GPP model will provide a better prediction. Standardisation of reflectance data means dividing the raw reflectance data by the integrated value across the wavelength [Vantrepotte et al., 2012]. There are two reasons for using the standardisation instead of, for example the log or square-root transformation for reflectance data. First, standardisation is a commonly used data-preprocessing operation for reflectance data, which is applied in the work of [Spyrakos et al., 2018], [Yang et al., 2021], and [Vantrepotte et al., 2012]. The second reason is that the standardised reflectance data can also be used to estimate the chlorophyll-a concentration. A linear regression model is build between the chlorophyll-a concentration and the standardised reflectance data at the wavelength band where the reflectance is strongly correlated to the chlorophyll-a concentration [Zhan et al., 2008]. Since the prediction of reflectance data will be used to estimate the water quality parameters such as chlorophyll-a concentration, modelling the standardised reflectance data can also achieve this aim. The standardisation is not applied to the reflectance data in the last chapter for two reasons. First, the residual plots for the univariate model, which is fitted to the reflectance data at wavelength band 412 nm in the last chapter, do not show evidence against the assumptions of constant variance and normality. Second, standardisation requires the reflectance data over wavelengths, and the application in the last chapter only uses the reflectance data for one wavelength band. Thus, standardisation is introduced in this chapter for the application of the reflectance data over multiple wavelength bands.

In standardisation, typically the integrated value is the area between the reflectance data points and the horizontal line equals 0. Since the Lake Garda dataset contains negative values for some wavelength bands, the integrated value is calculated between the reflectance data points and the

minimum of the reflectance for all wavelength bands. Since the reflectance data are observed at discrete wavelengths, the integrated value is calculated by the numerical integration in R. Suppose $\mathbf{y}_{it}$ is a vector of OLCI reflectance data at location i and time t, $y_{it}(\lambda)$ is the reflectance at location i and time t for wavelength λ.

$$A_{it} = \int y_{it}(\lambda) d\lambda,$$
$$y_{it}(\lambda)^{\text{standard}} = \frac{y_{it}(\lambda)}{A_{it}},$$

where the integrated value A_{it} represents the area between the reflectance curve and the reference line for each location and time point. The prior choices of the precision parameters are not changed for the standardised reflectance data since weakly informative priors are used for these precision parameters.

After applying the standardisation to the MODIS and OLCI reflectance data, Model 6.1 is fitted to the standardised reflectance data and the hyper-parameters are selected by the ten-fold cross-validation. The combination of hyper-parameters which minimises the RMSPE is {number of temporal basis functions=25, spatial decay parameters=288, number of knots for Gaussian predictive process=25}. After fitting the model with these hyper-parameters, the standardised residuals are plotted in Figure 6.8. The residuals plot shows that the residuals have approximately constant variance for different fitted values, and the points in the Q-Q plot are close to the straight line except for the light tails. Although there are still patterns of residuals versus fitted values and light tails in the Q-Q plot, the model fitting is better than the results for the reflectance data with the original scale. In the following work, the model comparison of the multivariate NSD-GPP model and the univariate NSD-GPP model will based on the application to standardised reflectance data. The spatial, temporal and spectral correlation of the residuals for standardised reflectance data are checked by empirical semi-variogram plots, and the results suggest there is no significant spatial, temporal and spectral correlation in the residuals for the majority of the plots.

Results of the model comparison

In this section, the predictive performance of the multivariate NSD-GPP model and the univariate NSD-GPP model are compared to demonstrate the effectiveness of the multivariate NSD-GPP model. After standardising the reflectance data over wavelengths and then selecting the hyper-parameters by the ten-fold cross-validation, the multivariate NSD-GPP model and the univariate NSD-GPP model are fitted and then the out-of-sample temporal predictions for OLCI reflectance data are compared. This process is repeated for five iterations, where different sets of days for OLCI sensors are selected to compute out-of-sample temporal prediction. The results of these five iterations are summarised in Table 6.2. The values of bias are not included in this table because the values are close to 0 up to four decimal places for both univariate and

multivariate models in these five iterations.

Band	**RMSPE**		**coverage**	
Iteration 1	**univariate**	**multivariate**	**univariate**	**multivariate**
412	0.00104	0.00095	0.964	0.963
443	0.00084	0.00069	0.975	0.971
490	0.00070	0.00069	0.985	0.957
665	0.00024	0.00025	1.000	1.000
681	0.00028	0.00028	1.000	0.999
all	0.00069	0.00063	0.985	0.978
Iteration 2	**univariate**	**multivariate**	**univariate**	**multivariate**
412	0.00074	0.00072	0.973	0.969
443	0.00046	0.00056	0.998	0.978
490	0.00049	0.00046	0.991	0.985
665	0.00020	0.00022	1.000	1.000
681	0.00023	0.00024	1.000	0.999
all	0.00047	0.00048	0.992	0.986
Iteration 3	**univariate**	**multivariate**	**univariate**	**multivariate**
412	0.00065	0.00065	0.987	0.982
443	0.00041	0.00048	0.999	0.998
490	0.00041	0.00040	0.999	0.997
665	0.00020	0.00022	1.000	1.000
681	0.00024	0.00024	0.999	0.999
all	0.00041	0.00043	0.997	0.995
Iteration 4	**univariate**	**multivariate**	**univariate**	**multivariate**
412	0.00072	0.00073	0.980	0.967
443	0.00049	0.00048	0.996	0.995
490	0.00045	0.00046	0.994	0.990
665	0.00027	0.00027	1.000	1.000
681	0.00030	0.00030	1.000	1.000
all	0.00047	0.00048	0.994	0.990
Iteration 5	**univariate**	**multivariate**	**univariate**	**multivariate**
412	0.00069	0.00069	0.984	0.976
443	0.00043	0.00047	0.998	0.996
490	0.00043	0.00045	0.996	0.988
665	0.00021	0.00023	1.000	1.000
681	0.00024	0.00026	1.000	1.000
all	0.00044	0.00045	0.996	0.992

Table 6.2: Summary of RMSPE and coverage of 95% credible interval for the univariate NSD-GPP model and multivariate NSD-GPP model. For each iteration, RMSPE and coverage are calculated for each wavelength band and for all five bands.

The first finding from this table is that RMSPE and coverage are generally stable over these five iterations for both univariate and multivariate models, except for Iteration 1. In Iteration 1,

the RMSPE are higher than the other four iterations for bands 412, 443 and 490 from both the univariate model and the multivariate model. The temporal predictions in Iteration 1 for these three bands are shown in Figure 6.9. From these plots, it is shown that the temporal predictions are away from the reflectance data in the prediction dataset for the three time points (the black points in the plots) near the right boundary of the time range. These plots suggest that it is difficult to predict the reflectance data near the boundary of the time range for both the univariate and multivariate models. Since the observed data (blue points) are only available on the left side of these three predicted time points (black points), it is difficult to extrapolate the values at these boundary points. For the other time points where the observed data are available at both sides, the interpolation at these time points is close to the values in the prediction dataset. If these three time points are removed when calculating the RMSPE and coverage, the results for Iteration 1 is shown in table 6.3. In this table, the results of the RMSPE is stable compared to the other four iterations. Considering the results in Table 6.2 and 6.3, the univariate model and the multivariate model achieve a similar temporal prediction quality among these five iterations.

The second finding is that the ratios of RMSPE and coverage between the univariate model and the multivariate model are close to 1 for each wavelength band, and this result is stable among these five iterations. The ratio of RMSPE between the univariate model and the multivariate model for band 443 is 0.82 in Iteration 2, which is the minimum ratio among these five iterations. The reason for this result is that the hyper-parameters for the multivariate model are tuned to minimize the overall RMSPE for the five bands, while the hyper-parameters for the univariate model are tuned to minimize the RMSPE for each band. Thus, the multivariate model provides larger values of RMSPE than the univariate model for some bands, but the overall RMSPE ratio is close to 1.

The coverages of 95% credible interval for each band are close to 1 from Table 6.2. To explore the reason for the large coverage, the temporal predictions for OLCI reflectance data from the univariate model and the multivariate model are plotted in Figure 6.10 at a randomly picked location and in a randomly picked iteration (Iteration 3) for each wavelength band. The temporal predictions capture the major temporal patterns of the reflectance data from the univariate model and the multivariate model. These plots show OLCI observations (blue points) outside the credible interval. Since the precision parameters of both models are estimated from the observed standardised reflectance data for all locations and time points, the precision parameters and hence the coverage probabilities are adversely affected by outliers that can be seen in Figure 6.10.

This comparison suggests that the multivariate model can achieve a similar predictive performance as the univariate model for the selected five bands for the Lake Garda dataset. Thus, the

multivariate NSD-GPP model can be used to make predictions of the OLCI reflectance data over wavelengths for the unobserved time points in the next section.

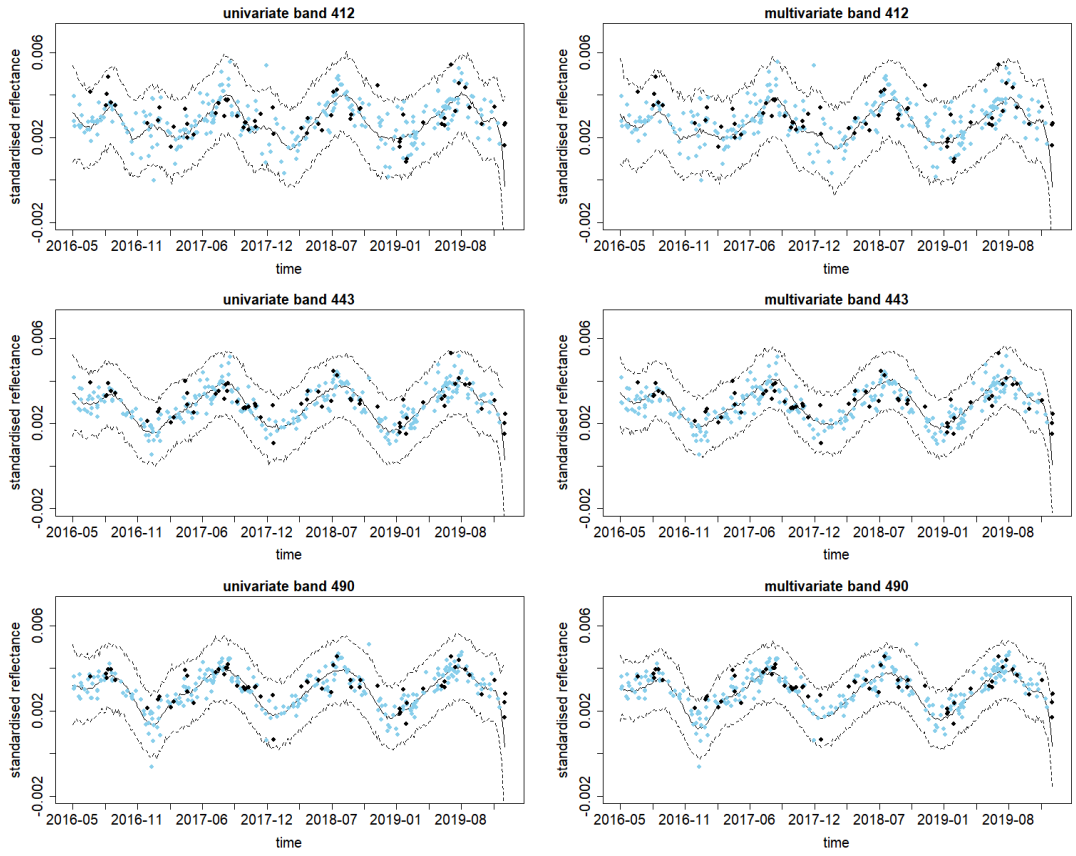


Figure 6.9: Temporal predictions for OLCI reflectance data from the univariate model and the multivariate model at a randomly picked location (near the northeast of the lake) and in Iteration 1 for band 412, 443, and 490. Sky blue points: observed OLCI standardised reflectance data; Black points: OLCI standardised reflectance data in the prediction dataset; black solid lines: temporal predictions; black dashed lines: 95% credible interval.

