UCL

UNIVERSITY COLLEGE LONDON

MPHIL THESIS

Statistical Downscaling of Air Quality Models Using Principal Fitted Components

Author:

Farha Ahmad Z. A. Alkuwari

A thesis submitted in fulfillment of the requirements for the degree of

Master of Philosophy in Statistics

March 9, 2018

I, Farha Ahmad Z. A. Alkuwari confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Signed: _____

Date: _____

Abstract

Statistical downscaling is a technique that is used to extract high-resolution information from regional scale variables produced by Chemical Transport Models (CTMs). The aim of this thesis is to shade light on the advantages of statistical downscaling in improving the forecasting ability of air quality models. Many statistical downscaling methods in geophysics often rely on dimension reduction techniques to reduce the spatial dimension of gridded model outputs without loss of essential spatial information. In this thesis we developed a new downscaling methodology that relies on using Principal Fitted Components (PFCs) to downscale an air quality model.

The main advantage of employing PFCs in downscaling relies in the fact that PFCs represent space-time variations associated with a particular location through the use of inverse regression. This means that PFCs will emphasize on location related regional information. We illustrate our proposed method by both simulation and application on ground level ozone over southeastern U.S region to downscale the Regional ChEmical TrAnsport Model (REAM). Both simulation and applications results indicate that PFC downscaling appears to yield more accurate forecasts.

Moreover, we accommodate the fact that covariance matrices that are used to compute PFCs might be unstable due to the fact that they have a relatively large dimension. This issue has motivated us to regularize the covariance matrices by thresholding prior to computing the PFCs and then proceed with the downscaling using thresholded PFCs. We illustrate the modified downscaling approach by simulation and application to ground level ozone. Simulation results suggest that employing thresholded PFCs in downscaling have improved the downscaling results, however, the application results do not agree with the simulation results.

Finally, we extend our PFC downscaling method to downscale an ensemble of air quality models. We propose a new two-stage dimension reduction approach to reduce the dimension of an ensemble. The proposed methodology reduces the spatial dimension in each ensemble member, and then the reduced variables are reduced further across the ensemble models. We illustrate our proposed methodology by simulation and application to downscale ground level ozone ensemble outputs in France. Both simulation and application results suggest that our proposed technique seem to show an adequate predictive performance.

"As we pass through this test with honor and dignity, I am addressing you to emphasize that Qatar needs every one of you to build its economy and protect its security. We require diligence, creativity, independent thinking, constructive initiatives and interest in ***academic achievement in all disciplines****, self-reliance and fighting indolence and dependency."*

-Sheikh Tamim Bin Hamad al-Thani

The Emir of the State of Qatar during his speach about the unjust blockade on Qatar, June 2017

Acknowledgments

First, I would like to thank my supervisor Prof. Serge Guillas for his support, I appreciate all his time and ideas. Also I would like to thank Prof. Valerie Isham my secondary supervisor for her constructive feedback throughout the first year of the research. I am really grateful to Dr. Yuhang Wang and Dr. Vivien Mallet for providing the data used in the research and for their useful ideas and comments.

I am deeply grateful to my boss Mr. Abdulaziz Alkhalifa for his constant support and for allowing me the time to complete my thesis. I am also thankful for the excellent example he has provided as a successful leader of our organization.

Also, I would like to thank my former colleagues at Qatar University Dr. Mahmoud Boutefnoucher, Dr. Adel Altayyeb, and Dr. Kassim Mitwondi who believed in me and kept me motivated during my studies.

I would like to express my sincere gratitude to my family for all their love and encouragement. Words cannot express how grateful I am to my amazing brothers and sisters, your prayers for me was what encouraged me thus far. I would like to express my appreciation to my friends, thank you for encouraging me during the most difficult times and for always being there for me.

Lastly, I would like to express my deepest gratitude and appreciation to our Emir. Sheikh Tamim Bin Hamad Al-Thani, for his insight that kept the moral of the people of Qatar high through such a critical time for our country. I was able to complete this thesis even though my country is facing an oppressive and an unjustified blockade by its neighbours. Honour and dignity demonstrated by the people of Qatar and its leadership supported me and filled me with determination to finish my thesis and contribute to Qatar's prosperity.

Contents

List of Tables

List of Figures

Chapter 1: Introduction

1.1 Background

The recent changes in climate conditions, emissions, and land use have had a significant impact on the environment. Unfortunately those impacts are not in the advantage of the quality of the air. For example, the increasing vehicle emissions along with the increasing temperature might result in elevated levels of ground level ozone, which is found to be harmful for human health and the environment. Many studies suggest that the recent change in climate conditions have affected the levels of air pollutants (e.g Weaver et al. [2009]). Six common air pollutants are well-documented - Ozone (O_3), Carbon Monoxide (CO), Sulfur Oxides (SO), Nitrogen Oxides (NOX), Lead (Pb) and Particulate Matter (PM) - typically PM2.5 and PM10. Out of the six, ground level ozone and particulate matter are most commonly spread in the air. As these pollutants are significantly harmful to human health, it is imperative that we take action to control their impact on our livelihood.

The World Health Organisation (WHO) regularly updates its air quality guidelines based on a systematic review of existing literature on the potential impact of air pollution on health. On the basis of those findings the institution makes recommendations and sets target guidelines for countries to follow [Krzyzanowski and Cohen, 2008]. Most countries in the world would typically adapt those targets to their own specific circumstances through legislation and enforcement agencies. For example, in the United States of America, the Environmental Protection Agency (EPA) sets air quality standards to apply the Clean Air Act [*The Plain English guide to the Clean Air Act, 2007*] - passed by the US congress in 1970 as a legal enforcement tool for monitoring and controlling pollution. The foregoing examples, and many others around the world, seek to help countries identify the levels at which pollutants concentrations are considered dangerous on human health.

The overall challenge is that not only different parts of our planet generate, and are affected by, pollution in different proportions but also that air quality modeling has continued to present management challenge - particularly in the Big Data era [Etzion and Aragon-Correa, 2016]. In recent years, there have been several initiatives by the international scientific community to obtain a universally acceptable spatial-temporal understanding of the underlying interaction rules among the main pollutants. Such understanding is expected to lead to successful intervention programmes as

well as lay out the path for new research directions.

