To be updated.

Julian Austin

Department of Mathematics, Statistics and Physics.
Newcastle University

This dissertation is submitted for the degree of
Doctor of Philosophy

April 2021

Dedication will be here.

Declaration

Declaration to be placed here.

Julian Austin
April 2021

Acknowledgements

Acknowledgements to be place here.

Abstract

Abstract will be written here.

Table of contents

List of figures

List of tables

Nomenclature

Roman Symbols

$\mathcal{GP}\left(m(\cdot), k(\cdot,\cdot)\right)$ Gaussian Process with mean function $m(\cdot)$ and covariance function $k(\cdot,\cdot)$

J_i Number of temporal observations for i^{th} functional observation.

N Number of spatial observations.

$\mathcal{S}$ Spatial Domain.

$\boldsymbol{s}$ Spatial location.

$\mathcal{T}$ Temporal Domain.

t Temporal location.

$\mathcal{X}$ Functional random variable.

y_{ij} Observed response at time t_j for i^{th} functional observation. See Equation 1.2.

Y Observed data set. See Equation 1.1.

Greek Symbols

χ Functional data (observation of $\mathcal{X}$).

ε_{ij} Noise process at time t_j for i^{th} functional observation. See Equation 1.2.

Subscripts

i Spatial index.

j Temporal index.

k Principal component index.

Acronyms / Abbreviations

CESM-LE CESM Large Ensemble, [27]

CESM Community Earth System Model.

EO Earth Observation.

FDA Functional Data Analysis.

NCAR National Centre for Atmospheric Research.

PS Pressure at reference height of 2m from CESM-LE, [27]

SAR Synthetic Aperture Radar.

TMQ Total vertically integrated precipitation from CESM-LE, [27]

TREFHT Temperature at reference hieght of 2m from CESM-LE, [27]

U10 Wind speed at height of 10m from CESM-LE, [27]

Chapter 1

Introduction

1.1 Earth observation

Many areas of science produce data on both a spatial and temporal scale. Take for example, the production of Earth Observation data. Earth Observation (EO) is the collection of information on the state of a physical, chemical or biological system of the planet. Typically EO data is acquired through some form of remote sensing in addition to perhaps some in-situ measurements. EO data is acquired to study a process either over a large area of land, a large time horizon, or both. For example, such EO studies include; land usage change in wetland environments in southern Spain, [37], crop production in the Netherlands, [28], and land deformation of the Tuscany region over a two year time period, [42]. In each case there is significant spatial and temporal dependency that is to be considered in the observed processes. For example, Raspini et al. use the temporal dependency in ground deformation signals to highlight areas of significant change in movement, [42]. They combine this with spatial maps to provide a monitoring bulletin for their area of interest. Of course to provide actionable insights from EO data requires an understanding of both the spatial and temporal dependency. As such, models that can handle both forms of dependency whilst maintaining parsimony are desired in the EO community.

An area where EO data is prominent is climatology. Climatology is the study of the atmosphere and weather patterns over time. In this case spatial and temporal dependency in the EO data used in various climatology studies is fairly evident. For example, consider the Community Earth System Model (CESM), [27], produced by the National Centre for Atmospheric Research (NCAR). Such a model provides simulations of various aspects of the Earth's climates for past, present and future time points. See Section 2.1 for a more detailed description of such a data set. Figure 1.1 provides an example of a subset of the CESM data focusing on the temperature in Kevlin (K) over the globe. As expected we can see a clear a temporal pattern emerging in various regions which relates to the seasons. For example, there is clear evidence for temporal correlation over the polar regions due to the gradual increase then reduction in temperature we see over these regions. Figure 1.1 also highlights clear spatial patterns with temperature varying as we move around the

globe. One particularly noticeable pattern is the localised change in temperature over north Africa between the months of May to September. For another example, consider the transition from sea to land. The change in temperature as we move from sea to land is often more abrupt than the same change in temperature we may see when moving the same distance over land. This is possibly evidence for the existence of a complex spatial and temporal process driving such a variable. Understanding such a complex process motivates a model which must take into account both spatial and temporal correlations jointly.

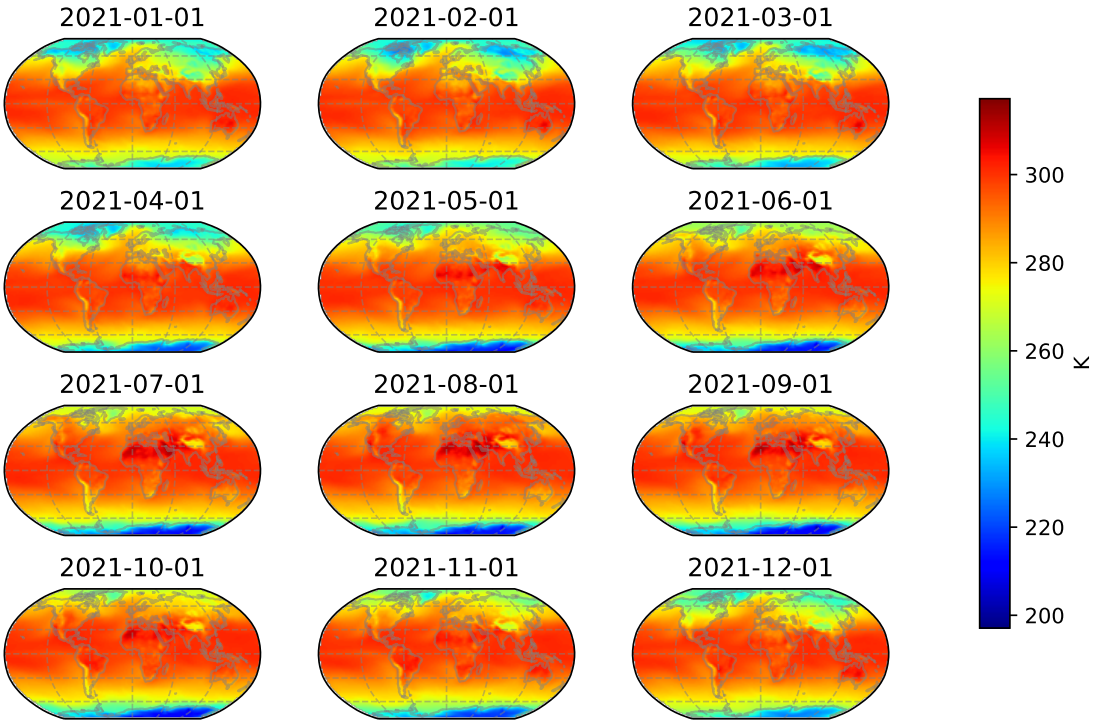


Fig. 1.1 Monthly average temperature over the globe in Kelvin (K) from a single simulation from CESM-LE. The figure illustrates both the temporal and spatial dependency that is observed in temperature across the globe. The figure is projected to the Robinson projection for illustration purposes.

The CESM data is generated through simulations of a complex model of the Earth (See Section 2.1 for more detail). However, EO data is also becoming more frequently generated through remote sensing. Remote sensing is the monitoring of a physical characteristic of an area through measuring its reflected and emitted radiation at a distance. Space borne remote sensing, typically achieved through the use of satellite based sensors, is becoming more prominent as a source of EO data and in particularly as a source of climatology data. The three studies above, [37, 42, 28], use space borne remote sensing to observe their process of interest. This is largely due too the increase in satellites launched which have been designed to capture various processes of the earth. Figure 1.2 highlights the rise in availability of a single type of remote sensing satellite. One particularly prominent remote sensing system is the European Space Agency's Sentinel Constellation, [3]. The Sentinel constellation of satellites provides a wide range of remote sensing sensors which are easily

accessible. The constellation provides capabilities to capture various physical characteristics through the many forms of sensors equipped to its satellites. These include Synthetic Aperture Radar (SAR), optical and multispectral sensors. For example, European Space Agency's Sentinel-4 from the Sentinel Constellation, [3], provides observation dedicated to air quality monitoring. As such the Sentinel constellation has been widely used in EO studies; the three studies above, [37, 28, 42] all utilise the Sentinel 1 SAR sensors for their observation source.

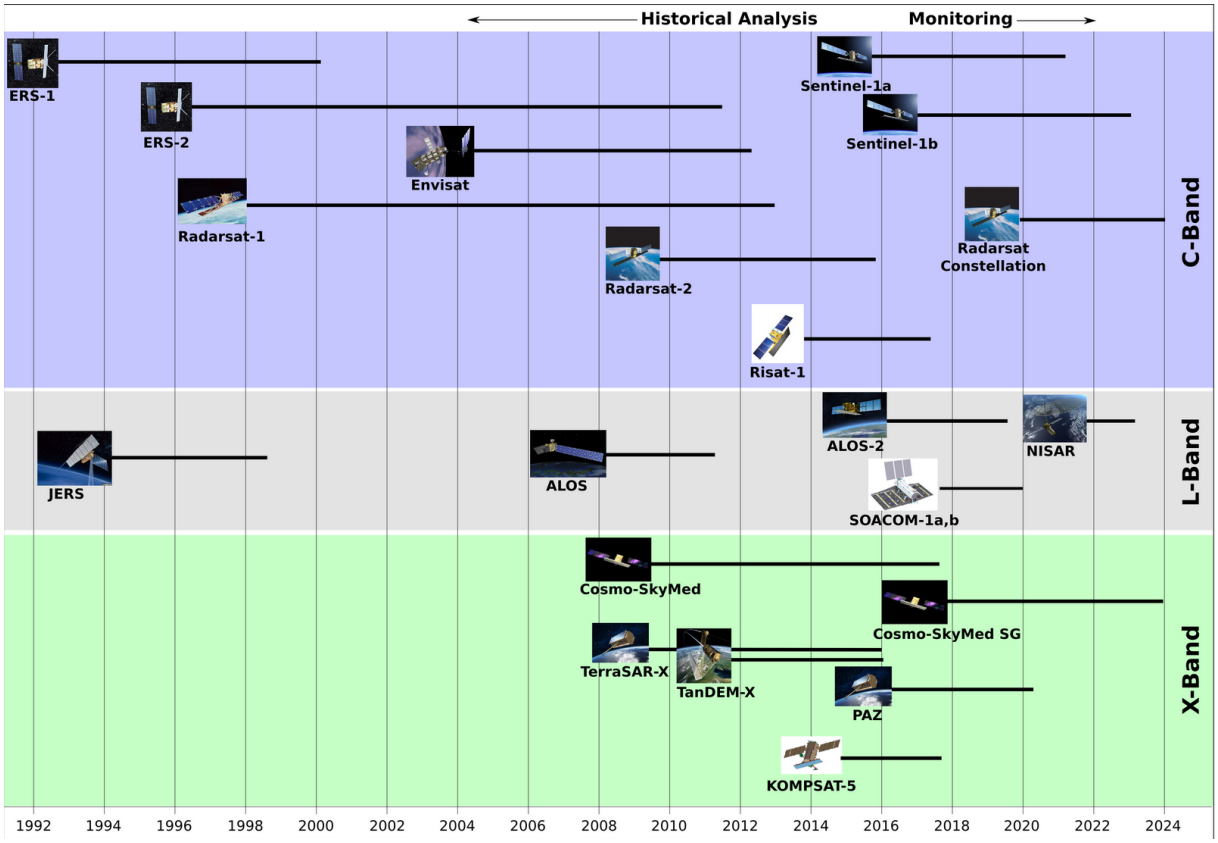


Fig. 1.2 A timeline of major satellite launches and operating periods for EO missions using SAR based sensors. This highlights the rapid rise in availability of such remote sensing capabilities, driven by the demand for observations to cover large spatial and temporal scales in areas such a climatology. The C, L, X band refers to the type of SAR sensor equipped to the satellite which are used in different applications. See [38] for details.

A prominent focus of the Sentinel satellite constellation is to provide repeated observations at relatively high frequency, [3]. This is in response to the rising demand for monitoring EO processes over time. This has been made possible by the development of remote sensing technologies which makes the revisit time possible. For example, the Sentinel 1 satellite constellation can provide revisit times of approximately five days for areas of Europe. Such short revisit times are advantageous as they give higher temporal resolution and thus studies can incorporate this additional information. For example, Raspini et al. utilise this in their study of land deformation change to identify anomalous regions, [42]. The increasing availability of high temporal frequency EO data such as those provided by the CESM-LE data set or the Sentinel satellite constellation thus drives a

demand for statistical models which can handle both high resolution spatial and temporal dependency.

EO data, such as those provided by the CESM model or through remote sensing, have an inherent spatial and temporal dependency. That is to say the underlying process driving both remotely sensed observations of the Earth and the CESM simulations will vary over the globe and also will be driven by the state of the system at prior time points. That is not to say the process is the same for both but rather that there is a commonality in that they could both be considered spatio-temporal processes. In addition there are more concrete similarities in the collection of such EO data. Typically, EO data are described on a lattice of points over space which is usually regular. This usually relates to the pixels of an image over the area of interest. Such a lattice is represented through a geodetic coordinate system which grounds a datum to a real world location. Finally, EO data typically have repeated observations through time over the same space. This usually relates to the repeated imaging of the same area of interest at multiple points in time.