Band	**RMSPE**		**coverage**	
Iteration 1	**univariate**	**multivariate**	**univariate**	**multivariate**
412	0.00072	0.00074	0.978	0.977
443	0.00051	0.00053	0.989	0.987
490	0.00050	0.00053	0.987	0.975
665	0.00024	0.00025	1.000	1.000
681	0.00027	0.00028	0.999	0.999
all	0.00048	0.00050	0.991	0.988

Table 6.3: Summary of RMSPE and coverage of 95% credible interval for the univariate NSD-GPP model and multivariate NSD-GPP model for Iteration 1 after removing the three time points near the right boundary of the time range.

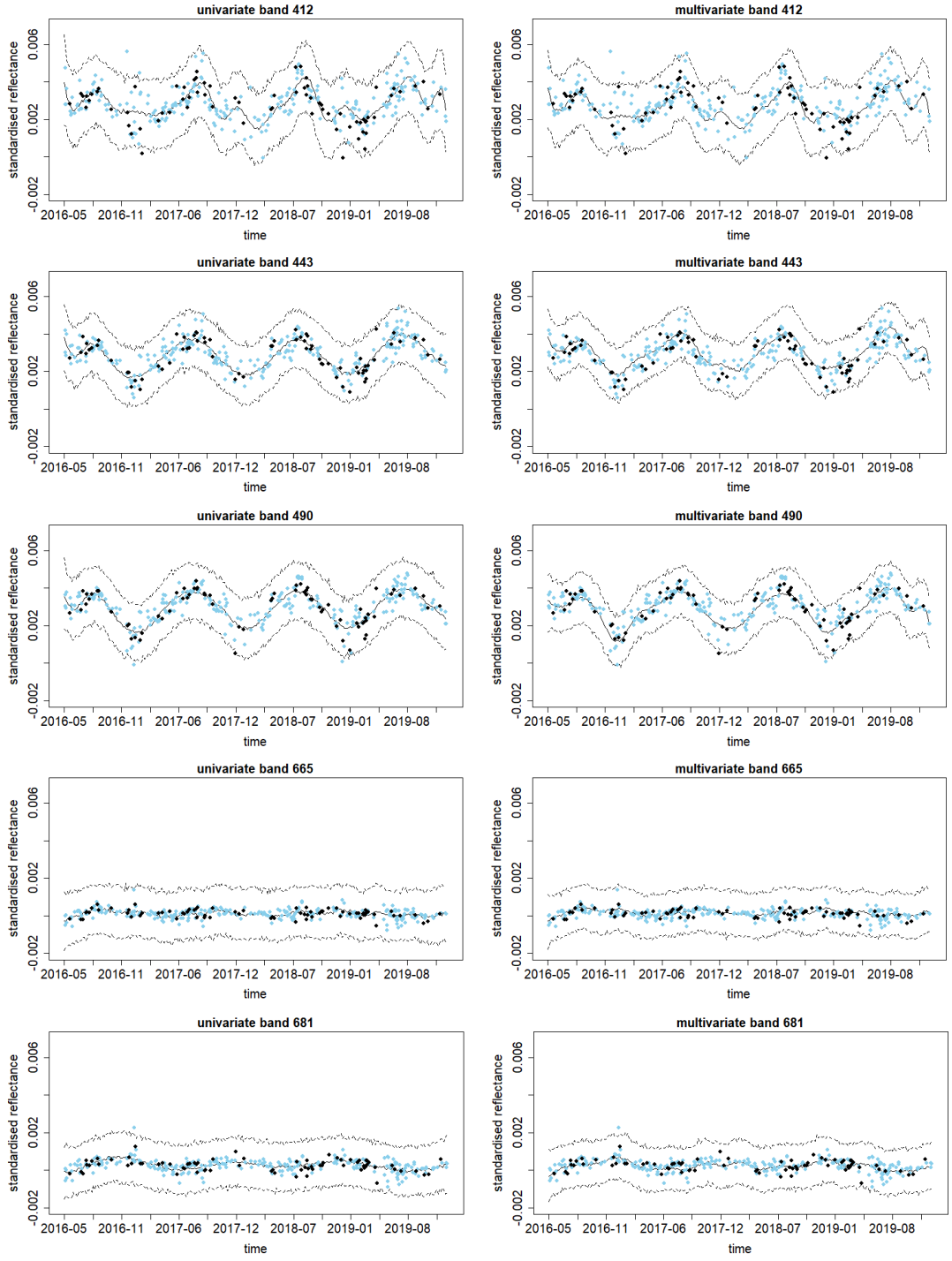


Figure 6.10: Temporal predictions for OLCI reflectance data from the univariate model and the multivariate model at a randomly picked location (near the northeast of the lake) and in Iteration 3 for each wavelength band. Sky blue points: observed OLCI standardised reflectance data; Black points: OLCI standardised reflectance data in the prediction dataset; black solid lines: temporal predictions; black dashed lines: 95% credible interval.

6.5.2 Prediction over time and wavelength using the multivariate NSD-GPP model

In this section, the multivariate NSD-GPP model is applied to make predictions for OLCI reflectance data for unobserved time points and wavelengths. The aim of this application is to illustrate the temporal and spectral prediction of the reflectance data, and the predictions at these wavelengths will be suitable for different types of retrieval algorithms as discussed in section 6.1.

In the above section, the temporal prediction of the multivariate NSD-GPP model is investigated for five iterations. In this section, the entire reflectance dataset from the MODIS and OLCI sensors is used to predict the reflectance at unobserved time points and wavelengths. The wavelength bands within 412 to 681 nm are selected for both sensors, which provide a similar overall coverage of wavelength bands for these two sensors. According to the bands in Table 6.1, there are ten bands for MODIS and nine bands for OLCI within 412 to 681 nm.

Then, the hyper-parameters of the multivariate NSD-GPP model will be selected for this dataset. In the above section, the number of spectral basis functions is four ,and the overall coverage of wavelength bands in this section is the same as in the previous section. Thus, the number of spectral basis functions in this section is also selected to be four. Then, the number of temporal basis functions, the number of knots for the Gaussian predictive process, and the spatial decay parameters are selected by the temporal cross-validation method described in Section 6.4.2. The combination of hyper-parameters which minimises the RMSPE is {number of temporal basis functions=25, spatial decay parameters=87, number of knots for Gaussian predictive process=25}. Then, the predictions are calculated using the method described in Section 6.4.2. The spectral prediction is calculated with a resolution of 1 nm between 412 nm and 681 nm, which is the observed wavelength range. This spectral resolution is selected because the reflectance data used in retrieval algorithms have a spectral resolution of 1 nm [Yang et al., 2022]. The spectral predictions at a randomly picked location near the centre of Lake Garda are shown in Figure 6.11 and Figure 6.12. Since these predictions are drawn independently from their posterior predictive distribution for each wavelength and time, the predictions are not smooth over wavelength in Figure 6.11 and Figure 6.12. From the plots in Figure 6.11, the multivariate NSD-GPP model can generally capture the spectral pattern of the OLCI reflectance data at the selected location and time points, and the observed reflectance data are within the 95% credible interval. The plots in Figure 6.12 demonstrate that the multivariate NSD-GPP model is able to predict over wavelengths on the days without OLCI observations. This application demonstrates the ability of the multivariate NSD-GPP model for spectral prediction over wavelengths, which can not be achieved by applying the univariate NSD-GPP separately for the observed wavelength bands.

The predicted values are plotted for all time points and wavelengths within the observed range

to investigate the spectral and temporal patterns of the OLCI reflectance data for Lake Garda. This plot shows that the reflectance data have a clear peak close to wavelength 500 nm, and this spectral pattern is similar across time. This spectral pattern suggests that the Chlorophyll–a concentration and turbidity are relatively low at this location, since the reflectance will have an additional peak near 700 nm for the water with high Chlorophyll–a concentration and turbidity [Goyens et al., 2022, Ahn and Park, 2020]. As discussed in Section 1, the predicted standardised reflectance data could be used to identify the Optical Water type for Lake Garda, and the retrieval algorithm for this water type could be selected based on the study of [Neil et al., 2019]. Then, the reflectance data at the required wavelength bands could be estimated by the multivariate NSD-GPP model as shown in Figure 6.12. Thus, the multivariate NSD-GPP model could fulfil the steps of the retrieval method suggested by Neil et al., and can potentially be used to estimate water quality parameters at the unobserved time points.

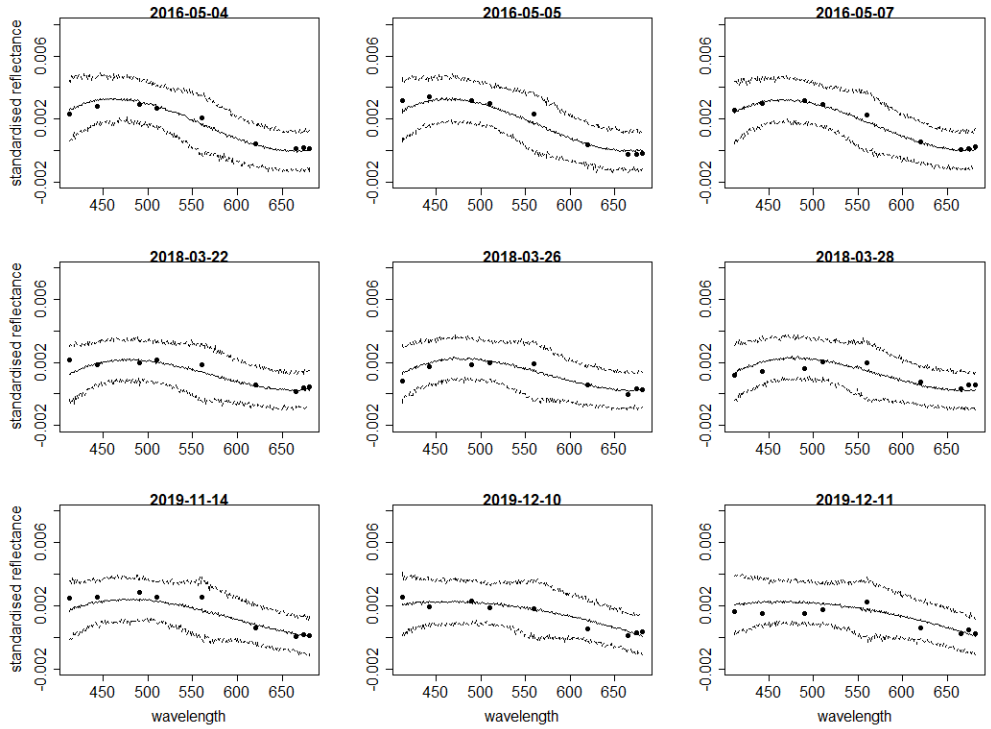


Figure 6.11: Spectral predictions for OLCI reflectance data from the multivariate NSD-GPP model at a randomly picked location (near centre of the lake) and 9 randomly picked days (near the start, middle, and end of the time points in the prediction dataset). Black points: OLCI standardised reflectance data in the prediction dataset at five selected wavelength bands; black solid lines: spectral predictions; black dashed lines: 95% credible interval.