1.2 Research Question and Objectives

This thesis seeks to contribute towards achieving that goal by setting a specific research question based on a number of objectives. Our research question is defined as: **How can the blending of dynamic and statistical models help in narrowing down global environmental phenomena to regional discordance?** The question sets out to provide fundamental insights as to whether the scientific community is capable of narrowing down the global environmental phenomena via regional discordance and scales - i.e., by blending dynamical and statistical models into robust non-linear methods for a better understanding of regional and global climate data. Its objectives are as follows

1. **To downscale a statistical air quality model using Principal Fitted Components (PFCs).**
    - By conducting data preparation and cleansing.
    - By reducing model grid dimensions using PFCs.
2. **To provide a comparative analysis of the PFC downscaling technique with other downscaling methods: Regression downscaling and Principal Components Regression Downscaling.**
    - By downscaling ground level ozone in Summer 2001 in the South-Eastern US using PFC regression.
    - By carrying out statistical downscaling using PFCs with a regularized covariance matrix.
    - By thresholding regularizing the covariance matrix of the data.
    - By downscaling ground level ozone in the South-Eastern US during summer using thresholded PFCs.
    - By cross-validating the performance of PFC model.
3. **To comparatively assess the predictive power of the PFC technique vis-a-vis alternative methods.**
    - By developing multiple versions of unsupervised model for identifying underlying structures in data.
    - By comparing multi-model performances on new data sampled from the global set.
    - By carrying out predictive modelling based on known classes in the sampled data.
    - By comparing the predictive power of selected models.

Based on the foregoing objectives, this thesis outlines the current need of improved atmospheric variables predictions and introduces a novel downscaling approach to downscale large scale model outputs using regression dimension reduction using Principal Fitted Components (PFCs). Moreover, it presents a modification to the PFC downscaling

technique to accommodate two scenarios - an unsupervised and supervised setting. Thus, the thesis emphasizes the benefits of statistical downscaling as a technique for improving air quality model forecasts and extracting reliable knowledge for intervention purposes. More precisely, it explores the best possible statistical downscaling techniques that can maintain as much regional information as possible from the downscaled climate data attributes. It also provides evaluation of the predictive performance of the downscaled models and, finally, the proposed downscaling method will be generalized to be applied to an ensemble of air quality models.

1.3 Rationale and Motivation

Gaining a thorough understanding of how our environmental surroundings impact our health and livelihood is fundamental to our sustainability, yet it is one of the most difficult challenges mankind faces, particularly as we become increasingly industrialised and mechanised. Recent enhancements in data harnessing, storage, transmission and manipulation mean that we are now better equipped than we have ever been to extract knowledge from high dimensional data on a wide range of phenomena. Hence, the motivation for this work arises from two perspectives - firstly, the need to understand and keep pace with the dynamics within our environmental surroundings - particularly the atmospheric factors that affect our sustainability. The second perspective derives naturally from the first - making sense of the environmental data deluge that surround us. Outdoor pollution has been associated with numerous diseases [Ostro, 2004] and in a study by Ezzati et al. [2004], the author's home country (Qatar) was placed in a highly vulnerable group - due to its geographical location and economic activities which underlines her interest in the area.

Knowledge extraction from data remains a huge area of ongoing research and as such, it naturally attracts research attention and motivation. Many numerical models have been developed to produce forecasts of atmospheric variables using some meteorological information - typically falling into two major categories - Chemical Transport Models (CTMs), also known as air quality models, and conventional statistical methods. As the name implies, the former uses meteorological data inputs to describe the spatial-temporal variability of the concentration of chemicals in the atmosphere [Rotman et al., 2004] which they achieve by solving the continuity equations for mass conservation of the chemicals in the atmosphere involving transport, chemical, emission and deposition processes in space and time [Jacob, 2004]. While the CTM rely on well-established physical principles, the confidence of their estimates is known to be unstable - often with variable-specific variations (e.g., conditional on temperature, precipitation etc.). Statistical methods have been adopted in modelling air quality but, like the CTMs, they are susceptible to variations due to underlying distributional assumptions and the random nature of data used for model training and testing.

1.4 Contribution to Knowledge

This thesis presents a novel approach to air quality downscaling using Principal Fitted Components (PFC). The proposed novel method is based on a two-step dimension reduction procedure which we call double dimension reduction (DDR) hence accounting for its acronym, PFC-DDR. Performance comparisons with PC-DDR and other conventional predictive methods confirm PFC-DDRs superiority. The thesis main idea was to come up with a model capable of capturing the chemical and physical processes while at the same time streamlining the CTM large resolutions to attain informative data at small-scale resolutions. The proposed methodology is described through a sequence of three interconnected components described in Chapter 5, and in Section 3.3.4 with part of its outlined implementation mechanics.

The thesis contribution is two-fold - the foregoing novel approach makes a positive contribution towards global sustainability and improving human livelihood and, secondly, it lays down practical foundations for knowledge extraction from data, typically via multiple sampling and testing. It is expected that the proposed method will form a basis for model enhancements in other applications and it is in this context that the thesis outlines potential future research directions. In particular, it exhibits great scope for extension into predictive modelling.

1.5 Thesis Structure

This thesis is organized as follows. Four sub-sections in Chapter 1 provide the background, research questions and objectives, study rationale and motivation as well as the synopsis of the thesis' contribution to knowledge. It is followed in Chapter 2 by the review of the literature relating to the problem under study in which a thorough coverage of the current state-of-affairs in environmental monitoring, tracking and prediction is provided. The chapter also provides justifications for the thesis work by highlighting existing gaps in our data-driven environmental understanding, data handling, sharing as well as the approaches, methods, techniques and frameworks. Chapter 3 focuses on the methodology - presenting four main subsections - data sources, modelling strategy, dimensional reduction and comparability and assessment of downscaling models. Generally, the four sub-sections outline the techniques and methods adopted in addressing the set questions via the objectives and data. Chapter 4 provides a comprehensive objective-based discussions of the implementation strategy and results. It has five subsections addressing EDA, simulation and downscaling; downscaling REAM model ground level ozone outputs; downscaling an air quality model with regularised covariance matrix; downscaling an ensemble air quality models and performance assessment using alternative predictive methods. Finally, concluding remarks are drawn in Chapter **??**, with two sub-sections the first focusing on the thesis summary of results, contributions and limitations while the second highlights potential future directions for research.

Chapter 2: Literature Review

2.1 Air Quality Modeling

One of the key constituents of atmospheric smog is the Ground Level Ozone O_3 - a colourless, reactive oxidant gas that is formed in the air by the photo-chemical reaction of sunlight and a variety of volatile organic compounds (VOCs) such as automobile exhaust, carbon monoxide (CO), and oxides of nitrogen (NO_x). In a twenty-day continuous hourly observation of more than 50 non-methane hydrocarbons in the Pearl River Delta (PRD) in Guangzhou, China, Wang et al. [2008] found a high correlation between NOx, CO, and VOCs which suggested that motor vehicle exhaust were the most dominant source. They further applied Principal component analysis (PCA) on the data from a local area and established that reactive alkenes were more depleted than alkanes and aromatics, on the basis of which they inferred that the air masses arriving at the local area under study were more aged in terms of photo-chemistry compared to the source area of Guangzhou. Approaches of this nature are fundamental in downscaling for purposes of getting deeper insights into local areas of study. What this study did not confirm, however, was the replicability of the models applied. It is difficult, for instance, to conclude that since the indicator of their focal area of study was higher than those of the source by about a factor of 3, then similar concentrations of ozone between source and target would yield the same factor. Further studies are therefore needed if we are to establish more robust findings.