The description of such spatio-temporal processes is well studied in statistics and a large amount of effort has been used to develop various models to suit them. A well known monograph which deals with such processes is that written by Cressie and Wikle, [9]. The monographs details various forms of spatio-temporal processes and typically focuses on the extension of spatial method to incorporate the additional temporal dimension. We discuss these methods in more detail in Section 1.3. Of particular importance in these methods is that temporal and spatial dimensions are treated distinctly as they are inherently different in the physical process. For example, one could consider a spatial point influencing its neighbours in all directions however a temporal point reasonably shouldn't influence its past. As such there is often a distinction in the method used to model the temporal and spatial aspects of the physical process.

An area of statistics which is often used to model data with temporal dependency is Functional Data Analysis (FDA). FDA is typically applied to analyse data which vary over a continuum, [41]. Time is one such continuum. EO data with high frequency temporal observations are therefore suitable candidates for FDA models. FDA is a relatively new branch of statistics and as such few studies have been presented which use FDA techniques on EO data. Liu et al. considers FDA techniques on periodic EO data, [31]. Similarly Hooker et al. considers FDA techniques to model the Harvard modified vegetation index sourced from EO data. In both the above studies they consider EO data as a collection of functional observations indexed by space where each functional observation represents a trajectory over time.

The monograph of Ramsay and Silverman provides a comprehensive introduction to the themes of FDA, [41]. FDA are often intuitive since viewing responses as being trajectories from an unknown smooth random function in some contexts closely matches the actual data generating process compared to a multivariate analysis. The use of such techniques therefore could be helpful in modelling such high frequency EO data in conjunction with the multivariate methods discussed by Cressie and Wikle in [9]. However, focus in the FDA literature to date has primarily revolved around independently observed functional

data. This is typically not the case in our motivating case of EO data where there is often obvious spatial dependency. Thus there is a need to describe functional data models which incorporate dependency among observations. In this work we consider developing such models for dependent functional data with a focus on application to EO data. We consider adapting well studied FDA methodologies and borrow techniques from spatio-temporal statistics to allow for spatially dependent observations. In Section 1.2 we make concrete our definition of functional data.

1.2 Functional representation

As mentioned in Section 1.1 EO data can be viewed as a collection of functional data. However there is a choice about how we interpret observations in this conversion. We may consider the data as a collection of functional observations with time being our functional dimension and space our collection dimension. For example, we may consider each spatial location from Figure 1.1 giving rise to a trajectory over time of which we have only observed 12 time points. Or we may consider the functional observations having a spatial domain and the collection dimension being time. That is we may consider each image in Figure 1.1 being a surface with observations only at pixel locations and the collection consists of a time series of such surfaces. The canonical presentation of functional data in FDA is to use time as the functional dimension, [41], and thus we use the below definition of functional data from this point of view.

1.2.1 Functional data

Multivariate data analysis usually revolves around the study of observations which are finite dimensional and is well studied. Modern data collection techniques can now create data which are extremely numerous and thus can often be viewed as functions.

For example, Ferraty and Vieu consider the case where we can observe a random variable at several times between some minimum and maximum time, $(t_{\min}, t_{\max})$, [14]. A single observation can then be considered as the collection $\{X(t_j); j = 1, 2, \ldots, J\}$ where J is the total number of temporal sample points and $X(t)$ is the response variable at time t. Unlike multivariate data we consider the case that the separation between observations becomes minimal. That is we consider the data as an observation from the continuous random process $\mathcal{X} = \{X(t); t \in (t_{\min}, t_{\max})\}$. We therefore propose as in [14] and [47] the following definition of a *functional variable*.

Definition 1.1 (Functional Variable)**.** *A random variable $\mathcal{X}$ is called a functional variable if it takes values in an infinite dimensional space (or functional space). An observation χ of $\mathcal{X}$ is called a functional data.*

Further to this, suppose we observe a collection of functional data (realisations of $\mathcal{X}$). Then we will denote this collection by the term *functional dataset*.

Definition 1.2 (Functional Dataset)**.** *A functional dataset, $\chi_1, \chi_2, \cdots, \chi_N$ is the collection of N realisations of functional variables $\mathcal{X}_1, \cdots, \mathcal{X}_N$ identically distributed to $\mathcal{X}$.*

The canonical way to present functional data and the subsequent methods is to use time as the continuous variable, [41, 14, 47], as described above. However, there is no such restrictions in either Definition 1.1 or Definition 1.2. In fact, another case is to consider the functional domain of the variables to be space. In our proposed methodologies we present when possible with respect to time due to the simplification it brings in notation. We will make explicit reference to when we change the domain of our functional data, for example if we consider space as our continuous domain.

We introduce the following notation for use in the remainder of this work. We consider our EO data set to be observed in some spatial domain which we denote by $\mathcal{S} \subset \mathbb{R}^2$ and temporal domain donated by $\mathcal{T} \subset \mathbb{R}$. Any observed dataset we can enumerate with one index over the spatial location and the other indexing the temporal locations.

We assume our dataset is comprised of N spatial locations and let $\boldsymbol{s}_i \in \mathcal{S}$ be the spatial location of the i^{th} observed functional variable. At each spatial location we suppose we observe J_i temporal observations and denote by $t_{ij} \in \mathcal{T}$ the j^{th} temporal observation of the i^{th} functional variable. Then our dataset can be summarised by Y where:

$$Y = \{y_{ij}; i = 1, 2, \cdots, N, j = 1, 2, \cdots, J_i\} \tag{1.1}$$

where y_{ij} is the response value of the i^{th} functional variable at time t_{ij} observed with error. That is we consider for each spatial location the discrete temporal observations being a sample from a realisation of a functional variable observed with error. That is:

$$y_{ij} = \chi_i(t_{ij}) + \varepsilon_{ij} \tag{1.2}$$

where, as in Definition 1.2, χ_i is a realisation of functional variable $\mathcal{X}_i$ for $i = 1, 2, \cdots, N$. We consider each functional variable as being identically distributed as $\mathcal{X}$. As is common in most observation models, we assume that we observe data with error. Typically one assumes that the error process $\{\varepsilon_{ij}; i = 1, 2, \cdots, N, j = 1, 2, \cdots, J_i\}$ is a white noise process with variance σ_ε^2.

In this case one considers the modelling of the EO dataset by ensuring smoothness of some kind over the temporal domain via its functional data representation. We can then consider building in spatial dependency by assuming a sampling correlation in our N functional data. An area where such spatial dependency has been long studied is multivariate spatio-temporal methods.

1.3 Spatio-Temporal methods,

In the statistical literature spatial and spatio-temporal models have been extremely well studied, especially due to the prevalence of geo-statistical applications. In the following we

briefly review some of the most commonly observed spatial and spatio-temporal statistical models in the multivariate analysis literature.

The monograph of [8] and references within provide a succinct summary of traditional methodologies in spatial statistics; many of which are applicable to EO data. Generally speaking, spatial data can be split into one of three categories; geo-statistical, area and point process data, [8]. In this work the EO data described in Section 1.1 are most suitably modelled using geo-statistical models. The canonical model used in geo-statistical setting is the Kriging model. The Kriging model is well described in [48]. Such models treat spatial data as samples from a random spatial process and that predictions for unknown values can be calculated from a weighted combination of known values in a neighbourhood of our unknown location utilising the correlation among neighbouring points. A prime example of the spatial Kriging model in use for EO data is given in [43]. Extensions to the basic Kriging technique have also been employed across a number of geo-statistical settings, including Co-Kriging involving extra covariate information for reconstruction, [59]. Kriging is well known in many fields through various names, in the FDA literature it is most often referred to as Gaussian processes regression. Shi and Choi describes in detail the concept of Gaussian processes in the context of functional regression, [47], and we discuss them more in Section 3.5.

As is detailed in [8] an key aspect to geo-statistical modelling is the specification of spatial dependency in the observed data. A common way for such specification is through parametric covariance or kernel functions. [8] details the traditional stationary parametric functions such as the Matèrn covariance. These commonly rely on the assumption of isotropy and stationarity in modelling which rarely holds in practice. Further literature has considered extensions of these and is in fact an active area of research. Schmidt and Guttorp compares a variety of methods for producing non stationary and heterogeneous covariance structures for the goal of spatial interpolation, [46]. They group the various methods of creating such structures into four categories; deformation, convolution, covariate, and stochastic partial differential equations. The deformation approach proposed by Sampson and Guttorp extends the anisotropic stationary covariances such as though describe in [8] by allowing for a non linear transformation to the space which creates a latent space where isotropy holds, [45]. The convolution approach proposed by Higdon uses a specific form of the covariance kernel which can be represented as a convolution between a convolution kernel and a white noise process. We discuss such an approach more in Chapter 5. The covariate based approach to constructing non stationary kernels tends to use an adaption to the convolution or deformation approaches with specific covariates, [46]. Finally the stochastic partial differential equation method proposed by Lindgren et al. construct non stationary covariances through formulating the resulting Gaussian process as the solution of a stochastic partial differential equation which guarantees the construction of a well defined covariance function, [30].

A natural extension to purely spatial modelling of spatio-temporal data is to include the temporal domain. Such models are known as spatio-temporal models. Spatio-temporal models are well discussed in the monograph [9] by Cressie and Wikle. Spatio-temporal

Kriging models are well suited to EO data; however such models are relatively scarce in the literature. Militino et al. considers the application of such modelling in the satellite remote sensing literature and reasons the lack of such modelling is primarily due to the added complexity such models produce in specifying valid space-time covariance functions, [35]. As such, one particular direction spatio-temporal modelling has considered is the creation of spatio-temporal covariance functions. Spatio-temporal covariance functions are discussed in [9]. Separability between spatial and temporal correlations is often a key assumption in some methods due to the reduction in computational complexity such models offer. In particular for EO data, [16] consider such selection of separable covariances and [12] consider such models for air pollution data. However, the separability assumption may be to restrictive for capturing complex spatio-temporal processes, [9]. As such [36, 15, 4] considers tests for when such assumptions hold. The work in [7, 17, 25] consider the construction of non separable covariance functions for when separability doesn't hold for use in spatio-temporal models.

1.4 Summary of Research

The motivation of this work is to provide a model designed for EO data which provides an explanation of both the spatial and temporal process in a parsimonious way. We present a novel method named Correlated Principal Analysis through Conditional Expectation (CPACE), that is designed for modelling EO data. The model builds upon existing FDA techniques to extend modelling from independently observed functional data to functional data which exhibits spatial correlation. The emphasis in this work is to utilise the FDA paradigm over the temporal domain to aid in the decomposition of the data; with the understanding that our data generating process is smooth across the temporal domain. Such a decomposition gives a parsimonious description of the data with respect to its temporal domain by describing its principal modes of variation. We then estimate a spatial correlation structure for each component using well known spatial statistical methods. The combination of the resulting estimated spatial covariance structures with the principal directions aims to capture the majority of temporal and spatial dependency observed in the data. We can then utilise the CPACE model to help predict response at unseen spatial and temporal locations, which is a keen area of interest in EO studies. We asses our model using various simulated data both with known correct data generating distribution and to simulations drawn from an incorrect data generating procedure. We apply our CPACE model to the select atmospheric variables from the CESM data set as an example application of the model to EO data.

In particular the work is structured as follows. In Chapter 2 we describe our example data sets which we use to illustrate the performance of the model. In Chapter3 we present the methodologies underpinning the CPACE models, these are typically well known FDA and spatio-temporal statistical methods. We also present the smoothing methodologies used to estimate the mean and covariance surfaces of our random functional variables. In Chapter 4 we present an interim model built on the combination of two well known existing

methodologies in the FDA literature with a focus on application to an EO data set. Such a model proposes an novel approach to modelling EOn data but helps to highlight the need of including both spatial and temporal effects in modelling such data. In addition, the proposed model in Chapter 4 gives an opportunity to explore EO data where the functional domain is space rather than time. We present the benefits and limitations of such an approach in practice in this chapter. In Chapter 5 we introduce the main contribution of this work; which is the CPACE model for correlated functional data. We describe the model in detail as well as providing asymptotic results for the model. In Chapter 6 we apply the CPACE model to simulated and real world data sets. Simulation results are presented with comparisons to various existing models with a focus on comparative ability to recover known data generating parameters. Applied results to real world data sets are included with comparisons to existing techniques with a focus on interpolation and forecasting abilities of the model. In Chapter 7 we highlight the practical difficulties in implementing the model with discussion on various techniques which are used to overcome the high dimensionality which is typical in the EO data. Finally, in Chapter 8 we draw the conclusions of the work and present area of further work.

Chapter 2

Data sets

In the following chapter we describe in detail our data which we will use as a source for assessing the performance of the models described within. We use a publicly available set of climate model simulations known as the CESM Large Ensemble (CESM-LE) data set, [27]. The CESM-LE data set provides a good example of EO data that is discussed in Section 1.1. The data set is publicly available from https://www.cesm.ucar.edu/projects/community-projects/LENS/data-sets.html.

2.1 Community Earth System Model - Large Ensemble

The CESM-LE data set is an extremely popular and significant data set in the climate research community. It was developed to enable the assessment of recent past and near future climate change in the presence of internal climate variability, [27]. It does so by providing forty simulations of a complex climate model where each simulation is subject to the same radiative forcing scenario but initialised from a slightly perturbed atmospheric state. As such the forty resultant simulations present the various trajectories the model might take due to internal climate variability of the model.