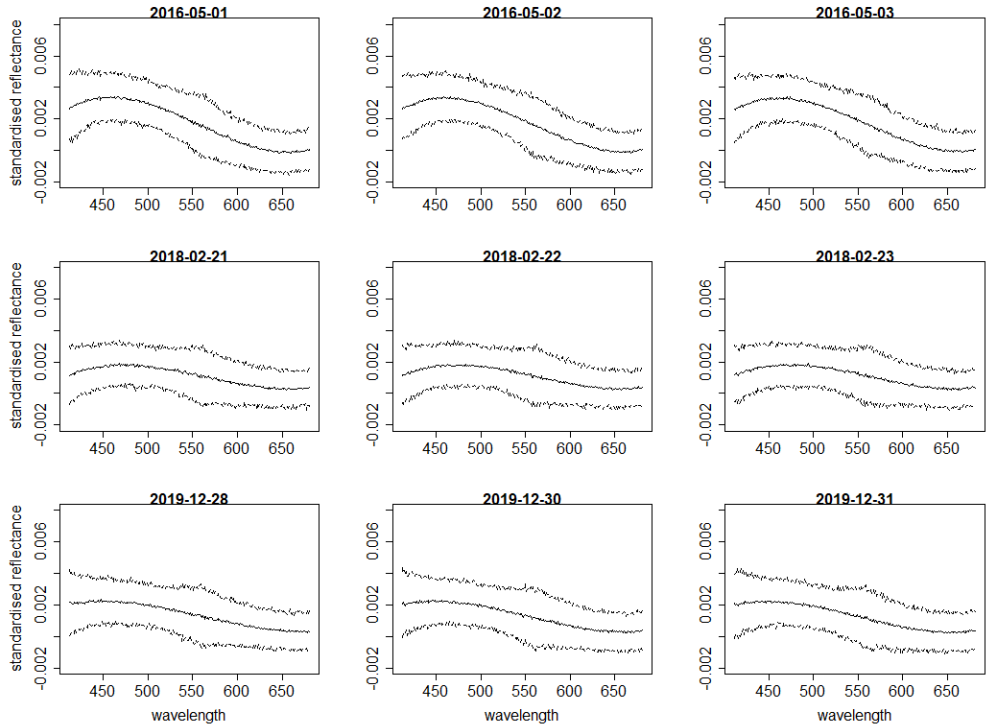


Figure 6.12: Spectral predictions for OLCI reflectance data from the multivariate NSD-GPP model at a randomly picked location (near centre of the lake) and 9 randomly picked days (near the start, middle, and end of the time points in the prediction dataset). Black solid lines: spectral predictions; black dashed lines: 95% credible interval.

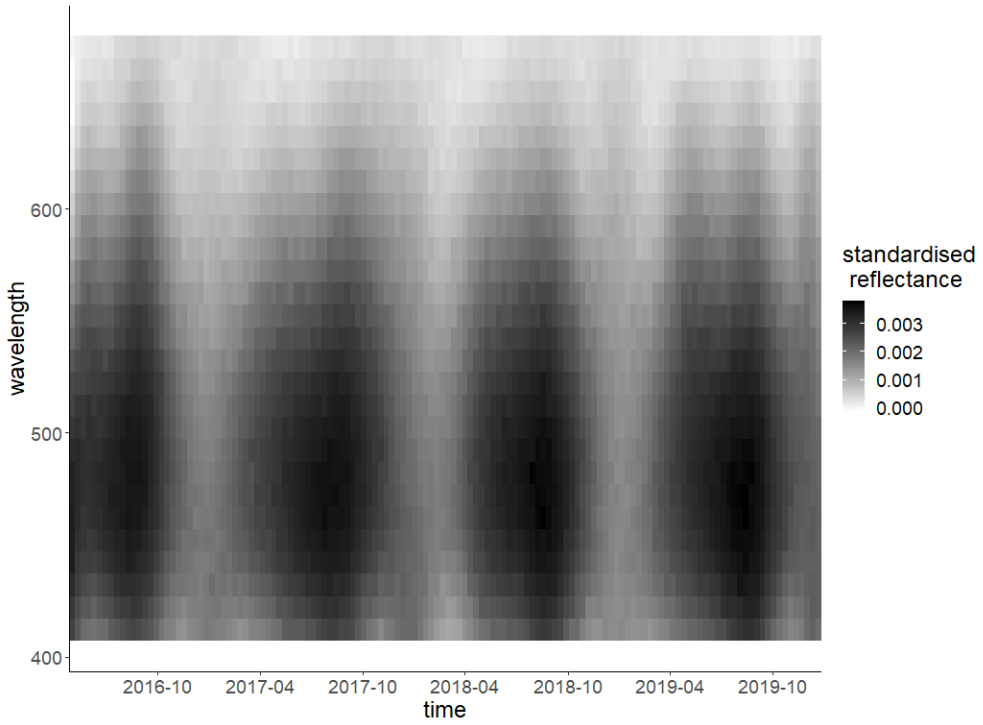


Figure 6.13: Predictions for OLCI reflectance data over time and wavelength from the multivariate NSD-GPP model at a randomly picked location (near centre of the lake).

6.6 Conclusion

In this chapter, a multivariate NSD-GPP model is developed to make predictions of reflectance data over wavelengths and time by fusing the data from two remote sensors. This model is motivated by the fusion problem for the Lake Garda reflectance dataset, which aims to predict OLCI reflectance data over wavelengths at unobserved time points and wavelengths. This model is an extension of the univariate NSD-GPP model proposed in the last chapter, and the multivariate model is developed to account for the change of spectral support of the two sensors and to allow for spectral prediction. In the multivariate NSD-GPP model, the reflectance data at each location are represented by a combination of the two-dimensional basis functions, and the relationship between the two sensors is built upon the corresponding coefficients. Since the basis functions are defined over the same range for the two sensors, this multivariate NSD-GPP model can fuse the reflectance data from the sensors with different spectral supports.

In the application to the Lake Garda dataset, the temporal prediction performance of the multivariate NSD-GPP model is compared to that of the univariate NSD-GPP model. The multivariate model achieves a similar RMSPE and coverage of its 95% credible interval for the selected wavelength bands, and the results are generally stable among the five iterations of temporal prediction. Since the reflectance data are assumed to be smooth over wavelengths, it is deduced that the multivariate NSD-GPP model could achieve a similar predictive performance as the univariate NSD-GPP model for the unobserved wavelength bands. Thus, the multivariate NSD-GPP model could be used to make spectral predictions for the unobserved wavelength bands at the selected locations and time points. From the application to the Lake Garda dataset, the multivariate NSD-GPP model can generally capture the spectral pattern of the OLCI reflectance data at the selected location and time points. The multivariate NSD-GPP model could predict the reflectance data over time and wavelengths, which could be used to identify the optical water type of the lake and estimate the reflectance data at required bands. Thus, this model will benefit the retrieval algorithms for water quality parameters, which assist in monitoring of lake water health. For future work, it is possible to relate the predicted reflectance data and water quality parameters using a two-stage approach. Villejo et al. (2023) proposed a two-stage approach data fusion approach, where the first stage is a Bayesian melding model which integrates the satellite images and computer simulations, and the second stage is used to link the health outcomes and the pollution exposures based on the spatial average of the first stage [Villejo et al., 2023].

There are two main limitations of the multivariate NSD-GPP model. First, the multivariate NSD-GPP model does not consider the change of spatial support problem because the provided Lake Garda data are aggregated into the same spatial resolution for the two sensors. This model could be extended to account for the change of spatial supports for the two sensors, and then make spatial predictions for reflectance data at the selected locations. Second, the computa-

tional time of the multivariate NSD-GPP model is longer than the total computational time of running the univariate NSD-GPP model for all wavelength bands. In the application to the Lake Garda dataset in Section 6.5.1, the multivariate NSD-GPP model takes 1.2 hours to estimate the parameters and then make temporal predictions. However, the total computational time of running the univariate NSD-GPP model for the five bands is about 1 hour with the same hyper-parameters as the multivariate model. In the multivariate NSD-GPP model, the precision parameters are defined for each wavelength band, which requires additional computational time to estimate these parameters. Future work could investigate the multivariate NSD-GPP model with fewer precision parameters, which assumes the precision parameters are the same over a range of wavelengths.