The ozone gas exists within two regions of the atmosphere - the stratosphere (approximately 15 to 50 kilometres above the Earth's surface) and the troposphere (approximately 0 to 15 kilometers above the Earth's surface). While both types of ozone have similar chemical compositions O_3, they have extremely different effect on the ecosystem. The upper stratospheric ozone forms a layer that protects the earth from the harmful ultraviolet rays that are produced by the sun while, on the other hand, the tropospheric ozone is considered to be a harmful pollutant that affects the environment and human health. Studies have shown that the exposure to elevated levels of ground-level ozone has major effects on human health (e.g. respiratory system problems) and the environment (e.g. the reduction in vegetation and crops). More details about the diverse effects of tropospheric ozone can be found in *US EPA (2006)* . Ozone concentrations present a major issue of concern and they are a subject of regulation all around the globe. Article 11 (1) (a) of the European Parliament Council Regulation stipulates that for any member state, the calculated level of hydrochloroflu-

orocarbons production over the "...period from 1 January 2010 to 31 December 2010 and in each 12-month period thereafter until 31 December 2013 does not exceed 35% of the calculated level of its production of hydrochlorofluorocarbons in 1997..." [*Regulation (EC) No 1005/2009 of The European Parliament and of the Council of 16 September 2009 on substances that deplete the ozone layer*] . Within the Gulf Co-operation Countries (GCC) member states, of which the author's home country, Qatar, is a member, the Unified Guiding Regulation of 2007 [*Environmental Statistics Annual Report 2013, Ministry of Development Planning and Statistics (Qatar)*] requires a full phase-out of the consumption of substances depleting the ozone layer and replacing them by safe substances conforming with national interests as outlined in the provisions of the Montreal Protocol and its amendments. It places strong regulatory conditions on the import, export, storage and recycling of any substances and appliances that are potentially harmful to the ozone layer.

In the US, the National Ambient Air Quality Standards requires that ozone levels not to exceed 0.075 parts per million (ppm) as a measure over an 8-hour averaged period. These regulations and many others elsewhere seek to achieve common goals centred around capturing the information in the environmental data attributes and using those outcomes for decision making purposes, designing and implementing interventions with the ultimate goal of attaining our sustainability. In other words, having accurate future ozone level forecasts will help in taking preventative actions when the forecasts suggest that ozone would reach an unhealthy level. Consequently, with ozone considered to be such a primary source of pollution, it is imperative that sound strategies are put in place to ensure that its concentration levels are appropriately monitored, tracked as well as being used as inputs in predictive modelling of related phenomena.

Air quality modelling has continued to gain great importance recently, as it becomes increasingly applicable as a tool for managing air pollution. As noted above, modelling atmospheric phenomena helps identify sources of air quality issues, say, and so it can further be helpful in setting novel strategies for monitoring and managing pollutants concentrations. Technically, such goals may be achieved via simulations under new, controlled conditions based on past and present conditions. On the basis of such simulations, future atmospheric concentrations and variations can be predicted. Recent developments in computing power - storage, performance and communication have resulted in the development of state-of-the-art air quality models with enormous capacity to simulate atmospheric variables and make projections. In the following exposition, we explore some of the general aspects of downscaling with respect to air quality modelling and we highlight potential novel paths to air quality modelling enhancement.

2.2 Downscaling in Air Quality Modelling

The strength of air quality models to produce simulations that well represent air pollutant emissions, is constrained by the fact that models such as Community Multi-scale Air Quality (CMAQ) and the Regional Chemical Transport

Model (REAM) or all those falling within the general domain of Chemical Transport Models (CTMs) have limitations. The CTMs are mathematical models that use numerical techniques to simulate physical and chemical processes of atmospheric variables (e.g. air pollutants). They use meteorological data as inputs to produce emission levels of the atmospheric variables. In addition, these models provide spatial-temporal distribution information of these elements. CTMs have been used for the purpose of measuring air quality (air quality models) and are built so they are able to simulate both primary pollutants (that are emitted into the air) and secondary pollutants (that are formed from the chemical reactions of other elements in the atmosphere). These models, as well as the conventional statistical methods, are susceptible to variations due to underlying distributional assumptions and the random nature of data used for model training and testing. Quite related to these, are challenges of scale which, like those arising from variability, they are directly associated with data sources.

In climate studies, Global Climate Models (GCMs) are a crucial tool to make inference on climate variables such as temperature, precipitation, ozone concentrations, etc. Although state of the art GCMs have been developed recently, they fail to give inference on a local level when needed, mainly because their projections are run at coarse spatial resolutions of about a few dozen miles which limits their capability to map on vital sub-grid scale features suchlike clouds and topography and therefore they cannot directly be used for local impact studies without being scaled down. The science of downscaling is considered to be relatively recent (the first reference to downscaling in literature was produced in the early 60's of the twentieth century by Klien [1963]). The main idea of downscaling is to apply a larger, regional, numerical model in higher spatial resolution in simulating local conditions as detailed in Wang et al. [2008]. In other words, downscaling is a technique that is used to extract high-resolution information from large/regional scale variables. The development of downscaling coincides with the occurrence of global climate models (GCMs), which are rather recent themselves [Benestad et al., 2008]. Benestad et al. [2008] defined downscaling as "*The process of making a link between the state of some variables representing a large space (referred to as the "large scale") and the state of some variable repressing a much smaller space (referred to as the "small scale")*". An example of a large scale variable can be the circulation pattern over a region while the small scale variable can be local precipitation at a given point. In downscaling, the link between the small scale variable and the large scale variable must be real and based on physical relationship that relate both variables to each other and it should not be because of coincidence or statistical convenience. The main objective of downscaling is to correct the spatial mismatch between the variables with different scales while synchronizing the time structure between them [Benestad et al., 2008]. Theoretical foundations and practical applications of the two main forms of downscaling techniques - dynamical and statistical are well-documented - see, for instance, Wang et al. [2008], Hay and Clark [2003] and Vrac et al. [2012]. The next sub-sections describe the two techniques and their suitability within the scope of our research question, objectives and potential extensions.