The model used to run the forty member ensemble is the Community System Earth Model version 1, [20], with the Community Atmosphere model version 5, [20], as the atmospheric component. The model is a fully coupled climate model which consists of a model for each of land, ocean, atmosphere and sea ice components of the climate. These are brought together with a coupler model. Figure 2.1 provides a simple overview as to how CESM model couples the various components. Such a model is capable of simulating various Land, Ocean, Atmosphere and Sea Ice variables of the climate, such as the wind speed, temperature or pressure. The CESM-LE produces simulations of such variables on the nominal 1 deg horizontal separation across the globe which induces our spatial resolution of the data. The ensemble produces variables at three different levels of temporal resolution between the years 1920 and 2100 for non-control simulations. The ensemble is able to produce variables at 6-hourly, daily, and monthly intervals.

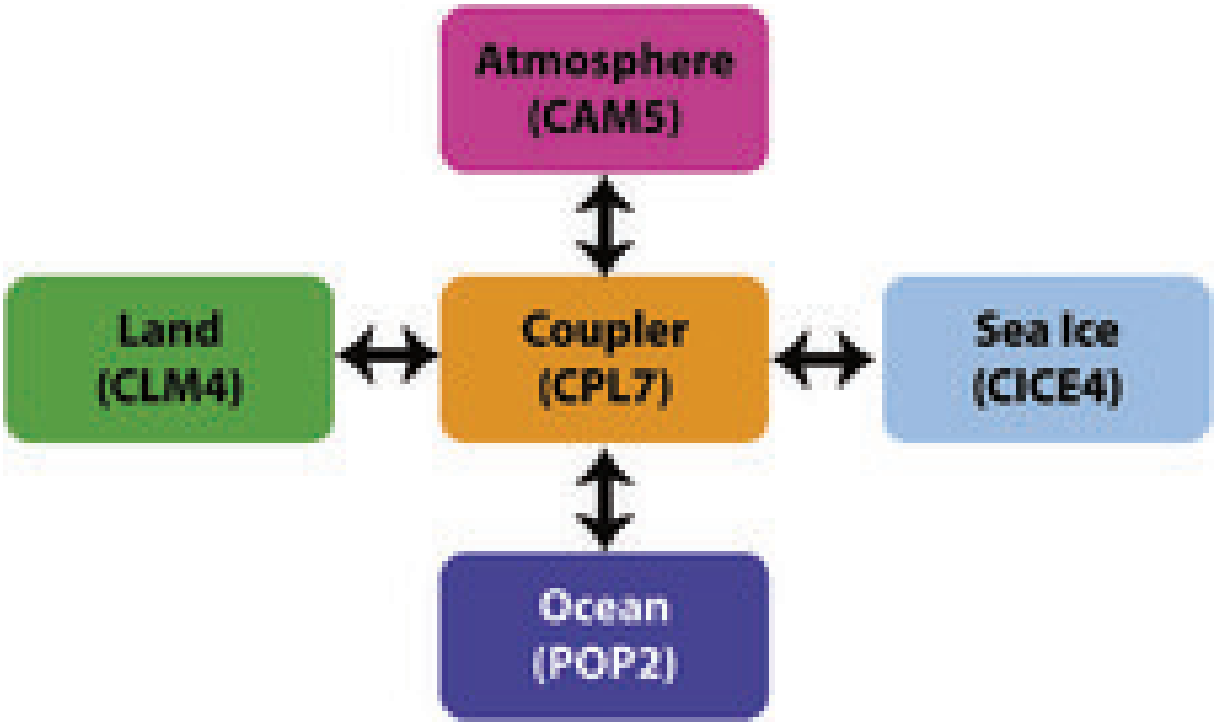


Fig. 2.1 The component models for the full CESM model, Figure from [27]. The individual component models are atmospheric (CAM5), ocean (POP2), land (CLM4), Sea Ice (CICE4), and coupler (CPL7). Details of which can be found in [27] and [20].

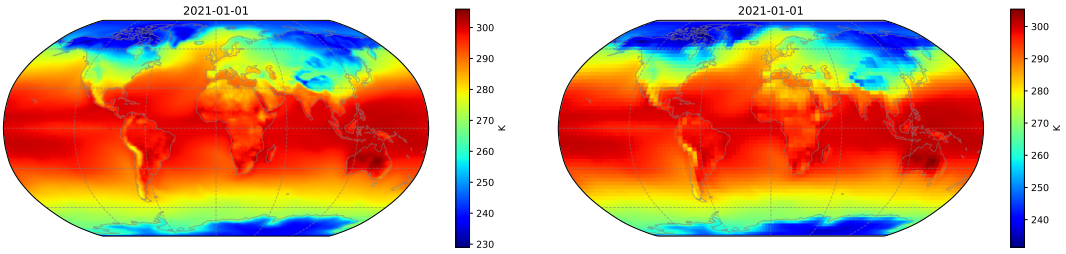
For this body of work we use the CESM-LE data by considering the forty members as separate simulations. Each simulation represents a single realisation of the various climate variables generated by the model described in [27] where the variation between realisations is coming from the internal climate variability. We apply a set of preprocessing to the raw data provided by the CESM-LE model as described below.

2.1.1 Preprocessing

The full CESM-LE data consists of a large number of climate variables on a relatively large spatial grid consisting of 192×276 locations and as such is a rather large data set. The main goal in our preprocessing is to reduce the size of the data through a series of variable selection, spatial resampling, and temporal sectioning. We reduce the data size by considering only a subset of the full data set by selecting four variables to study from the full model which are; temperature, pressure, wind speed, and precipitation. The next preprocessing step we use is a temporal cut. We consider only the output of the CESM-LE which occurs between December 2020 and January 2026. These time points were chosen such that the length of time gave reasonable ability to capture periodic elements but that the size of the data did not become too large. By using monthly frequency observations and this five year time horizon we have a sixty temporal observations for each spatial grid point and for each of our four variables considered.

The final step in our preprocessing pipeline is to reduce the spatial dimension. To do this we resample the model simulations to a smaller spatial grid for each variable of interest. Resampling is achieved by averaging values of neighbouring locations until our

desired resolution is achieved. In this case we resample until the spatial size of the data set is 64×96 which corresponds to a reduction factor of 3 from the original CESM-LE data. Figure 2.2 shows the comparison between the full and resampled spatial observation grid over the globe due to our preprocessing step The figure uses the temperature variable as at January 2021 as an illustrative example. Obviously using such an approach reduces the spatial resolution and thus our ability to see small scale spatial patterns. However, this is traded off against agility in terms of modelling due to the reduced data size. The reduction factor of 3 was chosen based on this trade off.



(a) Full resolution. (b) Reduced resolution.

Fig. 2.2 The resampled spatial grid of observation measurements across the globe. Notice the reduced spatial resolution in Figure 2.2b compared to that in Figure 2.2a due to the resampling causing some loss of fine spatial detail. Example used is the average monthly temperature in Kelvin (K) on January 2021.

2.1.2 Variables

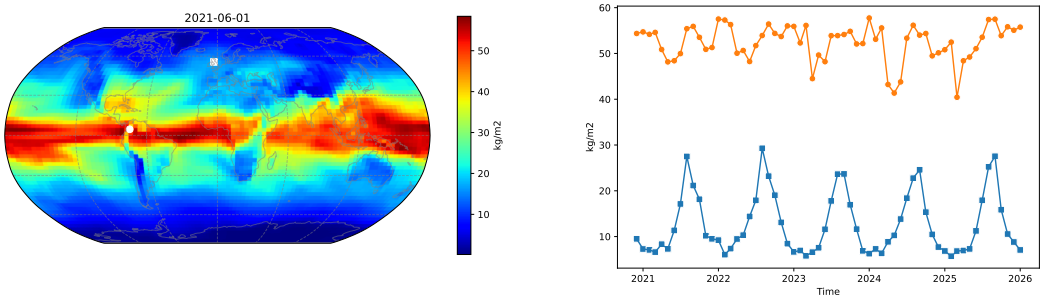
In the following section we focus our description to the four atmospheric model variables from the CESM-LE simulations that we will use in this body of work. These are; pressure, temperature, precipitation, and wind speed. We describe each variable in detail in their respective section and throughout this work we consider each as a separate EO data set. We aim to show, through the use of an example time point and spatial locations, the various spatial and temporal processes that exist in each of these variables. These make such a data set a credible EO data set to test our proposed CPACE model on.

Precipitation

The total (vertically integrated) precipitable water component abbreviated as TMQ is an atmospheric component output of the CESM-LE, [27]. The variable is given units of $\mathrm{kg\,m^{-2}}$ and is available monthly on the full spatial grid. The monthly precipitation is calculated as the average over time from the CESM-LE model six hourly output.

We can see clearly the spatial variability of the precipitation over the globe by considering the heat map of June 2021 monthly precipitation for a single simulation; which is shown in Figure 2.3a. As one would expect there is clear spatial correlation. For example, the tropics observe large amounts of precipitation whereas desert regions observe little. We

can also see some subtler differences in the spatial correlation structure. Figure 2.3a shows that bands of precipitation are evident over the globe. This indicates that precipitation is much more correlated over lines of latitude than lines of longitude. This may be an indicator of spatial anisotropy in the generating process. There is also a case of observing more complex, possibly non stationary, spatial processes as the correlation structure seems different between say North America and Indonesia. We can similarly observe clear temporal correlations in the precipitation variable of the CESM model. In particular, Figure 2.3b shows the time series of two locations; the United Kingdom (UK) and Colombia. Each exhibit clear periodic signals as wet seasons and dry seasons repeat each year. However, we can see a clear level difference between the UK and Colombia precipitation as well as differences in the range of precipitation. This highlights the fact that not only do we have temporal correlation but this correlation is dependent on location.



(a) TMQ as at June 2021.

(b) TMQ over time.

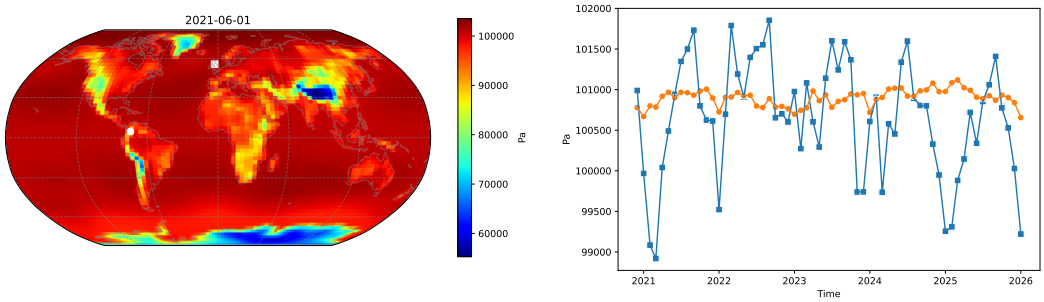
Fig. 2.3 Overview of the monthly average precipitation variable (TMQ) from CESM-LE ensemble member 1. Figure 2.3a highlights the spatial correlation present while Figure 2.3b highlights the temporal correlation at two distinct locations; namely Colombia and the United Kingdom. The orange circle and blue square markers mark these respectively on Figure 2.3b with the white markers highlighting their locations in Figure 2.3a. Note the level difference in temporal correlation structure between the two location, indicative of a spatio-temporal correlation process occurring.

Pressure

The surface pressure variable abbreviated as PS is another atmospheric component output of the CESM-LE, [27]. The component is given in Pascals (Pa) and represents the surface pressure at a height of 2m. It is available monthly on the full spatial grid and the monthly average is calculated as the average over time from the CESM-LE model six hourly output.

Figure 2.4 gives a brief insight to this variable. We can see the spatial correlation structure of Pressure over the globe in Figure 2.4a. One can clearly see areas of high and low pressure. For example, there is a significant difference between the low pressure zone over Antarctica and high pressure zone over Australia. It is interesting to note that we see a clear difference in the smooth structure over sea and a rougher structure over land variables. This again might motivate that a non stationary spatial process is driving such a variable. Considering the temporal variation displayed for Colombia and the UK in

Figure 2.4b; we can see definite structure over time, albeit different for each location. The UK exhibits much more variation in pressure than Colombia, however both do exhibit temporal correlation. Again, similar to the precipitation variable discussed in Section 2.1.2, this might suggest that modelling such a variable will need to consider both spatial and temporal correlations in conjunction with each other.



(a) PS as at June 2021. (b) PS over time.

Fig. 2.4 Overview of the monthly average pressure variable from CESM-LE ensemble member 1. Figure 2.4a highlights the spatial correlation present while Figure 2.4b highlights the temporal correlation at two distinct locations; namely Colombia and the United Kingdom. The orange circle and blue square markers mark these respectively on Figure 2.4b with the white markers highlighting their locations in Figure 2.4a. Notice the stark difference between the time series variance in the UK and Colombia.

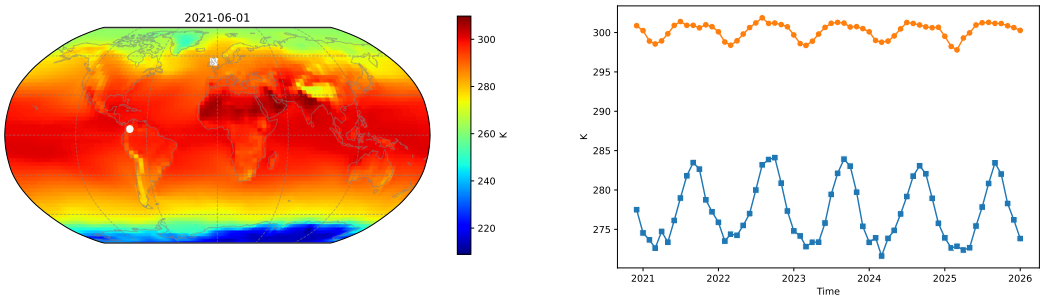
Temperature

The temperature variable abbreviated to TREFHT is an atmospheric component output of the CESM-LE, [27]. The variable refers to the average temperature in Kelvin (K) at the model reference height which is 2m above sea level. The average is available monthly with said average being calculated from the model six hourly output over the month.