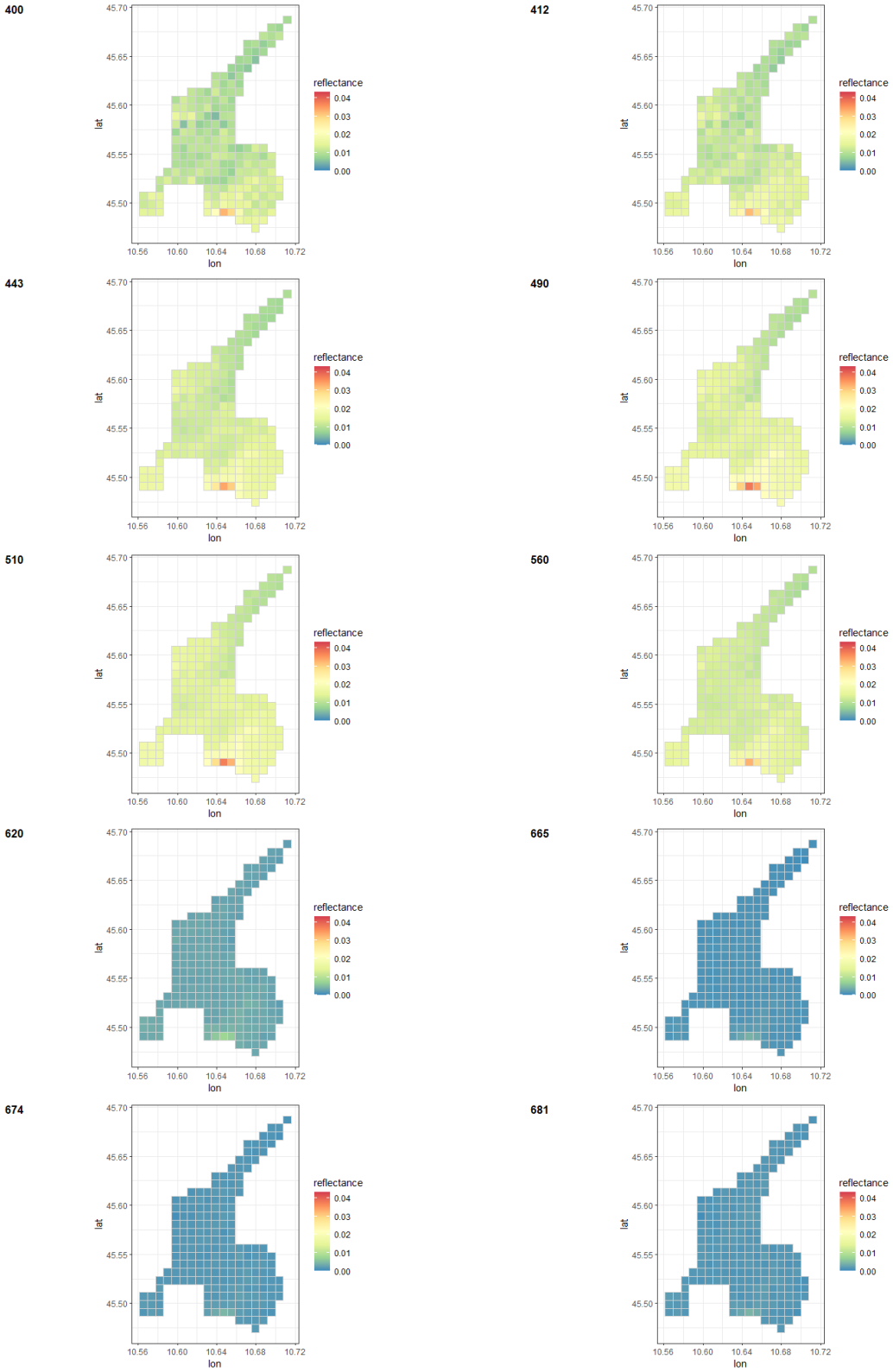


Figure 6.3: OLCI reflectance data over spatial locations for each wavelength band on a randomly selected day (16/07/2016).

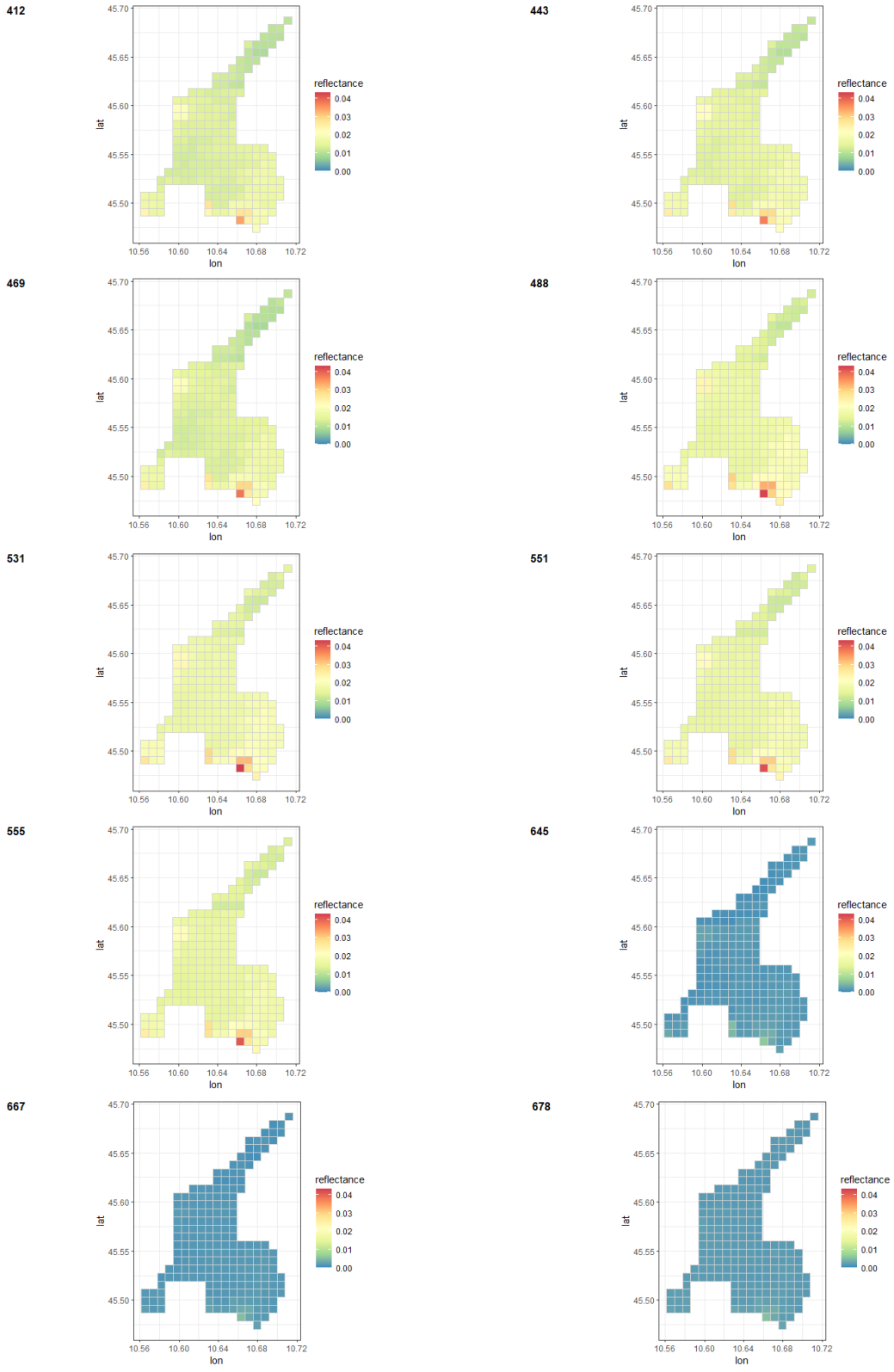


Figure 6.4: MODIS reflectance data over spatial locations for meach wavelength band on a randomly selected day (16/07/2016).

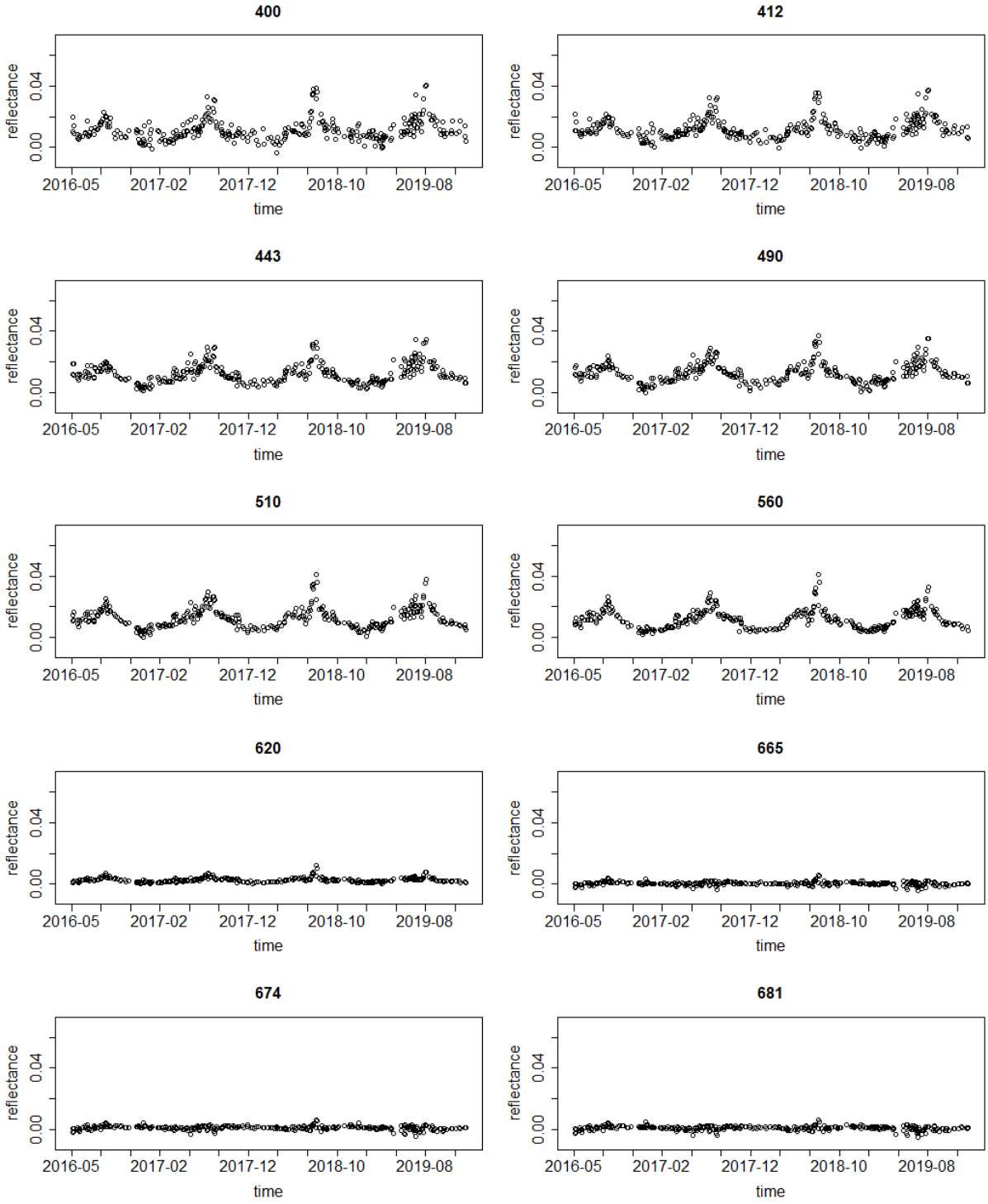


Figure 6.5: OLCI reflectance data over time for each wavelength band at a randomly selected location (near the northeast of the lake).