2.2.1 Dynamical Downscaling

To obtain high-resolution climate patterns from relatively coarse-resolution global climate models (GCMs) with, typically, a resolution of 150-300 km by 150-300 km, for making inferences at scales of 50 km or less we need to estimate the smaller-scale information. As graphically illustrated in Figure 2.1, *Dynamical Downscaling* reduces a global pattern to a local pattern and, as shown here, it involves nesting of a high resolution Regional Climate Model (RCM) within the region of an existing GCM which, typically, requires air quality simulations - with one major drawback being spatial resolution. RCMs extract regional information at a fine scale within the coarse scale information of a GCM [Jenkins and Barron, 1997] and, like GCMs, they are dynamical - the difference being that RCMs are defined on a specific small region on the atmosphere and contain physical information that is available within the GCM itself as well as in some regional specific (local) data. Thus, RCMs benefit from local and physical information to reproduce high resolution climate variables that is well represented physically (as it is consistent with the GCM). Despite the popularity of dynamical models, they have been associated with some major disadvantages as summarised below.

- Dynamical models are computationally demanding and expensive.
- It is difficult to relate the boundary conditions of the coarse GCM resolution to that of the local finer RCM.
- Since RCMs are derived from GCMs, they tend to inherit the systematic biases and errors that exist within the GCM.

Statistical Downscaling provides an alternative that is not only computationally inexpensive, but also provides a much finer resolution information at an even smaller scale than the RCM [Wang et al., 2010].

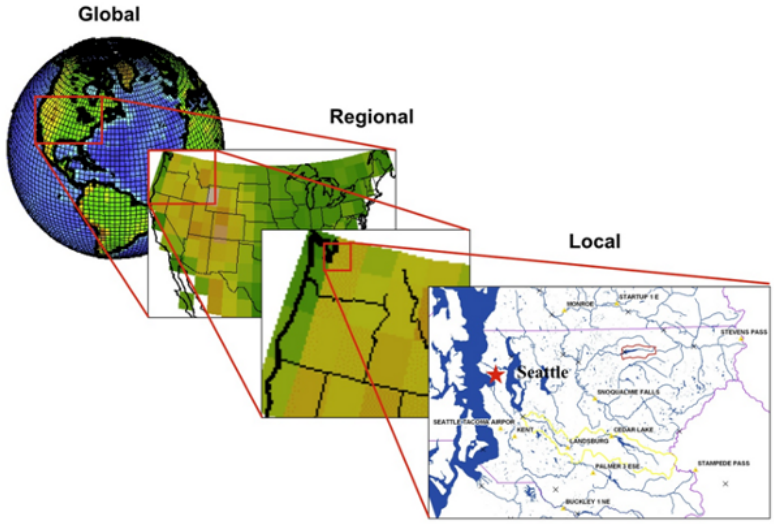

Figure 2.1: Downscaling regional large scale information to local scale information (http://www.earthsystemcog.org)

2.2.2 Statistical Downscaling

Statistical downscaling methods are based on statistical relationships between observed small-scale variables (as output) and large-scale variables (as inputs). They are becoming increasingly popular among the climate modelling community because they have shown satisfactory performance in obtaining climate information at a station level [Wang et al., 2010]. Their applications can take different forms - but generally, they assume the multivariate relationship that links the global to the local variables via a quantifiable function of the form

Local Scale Climate Observation (Response) = F(Large Scale Climate Outputs)

The function feeds the large-scale variables into the model to estimate the corresponding small-scale output, forming relationships that are used to derive local data attributes from the global GCM data model. A graphical illustration of this relationship is shown in Figure 2.2 and, as noted above, it is computationally inexpensive and so it is quite appropriate when computational resources for locations are limited [Wilby et al., 2004]. Its process consists of two main steps: Specification of the statistical relationships between the global and local climate variables and the generation of local station-level outputs using previously specified functions. Inevitably, the selected linkage function should be as accurate as possible in representing the relationship between the variables of interest and the set of chosen climate predictors. Furthermore, the selected predictors must be carefully chosen such that they best describe the climate variables of interest. Both points are fundamentally important because they directly impinge on the downscaling model performance. In other words, statistical downscaling faces all classical challenge of addressing the issue of variability.

Classification of Statistical Downscaling

Classification of statistical downscaling is described in Wilby et al. [2004]; Giorgi et al. [2001] and Chen et al. [2010] with the main categories being **Weather Data Generators, Weather Classifications** and **Regression Methods**. Apparently, depending on the purpose, nature, or scope, other classifications can be made. Based on our research question and objectives, we can consider classifying statistical downscaling in a much broader, yet more practical, sense - i.e., unsupervised and supervised models which, essentially, relates to each of the three categories.

1. **Weather Generators:** According to Wilks and Wilby [1999], these are statistical models of observed weather sequences which are designed to reproduce attributes of local climate parameters such as mean and variance. Based on those parameters the models can generate long-term weather time series of meteorological variables for modelling, say, daily precipitation and other climate variables such as temperature, relative humidity and solar radiation. The most commonly used weather generator methods are Markov Chains [Wilks, 1999], spell length methods [Racsko et al., 1991] as well as all forms of stochastic processes. Weather generation approaches

become amenable to statistical downscaling when the model parameters are conditioned on large-scale atmospheric variables. In this case the resulting weather generator models can generate future climate information using larger scale information obtained through simulation. They are therefore relevant to our work in that they provide scope producing long time series of future climate change information which may form a good source for both training and testing large ensembles and in doing so help mitigate the effect of data randomness.

2. **Weather Classifications:** This category of weather classification methods places records into states (classes) based on their measure or measures of homogeneity. Usually, this is achieved using a classification technique such as cluster analysis - for unlabelled data and classification or regression for labelled data. Consequently, the local climate variable is related to the weather structured or classes and replicated under modified climate scenarios using resampling or regression functions [Wilby et al., 2004]. In other words, the relationships between the large-scale data attributes such as atmospheric surface pressure, precipitation, wind speed etc are established with or without a class label and the task is to either identify inherent patterns in the data or allocate new cases to known classes. The two scenarios set the scene for unsupervised and supervised modelling. A disadvantage of this downscaling method is not only that the characteristics of the structures are affected by spatial-temporal variations, but also that model parameters are likely to influence the results. There are many examples of methods applied in this group - see, for instance, hybrid approaches in Frey-Buness et al. [1995] and comparative methods as used in Chen et al. [2010]. As the ultimate purpose of downscaling is to scale down the global environmental impact on a local area, it means that learning rules from data must yield robust outcomes that are capable of being replicated on other local regions. To achieve this goal, randomness in training, validation and testing data must be minimised - that is the fundamental approach adopted by this work.