Quite clearly the temperature exhibits spatial correlation across the globe and periodic signals through time as area move from winter to summer. Figures 2.5a, 2.5b highlight this for the spatial and temporal correlation respectively. Clearly the temperature at June increases as we move closer to the equator and decreases at the poles. Similarly to the precipitation, we observe that the spatial correlation structure is clearly anisotropic. Correlation is much more pronounced over longitude than latitude. Looking more deeply at Figure 2.5 we can observe more complex correlation structures, as mentioned in Section 1.1. For example, Asia exhibits a localised area of low temperature right next to an area of relatively high temperature. This is very different to the extremely smooth variation that we see over the larger oceans such as the Atlantic. Again similar to the previous variables, this is an indication that there may be a non-stationary spatial process helping to drive this variable.

Figure 2.5b gives an insight into the temporal correlation structure in the variable for two locations; namely Colombia and the UK. There is strong evidence of a periodic signal driving both but clearly there is a level shift and change in amplitude for the two locations.

This is again similar to the precipitation variable discussed in Section 2.1.2 and motivates the idea that there is a clear spatio-temporal process driving the variable rather than just either temporal or spatial process.



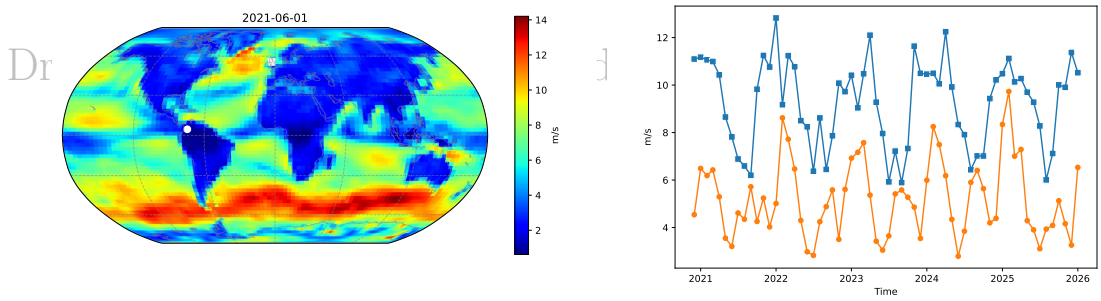
(a) TREFHT as at June 2021. (b) TREFHT over time.

Fig. 2.5 Overview of the monthly average temperature variable from CESM-LE ensemble member 1. Figure 2.5a highlights the spatial correlation present while Figure 2.5b highlights the temporal correlation at two distinct locations; namely Colombia and the UK. The orange circle and blue square markers mark these respectively on Figure 2.5b with the white markers highlighting their locations in Figure 2.5a.

Wind speed

The wind speed variable abbreviated to U10 is another atmospheric component output of the CESM-LE model. The variable refers to the average wind speed in $\mathrm{m\,s^{-1}}$ at a height of 10m above sea level. Again the variable is available on the full spatial grid and is available as a monthly average over time.

We visualise the spatial correlation of the variable in Figure 2.6a by considering a snap shot of the monthly average wind speed in June 2021. We can see, in contrast to the previous three variables, that this has a much rougher spatial correlation structure over the sea. In fact it is interesting to observe the distinct difference in variability over the sea compared to that over the land. This may suggest that we have two types of correlation structure existing for this variable, one for the land and one for the sea. In this case the model for the whole variable will clearly need to include a non-stationary spatial component to capture such a phenomena. Wind speed, alike the other studied model variables, also exhibit temporal correlation. This is illustrated for two locations, Colombia and the UK, in Figure 2.6b. The temporal correlation is much less pronounced for this variable than compared to the others. Visually there is perhaps evidence for a periodic signal for the UK location. However, yet again we do see a clear level shift between the two location. This again suggests that the spatial coordinate clearly impacts the observed function of wind speed over time. Similar to the other variables this suggests a model which includes both space and time as drivers for the process.



(a) U10 as at June 2021.

(b) U10 over time.

Fig. 2.6 Overview of the monthly average wind variable from CESM-LE ensemble member 1. Figure 2.6a highlights the spatial correlation present while Figure 2.6b highlights the temporal correlation at two distinct locations; namely Colombia and the UK. The orange circle and blue square markers mark these respectively on Figure 2.6b with the white markers highlighting their locations in Figure 2.6a.

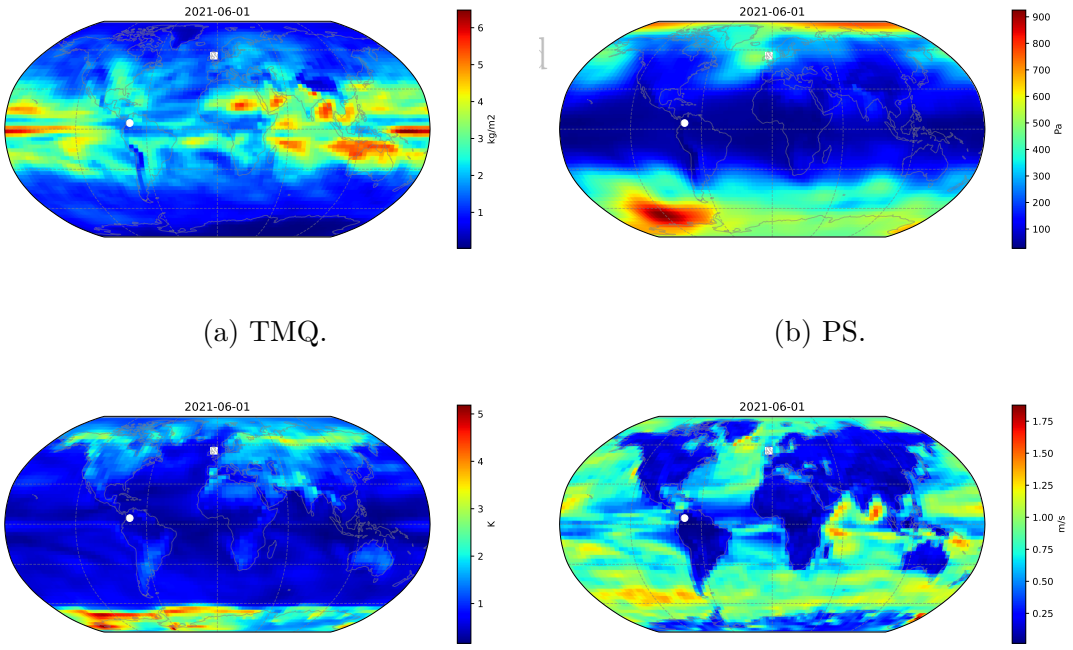
2.1.3 Simulations

For each variable discussed in Section 2.1.2 the CESM-LE data provides forty simulations; one from each ensemble member. We have illustrated the spatial and temporal correlations in the four variables in the Figures 2.3, 2.4, 2.5, and 2.6 for a single simulation. However we also have variability between simulations and it is useful to illustrate this as it may provide insight into where we may expect difficulty in modelling. It is important that any model developed for such data should be able to account for this variability in the data generating process. Figure 2.7 displays a snap shot of the standard deviation of the respective variables in June 2021.

Figure 2.7 gives an indication of how drastically each simulation differs for each variable. We would like to propose a model that can accommodate all the different scenarios presented in the various simulations. Thus our model must be able to adapt to the regions of high standard deviation. From Figure 2.7a we can see that each simulation varies significantly in the tropics, but mostly around Indonesia. As such we might be particularly interested in our model performance in this area for the precipitation variable. For the pressure variable there is an increase in standard deviation for the poles and particularly to the south west of Antarctica, see Figure2.7b. There is a similar result for the temperature variable in Figure 2.7c. Therefore we will be interested in model performance in this area for these variables. The wind speed variable has many more localised areas with high variance from simulations. For example, the localised high variance of the coast of India and eastern Africa. These may pose a significant challenge to accommodate in a model for this variable.

Similar to the spatial variability in simulations we can consider the variability over time. Figure 2.8 highlight this variation over the forty simulations from the CESM-LE data set.

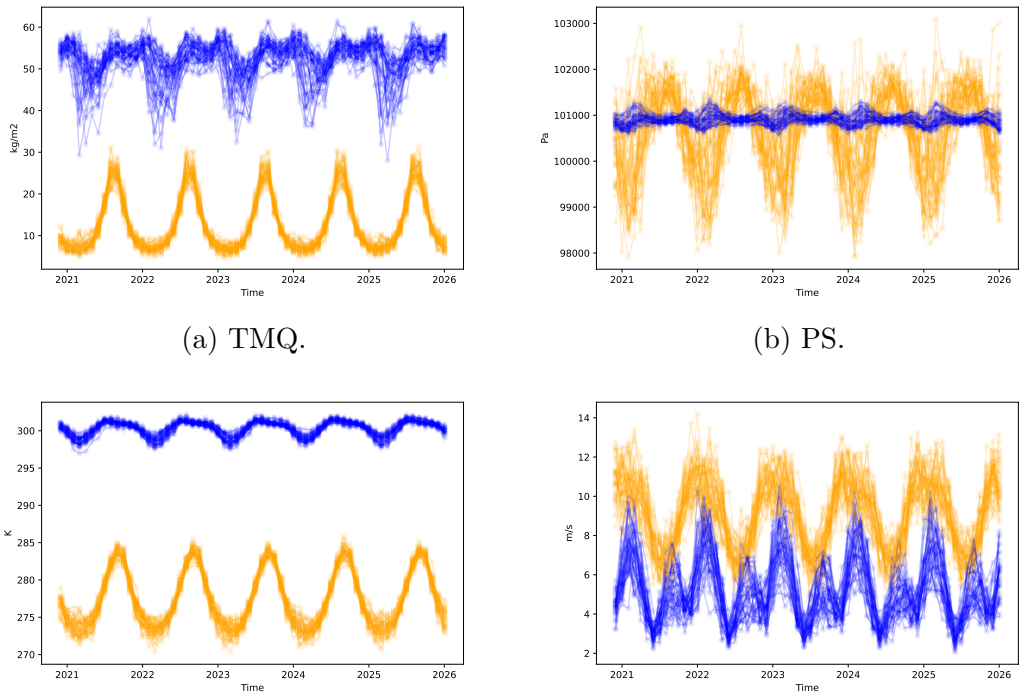
We can see clearly from Figure 2.8c that the temperature variable function over time shows little variation between simulations. However, variability does tend to increase in



(a) TMQ. (b) PS.

(c) TREFHT. (d) U10.

Fig. 2.7 Standard deviation of the four variables considered at June 2021 for the forty simulations present in the CESM-LE data set. The locations of the UK and Colombia are shown by the white square and circle markers respectively. These points are used as examples in Figure 2.8.



(a) TMQ. (b) PS.

(c) TREFHT. (d) U10.

Fig. 2.8 Four variables over observation period December 2020 to January 2026 for the forty simulations present in the CESM-LE data set at two locations; namely Colombia and UK. These are represented by the blue and orange colours respectively.

the troughs of these functions. Such a phenomena is more pronounced for the Colombia location. It may be interesting to assess performance of proposed models of the temperature variable with regard to this. The precipitation functions vary differently for the Colombia and UK locations. From the UK time series we see that although variability is high we can observe a periodic signal. It may be interesting to asses whether any model for this variable is able to pick up such a periodic signal, given that a single simulation may not show great periodicity. Pressure similarly exhibits very different function variability between locations, with the Colombia functions showing large changes in pressure compared to the changes observed between simulations at the UK location. Again this indicates that there is a clear spatial component to the process driving such simulations.

The above illustration of variability in simulations gives an indication about the difficulties that any model must overcome to describe such variables and gives testament to the use of such a data set to test our proposed model for EO data.

Chapter 3

Background Methodologies

In the following chapter we consider the various statistical methodologies upon which we build our CPACE model. This chapter is structured as follows. First we focus on common FDA techniques applicable to EO data. We follow this by discussing smoothing methodologies which are use in the FDA techniques. Finally we discuss Gaussian processes that are used in the CPACE framework to model correlation.

3.1 Functional principal components analysis

A commonly used technique in multivariate statistics is Principal Components Analysis (PCA), [53]. The technique finds dominant directions of variation and helps to achieve dimensionality reduction. This offers a parsimonious way to view data which is driven by the data themselves. The equivalent technique when the data are functional in nature is known as Functional Principal Components Analysis (FPCA), [41]. The basic concepts of which were studied in the mid twentieth century. The work of Karhunen and independently Loève paved the basic foundations of the technique in the FDA literature, [26, 32]. The FPCA technique essentially stems from representing the random function $\mathcal{X}(t)$ as an infinite linear combination of orthogonal functions. Such a representation is now known as the Karhunen-Loève theorem after its discoverers.