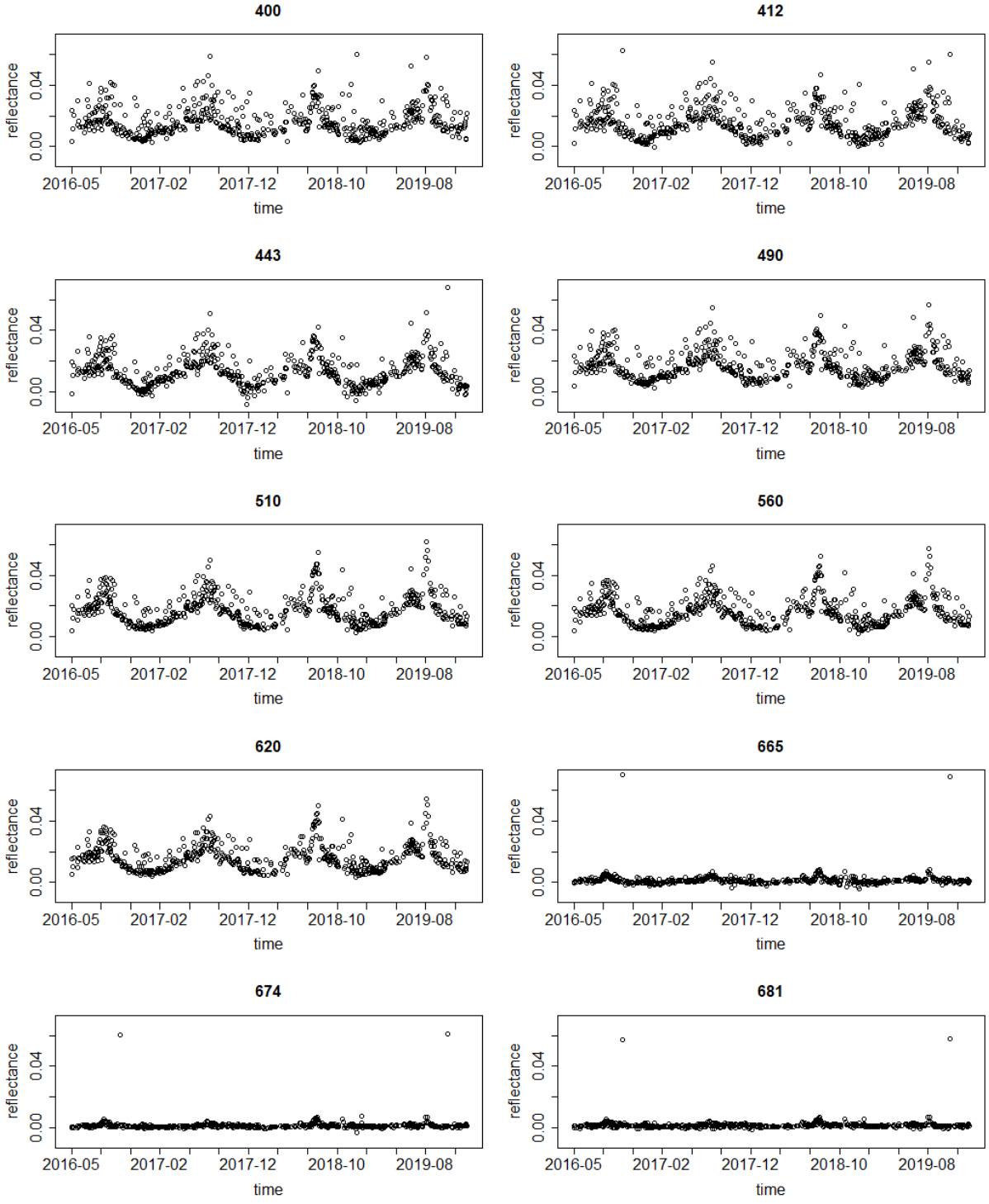


Figure 6.6: MODIS reflectance data over time for each wavelength band at a randomly selected location (near the northeast of the lake).

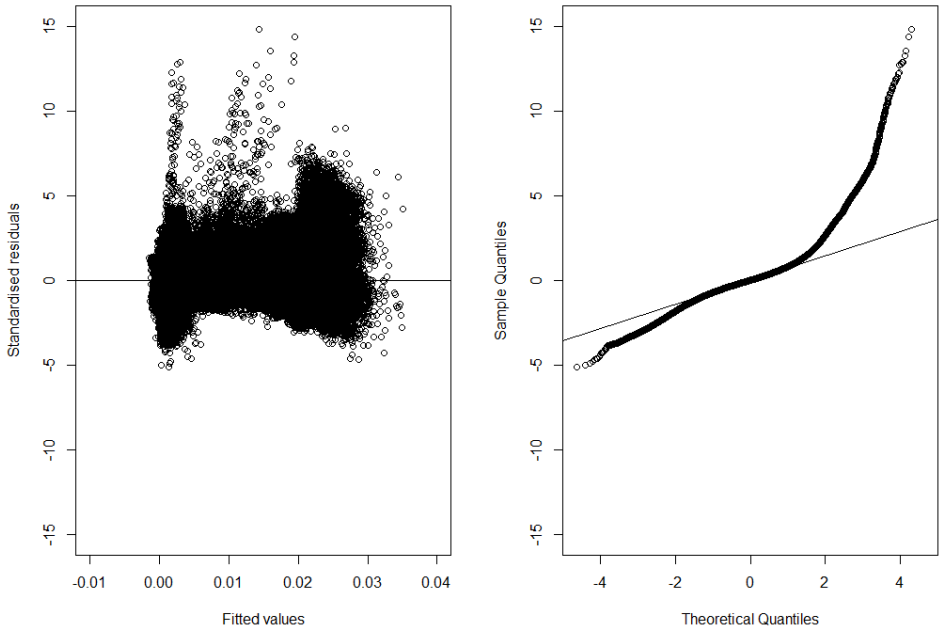


Figure 6.7: Standardised residuals vs fitted values plot and Q-Q plot of Model 6.1, fitted to reflectance data with the original scale.

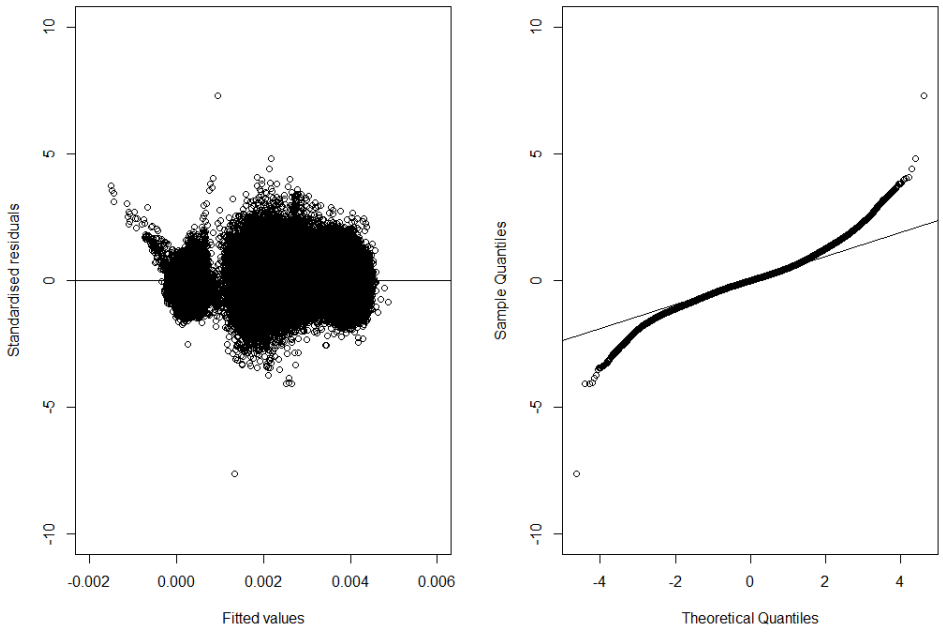


Figure 6.8: Standardised residuals vs fitted values plot and Q-Q plot of Model 6.1, fitted to standardised reflectance data.

Chapter 7

Conclusion

This thesis has focused on developing and proposing a statistical methodology to fuse lake water reflectance from the two satellite sensors with possibly different spectral and temporal supports. This work is motivated by the practical problem of estimating the reflectance at the unobserved wavelengths and time points using data from the two sensors. Lake water plays a vital role in protecting global biosphere and conserving biodiversity. Thus, lake water monitoring is crucial for decision-makers to identify and establish action plans for the areas in need [Behmel et al., 2016]. Water quality parameters such as chlorophyll-a, suspended matter concentration, and coloured dissolved organic matter (CDOM) can be estimated by in-situ or remote-sensing techniques. Remote sensors collect reflectance data and estimate water quality parameters by retrieval algorithms, while the in-situ data are estimated from water samples collected directly from the water bodies. Remote sensing techniques have an advantage in the high spatial-temporal resolution of data, but they have a disadvantage in accuracy compared to in-situ data.

In this thesis, data from three different types of remote sensors have been explored, which are the MERIS, MODIS and OLCI sensors and they have have possibly different spatial, temporal and spectral supports. For example, the MODIS sensor has more observations in the temporal dimension, but the MERIS and OLCI sensors are assumed to be more suitable for inland water. Thus, combining reflectance data from MODIS and MERIS, or MODIS and OLCI for their overlapping years will maximise the benefit of their advantages. Data fusion models are used to merge information from different data sources to achieve a refined estimation of the target values [Utazi et al., 2022]. Spatio-temporal models, such as the NSD model, enable data fusion with different spatio-temporal supports but are computationally demanding for large datasets. Additionally, multivariate spatial-temporal fusion models such as the models proposed by Rundel et al. (2015) and Gong et al. (2021) require the same set of variables for the two data sources, which are not suitable for fusing the reflectance from the two sensors with different spectral supports.

Considering data fusion in this context, the research aims of this thesis was as follows. The first aim is to compare existing state-of-the-art data fusion methods, which are the nonparametric statistical downscaling (NSD) model and the fixed rank kriging (FRK) model, when the two sensors provide data with the same spatio-temporal supports. This comparison is used to identify the advantages and disadvantages of the existing models, which indicates possible developments for the application of reflectance data modelling. This comparison suggests that the NSD model has a limitation of computational efficiency while the FRK model is not suitable for data sources with different temporal supports. Thus, the second research aim was to improve the computational efficiency of an existing data fusion model, and propose a model can achieve a similar predictive performance as the original model. The third research aim was to develop a multivariate spatio-temporal data fusion model for the reflectance with different spectral and temporal supports from the two sensors. In this thesis, a computationally efficient spatial-temporal data fusion model was developed in Chapter 5 and extended this model for the multivariate response to solve the change of temporal and spectral supports problem in Chapter 6. The reflectance data from the MERIS, MODIS and OLCI sensors at Lake Garda were used to illustrate the predictive performance and the computational efficiency of the data fusion models in this thesis.

7.1 Comparison of existing models

In Chapter 4, the nonparametric statistical downscaling (NSD) model and the fixed rank kriging (FRK) model were applied to the Lake Garda reflectance data set to compare their prediction performance and computational efficiency. The NSD model assumes that the discrete data at each location from each data source are observations of smooth functions over time and that the coefficients of these smooth functions are modelled as spatially correlated via a covariance matrix. The NSD model enables data fusion with different spatio-temporal supports, and it allows the observed data with missingness in both spatial and temporal dimensions. The fixed rank kriging (FRK) model assumes a linear relationship between the data from the two sources, and it uses a spatio-temporal random effect term to capture the remaining spatio-temporal correlations. However, the FRK model requires the data from the two sources to have the same temporal support, and it does not allow for missingness in the spatial and temporal dimensions based on its current version in R. Since the NSD model has not been applied to fuse the reflectance data elsewhere before to the best of our knowledge, the predictive performance of the NSD model is compared here to the FRK model.

The reflectance from the MERIS and MODIS sensors for wavelength 412 nm at Lake Garda was used to compare the performance of these two models, and the data are aggregated into monthly averages at each pixel to prevent missingness in the dataset. There are 211 pixels for

Lake Garda, 21 pixels (approximately 10 percent of the total pixels) are randomly selected as the test dataset and the remaining 190 pixels are treated as the training dataset. The results show that the prediction ability of the NSD model is slightly better than that of the FRK model, but the NSD model requires a longer computational time than the FRK model. Additionally, the NSD model can be applied to the reflectance data with possibly different spatio-temporal supports and missingness in these two dimensions. Thus, the NSD model is selected as a starting point for the developments in this thesis, and the developments are made to improve the computational efficiency in Chapter 5 and account for the change of spectral supports of the reflectance in Chapter 6.