3. **Regression Methods:** It is the underlying distributional assumptions that are mainly at fault here - that is, assuming that specific relationships between the large scale predictors (climate model outputs) and the small scale variables (response) hold and that they will remain so in time and space [Wilby et al., 1998]. The advantage that regression methods provide lies in their mathematical foundations and simplicity which make them readily applicable assuming that the underlying assumptions hold. Apparently, when these assumptions are violated, the model performs poorly. Our work adapts regression principles with this level of understanding.

Guidelines for Statistical Downscaling

Giorgi et al. [2001] gave extensive review on the major issues of statistical downscaling. A guideline summary of the main issues that need to be accounted for in statistical downscaling are as follows:

- **Choice of Statistical Method:**

  The chosen method mainly depends on the nature of the data. For example, if the data are normally distributed then multiple regression can be employed. On the other hand, if the data are more noisy or discrete then more complex techniques should be used.

- **Choice of Predictors:**

  Similar to any statistical approach, the selected predictors should be physically related to the response. The relationship between the variables are determined by the scientific knowledge and the expertise of the modellers. The predictors must be chosen so that they best represent the climate variable of interest. Hence, the omission of crucial predictors (either because the data is not available or because the variable is correlated with other predictors) might lead to loss of important information or causes a fall in the model performance.

- **Ability to Handle Extremes:**

  The majority of downscaling methods are not capable of handling extremes as well as it does when handling other measures such as means. However, some studies are investigating possible solutions to this issue [The Statistical and Regional dynamical Downscaling of Extremes for European regions project ”STARDEX”, see: *http://www.cru.uea.ac.uk/projects/stardex/*].

- **The Nature of the Study Region:**

  Wilby et al. [2004] have argued that tropical regions, such as India, might be more complicated to deal with when downscaling. This is because these regions are affected by the ocean, hence, ocean related variables must be considered. This indicates that a relatively large set of predictors is needed to be taken in to consideration.

- **Downscaling Model Evaluation:**

  Statistical downscaling is a data-oriented technique. The model is established using a sufficient size of training (historical) data. To be able to reach reliable results, the model adequacy needs to be validated by using an independent dataset to evaluate the model skill.

Further details on downscaling guidelines and issues are provided by Giorgi et al. [2001] and Wilby et al. [2004].

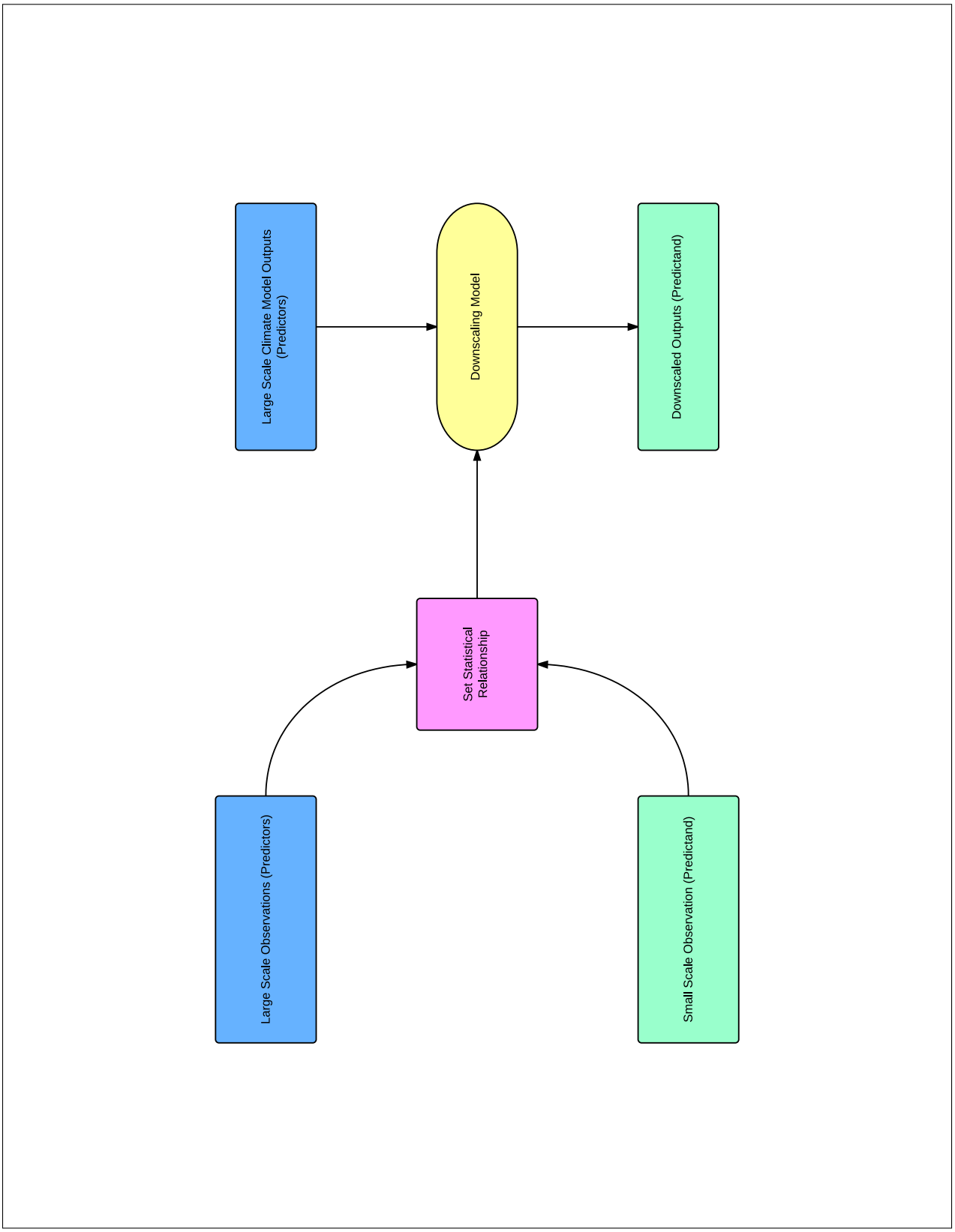

Figure 2.2: An illustrative plot of Statistical Downscaling

2.3 Comparability of Air Quality Downscaling Methods

As noted above, downscaling extracts high-resolution information from large/regional scale variables produced by numerical models. Many downscaling methods are available but pertain mostly to the climate modelling community. Statistical downscaling is based on developing a statistical relationship between observed small-scale variables (predictands) and large-scale variables (predictors) from a numerical model. The main advantage of statistical downscaling methods is that they are computationally inexpensive and appropriate when computational resources are limited [Wilby et al., 2004].