3.1.1 Formulation

The formulation of FPCA begins by assuming that $\mathcal{X}(t)$, $t \in \mathcal{T}$ is a square integrable stochastic process over some domain $\mathcal{T}$. Let the mean and the covariance of the stochastic process $\mathcal{X}$ be denoted by $\mu(t)$ and $G(s,t)$ respectively, where:

$$\mu(t) = \mathbb{E}(\mathcal{X}(t)) \tag{3.1}$$

$$G(s,t) = \mathrm{Cov}(\mathcal{X}(s), \mathcal{X}(t)) \tag{3.2}$$

Associated with the covariance surface $G(s,t)$ we have the linear operator T_G defined by:

$$T_G : L^2(\mathcal{T}) \to L^2(\mathcal{T}) \tag{3.3}$$

$$T_G : f \mapsto T_G f = \int_{\mathcal{T}} G(s,\cdot)\, f(s) ds \tag{3.4}$$

As T_G is a linear operator we can consider its eigenvalues and eigenfunctions which we will denote by λ_k and ϕ_k respectively (following convention set out in [58]) for $k = 1, 2, \cdots$. These are defined as the solutions to the Fredholm integral equations of the second kind, [58]:

$$\langle G(\cdot,t), \phi_k \rangle = \lambda_k \phi_k(t) \tag{3.5}$$

where $\langle f, g \rangle = \int_{\mathcal{T}} f(s)g(s)ds$ is the inner product in the space $L^2(\mathcal{T})$. Then by the Karhunen-Loève theorem one can express the centred process through the eigenvalues and eigenfunctions of the linear operator associated to the covariance surface, [26, 32]. That is:

$$\mathcal{X}(t) - \mu(t) = \sum_{k=1}^{\infty} \xi_k \phi_k(t) \tag{3.6}$$

where ξ_k is the k^{th} principal component associated to the eigenfunction ϕ_k. The Karhunen-Loève theorem assures us this L^2 convergence is uniform in t. The principal components are given by the following:

$$\xi_k = \langle \mathcal{X} - \mu, \phi_k \rangle \tag{3.7}$$

Further to this decomposition the Karhunen-Loève theorem means we have that the principal components are independent from each other, centred, and have variance equal to their associated eigenvalue, [26, 32]. That is:

$$\mathbb{E}(\xi_k) = 0 \tag{3.8}$$

$$\text{Var}(\xi_k) = \lambda_k \tag{3.9}$$

$$\mathbb{E}(\xi_k \xi_l) = 0, \text{ for } k \neq l \tag{3.10}$$

3.1.2 Interpretation

As with the multivariate principal components analysis the interpretation of the eigenvectors is often useful in exploratory analysis of data. The functional principal components analysis is of a similar form to the multivariate case and as such the same interpretation of the eigenfunctions is often employed. The first eigenfunction $\phi_1(t)$ encapsulates the dominant mode of variation in $\mathcal{X}(t)$ by construction since:

$$\phi_1 = \underset{\|\phi\|=1}{\arg\max}\, \text{Var}(\langle \mathcal{X} - \mu, \phi \rangle) \tag{3.11}$$

Similarly, the k^{th} eigenfunction is the dominant mode of variation which is orthogonal to the preceding $k-1$ components. Therefore exploring the first few eigenfunctions often gives a parsimonious way to view the variation in the data. Alike multivariate PCA, it is often that the structure of the eigenfunctions replicates some observed physical process. As such the FPCA decomposition is often used widely as a tool for data exploration.

In addition to this, we can use the fact that subsequent eigenfunctions capture less and less variation of the data as a form of dimensionality reduction, alike PCA, [53]. In this sense we can consider truncating the full representation given in Equation (3.6) to the K leading eigenfunctions which gives an approximation to the full process which we will denote by $\mathcal{X}^K(t)$ where:

$$\mathcal{X}^K(t) = \mu(t) + \sum_{k=1}^{K} \xi_k \phi_k(t) \tag{3.12}$$

The approximation of $\mathcal{X}$ by $\mathcal{X}^K$ converges as:

$$\mathbb{E}\left(\langle \mathcal{X} - \mathcal{X}^K, \mathcal{X} - \mathcal{X}^K \rangle\right) = \sum_{k>K}^{\infty} \lambda_k \to 0 \text{ as } K \to \infty \tag{3.13}$$

As such using the leading principal components for reconstruction has the effect of capturing the main modes of variation of the data and ignoring smaller modes of variation. Choosing the number of principal components is then up to the practitioner as in multivariate PCA, [53]. Ramsay and Silverman discuss in length the comparison of PCA to FPCA including commentary on the optimal choice of the number of principal components, [41, Chapter 8]. The practical implementation of FPCA then involves estimating various components. In particular estimation of; the mean function $\mu(t)$, the covariance surface $G(s,t)$, the K eigenfunctions and eigenvalues $\phi_k(t)$, λ_k respectively, and the principal components ξ_k for each realisation of the process $\mathcal{X}$ we observe.

3.2 Principal analysis through conditional expectation

Assuming for now we have a sufficient method for estimating the mean and covariance surfaces which we will denote by $\hat{\mu}(t)$ and $\hat{G}(s,t)$ respectively. We discuss in more detail the estimation of these components in Section 3.3. Prior to the introduction of the Principal Analysis through Conditional Expectation (PACE) methodology in [58] FPCA decomposition was restricted due to the need for approximating the integrals in Equations (3.7). As such it was often a requirement that the functional data were observed on a dense regular grid which meant that the principal components could be reliably estimated though some numerical integration scheme, [41, Chapter 8]. This very much restricted the application of the FPCA technique, however Yao et al. introduced the PACE method for over coming such an obstacle using conditional expectations for sparsely

observed functional data, [58]. At the same time the technique of [58] accommodates for observation error.

Traditionally Equation (3.7), used for estimating the principal component scores for the i^{th} realisation, is approximated through sums. Substituting y_{ij} for $\mathcal{X}(t_{ij})$, $\hat{\mu}(t_{ij})$ for $\mu(t_{ij})$, and $\hat{\phi}_k(t_{ij})$ for $\phi_k(t_{ij})$ we obtain the estimate $\xi_i^S = \sum_1^{J_i} \left(y_{ij} - \hat{\mu}(t_{ij})\right) \hat{\phi}_k(t_{ij}) \left(t_{ij} - t_{i(j-1)}\right)$, [58]. Where y_{ij} is as described in Equation (1.2) and setting $t_{i0} = 0$. However such an estimate breaks for the case that observations are sparse. Similarly such an approximation will be biased when the error processes from Equation (1.2), ε_{ij}, is non-zero. Yao et al. overcomes this by first assuming that the model is as follows, [58]:

$$y_{ij} = \chi_i + \varepsilon_{ij} \tag{3.14}$$

$$= \mu(t_{ij}) + \sum_{k=1}^{\infty} \xi_{ik}\phi_k(t_{ij}) + \varepsilon_{ij} \tag{3.15}$$

with ε_{ij} being jointly Gaussian with ξ_{ik}. We also require the noise process satisfies:

$$\mathbb{E}\left(\varepsilon_{ij}\right) = 0 \tag{3.16}$$

$$\text{Var}\left(\varepsilon_{ij}\right) = \sigma_\varepsilon^2 \tag{3.17}$$

The number of measurements of the i^{th} subject is considered random which reflect sparse functional data. Such a description follows naturally from our dataset description, given in Equation (1.2), by using the FPCA decomposition structure of $\mathcal{X}$ as discussed in Section 3.1. Following [58] we define the subsequent vector notations:

$$\boldsymbol{Y}_i = \left(y_{i1}, y_{i2}, \cdots, y_{iJ_i}\right)^\top \tag{3.18}$$

$$\boldsymbol{\phi}_{ik} = \left(\phi_k(t_{i1}), \phi_k(t_{i2}), \cdots, \phi_k(t_{iJ_i})\right)^\top \tag{3.19}$$

$$\boldsymbol{\mu}_i = \left(\mu(t_{i1}), \mu(t_{i2}), \cdots, \mu(t_{iJ_i})\right)^\top \tag{3.20}$$

$$\boldsymbol{t}_i = \left(t_{i1}, t_{i2}, \cdots, t_{iJ_i}\right)^\top \tag{3.21}$$

With such a model and assumptions, as stated in [58], the best prediction of the principal component scores for the i^{th} subject is given by:

$$\tilde{\xi}_{ik} = \mathbb{E}\left(\xi_{ik} | \boldsymbol{Y}_i, \boldsymbol{t}_i\right) = \lambda_k \boldsymbol{\phi}_{ik}^\top \boldsymbol{\Sigma}_{\boldsymbol{Y}_i}^{-1} \left(\boldsymbol{Y}_i - \boldsymbol{\mu}_i\right) \tag{3.22}$$

where $\boldsymbol{\Sigma}_{\boldsymbol{Y}_i} = \text{Cov}\left(\boldsymbol{Y}_i, \boldsymbol{Y}_i\right)$. The estimate for the principal component score can then be found by substituting in estimates for the various components in Equation (3.22). That is:

$$\hat{\xi}_{ik} = \hat{\mathbb{E}}\left(\xi_{ik} | \boldsymbol{Y}_i, \boldsymbol{t}_i\right) = \hat{\lambda}_k \hat{\boldsymbol{\phi}}_{ik}^\top \hat{\boldsymbol{\Sigma}}_{\boldsymbol{Y}_i}^{-1} \left(\boldsymbol{Y}_i - \hat{\boldsymbol{\mu}}_i\right) \tag{3.23}$$

The covariance matrix $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{Y}_i}$ is formed with $(l, m)^{\text{th}}$ element:

$$\left[\hat{\boldsymbol{\Sigma}}_{\boldsymbol{Y}_i}\right]_{lm} = \hat{G}(t_{il}, t_{im}) + \hat{\sigma}_\varepsilon^2 \delta_{lm} \tag{3.24}$$

where $\hat{\sigma}_\varepsilon^2$ is the estimated variance of the noise process. The estimation method for this is discussed in Section 3.3. Yao et al. also provide asymptotic properties of such an estimator along with asymptotic confidence bands where the mean and covariance surfaces are estimated with local linear smoothers, [13].

The conditional expectation technique describe above from [58] alleviates the issue of poor integral approximation from sparsely observed data when the estimated covariance surface is a relatively good fit to the true covariance surface. This is a somewhat better condition as it allows one to pool data from different observed subjects to estimate such a surface and thus the requirement of dense data per subject is relaxed to having dense data from the collection over all subjects. We discuss a particular method for estimating such surfaces in Section 3.3.

3.3 Penalised regression splines

Smoothing models underpin much of FDA. FDA uses the smoothness of observations over a continuous domain to help inform and model observed data, [41]. Typically, as described in Section 1.2, data is only observed discretely. Therefore with most FDA methodology there must be a conversion from discretely observed data and the continuous functional variable that generates it. This is particularly the case for our EO data since we have discrete observations specified by our data model given in Equation (1.1) which we assume is generated by observations of continuous functions given by our models in Equation (1.2). As such many models for obtaining such a smooth of the data have been studied, such as kernel smoothing, polynomial regression, and local linear smoothing, [41, Chapter 4]. In this section we consider the well studied technique of obtaining smooths of discrete data through penalised regression splines, [44]. We will use such a method to estimate the mean and covariance surfaces present in the PACE methodology as described in Section 3.2. We first describe the components that form the foundations of a regression spline, the spline basis.

3.3.1 Basis Splines

One of the components of a penalised regression splines is the basis functions used in the regression. As the name suggests regression splines uses spline functions as the regression basis. Spline functions of order d, which are well documented in the monograph of De Boor, is a piecewise polynomial function of degree $d-1$, [11]. In the case of a spline function of order d, $S : \mathcal{T} \to \mathbb{R}$, over a univariate domain $\mathcal{T} = [a, b] \subset \mathbb{R}$ we have:

$$S : t \mapsto S(t) = \begin{cases} P_0(t) \text{ if } \tau_0 < t \leq \tau_1, \\ P_1(t) \text{ if } \tau_1 < t \leq \tau_2, \\ \vdots \\ P_{m-1}(t) \text{ if } \tau_{m-1} < t \leq \tau_m, \end{cases} \tag{3.25}$$

where $P_i : [\tau_i, \tau_{i+1}] \to \mathbb{R}$ are polynomial functions of degree $d-1$. The vector of points $\boldsymbol{\tau} = (\tau_0, \tau_1, \cdots, \tau_m)$ is known as the knot vector for the spline and must satisfy $a = \tau_0 < \tau_1 < \cdots < \tau_m = b$. By specifying that the piecewise polynomials must share the same derivative order up to a degree we can ensure continuity of relative smoothness over the knot points and the whole spline function. We specify the continuity at each point in our knot vector by the continuity vector $\boldsymbol{r} = (r_0, \cdots, r_m)^\top$ where r_i specifies that P_i and P_{i+1} share common derivative values at point τ_i for derivatives up to order r_i. The spline type can be specified completely by specifying the knot locations and the continuity vector, [11]. In fact, one can extend our definition of the knot vector to incorporate both the knot and continuity vector into one. This is known as the extended knot vector, which will completely specify the spline type, [11]. We define the extended knot vector as the vector of knot points which repeats the i^{th} knot vector exactly $n - r_i$ times. That is:

$$(\tau_0, \cdots, \tau_0, \tau_1, \cdots, \tau_1, \cdots, \tau_{m-1}, \cdots, \tau_{m-1}, \tau_m \cdots, \tau_m)$$

We denote the spline functions of order d with extended knot vector by $S_{d,\boldsymbol{\tau}}$.