The first drawback of this comparison is that it is based on the results of one training dataset and one prediction dataset, and the locations of pixels can influence the results in these sets. The k-fold cross-validation method repeats this process k times, and the model performance is assessed based on the average of these k iterations. Thus, the k-fold cross-validation method is used in the model comparison in Chapters 5 and 6. The second drawback of this comparison is that it focuses on the spatial prediction of these two models, while the temporal prediction is not investigated. Since the remote sensors in this thesis are not available at complete daily resolution because of their revisit time periods, predicting the reflectance at the unobserved time will provide a more complete dataset. Thus, the temporal prediction of the data fusion models is investigated in Chapters 5 and 6. A limitation of this comparison is that it only includes two spatial-temporal fusion models. There are other spatio-temporal fusion models, such as the fusion model proposed by Nguyen et al. (2014), which assumes a common underlying spatio-temporal process for the observations from different sources [Nguyen et al., 2014]. Thus, comparing more data fusion models in future studies will provide a better understanding of their advantages and disadvantages.

7.2 Computationally efficient data fusion models

The nonparametric statistical downscaling model with Gaussian predictive process (NSD-GPP) was proposed in Chapter 5 to improve the computational efficiency of the NSD model for fusing the reflectance data. This is because the parameters in the NSD model are estimated by Gibbs sampling, which involves the inversion of the matrix with the dimension of the number of the locations at each iteration. A Gaussian predictive process is one approximation method for representing the spatial varying coefficients in the NSD model, and it requires the inversion of a matrix with smaller dimensions in the Gibbs sampling process, which thus reduces the computational time for the parameter estimation. The predictive performance and computational efficiency of the NSD model and NSD-GPP model are compared through simulation and using satellite reflectance data from Lake Garda.

It was shown that in the application to the Lake Garda dataset, the NSD-GPP model with the number of knots equal to 10% of the total number of the locations could achieve a similar RMSPE as the NSD model but use about 30% of the computational time for both spatial cross-validation and temporal cross-validation studies. A simulation study with different specifications of the spatial correlation and spatial variation was conducted, and the results suggest that increasing the number of knots in the NSD-GPP model will decrease the RMSPE, and this effect is more obvious for the dataset with strong spatial correlation and large spatial variation. Considering the computational time and the RMSPE, the NSD-GPP model is used in the further development in this thesis.

One limitation of the study in Chapters 5 and 6 is that the change of spatial supports of remote sensors is not investigated. Although the reflectance data used in this thesis are aggregated into the same spatial supports for the remote sensors, the change of spatial supports is a common problem in fusing remote sensing data. Assessing the model performance for the data sets with different spatial supports will be useful in understanding the effectiveness of the proposed NSD-GPP models in this thesis. Another limitation of the NSD-GPP models is the selection of the number of knots. In this thesis, equally spaced knots are used, and the number of knots is determined by the cross-validation method for minimising the root-mean squared prediction error. In the work of Tokdar (2011), the locations of the knots are treated as additional model parameters, and the number of knots is selected to meet the approximation accuracy of the predictive process compared to the original spatial process [Tokdar, 2011]. This method provides an automatisation of the fitting of the models with a Gaussian predictive process, which can be used to select the knots for the NSD-GPP models in future studies. However, this automatisation will increase the computational time because of the additional parameters of knots locations and the change of distance matrix in each iteration of MCMC when the knots locations are updated.

Future study could explore the nearest neighbour Gaussian process (NNGP) as an alternative to save computing time for the NSD model. The NNGP only considers the spatial dependency of the neighbours for each location, which increases the sparsity of the covariance matrix and makes it computationally efficient to manipulate this matrix [Datta et al., 2016].

7.3 Multivariate modelling for fusion over wavelength

The multivariate NSD-GPP model was proposed in Chapter 6 to enable data fusion from the two sensors with a multivariate wavelength dimension and solve the change of temporal and spectral support problem of the two sensors. The two-dimensional B-spline basis functions were used to represent the reflectance over both time and wavelength at each location. Since the same set of

basis functions was used for both sensors, this multivariate model solved the change of temporal and spectral supports problem. The precision matrices for both sensors were assumed to be block diagonal matrices, where a different precision parameter was used for each wavelength. These two modifications on the NSD-GPP model allowed data on multiple wavelengths to be modelled jointly, and this multivariate NSD-GPP model was used to estimate the reflectance at any wavelength and time point within the observed data range.

The temporal predictive performance of the multivariate NSD-GPP model was compared to the univariate NSD-GPP model. The univariate NSD-GPP model was fitted to the reflectance data at multiple wavelengths separately, while the multivariate NSD-GPP model was fitted to the reflectance data for all wavelengths simultaneously. For the temporal prediction, the multivariate model achieved a similar RMSPE and coverage of 95% credible interval as the univariate model for the selected wavelengths. Additionally, the multivariate NSD-GPP model was used to make predictions for unobserved time points and wavelengths, which could not be achieved by applying the univariate NSD-GPP separately for the observed wavelength bands.

One limitation of the multivariate NSD-GPP model is the structure of the residuals, which are assumed to be independent across time and wavelength. This structure assumes that the temporal and spectral variations in reflectance data are captured mainly by the two-dimensional basis functions. It would be beneficial to explore more complex correlation structures, such as the autoregressive structure between two adjacent time points or wavelengths, which are better for capturing the remaining temporal and spectral correlations in the residuals. However, the autoregressive structure requires the observations to be equally spaced in time and wavelength, which may not hold for the reflectance data in this thesis. Thus, the Irregular Autoregressive model including the distance between two adjacent points could fit the residuals with irregular observations in time or wavelength dimension [Eyheramendy et al., 2018]. However, these complex correlation structures require additional computational time to estimate the additional parameters representing the correlations between time points or wavelengths.

Another limitation is that the multivariate NSD-GPP model is not compared to other state-of-the-art data fusion models. For example, the unmixing-based models which decompose the spatial-spectral data $\mathbf{X} \in \mathbb{R}^{L \times N}$ at each time point into an endmember matrix $\mathbf{E} \in \mathbb{R}^{L \times P}$ and abundance matrix $\mathbf{A} \in \mathbb{R}^{P \times N}$, where N is the number of locations, L is the number of wavelength bands, and P is the number of endmembers [Gevaert and García-Haro, 2015, Zhou et al., 2022]. Then, the relationship between the two data sources and the temporal dependency is also modelled upon these endmember and abundance matrices. These unmixing-based models only consider the temporal dependency for two consecutive time points, which makes it difficult to capture the long-term and yearly temporal patterns. Thus, future studies could compare the

multivariate NSD-GPP model and the unmixing-based models for the dataset with a relatively smaller range in the temporal dimension.

The third limitation of is the computational challenge of selecting parameters. Since the multivariate NSD-GPP model requires cross-validation to select the parameters, such as the number of basis functions and the spatial decay parameters, it is computationally demanding to fit the model multiple times in the cross-validation process. In practice, it is possible to fit the univariate NSD-GPP model separately for each wavelength band, as in Chapter 7, which could be used to select the possible range of the optimal number of basis functions for the multivariate model.

7.4 General limitations and future studies

This section discusses the general limitations of this thesis and potential future studies. The first limitation of this study is that the estimates are not compared to ground truth data. In this thesis, the data fusion models are used for reflectance data from satellite sensors since the in-situ reflectance data are not available for this dataset. However, these methods could also be used to fuse remote sensing data with ground truth, such as in-situ data, which could have different spatial resolutions since the in-situ data are limited in spatial dimension.

The second limitation of this study is that it only used the Lake Garda reflectance dataset as an illustration to compare the performance of the data fusion models developed in this chapter. It would be interested to apply the multivariate NSD-GPP model to lakes with different optical water types, such as the clear water, turbidity water and Phytoplankton-dominated water [Liu et al., 2021]. This application can be used to investigate the spectral prediction performance of the multivariate NSD-GPP model for datasets with different spectral patterns. In this thesis, the exponential covariance function is used to model the spatial correlation of the spatially varying coefficients in the NSD-GPP model. For the lakes with more complex spatial structure, the Matérn family covariance function is more flexible with different types of spatial patterns. Thus, exploring different spatial correlation structures is required to generalise the NSD-GPP model and the multivariate NSD-GPP model to different lakes.

The third limitation is that this thesis only focused on the data fusion of two data sources. The fusion models developed in this thesis are not suitable for three or more data sources because these models are based on the downscaling model, which regresses one data source on the other. There are different types of sensors for reflectance data, such as hyperspectral satellites with hundreds of bands and non-Satellite Remote Sensing Data from multi-spectral cameras, infrared sensors and lidar [Yang et al., 2022]. Thus, fusing the data from more than three data sources will be a potential future study. A possible solution is to assume a common underlying process for all

these data sources and then estimate this underlying process conditional on all these observations [Nguyen et al., 2014].

The fourth limitation is that the prediction of reflectance data in this thesis focused on the unobserved data points within the data range, and the predictions into the future time points or to wavelengths outside the data range or spatial prediction outside the data range are not explored. In Chapter 6, predicting the reflectance at the time points near the boundary of the data range results in a larger RMSPE than the other time points. Thus, future work will explore the extrapolation of data fusion models outside the data range. For example, the long-term trend in temporal dimension can be captured by a linear regression model. Then, periodic basis functions, such as Fourier basis functions, can be used to model the seasonal patterns in the detrended data. Finally, the combined model with both long-term trends and seasonal patterns can be used to predict future time points [Jose, 2022].

It was discussed in Chapter 1 that the water quality parameters can be estimated by a function of reflectance data at multiple wavelength bands [Yang et al., 2022]. Thus, it is possible to estimate the reflectance data and infer the water quality parameters through their functional relationships. Although the data fusion models are only applied to the reflectance data in this thesis, it is possible to use these models to other types of data. The NSD-GPP model developed in Chapter 5 is a spatial-temporal model for fusing the data from two sources, which can be applied to spatial-temporal data such as temperature and disease outbreaks. The multivariate NSD-GPP model developed in Chapter 6 can be applied to the multi-spectral image data such as the agricultural fields, ice cover or forest cover. Another possible future study is improving the computational efficiency of the multivariate NSD-GPP model. The newer Sentinel satellites have a spatial resolution of 10m $\times$ 10m [Salgueiro Romero et al., 2020], while the reflectance data used in this thesis have a spatial resolution of 1km $\times$ 1km. The finer spatial resolution of the newer satellites will provide more observations in the spatial dimension, which is computationally expensive using the multivariate NSD-GPP model. One possible solution is using the integrated nested Laplace approximation (INLA) to construct approximations of the posterior marginals [Blangiardo and Cameletti, 2015], which is faster than using the MCMC simulations to estimate the parameters in the multivariate NSD-GPP model.