Statistical downscaling techniques have been used in the literature to model and forecast ground-level ozone (Sahu et al. [2007]) and despite having found to be satisfactory in predicting ozone, they do not capture the chemical and physical processes as well as CTMs. Using Bayesian approach to downscale air quality has gained momentum recently (Berrocal et al. [2009], Berrocal et al. [2010], and Berrocal et al. [2011]). Usually, this approach is prioritised due to its mechanics that effectively allows it to convert currently available information *(a priori)* into predictions *(posteriori)* and recursively using the latter as newly gained knowledge to generate novel predictions.

2.3.1 Regression-based Techniques

Regression-based downscaling is a widely applied downscaling method in practice. It formalizes mathematically the relationship between large-scale predictors and the small-scale predictand. Generally, CTMs produce simulations of climatological variables at large resolution, but as computing power increases CTMs' capability to produce forecasts on a relatively small resolution increases (Eder et al. [2009], Lee et al. [2009] and Ngan et al. [2012]). Their forecasts are averaged over a grid cell and although some provide forecasts at a fine grid cell resolution, in practice, point-specific information is required because of potential spatial variations.

One approach that can improve CTM forecasts is to combine the actual observations (usually produced by monitoring stations) with the CTM outputs. However, the model outputs and the actual ozone observations have different scales and so to overcome the issue of differences in scale, downscaling the model output to a point resolution can be conducted. This approach was adopted by Guillas et al. [2008], where the authors applied a two-step regression technique to downscale an air quality model in order to improve local forecasts at monitoring stations from the Atlanta area. This improved ozone forecasts by up to 25% compared to the direct use of the numerical model.

The nature of atmospheric variables (such as ozone) is highly affected by regional conditions. CTMs produce gridded outputs, and downscaling by relating the local variable with its corresponding gridded outputs might yield an inaccurate forecast. To overcome this, one might consider using multiple regression and take all available gridded outputs

into account when determining the downscaling model. Many downscaling studies often involves the use of multiple regression to downscale air quality models [Murphy, 1999], but usually when the number of predictors is relatively large, this may produce unstable results when the predictors are correlated (multicollinearity) and might raise other model fitting issues such over-fitting. To overcome these problems, dimensional reduction techniques such as Principal Component Analysis (PCA) can be used to reduce the dimensionality of predictors (while maintaining essential information from the original date) and to eliminate multicollinearity.

2.3.2 Dimensional Reduction Techniques

A number of studies in the climate literature make use of PCA downscaling, see, for instance, Kim et al. [1984], Kidson and Thompson [1998], Hessami et al. [2008]. Other approaches include combining predictor variables as in Benestad [2002*b*] who downscaled temperatures over Northern Europe by jointly reducing dimensions of several spatial predictors. Even in the foregoing cases, the extracted components or number of combined variables are strictly a function of the data used in the modelling process. Further, the optimum number of components and their interpretations are both subject to interpretation, which inevitably entails variability. Although the application of PCA downscaling have shown satisfactory results in literature, there are concerns regarding the use of PC as predictors in a regression model. First, PCs are obtained from the predictors only without reference to the response variable. This makes PCs informative when the objective of the study is to identify why and how the predictors are related amongst each other (interdependence relationship). Second, PCs are not invariant (remains unchanged under some transformation) or equivariant (changes in a convenient way under some transformations) under full rank linear transformations of the predictors. This raises a problem when the predictors are in different scales. Hence, to overcome this issue the regression dimension technique should take into account these concerns. One approach will be to use Principal Fitted Components (PFCs). PFCs as a dimension reduction approach was first presented by Cook [2007]. PFCs have two major advantages over PCs when used as a dimension reduction in regression. They can be tailored to the value of the response if the response value is known and they are equivariant under full rank transformation of the predictors.

To our knowledge, there is no application of Principal Fitted Components (PFC) regression [Cook, 2007] as a downscaling method in the literature. This work presents the first of such application, where we downscale an air quality model using a novel statistical downscaling approach that relies on enhanced mechanics of the PFC. Like its predecessor, the PCA, the PFC reduces the dimension of predictors in a regression model, but the reduction is done with reference to the values of the predictand.

We downscale an air quality model using PFC regression and compare the predictive ability of this technique to other downscaling techniques that are widely used in literature: multiple linear regression and PC regression. In the geophysical sciences, PCs are also called Empirical Orthogonal Functions EOFs [Lorenz, 1956]. Indeed, EOFs are spatial

PCs corresponding to the space-time variations of a quantity over a specific region. We introduce here a new kind of EOFs: Fitted Empirical Orthogonal Functions (F-EOFs). F-EOFs are spatial Principal Fitted Components (PFCs) that represent space-time variations over a region but are associated with a particular location through the use of inverse regression. The general method was developed by Cook [2007]; it consists of obtaining principal components with reference to the response variable. PFCs are computed by performing PCA using the covariance matrix of fitted values that results from the inverse regression of the predictors on a vector valued function of the response. The simulation studies in Cook [2007] indicate that PFCs outperformed PCs as a regression dimension reduction technique and that PFC regression models exhibit better predictive ability than the Ordinary Least Squares (OLS) and PC regression models. In a follow up study Cook and Forzani [2008] presented a comprehensive theory of PFCs and further explained the advantages PFCs has (as a regression dimension reduction method) over PCs. The authors also identified the relationship between the PFC regression model and other methods (i.e. sliced inverse regression, partial and ordinary least squares, and seeded reductions). Johnson [2008] analyzed the properties of PFCs and derived some theoretical properties. He studied Cook's simulation results and argued that PFCs outperformed PCs under Cook's model assumptions. Finally, Cook and Li [2009] extended the PFC methodology to regressions with categorical predictors or a mixture of categorical and continuous predictors.

The covariance matrix is an essential estimate in many multivariate dimension reduction statistical methods and, in most applications, the target is not to estimate the covariance matrix itself, rather than it is required to build up other estimators and hence carry out the analysis. The matrix is also the key tool used in dimension reduction using principal components and, additionally, it is used to explore the independence and conditional dependence between the variables in exploratory data analysis (EDA). In many applications data may not have a natural order. In this case, the covariance matrix estimation techniques must be invariant to variable permutations.