The Basis splines are more commonly referred to as B-splines, [29]. B-splines are basis functions for splines of the same order defined over the same knots. They are typically defined recursively, [29, 11]. The classic algorithm for the recursive construction is known as the Cox-de Boor recursion formula, [11], and is given as follows. Given a knot vector $(\tau_0, \cdots, \tau_0, \tau_1, \cdots, \tau_1, \cdots, \tau_{m-1}, \cdots, \tau_{m-1}, \tau_m, \cdots, \tau_m)^\top$ the B-spline of order 1 is given by:

$$B_{i,1}(t) = \begin{cases} 1, \text{ for } \tau_i \leq t < \tau_{i+1} \\ 0, \text{ otherwise.} \end{cases} \tag{3.26}$$

The higher order B-splines are defined by recursion as:

$$B_{i,q+1}(t) = w_{i,p}(t) B_{i,q}(t) + [1 - w_{i+1,q}(t)] B_{i+1,q}(t) \tag{3.27}$$

where $w_{i,q}$ is a weighting for the i^{th} B-spline of order d given by:

$$w_{i,q}(t) = \begin{cases} \frac{x - \tau_i}{\tau_{i+q} - \tau_i}, \text{ for } \tau_{i+q} \neq \tau_i \\ 0, \text{ otherwise.} \end{cases} \tag{3.28}$$

A B-spline basis system of size Q can then be considered by choosing the extended knot vector $\boldsymbol{\tau}$ and specifying the order, d, of the B-spline functions, and is given by the collection:

$$\{B^{\boldsymbol{\tau}}_{d,q}(t)\}_{q=1}^{Q} \tag{3.29}$$

where Q is the number of basis functions to use in the system, $\boldsymbol{\tau}$ is the extended knot vector, and $B^{\boldsymbol{\tau}}_{d,q}$ is the q^{th} B-spline of order d defined by Equation (3.27) for our knot vector $\boldsymbol{\tau}$.

3.3.2 Regression splines

As discussed in Section 3.2 the PACE methodology requires estimation of both the mean function, $\mu(t)$, and covariance surface $G(s,t)$. Estimating such functions is a problem due to their infinite dimensional nature. A well studied and effective method for representing such functions is the use of a basis function expansion, [41]. That is representing the target surface using a linear combination of known basis functions. In this work we will utilise the B-spline basis function as discussed in Section 3.3.1. The B-spline system is exceptionally popular due to its ease of computation and ability to reconstruct many surfaces, [11]. Such ease of computation makes it feasible to not only create large basis systems but also alleviates many fitting procedures as we can re-evaluate the basis system at various points with ease. Such properties are very useful when using such a basis for regression models. Other common basis systems include the Fourier, Monomial, and Polynomial basis systems. See [41] for details of these basis systems in the functional framework. In the following we present the approach for estimating an arbitrary realisation of our functional random variable $\chi_i(t)$ over domain $\mathcal{T}$ and discuss how we extend the same concept to a two dimensional surface over $\mathcal{T} \times \mathcal{T}$ in Section 3.3.2.

We assume that our function can be represented using an order d B-spline basis system with knot vector $\boldsymbol{\tau}$:

$$\chi_i(t) = \sum_{q=1}^{Q} c_q B_{d,q}^{\boldsymbol{\tau}}(t) \tag{3.30}$$

$$= \boldsymbol{c}^\intercal \boldsymbol{B}_d^{\boldsymbol{\tau}}(t) \tag{3.31}$$

$$\tag{3.32}$$

where $\boldsymbol{c} = (c_1, \cdots, c_K)^\intercal$, $\boldsymbol{B}_d^{\boldsymbol{\tau}}(t) = \intercal\Big(B_{d,1}^{\boldsymbol{\tau}}(t), B_{d,2}^{\boldsymbol{\tau}}(t), \cdots, B_{d,Q}^{\boldsymbol{\tau}}(t)\Big)$, and Q is the dimension of the expansion. If such basis functions have nice properties like being easy to compute then such a representation for χ_i given by Equation (3.31) can be extremely useful since most problems can be reduced to involving only the finite dimensional vector $\boldsymbol{c} \in \mathbb{R}^Q$.

Our representation of χ_i using a basis system then becomes the problem of choosing the coefficients $\boldsymbol{c}^\intercal$ using only our set of observations of $\boldsymbol{Y}_i$ which are observed with error. The most common method for fitting a basis system to discretely observed data is by choosing the coefficients of the expansion c_q given in Equation 3.31 by minimising the criterion, [6]:

$$\text{SSE}_{\boldsymbol{Y}_i}(\boldsymbol{c}) = \|\boldsymbol{Y}_i - \boldsymbol{B}\boldsymbol{c}^\intercal\|^2 \tag{3.33}$$

where $\boldsymbol{B} = (\boldsymbol{B}_d^{\boldsymbol{\tau}}(t_{i1}), \boldsymbol{B}_d^{\boldsymbol{\tau}}(t_{i2}), \cdots, \boldsymbol{B}_d^{\boldsymbol{\tau}}(t_{iJ_i}))^\intercal$ is the $J_i \times Q$ matrix of the basis system evaluated at observed time points corresponding to the J_i length observation vector $\boldsymbol{Y}_i$. Minimising such a criterion is given by, [6]:

$$\hat{\boldsymbol{c}} = \left(\boldsymbol{B}^\intercal \boldsymbol{B}\right)^{-1} \boldsymbol{B}^\intercal \boldsymbol{Y}_i \tag{3.34}$$

The simple least squares approximation is a well studied and standard approach. See [6] for a through introduction to the concept. Such a methodology is often suitable for situations where our error process $\varepsilon(t)$ is a white noise process. Such a process for the noise is often unrealistic; as such a simple adjustment to the least squares criterion in Equation 3.33 can be used to allow for correlation among the observation errors:

$$\text{SSE}_{\boldsymbol{Y}_i,\boldsymbol{W}}(\boldsymbol{c}) = \|\boldsymbol{W}^{\frac{1}{2}}\left(\boldsymbol{Y}_i - \boldsymbol{B}\boldsymbol{c}^\intercal\right)\|^2 \tag{3.35}$$

where $\boldsymbol{W}$ is a weighting matrix for the observations. Ideally the matrix will be the inverse of the variance-covariance matrix of the observations. Minimising the adjusted criterion is given by, [6]:

$$\hat{\boldsymbol{c}} = \left(\boldsymbol{B}^\intercal\boldsymbol{W}\boldsymbol{B}\right)^{-1}\boldsymbol{B}^\intercal\boldsymbol{W}\boldsymbol{y} \tag{3.36}$$

The estimate with least squares fitting can then be substituting $\hat{\boldsymbol{c}}$ for the $\boldsymbol{c}$ in Equation3.31, [6]. That is:

$$\hat{\chi}_i(t) = \hat{\boldsymbol{c}}^\intercal\boldsymbol{B}_d^\tau(t) \tag{3.37}$$

The selection of the knot vector is well studied and the classical choice is to choose a knot vector where knots are located at the sampling points, [11].

An issue with the classical least squares fitting using a basis system expansion is the choice of number of basis functions, [41]. We are constrained to choose Q to be less than or equal to the number of observations, J_i. This is because more than J_i basis functions would results in Equation 3.36 being ill defined since the matrix $\boldsymbol{B}$ would have linearly dependent columns. However, we still have the choice to choose Q between 1 and J_i. Exactly which value for Q to choose is unknown and results in bias - variance trade off in the estimator, [41]. A large number of basis functions reduces bias in the estimator $\hat{\chi}_i(t)$, but the variance of this estimator may be unacceptably high. Conversely, a lower number of basis functions will result in high bias of the estimator but low variance. The bias-variance trade off is well studied and there is a vast literature on the methodology of choosing the number of basis functions. However, there is no gold standard and often the choice is made in an ad hoc fashion, [41]. Such an issue motivates modifying the fitting criterion which determines $\hat{\boldsymbol{c}}$ in Equation (3.36).

Penalties

Ideally, we want to penalise estimators which have high variance, that occur naturally when we have a large number of basis functions, but keep bias low. The naive choice of just reducing the number of basis functions, known as regression splines, fails in this respect, [44]. One such approach to do this is to reduce the number of basis functions in conjunction with a penalty, known as penalised regression splines, [44]. Such an approach was first used in [39] who applied such a technique to ill posed inverse problems. [44] discusses various other spline smoothing techniques as well as the penalised regression splines.

Penalised regression spline models adjust the fitting criterion in Equation (3.36) to, [44]:

$$\text{PSSE}_{\boldsymbol{Y}_i,\boldsymbol{W},\lambda}(\boldsymbol{c}) = \|\boldsymbol{W}^{\frac{1}{2}}(\boldsymbol{Y}_i - \boldsymbol{B}\boldsymbol{c})\|^2 + \omega \boldsymbol{c}^\intercal \boldsymbol{P}\boldsymbol{c} \tag{3.38}$$

where $\boldsymbol{P}$ is formed with $(l,m)^{\text{th}}$ element $[\boldsymbol{P}]_{lm} = \langle L(\boldsymbol{B}_l), L(\boldsymbol{B}_m)\rangle$ and ω is a parameter which controls the regularisation trade off. L is some linear differential operator. Typically, one chooses L to be the required smoothness of the target function and examples include simple first or second derivatives, [44].

Analytically minimising the PSSE criterion in Equation (3.38) can be found via, [44]:

$$\hat{\boldsymbol{c}} = \left(\boldsymbol{B}^\intercal \boldsymbol{W}\boldsymbol{B} + \omega \boldsymbol{P}\right)^{-1} \boldsymbol{B}^\intercal \boldsymbol{W}\boldsymbol{Y}_i \tag{3.39}$$

Essentially such a penalisation term determines that there should be a trade off between the bias which corresponds to the first term in Equation (3.38) and the variance which is the second term. This trade off is controlled by the regularisation parameter ω. The advantage of this method is that we can now let Q, our number of basis functions, be large without worrying of over fitting as the penalty term in Equation (3.38) will penalise functions with high variability in terms of the differential operator L.

The choice of differential operator is a well studied problem also. A common choice is the first or second order differential, denoted by D^1 and D^2 respectively, as this specifies a reasonable level of smoothness in the target function, [44]. However, often more complex terms are used to facilitate known properties of the target functions, such a letting L be the harmonic acceleration operator which forces a periodic form of the target functions. More care must be taken when extending the linear differential operator to higher dimensions which is discussed in Section 3.3.2. Additionally, in the case of B-spline basis system these penalty matrices are typically evaluated using a form of numerical integration, [41].

Penalised regression splines moves our problem of selecting Q to choosing our regularisation parameter, ω. Such a parameter influences the strictness with which we expect our target function to be smooth as defined by the operator L. Choosing such a parameter is a problem that is present not only in spline smoothing but other penalised regression approaches, [33]. A popular method for choosing such a parameter is the Generalised Cross Validation (GCV). GCV, introduced by Wahba in [49], is a well studied method which has good asymptotic properties as the number of observations tends to infinity, [51, 50]. GCV chooses ω as the minimiser of the GCV criterion $V(\omega)$ which is given by, [51]:

$$V(\omega) = \frac{J_i^{-1}\|(\boldsymbol{I} - \boldsymbol{A})\boldsymbol{Y}_i\|^2}{\left[J_i^{-1}\text{tr}(\boldsymbol{I} - \boldsymbol{A})\right]^2} \tag{3.40}$$

where $\boldsymbol{A}$ is the influence matrix defined by:

$$\boldsymbol{A} = \boldsymbol{B}\left(\boldsymbol{B}^\intercal \boldsymbol{W}\boldsymbol{B} + \omega \boldsymbol{P}\right)^{-1} \boldsymbol{B}^\intercal \boldsymbol{W} \tag{3.41}$$

The GCV method can then be minimised for ω using a numerical minimisation routing. For large J_i it is known that the GCV criterion performs well in recovering a regularisation

parameter which minimises variance while maintaining low bias in the reconstruction of the target function, [50]. For the case of low J_i the GCV method may not be reliable as such method to extend the GCV criterion have been considered. The modified GCV criterion, which adds a further modifier to the denominator in Equation (3.40) by multiplying the trace of the influence matrix by a factor, [10]. The modified GCV approach effectively increases the cost associated with each effective parameter in the curve which reduces the chance of choosing an ω which under smooths the data, [10]. A similar but separate approach to adjusting the GCV is robust GCV, introduced by Lukas. Lukas uses a weighted sum of the GCV function with a term which penalises ω values that are close to zero, [33]. The performance of such methods are discussed in [34].

Choosing a basis system, a criterion to choose the regularisation parameter, and a differential operator then fully specifies the penalised regression spline approach. In the case of one dimensional functions the procedure applies as above. As such, we can estimate our mean function $\mu(t)$ through the use of a penalised regression splines where our observation points for the mean function are the pooled mean across subjects of the union of observed time points for all curves. However for multiple dimensions, particularly the case when we wish to smooth the covariance surface, we must make some adjustments to the approach described above.