Bibliography

[Ahn and Park, 2020] Ahn, J.-H. and Park, Y.-J. (2020). Estimating water reflectance at near-infrared wavelengths for turbid water atmospheric correction: A preliminary study for goci-ii. *Remote Sensing*, 12(22):3791.

[Aulló-Maestro et al., 2017] Aulló-Maestro, M. E., Hunter, P., Spyrakos, E., Mercatoris, P., Kovács, A., Horváth, H., Preston, T., Présing, M., Torres Palenzuela, J., and Tyler, A. (2017). Spatio-seasonal variability of chromophoric dissolved organic matter absorption and responses to photobleaching in a large shallow temperate lake. *Biogeosciences*, 14(5):1215–1233.

[Banerjee et al., 2008] Banerjee, S., Gelfand, A. E., Finley, A. O., and Sang, H. (2008). Gaussian predictive process models for large spatial data sets. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 70(4):825–848.

[Behmel et al., 2016] Behmel, S., Damour, M., Ludwig, R., and Rodriguez, M. (2016). Water quality monitoring strategies—a review and future perspectives. *Science of the Total Environment*, 571:1312–1329.

[Belyaev et al., 2014] Belyaev, M., Burnaev, E., and Kapushev, Y. (2014). Exact inference for gaussian process regression in case of big data with the cartesian product structure. *arXiv preprint arXiv:1403.6573*.

[Berrocal et al., 2010a] Berrocal, V. J., Gelfand, A. E., and Holland, D. M. (2010a). A bivariate space-time downscaler under space and time misalignment. *The annals of applied statistics*, 4(4):1942.

[Berrocal et al., 2010b] Berrocal, V. J., Gelfand, A. E., and Holland, D. M. (2010b). A spatio-temporal downscaler for output from numerical models. *Journal of agricultural, biological, and environmental statistics*, 15:176–197.

[Bielski and Toś, 2022] Bielski, A. and Toś, C. (2022). Remote sensing of the water quality parameters for a shallow dam reservoir. *Applied Sciences*, 12(13):6734.

[Blangiardo and Cameletti, 2015] Blangiardo, M. and Cameletti, M. (2015). *Spatial and spatio-temporal Bayesian models with R-INLA*. John Wiley & Sons.

[Boaz et al., 2019] Boaz, R., Lawson, A., and Pearce, J. (2019). Multivariate air pollution prediction modeling with partial missingness. *Environmetrics*, 30(7):e2592.

[Carrea et al., 2022] Carrea, L., Crétaux, J.-F., Liu, X., Wu, Y., Bergé-Nguyen, M., Calmettes, B., Duguay, C., Jiang, D., Merchant, C., Mueller, D., Selmes, N., Simis, S., Spyrakos, E., Stelzer, K., Warren, M., Yesou, H., and Zhang, D. (2022). Esa lakes climate change initiative (lakes_cci): lake products, version 2.0.2. *NERC EDS Centre for Environmental Data Analysis*.

[Carrea et al., 2023] Carrea, L., Crétaux, J.-F., Liu, X., Wu, Y., Calmettes, B., Duguay, C. R., Merchant, C. J., Selmes, N., Simis, S. G., Warren, M., et al. (2023). Satellite-derived multivariate world-wide lake physical variable timeseries for climate studies. *Scientific Data*, 10(1):30.

[Cressie and Johannesson, 2008] Cressie, N. and Johannesson, G. (2008). Fixed rank kriging for very large spatial data sets. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 70(1):209–226.

[Dash and Ogutu, 2016] Dash, J. and Ogutu, B. O. (2016). Recent advances in space-borne optical remote sensing systems for monitoring global terrestrial ecosystems. *Progress in Physical Geography*, 40(2):322–351.

[Datta et al., 2016] Datta, A., Banerjee, S., Finley, A. O., and Gelfand, A. E. (2016). Hierarchical nearest-neighbor gaussian process models for large geostatistical datasets. *Journal of the American Statistical Association*, 111(514):800–812.

[De Boor and De Boor, 1978] De Boor, C. and De Boor, C. (1978). *A practical guide to splines*, volume 27. springer-verlag New York.

[Diggle and Ribeiro, 2007] Diggle, P. and Ribeiro, P. J. (2007). *Model-Based Geostatistics*. Springer.

[Directive, 2000] Directive, W. F. (2000). Eu water framework directive. *EC Directive*, 60.

[Eilers and Marx, 1996] Eilers, P. H. and Marx, B. D. (1996). Flexible smoothing with b-splines and penalties. *Statistical science*, 11(2):89–121.

[Eilers et al., 2015] Eilers, P. H., Marx, B. D., and Durbán, M. (2015). Twenty years of p-splines. *SORT: statistics and operations research transactions*, 39(2):0149–186.

[Epperson, 1987] Epperson, J. F. (1987). On the runge example. *The American Mathematical Monthly*, 94(4):329–341.

[Eyheramendy et al., 2018] Eyheramendy, S., Elorrieta, F., and Palma, W. (2018). An irregular discrete time series model to identify residuals with autocorrelation in astronomical light curves. *Monthly Notices of the Royal Astronomical Society*, 481(4):4311–4322.

[Fang et al., 2006] Fang, J., Wang, Z., Zhao, S., Li, Y., Tang, Z., Yu, D., Ni, L., Liu, H., Xie, P., Da, L., et al. (2006). Biodiversity changes in the lakes of the central yangtze. *Frontiers in Ecology and the Environment*, pages 369–377.

[Febrero-Bande and De La Fuente, 2012] Febrero-Bande, M. and De La Fuente, M. O. (2012). Statistical computing in functional data analysis: The r package fda. usc. *Journal of statistical Software*, 51:1–28.

[Fuentes and Raftery, 2005] Fuentes, M. and Raftery, A. E. (2005). Model evaluation and spatial interpolation by bayesian combination of observations with outputs from numerical models. *Biometrics*, 61(1):36–45.

[Gelman et al., 2014] Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2014). *Bayesian data analysis*. Chapman and Hall/CRC.

[Gevaert and García-Haro, 2015] Gevaert, C. M. and García-Haro, F. J. (2015). A comparison of starfm and an unmixing-based algorithm for landsat and modis data fusion. *Remote sensing of Environment*, 156:34–44.

[Geweke, 1992] Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to the calculations of posterior moments. *Bayesian statistics*, 4:641–649.

[Giardino et al., 2014] Giardino, C., Bresciani, M., Cazzaniga, I., Schenk, K., Rieger, P., Braga, F., Matta, E., and Brando, V. E. (2014). Evaluation of multi-resolution satellite sensors for assessing water quality and bottom depth of lake garda. *Sensors*, 14(12):24116–24131.

[Gilani et al., 2019] Gilani, O., Berrocal, V., and Batterman, S. A. (2019). Nonstationary spatiotemporal bayesian data fusion for pollutants in the near-road environment. *Environmetrics*, 30(7):e2581.

[Gong et al., 2021a] Gong, M., Miller, C., Scott, M., O'Donnell, R., Simis, S., Groom, S., Tyler, A., Hunter, P., and Spyrakos, E. (2021a). State space functional principal component analysis to identify spatiotemporal patterns in remote sensing lake water quality. *Stochastic Environmental Research and Risk Assessment*, 35:2521–2536.

[Gong et al., 2021b] Gong, W., Reich, B. J., and Chang, H. H. (2021b). Multivariate spatial prediction of air pollutant concentrations with inla. *Environmental research communications*, 3(10):101002.

[Goyens et al., 2022] Goyens, C., Lavigne, H., Dille, A., and Vervaeren, H. (2022). Using hyperspectral remote sensing to monitor water quality in drinking water reservoirs. *Remote Sensing*, 14(21):5607.

[Guhaniyogi et al., 2011] Guhaniyogi, R., Finley, A. O., Banerjee, S., and Gelfand, A. E. (2011). Adaptive gaussian predictive process models for large spatial datasets. *Environmetrics*, 22(8):997–1007.

[Hall and Llinas, 2001] Hall, D. and Llinas, J. (2001). *Multisensor data fusion*. CRC press.

[Hastie et al., 2009] Hastie, T., Tibshirani, R., Friedman, J., Hastie, T., Tibshirani, R., and Friedman, J. (2009). Kernel smoothing methods. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, pages 191–218.

[Hijmans et al., 2015] Hijmans, R. J., Van Etten, J., Cheng, J., Mattiuzzi, M., Sumner, M., Greenberg, J. A., Lamigueiro, O. P., Bevan, A., Racine, E. B., Shortridge, A., et al. (2015). Package 'raster'. *R package*, 734:473.

[Huang et al., 2018] Huang, G., Lee, D., and Scott, E. M. (2018). Multivariate space-time modelling of multiple air pollutants and their health effects accounting for exposure uncertainty. *Statistics in medicine*, 37(7):1134–1148.

[Huete, 2004] Huete, A. (2004). Remote sensing for environmental monitoring. *Environmental monitoring and characterization*, pages 183–206.

[Irvine et al., 2007] Irvine, K. M., Gitelman, A. I., and Hoeting, J. A. (2007). Spatial designs and properties of spatial correlation: effects on covariance estimation. *Journal of agricultural, biological, and environmental statistics*, 12:450–469.

[Jolliffe, 2002] Jolliffe, I. T. (2002). *Principal component analysis for special types of data*. Springer.

[Jona Lasinio et al., 2013] Jona Lasinio, G., Mastrantonio, G., and Pollice, A. (2013). Discussing the "big n problem". *Statistical Methods & Applications*, 22:97–112.

[Jose, 2022] Jose, J. (2022). Introduction to time series analysis and its applications. *Christ University, Bangalore*.

[Joseph, 2005] Joseph, G. (2005). *Fundamentals of remote sensing*. Universities Press.

[Kiefer et al., 2015] Kiefer, I., Odermatt, D., Anneville, O., Wüest, A., and Bouffard, D. (2015). Application of remote sensing for the optimization of in-situ sampling for monitoring of phytoplankton abundance in a large lake. *Science of the Total Environment*, 527:493–506.

[Kravitz et al., 2020] Kravitz, J., Matthews, M., Bernard, S., and Griffith, D. (2020). Application of sentinel 3 olci for chl-a retrieval over small inland water targets: Successes and challenges. *Remote Sensing of Environment*, 237:111562.

[Kutser et al., 2016] Kutser, T., Paavel, B., Verpoorter, C., Ligi, M., Soomets, T., Toming, K., and Casal, G. (2016). Remote sensing of black lakes and using 810 nm reflectance peak for retrieving water quality parameters of optically complex waters. *Remote Sensing*, 8(6):497.

[Lindgren and Rue, 2015] Lindgren, F. and Rue, H. (2015). Bayesian spatial modelling with r-inla. *Journal of statistical software*, 63(19).

[Liu et al., 2021] Liu, X., Steele, C., Simis, S., Warren, M., Tyler, A., Spyrakos, E., Selmes, N., and Hunter, P. (2021). Retrieval of chlorophyll-a concentration and associated product uncertainty in optically diverse lakes and reservoirs. *Remote Sensing of Environment*, 267:112710.

[Mavrakakis and Penzer, 2021] Mavrakakis, M. C. and Penzer, J. (2021). *Probability and statistical inference: from basic principles to advanced models*. Chapman and Hall/CRC.