In many applications involving high dimensional data, the resulting covariance matrix is also high-dimensional and so reliability becomes a function of the sample size - that is, the sample covariance matrix is not a good estimator of the population covariance matrix if the ratio $\frac{p}{n}$ is not small, where p denotes the number of variables and n is the size of the sample. This has been explained extensively in the literature of matrix theory, see Johnstone [2001], El Karoui [2007], Furrer and Bengtsson [2007], and Paul [2007]. It was already pointed out by Dempster [1969] that estimators that were obtained as a scalar multiples of the sample covariance matrix have a tendency to distort the eigenstructure of the population covariance matrix if the ratio $\frac{p}{n}$ is large. Thus using it as a dimensional-reduction tool in techniques like PCA or PFC requires that the estimated sample covariance matrix in use to be reliable and a good representative of the population covariance. Thus, with high-dimensional data we should always look to improve the covariance matrix estimates to become more reliable estimators of the population covariance. Hence, an alternative and more reliable estimator of the sample covariance matrix is needed when we are dealing with high dimensional data.

Many alternative estimators of the covariance matrix have been proposed in the literature for high dimensional prob-

lems. One class of estimators depends on the fact that the data in hand have an ordered nature. In other words, this particular class of estimators is used when variables that are close are strongly correlated and variables that far a part are weakly correlated. Examples of such data are: longitudinal data, time series, spatial data, and many more. Estimators for such data have been discussed extensively. For instance, Wu and Pourahmadi [2003] proposed a non-parametric estimate of the covariance matrix of longitudinal data. Their estimator was constructed using the coefficient of a linear regression of each variable on its predecessors. This has helped in maintaining the positive definiteness of the estimated matrix. A widely used technique for estimating the covariance matrix of an ordered data is by banding or tapering the sample covariance matrix. This method has been discussed in Bickel and Levina [2004], Bickel and Levina [2008], El Karoui [2007], Furrer and Bengtsson [2007], and Rothman et al. [2009].

Bickel and Levina [2008] proposed a method to estimate the covariance in such cases by significantly thresholding small elements of the sample covariance matrix to zero, thus, yielding a sparse and more improved estimates. The upper hand of regularizing the covariance matrix by thresholding lies in its simplicity and cost effectiveness in terms of computational time. It also has some theoretical advantages over other techniques with its major downside being that it does not guarantee positive definiteness of the obtained estimate.

Other techniques depend on providing an estimator using the inverse of the covariance matrix rather than the covariance matrix itself. These methods are mainly based on regularizing the Cholesky factor of the inverse of the covariance matrix. They make use of the fact that the entries of the Cholesky factor has a regression interpretation. This allows the use of regression regularization tools such as ridge regression and Lasso penalties, see Wu and Pourahmadi [2003] and Huang et al. [2006]. Bayesian regularization of the Cholesky factor of the inverse also have been proposed by Smith and Kohn [2002] while Wong et al. [2003] have discussed a bayesian technique for regularizing the inverse of the covariance matrix. In the exposition, we outline its key characteristics, classification, underlying assumptions and general guidelines.

2.4 Handling Variability and Uncertainty

Air quality models have shown satisfactory results in forecasting air quality variables, however, the uncertainties that rise in these models limit the reliability of the resulted forecasts. One possible source of the model uncertainty lies within the model's input parameters such as, emissions, land use, meteorological fields, *etc* [Hanna et al., 1998]. This is because forecasts are created based on previous observations that are fed to the model to produce the best forecast. This means that the slightest error in the initial input of the model could lead to a larger error in the forecast. Thus, there will always be a limit to the forecasting reliability of the air quality model. Another source of model uncertainty is the deficiency in capturing the physical and chemical parameterization, such as, chemical mechanism, biogenic

emissions, *etc* [Mallet and Sportisse, 2006]. The natural and random fluctuations in the atmosphere is another cause of uncertainties. The latter source of uncertainty is impossible to control. Due to these nontrivial uncertainties, forecasts that are produced by air quality models need to be assessed and cross validated to estimate their reliability.
Several techniques have been used to assess the uncertainties of air quality models. For example, Schmidt [2002] estimated local uncertainties by differentiating model outputs with respect to the model input; Hanna et al. [2001] used standard probability density functions and Solazzo et al. [2012] and Vautard et al. [2009] used ensemble modelling methods. The scientific community focusing on climatology has not had many studies circumventing issues of data randomness. The very nature of climatology is spatial-temporal and yet literature is devoid of modelling methods that are robust to variations in this context. Thus, in the light of existing literature and knowledge, this work highlights the way forward by first setting the scene for point-specific forecasts based on functional relationships that derive from combined ideas of statistical and CTM downscaling nature. Moreover, there exist a great deal of uncertainty within CTM simulations due to some input and model parameterization resources. An ensemble of CTM models can be used to overcome the uncertainty issue. This would raise another concern, which is the need for a single combined local forecast that well represents the atmospheric variable of interest.

2.4.1 Building Ensemble Models

An ensemble is a set of numerical models that slightly differ from each other in their initial conditions or formulations. Its relevance to the typically chaotic and highly variable atmosphere is therefore self-explanatory. In addition to data randomness, single air quality models have two possible sources of uncertainty - the initial conditions of the model (usually chosen under specific assumptions which are likely to be violated, such as the natural state of atmospheric conditions) and the construction of the model itself. This uncertainty is more pronounced in standard air quality models which are usually constructed *a priori* based on mathematical equations that were solved numerically. Ensemble modelling circumvents this problem as they are created using models with different initial conditions and formulations, hence they are capable of *averaging out* a range of uncertainties by taking into account all sources of uncertainty. The reasoning here is that the variations amongst the ensemble members predictions can be used as an uncertainty measurement of the forecasts. Ensembles can be built either by different forecast models which have the same modelling platform or with perturbations of input parameters of a single model [Damien and Mallet, 2010]. A set of ensemble forecasts produced using different forecasting models is referred to as *multi-model ensemble forecasting* [Zhou and Du, 2010] while, on the other hand, a set obtained based on various initial conditions or different physical or chemical parameters of single forecasting model, the ensemble is said to be *stochastic*.

2.4.2 Addressing Distribution-related Issues in Ensemble Modelling

It is rather obvious that direct comparison of the set of ensemble forecasts with the actual observations is not a straightforward task. This issue can be resolved when ensemble members results are combined in a new single prediction value which is expected to have a lower prediction error as each member's errors, hopefully, cancel each other to some extend [Vautard et al., 2009] - we come back to the implementation of this approach in Section 3.3.4.