Extension to higher dimensions

There are two issues when extending the penalised regression spline to higher dimensions; extending the basis system and extending the penalty specification. To alleviate the first we must specify a basis system which can cover multiple dimensions. In fact there are many such systems, [51]. One such popular approach when we have regular data for FDA is using a tensor product B-spline system, [57]. We describe the extension to two dimensional surface, but the same extension will work for higher dimensional surfaces. Consider a two dimensional surface, $\sigma(s,t)$, which we represent by the tensor product spline given by, [57]:

$$\sigma(s,t) = \sum_{1 \leq q_1, q_2 \leq \bar{Q}} c_{q_1,q_2} \boldsymbol{B}_{d_1}^{\boldsymbol{\tau}_1}(s) \boldsymbol{B}_{d_2}^{\boldsymbol{\tau}_2}(t) \tag{3.42}$$

where $\boldsymbol{B}_{d_i}^{\boldsymbol{\tau}_i}$ is the B-spline basis system for the i^{th} dimension for $i = 1, 2$. For notational simplicity we assume the dimension of each marginal basis system is the same, $\bar{Q}$. However, in general this need not be the case. $\boldsymbol{C} \in \mathbb{R}^{\bar{Q} \times \bar{Q}}$ is a coefficient matrix to be determined. Equation (3.42) can be written more succinctly using a Kronecker product as, [57]:

$$\sigma(s,t) = \bar{\boldsymbol{B}}^{\mathsf{T}}(s,t) \operatorname{Vec}(\boldsymbol{C}) \tag{3.43}$$

where $\bar{\boldsymbol{B}}(s,t) = \boldsymbol{B}_{d_2}^{\boldsymbol{\tau}_2}(t) \otimes \boldsymbol{B}_{d_1}^{\boldsymbol{\tau}_1}(s)$ and $\operatorname{Vec}(\cdot)$ is an operator which stacks the columns of a matrix into a vector. We use the $\bar{\cdot}$ notation to make explicit that this basis is over multiple dimensions.

The same methods now follow as in the non penalised univariate case with this Kronecker basis system, [57]. However, we must still adjust the penalty matrix in Equation (3.39) to account for smoothness across multiple dimensions.

Using the tensor product basis system as described above one might consider specifying that the smoothness of the surface in smooth in both dimensions. Indeed, one such approach to extending the penalty specification which was introduced by Wood is to consider setting penalties on the marginal basis separately and to combine them by a weighted sum, [55]. Such an approach known as tensor product penalties is well studied in the linear generalised additive model setting, [54]. A two dimensional penalty matrix $\bar{\boldsymbol{P}}$ may be described as follows:

$$\bar{\boldsymbol{P}} = \omega_1 \boldsymbol{P}_1 \otimes \boldsymbol{I}_2 + \omega_2 \boldsymbol{I}_1 \otimes \boldsymbol{P}_2 \tag{3.44}$$

where $\boldsymbol{P}_i$ is marginal penalty over a single basis dimension as described in Equation (3.38), I_i is the identity matrix of dimension of the i^{th} dimension basis, and ω_i is the marginal regularisation parameter for $i = 1, 2$. The properties of such a smoothness penalty are discussed in detail in [55] with the main points being such a penalty is both scale invariant and low rank. In addition, [56] studies the use of such a penalty for the case of unevenly distributed data. The additional complication is that we now have multiple smoothness parameters ω_i, one for each dimension of the surface to be smoothed. In this case the GCV methodology described above, can still be applied but now minimisation occurs with respect to the vector $\boldsymbol{\omega} = (\omega_1, \omega_2)^\intercal$. Implementation details of such can be found in [54].

We can now use the above approach to estimate our covariance surface, denoted by $\hat{G}(s,t)$, for use in PACE methodology, [58]. The discrete observations for the covariance surface to be smoothed are gathered by pooling individual observed covariances from across subjects, which is discussed in detail in both [58, 57]. Xiao provides asymptotic properties of such an approach to the covariance surface of independent functional data which are on par to the asymptotic results of other smoothers used in [58] for the PACE methodology, [57].

3.4 Functional time series

As discussed in Section 1.1 EO data is often both spatially and temporally correlated. This two types of correlation is often considered separately. An area in FDA which has considered a similar case where functional data is observed and observations are correlated is functional time series, [2]. Typically, functional observations are naturally indexed by some time of observation and correlation may occur between observations. Hence we may build up a time series of functional observations. Functional time series models are some of the first in the FDA literature to start to consider correlated functional observations. Although they limit themselves to temporal correlation many of the ideas can be considered for extensions to higher dimension correlation and so we discuss a few of the more popular methodologies in this section.

We focus on a technique introduced by Hyndman and Shang in [24] to forecast functional time series. Such a method is of interest as it expands methodology on how to use existing forecasting techniques in a functional setting. In particular [24] uses the FPCA decomposition described in Section 3.1 to decompose functional observations and then uses independent forecasting of each principal component scores using standard multivariate techniques.

Hyndman and Shahid Ullah in [23] suggests to assume the principal component scores, ξ_{ik}, follow independent univariate time series models. Then, conditioning on the observed data $\boldsymbol{Y}$ given in Equation (1.1) and the set of principal components $\boldsymbol{\phi}(t) = (\phi_1(t), \phi_2(t), \cdots, \phi_K(t))$ they obtain the h-step ahead forecast of $\chi_{i+h|i}(t)$ as, [23]:

$$\hat{\chi}_{i+h|i} = \mathbb{E}\left(\chi_{i+h}(t)|\boldsymbol{Y}, \boldsymbol{\phi}\right) = \hat{\mu}(t) + \sum_{k=1}^{K} \hat{\xi}_{i+h|i,k}\phi_k(t) \tag{3.45}$$

where $\hat{\xi}_{i+h|i,k}$ denotes the h-step ahead forecast of the k^{th} principal component score. The method for which $\hat{\xi}_{i+h|i,k}$ is obtained can be any univariate time series method. Such methods are extremely well studied and discussed in the monograph of Hyndman and Athanasopoulos in [21]. Hyndman and Booth highlight that the forecast, given by Equation (3.45), is relatively insensitive to the choice of number of components in the principal decomposition provided it is sufficiently large. The variance of such a method can also easily be obtained through the sum of the component variances, [22]. The component variance of the forecast principal component scores are generally readily available from many time series models, [21]. The above forecasting methodology initially described in [23] used normal FPCA procedure with outliers weighted to zero, however this was reconsidered in [24] to include a geometric weighting to the principal components to allow for changes in the function over time.

Such a methodology motivates the construction of the CPACE model described in Chapter 5 for correlated functional data by considering the case where the principal components scores obey univariate correlated models not just time series. To describe some of these such models we use the concept of a Gaussian Process. We give background to this in the following section.

3.5 Gaussian process regression

The above section of functional time series shows there is scope for placing a model on the principal component scores to allow for correlation among functional observations. The natural progression to such work is to consider what options are available when we have more complex correlation structure or higher dimensional domain. For example, in the case of EO data discussed in Section 1.1, we have functional observations over a spatial domain indexed by some coordinate, $\boldsymbol{s} \in \mathcal{S}$. For this the univariate time series methods discussed in [24] are not suitable and we look to Gaussian processes as one possible solution to model principal component scores which are indexed by space. A such we discuss the basic concept of a Gaussian process in the following.

To describe a Gaussian process we first discuss the concept of a stochastic process. A real valued stochastic process is a collection of real random variables defined on the same probability space $(\Omega, \mathcal{F}, \mathcal{P})$ where Ω is a sample space, $\mathcal{F}$ is a σ-algebra, and $\mathcal{P}$ is a probability measure; and the random variables, indexed by some set $\mathcal{S}$ are all real valued. More details of such constructions can be found in [5]. A stochastic process can then be written as the collection:

$$\{\xi(\boldsymbol{s}, w) | \boldsymbol{s} \in \mathcal{S}\}$$

where $w \in \Omega$. A sample function of the stochastic process is the mapping, for a point $w \in \Omega$:

$$\xi(\cdot, w) : \mathcal{S} \to \mathbb{R}$$

A Gaussian process is a stochastic process which is parametrised by a mean function $m : \mathcal{S} \to \mathbb{R}$ where $m(\boldsymbol{s}) = \mathbb{E}(\xi(\boldsymbol{x}))$ and its covariance function:

$$k : \mathcal{S}^2 \to \mathbb{R}$$
$$k : \boldsymbol{s} \mapsto \mathrm{Cov}(\xi(\boldsymbol{s}), \xi(\boldsymbol{s}))$$

where for any finite collection of points, $\boldsymbol{s}_1, \boldsymbol{s}_2, \cdots, \boldsymbol{s}_n \in \mathcal{S}$, the joint distribution of $\boldsymbol{\xi}_n = (\xi(\boldsymbol{s}_1), \xi(\boldsymbol{s}_2), \cdots, \xi(\boldsymbol{s}_n))^\intercal$ is a multivariate normal distribution with mean vector $\boldsymbol{m}_n = (m(\boldsymbol{s}_1), m(\boldsymbol{s}_2), \cdots, m(\boldsymbol{s}_n))^\intercal$ and covariance matrix $\boldsymbol{K}_n$ whose $(l, m)^{\text{th}}$ entry is given by $k(\boldsymbol{s}_l, \boldsymbol{s}_m)$, [47]. As such, Gaussian processes are a natural way of defining a prior distribution over spaces of functions, which are the parameter spaces for Bayesian non linear regression models. In this work, we will denote such a Gaussian process by $\mathcal{GP}$ and write:

$$\xi(\cdot) \sim \mathcal{GP}(m(\cdot), k(\cdot, \cdot)) \tag{3.46}$$

One aspect of Gaussian process regression models is that under Gaussian assumptions they have a nice closed form for prediction. Let $S = \{\boldsymbol{s}_1, \boldsymbol{s}_2, \cdots, \boldsymbol{s}_n\}$ denote the design matrix of the regression, and $\boldsymbol{\xi}$ denote the corresponding target vector. Then conditioning the joint Gaussian prior distribution on the observations gives the posterior at prediction points S_*, [52]:

$$\boldsymbol{\xi}_* | S_*, S, \boldsymbol{\xi} \sim \mathcal{N}\left(K(S_*, S)K(S, S)^{-1}\boldsymbol{\xi}, K(S_*, S_*) - K(S_*, S)K(S, S)^{-1}K(S, S_*)\right) \tag{3.47}$$

where $\boldsymbol{\xi}_*$ is our posterior process evaluated at the collection of prediction points, and $K(\cdot, \cdot)$ is the covariance matrix formed by evaluating the covariance function $K(\cdot, \cdot)$ at all pairs of inputs. This can be extended easily to noisy observations by adjusting the observed covariance $K(S, S)$ to include a diagonal component which is the effect of model observation error. See [52] for details.

One key aspect of the Gaussian process is the covariance function $k(\cdot, \cdot)$. The covariance function characterises various smoothness properties such as the sample path continuity and its differentiability, [52]. As such the choice of $k(\cdot, \cdot)$ heavily influences the prediction mean and covariance as described in Equation (3.47). There are various common forms

of the covariance function but all must have the intrinsic property of being non-negative definite. The covariance function is of such importance in Gaussian process modelling and spatial statistics that it has been widely studied. See [52, Chapter 4] for a detailed introduction to various covariance functions. We expand on Section 1.3 to briefly introduce our form of covariance function which we will consider throughout this work.

Of the many different covariance functions employed in Gaussian processes stationary covariance functions are most commonly employed due to their simplicity and ease of construction, [8]. One such commonly used covariance function is the Matèrn covariance function which is given by, [1]:

$$C_\nu(d) = \sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\sqrt{2\nu} d\right)^\nu K_\nu \left(\sqrt{2\nu} d\right) \tag{3.48}$$

where Γ is the gamma function, K_ν is the modified Bessel function of the second kind, ν is a shape parameter of the kernel, and d is the possibly anisotropic separation between two vectors $\boldsymbol{s}, \boldsymbol{s}'$, [1]. The covariance kernel $k(\boldsymbol{s}, \boldsymbol{s}')$ is then simply $C_\nu(d(\boldsymbol{s}, \boldsymbol{s}'))$.

The issue with stationary covariance forms is that they are often quite restrictive in the sense that the correlation structure cannot vary across the domain. For example, this might be a too restrictive assumption in the case of climate data where correlation structure might be quite different in different parts of the globe. One particular way to extend the stationary Matèrn kernel to be non-stationary is proposed in [40]. Paciorek and Schervish propose a method to knit together multiple stationary correlation functions such that the resultant function is non-stationary.

They provide a form of non-stationary covariance function $k^{NS}(\cdot, \cdot)$ from stationary covariance function $k^S(\cdot, \cdot)$ as follows, [40]:

$$k^{NS}(\boldsymbol{s}, \boldsymbol{s}') = |\Sigma_{\boldsymbol{s}}|^{\frac{1}{4}} |\Sigma_{\boldsymbol{s}'}|^{\frac{1}{4}} \left|\frac{\Sigma_{\boldsymbol{s}} + \Sigma_{\boldsymbol{s}'}}{2}\right|^{-\frac{1}{2}} k^S(Q(\boldsymbol{s}, \boldsymbol{s}')) \tag{3.49}$$

where $Q(\boldsymbol{s}, \boldsymbol{s}') = (\boldsymbol{s} - \boldsymbol{s}')^\intercal \left(\frac{\Sigma_{\boldsymbol{s}} + \Sigma_{\boldsymbol{s}'}}{2}\right)^{-1} (\boldsymbol{s} - \boldsymbol{s}')$ and $\Sigma_{\boldsymbol{s}} = \Sigma(\boldsymbol{s})$ is the covariance matrix of the Gaussian kernel centred at $\boldsymbol{s}$. How $\Sigma_{\boldsymbol{s}}$ varies across the domain specifies how non-stationary the full covariance kernel is.

With almost all covariance functions and especially non-stationary covariances there are typically hyper parameters which must be estimated from the data. For example in the Matèrn covariance we have the shape parameter ν and any length scale parameters defined in the distance function $d(\cdot, \cdot)$. These are typically estimated though maximum likelihood estimation, [52], however fully Bayesian estimation can also be achieved through some Markov Chain Monte Carlo (MCMC) scheme, [40].