[McBride et al., 2019] McBride, C. G., Rose, K. C., Kumagai, M., Weathers, K. C., Hanson, P. C., and Jennings, E. (2019). Automated high-frequency monitoring and research. In *Lake Restoration Handbook: A New Zealand Perspective*, pages 419–461. Springer.

[Morris, 2015] Morris, R. (2015). Spectrophotometry. *Current Protocols Essential Laboratory Techniques*, 11(1):2–1.

[Neil et al., 2019] Neil, C., Spyrakos, E., Hunter, P. D., and Tyler, A. N. (2019). A global approach for chlorophyll-a retrieval across optically complex inland waters based on optical water types. *Remote Sensing of Environment*, 229:159–178.

[Nguyen et al., 2012] Nguyen, H., Cressie, N., and Braverman, A. (2012). Spatial statistical data fusion for remote sensing applications. *Journal of the American Statistical Association*, 107(499):1004–1018.

[Nguyen et al., 2014] Nguyen, H., Katzfuss, M., Cressie, N., and Braverman, A. (2014). Spatio-temporal data fusion for very large remote sensing datasets. *Technometrics*, 56(2):174–185.

[Nychka et al., 2016] Nychka, D., Hammerling, D., Sain, S., Lenssen, N., Smirniotis, C., and Iverson, M. (2016). Latticekrig: Multiresolution kriging based on markov random fields. *R package version*, 6.

[Odermatt and Gege, 2022] Odermatt, D. and Gege, P. (2022). Lake colors: interpreting apparent optical properties. *Encyclopedia of Inland Waters*, 1:474–489.

[O'Reilly and Werdell, 2019] O'Reilly, J. E. and Werdell, P. J. (2019). Chlorophyll algorithms for ocean color sensors-oc4, oc5 & oc6. *Remote sensing of environment*, 229:32–47.

[O'Neil et al., 2012] O'Neil, J. M., Davis, T. W., Burford, M. A., and Gobler, C. J. (2012). The rise of harmful cyanobacteria blooms: the potential roles of eutrophication and climate change. *Harmful algae*, 14:313–334.

[Palmer et al., 2015] Palmer, S. C., Kutser, T., and Hunter, P. D. (2015). Remote sensing of inland waters: Challenges, progress and future directions. *Remote sensing of Environment*, 157:1–8.

[Pauchard, 2017] Pauchard, N. (2017). Access and benefit sharing under the convention on biological diversity and its protocol: what can some numbers tell us about the effectiveness of the regulatory regime? *Resources*, 6(1):11.

[Peng et al., 2021] Peng, K., Wang, Q., Tang, Y., Tong, X., and Atkinson, P. M. (2021). Geographically weighted spatial unmixing for spatiotemporal fusion. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–17.

[Ploton et al., 2020] Ploton, P., Mortier, F., Réjou-Méchain, M., Barbier, N., Picard, N., Rossi, V., Dormann, C., Cornu, G., Viennois, G., Bayol, N., et al. (2020). Spatial validation reveals poor predictive performance of large-scale ecological mapping models. *Nature communications*, 11(1):4540.

[Ramsay and Silvermann, 2005] Ramsay, J. and Silvermann, B. (2005). Functional data analysis. springer series in statistics.

[Ramsay et al., 2009] Ramsay, J. O., Hooker, G., and Graves, S. (2009). *Functional Data Analysis with R and MATLAB*. Springer Publishing Company, Incorporated, 1st edition.

[Rast et al., 1999] Rast, M., Bezy, J., and Bruzzi, S. (1999). The esa medium resolution imaging spectrometer meris a review of the instrument and its mission. *International Journal of Remote Sensing*, 20(9):1681–1702.

[Reaka-Kudla et al., 1996] Reaka-Kudla, M. L., Wilson, D. E., and Wilson, E. O. (1996). *Biodiversity II: understanding and protecting our biological resources*. Joseph Henry Press.

[Richman, 1986] Richman, M. B. (1986). Rotation of principal components. *Journal of climatology*, 6(3):293–335.

[Roberts et al., 2017] Roberts, D. R., Bahn, V., Ciuti, S., Boyce, M. S., Elith, J., Guillera-Arroita, G., Hauenstein, S., Lahoz-Monfort, J. J., Schröder, B., Thuiller, W., et al. (2017). Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography*, 40(8):913–929.

[Rodgers, 1990] Rodgers, C. D. (1990). Characterization and error analysis of profiles retrieved from remote sounding measurements. *Journal of Geophysical Research: Atmospheres*, 95(D5):5587–5595.

[Rundel et al., 2015] Rundel, C. W., Schliep, E. M., Gelfand, A. E., and Holland, D. M. (2015). A data fusion approach for spatial analysis of speciated pm2. 5 across time. *Environmetrics*, 26(8):515–525.

[Sahu et al., 2006] Sahu, S. K., Gelfand, A. E., and Holland, D. M. (2006). Spatio-temporal modeling of fine particulate matter. *Journal of Agricultural, Biological, and Environmental Statistics*, 11:61–86.

[Salgueiro Romero et al., 2020] Salgueiro Romero, L., Marcello, J., and Vilaplana, V. (2020). Super-resolution of sentinel-2 imagery using generative adversarial networks. *Remote Sensing*, 12(15):2424.

[Salomonson et al., 2006] Salomonson, V. V., Barnes, W., and Masuoka, E. J. (2006). Introduction to modis and an overview of associated activities. *Earth Science Satellite Remote Sensing: Vol. 1: Science and Instruments*, pages 12–32.

[Shen et al., 2016] Shen, H., Meng, X., and Zhang, L. (2016). An integrated framework for the spatio–temporal–spectral fusion of remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 54(12):7135–7148.

[Soomets et al., 2020] Soomets, T., Uudeberg, K., Jakovels, D., Brauns, A., Zagars, M., and Kutser, T. (2020). Validation and comparison of water quality products in baltic lakes using sentinel-2 msi and sentinel-3 olci data. *Sensors*, 20(3):742.

[Spyrakos et al., 2018] Spyrakos, E., O'Donnell, R., Hunter, P. D., Miller, C., Scott, M., Simis, S. G., Neil, C., Barbosa, C. C., Binding, C. E., Bradt, S., et al. (2018). Optical types of inland and coastal waters. *Limnology and Oceanography*, 63(2):846–870.

[Tiberti et al., 2021] Tiberti, R., Caroni, R., Cannata, M., Lami, A., Manca, D., Strigaro, D., and Rogora, M. (2021). Automated high frequency monitoring of lake maggiore through in situ sensors: System design, field test and data quality control. *J. Limnol*, 80(2):1–19.

[Tokdar, 2011] Tokdar, S. T. (2011). Adaptive gaussian predictive process approximation. *arXiv preprint arXiv:1108.0445*.

[Tyler et al., 2016] Tyler, A. N., Hunter, P. D., Spyrakos, E., Groom, S., Constantinescu, A. M., and Kitchen, J. (2016). Developments in earth observation for the assessment and monitoring of inland, transitional, coastal and shelf-sea waters. *Science of the Total Environment*, 572:1307–1321.

[Utazi et al., 2022] Utazi, C. E., Jochem, W. C., Gacic-Dobo, M., Murphy, P., Sahu, S. K., Danovaro-Holliday, C. M., and Tatem, A. J. (2022). Bayesian hierarchical modelling approaches for combining information from multiple data sources to produce annual estimates of national immunization coverage. *arXiv preprint arXiv:2211.14919*.

[Vantrepotte et al., 2012] Vantrepotte, V., Loisel, H., Dessailly, D., and Mériaux, X. (2012). Optical classification of contrasted coastal waters. *Remote Sensing of Environment*, 123:306–323.

[Villejo et al., 2023] Villejo, S. J., Illian, J. B., and Swallow, B. (2023). Data fusion in a two-stage spatio-temporal model using the inla-spde approach. *Spatial Statistics*, 54:100744.

[Wahba, 1975] Wahba, G. (1975). Smoothing noisy data with spline functions. *Numerische mathematik*, 24(5):383–393.

[Wilkie et al., 2019] Wilkie, C. J., Miller, C. A., Scott, E. M., O'Donnell, R. A., Hunter, P. D., Spyrakos, E., and Tyler, A. N. (2019). Nonparametric statistical downscaling for the fusion of data of different spatiotemporal support. *Environmetrics*, 30(3):e2549.

[Wood, 2017] Wood, S. N. (2017). *Generalized additive models: an introduction with R*. CRC press.

[Wramner et al., 2015] Wramner, P., Philipson, P., Brockmann, C., Gangkofner, U., Odermatt, D., Thulin, S., and Brito, J. (2015). Diversity ii - supporting the convention on biological diversity. `http://www.diversity2.info`.

[Wynne et al., 2021] Wynne, T. T., Mishra, S., Meredith, A., Litaker, R. W., and Stumpf, R. P. (2021). Intercalibration of meris, modis, and olci satellite imagers for construction of past, present, and future cyanobacterial biomass time series. *Remote Sensing*, 13(12):2305.

[Xie et al., 2016] Xie, D., Zhang, J., Zhu, X., Pan, Y., Liu, H., Yuan, Z., and Yun, Y. (2016). An improved starfm with help of an unmixing-based method to generate high spatial and temporal resolution remote sensing data in complex heterogeneous regions. *Sensors*, 16(2):207.

[Yang et al., 2022] Yang, H., Kong, J., Hu, H., Du, Y., Gao, M., and Chen, F. (2022). A review of remote sensing for water quality retrieval: progress and challenges. *Remote Sensing*, 14(8):1770.

[Yang et al., 2021] Yang, T., Lu, J., Liao, F., Qi, H., Yao, X., Cheng, T., Zhu, Y., Cao, W., and Tian, Y. (2021). Retrieving potassium levels in wheat blades using normalised spectra. *International Journal of Applied Earth Observation and Geoinformation*, 102:102412.

[Zammit-Mangion and Cressie, 2017] Zammit-Mangion, A. and Cressie, N. (2017). Frk: An r package for spatial and spatio-temporal prediction with large datasets. *arXiv preprint arXiv:1705.08105*.

[Zeng and Binding, 2021] Zeng, C. and Binding, C. E. (2021). Consistent multi-mission measures of inland water algal bloom spatial extent using meris, modis and olci. *Remote Sensing*, 13(17):3349.

[Zhan et al., 2008] Zhan, J., Zha, Y., and Li, L. (2008). Impact of normalization intervals for hyperspectral reflectance on estimation of chlorophyll-a in lake tai, china. In *2008 2nd International Conference on Bioinformatics and Biomedical Engineering*, pages 3014–3017. IEEE.

[Zhang et al., 2010] Zhang, M., Tang, J., Dong, Q., Song, Q., and Ding, J. (2010). Retrieval of total suspended matter concentration in the yellow and east china seas from modis imagery. *Remote Sensing of Environment*, 114(2):392–403.

[Zhou et al., 2022] Zhou, J., Sun, W., Meng, X., Yang, G., Ren, K., and Peng, J. (2022). Generalized linear spectral mixing model for spatial–temporal–spectral fusion. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–16.