Functional Data Analysis for Earth Observation

Julian Austin

Department of Mathematics, Statistics and Physics.
Newcastle University

This dissertation is submitted for the degree of
Doctor of Philosophy

May 2022

Functional Data Analysis for Earth Observation

Julian Austin

Earth observation data, that is data observed over the surface of the earth, is often characterised by its spatial and temporal dependency. Such datasets are being collected more frequently and over larger spatial domains as remote sensing and in-situ collection methodologies become more sophisticated. However, they often include large amounts of missing observations. There is a high demand for models which can help interpret and interpolate to aid in the use of these datasets for a vast array of disciplines. Often the most challenging aspect of such data is how to interpolate missing observations. In this work, we consider such datasets from a functional data perspective. In particular, we focus on methods which can help explain the datasets variation in a parsimonious way whilst maintaining predictive accuracy for missing observations.

We begin by discussing the current methods available to earth observation datasets from both a spatio-temporal and functional data perspective. Following this, we introduce an interim functional time series model, based on a functional data decomposition which considers the spatial dimension as our functional domain. We discuss the consequences of taking this approach from a practical perspective.

Finally, we develop a novel framework which treats the temporal dimension as the functional domain. We maintain parsimony by basing this model on the main modes of variation using a functional principal components analysis and incorporate spatial dependency between functional observation using a structured Gaussian process. We present the validity of this methodology under spatial correlation of the observed data and evidence the ability of this framework using various spatial dependency models on both a simulation and real world study. We show that such a model performs well on sparsely observed datasets and also highlight the approaches used to make the model applicable to large datasets.