Chapter 4

Dynamic functional time series modelling

Chapter 5

Correlated principal analysis through conditional expectation

Chapter 6

Application of CPACE model

Chapter 7

Implementation of CPACE model

Chapter 8

Conclusions and further work

References

[1] Abramowitz, M. and Stegun, I. A., editors (2013). *Handbook of mathematical functions: with formulas, graphs, and mathematical tables*. Number 1 in Dover books on mathematics. Dover Publ, New York, NY.

[2] Aguilera, A. M., Ocaña, F. A., and Valderrama, M. J. (1999). Forecasting time series by functional PCA. Discussion of several weighted approaches. *Computational Statistics*, 14(3):443–467.

[3] Aschbacher, J. and Milagro-Pérez, M. P. (2012). The European Earth monitoring (GMES) programme: Status and perspectives. *Remote Sensing of Environment*, 120:3–8.

[4] Aston, J. A. D., Pigoli, D., and Tavakoli, S. (2017). Tests for separability in nonparametric covariance operators of random surfaces. *The Annals of Statistics*, 45(4):1431–1461.

[5] Billingsley, P. (1995). *Probability and Measure*. Wiley. Google-Books-ID: z39jQgAACAAJ.

[6] Björck, A. (1996). *Numerical methods for least squares problems*. SIAM, Philadelphia.

[7] Cressie, N. and Huang, H.-c. (1999). Classes of Nonseparable, Spatio-temporal Stationary Covariance Functions. *Journal of the American Statistical Association*, 94:1330–1340.

[8] Cressie, N. A. C. (2010). *Statistics for spatial data*. Wiley, New York. OCLC: 1039155476.

[9] Cressie, N. A. C. and Wikle, C. K. (2011). *Statistics for spatio-temporal data*. Number 1 in Wiley series in probability and statistics. Wiley, Hoboken, N.J.

[10] Cummins, D. J., Filloon, T. G., and Nychka, D. (2001). Confidence Intervals for Nonparametric Curve Estimates: Toward More Uniform Pointwise Coverage. *Journal of the American Statistical Association*, 96(453):233–246. Publisher: [American Statistical Association, Taylor & Francis, Ltd.].

[11] De Boor, C. (2001). *A practical guide to splines: with 32 figures*. Number v. 27 in Applied mathematical sciences. Springer, New York, rev. ed edition.

[12] Deb, S. and Tsay, R. S. (2017). Spatio-temporal models with space-time interaction and their applications to air pollution data. *arXiv:1801.00211 [stat]*. arXiv: 1801.00211.

[13] Fan, J., Gijbels, I., Hu, T.-C., and Huang, L.-S. (1996). A STUDY OF VARIABLE BANDWIDTH SELECTION FOR LOCAL POLYNOMIAL REGRESSION. *Statistica Sinica*, 6(1):113–127. Publisher: Institute of Statistical Science, Academia Sinica.

[14] Ferraty, F. and Vieu, P. (2006). *Nonparametric Functional Data Analysis*. Springer Science+Business Media, Inc., New York. OCLC: 318296135.

[15] Fuentes, M. (2006). Testing for separability of spatial–temporal covariance functions. *Journal of Statistical Planning and Inference*, 136(2):447–466.

[16] George, B. and Aban, I. (2015). Selecting a Separable Parametric Spatiotemporal Covariance Structure for Longitudinal Imaging Data. *Statistics in medicine*, 34(1):145–161.

[17] Gneiting, T. (2002). Nonseparable, Stationary Covariance Functions for Space-Time Data. *Journal of the American Statistical Association*, 97(458):590–600.

[18] Higdon, D. (2002). Space and Space-Time Modeling using Process Convolutions. In Anderson, C. W., Barnett, V., Chatwin, P. C., and El-Shaarawi, A. H., editors, *Quantitative Methods for Current Environmental Issues*, pages 37–56, London. Springer.

[19] Hooker, G., Roberts, S., and Shang, Lin, H. (2015). Maximal autocorrelation factors for function-valued spatial/temporal data. In *Weber, T., McPhee, M.J. and Anderssen, R.S. (eds) MODSIM2015, 21st International Congress on Modelling and Simulation*. Modelling and Simulation Society of Australia and New Zealand.

[20] Hurrell, J. W., Holland, M. M., Gent, P. R., Ghan, S., Kay, J. E., Kushner, P. J., Lamarque, J.-F., Large, W. G., Lawrence, D., Lindsay, K., Lipscomb, W. H., Long, M. C., Mahowald, N., Marsh, D. R., Neale, R. B., Rasch, P., Vavrus, S., Vertenstein, M., Bader, D., Collins, W. D., Hack, J. J., Kiehl, J., and Marshall, S. (2013). The Community Earth System Model: A Framework for Collaborative Research. *Bulletin of the American Meteorological Society*, 94(9):1339–1360. Publisher: American Meteorological Society Section: Bulletin of the American Meteorological Society.

[21] Hyndman, R. J. and Athanasopoulos, G. (2018). *Forecasting: principles and practice*. OTexts. Google-Books-ID: _bBhDwAAQBAJ.

[22] Hyndman, R. J. and Booth, H. (2008). Stochastic population forecasts using functional data models for mortality, fertility and migration. *International Journal of Forecasting*, 24(3):323–342.

[23] Hyndman, R. J. and Shahid Ullah, M. (2007). Robust forecasting of mortality and fertility rates: A functional data approach. *Computational Statistics & Data Analysis*, 51(10):4942–4956.

[24] Hyndman, R. J. and Shang, H. L. (2009). Forecasting functional time series. *Journal of the Korean Statistical Society*, page 13.

[25] Iaco, S. D., Myers, D. E., and Posa, D. (2002). Nonseparable Space-Time Covariance Models: Some Parametric Families. *Mathematical Geology*, 34(1):23–42.

[26] Karhunen, K. (1946). Zur Spektraltheorie stochastischer Prozesse. *Ann. Acad. Sci. Finnicae, Ser. A*, 1:34.

[27] Kay, J. E., Deser, C., Phillips, A. S., Mai, A., Hannay, C., Strand, G., Arblaster, J. M., Bates, S. C., Danabasoglu, G., Edwards, J. C., Holland, M. M., Kushner, P. J., Lamarque, J.-F., Lawrence, D. M., Lindsay, K., Middleton, A., Munoz, E., Neale, R. B., Oleson, K. W., Polvani, L. M., and Vertenstein, M. (2015). The Community Earth System Model (CESM) large ensemble project: a community resource for studying climate change in the presence of internal climate variability. *Bulletin of the American Meteorological Society*, 96(8):1333–1349. Publisher: American Meteorological Society.

[28] Khabbazan, S., Vermunt, P., Steele-Dunne, S., Ratering Arntz, L., Marinetti, C., van der Valk, D., Iannini, L., Molijn, R., Westerdijk, K., and van der Sande, C. (2019). Crop Monitoring Using Sentinel-1 Data: A Case Study from The Netherlands. *Remote Sensing*, 11(16):1887. Number: 16 Publisher: Multidisciplinary Digital Publishing Institute.

[29] Knott, G. D. (2000). *Interpolating Cubic Splines*. Number 1 in Progress in Computer Science and Applied Logic. Birkhäuser Basel.

[30] Lindgren, F., Rue, H., and Lindström, J. (2011). An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(4):423–498. _eprint: https://rss.onlinelibrary.wiley.com/doi/pdf/10.1111/j.1467-9868.2011.00777.x.

[31] Liu, C., Ray, S., Hooker, G., and Friedl, M. (2012). Functional factor analysis for periodic remote sensing data. *The Annals of Applied Statistics*, 6(2):601–624.

[32] Loève, M. (1946). Fonctions aléatoires à décomposition orthogonale exponentielle. *La Revue Scientifique*, 84:159–162.

[33] Lukas, M. A. (2006). Robust generalized cross-validation for choosing the regularization parameter. *Inverse Problems*, 22(5):1883–1902. Publisher: IOP Publishing.

[34] Lukas, M. A., Hoog, F. R. D., and Anderssen, R. S. (2012). Performance of Robust GCV and Modified GCV for Spline Smoothing. *Scandinavian Journal of Statistics*, 39(1):97–115. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1467-9469.2011.00736.x.

[35] Militino, A. F., Ugarte, M. D., and Pérez-Goya, U. (2018). An Introduction to the Spatio-Temporal Analysis of Satellite Remote Sensing Data for Geostatisticians. In Daya Sagar, B., Cheng, Q., and Agterberg, F., editors, *Handbook of Mathematical Geosciences: Fifty Years of IAMG*, pages 239–253. Springer International Publishing, Cham.

[36] Mitchell, M. W., Genton, M. G., and Gumpertz, M. L. (2006). A likelihood ratio test for separability of covariances. *Journal of Multivariate Analysis*, 97(5):1025–1043.

[37] Muro, J., Canty, M., Conradsen, K., Hüttich, C., Nielsen, A. A., Skriver, H., Remy, F., Strauch, A., Thonfeld, F., and Menz, G. (2016). Short-Term Change Detection in Wetlands Using Sentinel-1 Time Series. *Remote Sensing*, 8(10):795. Number: 10 Publisher: Multidisciplinary Digital Publishing Institute.

[38] Oliver, C. (2004). *Understanding Synthetic Aperture Radar Images*. Scitech Publishing, Raleigh, NC.

[39] O'Sullivan, F. (1986). A Statistical Perspective on Ill-Posed Inverse Problems. *Statistical Science*, 1(4):502–518. Publisher: Institute of Mathematical Statistics.

[40] Paciorek, C. J. and Schervish, M. J. (2006). Spatial Modelling Using a New Class of Nonstationary Covariance Functions. *Environmetrics*, 17(5):483–506.

[41] Ramsay, J. O. and Silverman, B. W. (2010). *Functional data analysis*. Springer Science+Business Media, New York (N.Y.).

[42] Raspini, F., Bianchini, S., Ciampalini, A., Del Soldato, M., Solari, L., Novali, F., Del Conte, S., Rucci, A., Ferretti, A., and Casagli, N. (2018). Continuous, semi-automatic monitoring of ground deformation using Sentinel-1 satellites. *Scientific Reports*, 8(1):7253. Number: 1 Publisher: Nature Publishing Group.

[43] Rossi, R. E., Dungan, J. L., and Beck, L. R. (1994). Kriging in the shadows: Geostatistical interpolation for remote sensing. *Remote Sensing of Environment*, 49(1):32–40.

[44] Ruppert, D., Wand, M. P., and Carroll, R. J. (2003). *Semiparametric Regression*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge.

[45] Sampson, P. D. and Guttorp, P. (1992). Nonparametric Estimation of Nonstationary Spatial Covariance Structure. *Journal of the American Statistical Association*, 87(417):108–119. Publisher: [American Statistical Association, Taylor & Francis, Ltd.].

[46] Schmidt, A. M. and Guttorp, P. (2020). Flexible spatial covariance functions. *Spatial Statistics*, 37:100416.

[47] Shi, J. Q. and Choi, T. (2011). *Gaussian Process Regression Analysis for Functional Data*. CRC Press. Google-Books-ID: DkgdN6dRAicC.

[48] Stein, M. L. (1999). *Interpolation of Spatial Data Some Theory for Kriging*. OCLC: 1184292315.

[49] Wahba, G. (1977). Practical Approximate Solutions to Linear Operator Equations when the Data are Noisy. *SIAM Journal on Numerical Analysis*, 14(4):651–667. Publisher: Society for Industrial and Applied Mathematics.

[50] Wahba, G. (1985). A Comparison of GCV and GML for Choosing the Smoothing Parameter in the Generalized Spline Smoothing Problem. *The Annals of Statistics*, 13(4):1378–1402. Publisher: Institute of Mathematical Statistics.

[51] Wahba, G. (1990). *Spline Models for Observational Data*. CBMS-NSF Regional Conference Series in Applied Mathematics. Society for Industrial and Applied Mathematics.

[52] Williams, C. K. and Rasmussen, C. E. (2006). *Gaussian processes for machine learning*, volume 2. MIT press Cambridge, MA.

[53] Wold, S., Esbensen, K., and Geladi, P. (1987). Principal component analysis. *Chemometrics and intelligent laboratory systems*, 2(1-3):37–52. Publisher: Elsevier.

[54] Wood, S. N. (2006a). *Generalized Additive Models: An Introduction with R*. CRC Press LLC, Philadelphia, PA, UNITED STATES.

[55] Wood, S. N. (2006b). Low-rank scale-invariant tensor product smooths for generalized additive mixed models. *Biometrics*, 62(4):1025–1036.

[56] Wood, S. N. (2017). P-splines with derivative based penalties and tensor product smoothing of unevenly distributed data. *Statistics and Computing*, 27(4):985–989. arXiv: 1605.02446.

[57] Xiao, L. (2020). Asymptotic properties of penalized splines for functional data. *Bernoulli*, 26(4):2847–2875. Publisher: Bernoulli Society for Mathematical Statistics and Probability.

[58] Yao, F., Müller, H.-G., and Wang, J.-L. (2005). Functional Data Analysis for Sparse Longitudinal Data. *Journal of the American Statistical Association*, 100(470):577–590. Publisher: [American Statistical Association, Taylor & Francis, Ltd.].

[59] Zhang, C., Li, W., and Travis, D. J. (2009). Restoration of clouded pixels in multispectral remotely sensed imagery with cokriging. *International Journal of Remote Sensing*, 30(9):2173–2195. Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/01431160802549294.