One simple and widely used ensemble combination approach in air quality studies is the ensemble mean, where the average of the ensemble members is used for straightforward comparison with the actual air quality measurements. For obvious reasons, using the ensemble mean has produced good predictive performance in several studies. For example, Delle Monache and Stull [2003] used the average of an air quality multi-model ensemble to improve the forecasts of ozone in western Europe. They compared the performances of averaged ensembles with individual models and concluded that the ensembles outperform single models by significant margins. There are many other studies with similar conclusions - see, for instance, McKeen et al. [2007] and Van Loon et al. [2007*b*]. However, it is important to note that, although the ensemble averaging has shown good results, much of its performance can be attributed to the distributional behaviour of the data. For instance, the mean is sensitive to extreme values within ensemble members and if there are outliers in the data which could not be detected, results are not going to be reliable. To overcome the foregoing issue, some studies have adopted the use of the ensemble median - with good predictive performance reported by Galmarini et al. [2004] and Monteiro et al. [2013]. However, using the median to combine ensemble forecasts comes with the assumption that ensemble members are equally weighted and because of this, the median might not be a very good criterion for the ensemble.

One way of getting around the foregoing ensemble-related issues would be to assign varying weights to each member. There are various ways in literature by which weights can be determined. One way is to use regression methods to obtain weight coefficients for each model within the ensemble such that the linear combination of a weighted ensemble members is used to investigate the forecasting ability of the ensemble. Krishnamurti et al. [2000] used the multiple regression approach to obtain weight coefficients from a *super-ensemble* - i.e., an ensemble built from a set of multi-model ensembles that have been adjusted according to their biases - of precipitation forecasts. The authors regressed each ensemble member forecast within against the observation. Then they obtained varying weights by minimizing the least square errors between the model and the observation. This was applied at all geographical locations (grid points) and the results indicated that the weighted ensemble outperformed all single models in forecasting precipitation. Pagowski et al. [2005] replicated the technique on an ensemble of seven air quality models to improve ozone predictions over eastern US and southern Canada and the results concluded that the weighted ensemble outperformed the average ensemble and any single model therein. Again, it is important to note that this technique is distribution-dependent.

Bayesian methods have also been used to combine ensemble members - most notably the Bayesian Model Averaging (MBA) method introduced by Raftery et al. [2005]. The BMA technique combines different models within an ensemble by assigning weights to each model based on their probabilistic information [Monteiro et al., 2013]. The weights are obtained based on the probability density functions (PDFs) of the models [Raftery et al., 2005] and based on a pre-specified training period. The weights are assigned to each model such that the models with better forecasts of the measure of interest get the highest weights [Monteiro et al., 2013]. Theoretical details of the BMA method are provided in Raftery et al. [2005]. The method has been widely adopted in many studies to combine ensemble forecasts and calibrate forecasting abilities of models - see, for instance, Raftery et al. [2005], Duan et al. [2007], Sloughter et al. [2010] and Sloughter et al. [2012]). Chandler [2013] presented a Bayesian framework for combining multi-model outputs which emphasized the strength and weaknesses of each model by assigning varying weights to each model taking into account prior knowledge of existing conditions as well as available information in the historical data attributes. The framework accounts for biases between the models and real climate data and it forms a basis for developing statistical models between the multi-model and the actual observations by combining the weighted models.

2.4.3 Assessing Uncertainty in Ensembles by Downscaling

Despite the high level of sophistication of air quality ensembles in addressing uncertainties in the models, spatial resolution remains an issue of concern. The models produce forecasts at coarse spatial resolution and since locations are unlikely to have the same amount of atmospheric variable concentrations, a higher spatial resolution forecast is needed. The method of combining the ensemble forecasts have succeeded in improving the predictive performance of the models, however, the combined forecast is produced over a large scale grid cell and in application forecasts at a specific location is required. In order to improve ensemble predictions and overcome the spatial resolution differences between what is actually needed and what the ensemble produces, statistical downscaling can be applied.

So far we have seen how statistical downscaling aims to improve air quality models forecasts by reproducing them at a higher spatial resolution using a statistical model. The main advantage ensemble downscaling has over downscaling a single model is the fact that the former takes into account models uncertainty (due to parameterization and input data). Hence, downscaled ensemble predictions maintain all available atmospheric information and have a finer spatial resolution at the same time. But as we discussed, elements of uncertainty still remain within the ensemble model constructs.

Downscaling of air quality ensembles have been discussed extensively in literature. A widely used ensemble technique is regression, see Feddersen and Andersen [2005], Kryzhov [2012], and Min et al. [2011]. Bayesian methods have also been employed in ensemble downscaling, see Coelho et al. [2006]. A combination of regression and Bayesian techniques in downscaling ensembles have shown good results however there is always room for improvement. Our

work's contribution to knowledge picks up from these achievements to develop a robust downscaling model with minimal sensitivity to variability.

Ensemble downscaling using a combined ensemble forecasts (*e.g.* ensemble mean) or all ensemble members is quite a popular approach. When downscaling a combined ensemble output in this way important information based on the parameterization of each model might not be accounted for very well. An alternative way would be to consider all ensemble members outputs when downscaling, while this method takes into account uncertainties, it raises the issue of multi-collinearity among predictors. To overcome these issues, dimension reduction methods can be used to ensure that no important information on uncertainty is lost. Some studies have used PCA to reduce the dimension of the ensemble members and then carry out the downscaling. For example, Benestad [2002*b*] used common EOFs to downscale multi-model ensemble of temperature scenarios over northern Europe, he employed the EOF method as described in Benestad [2001]. The same technique was employed in Benestad [2002*a*] to downscale a multi-model of temperature and precipitation. To our knowledge there is not extensive work about the use of dimension reduction methods when downscaling ensembles. In the next chapter we outline our general methodology to addressing the research question based on the objectives as listed in Section 1.2. More specifically, the chapter outlines the mechanics of our contribution to knowledge which build upon the methods discussed above.

Chapter 3: Data Sources and Methodology

This chapter outlines the path towards answering the research question in Section 1.2. It provides a comprehensive description of the data sources and study areas; the adopted modelling strategy and the modelling methods used. It sets off from the premises that results from downscaled air quality models are susceptible to variability and uncertainty that arise from both data and the modelling process - hence the need for plausibility of both data and models. Cook [2007] recounts the arguments on how plausible it is to use the response variable to develop the predictors between prominent statisticians - R. A. Fisher, F. Mosteller and J. W. Tukey being skeptical of nature's malice to relate the response to the least important Principal Components and D. R. Cox, H. Hotelling, D. M. Hawkins and L. P. Fatti being open to the idea that in nature it is helpful to use the response to choose the predictors. Our adopted methodology refrains from the foregoing philosophical debate and focuses on the position that variability, as discussed above, stems from two sources - data randomness and the modelling techniques employed. Thus, developing robust models for air quality inevitably requires guaranteeing three vital aspects of modelling - data quality, model resilience and concept drifting of the two. In other words, we need adaptability of modelling approaches to changing spatial-temporal conditions.
The methodology is developed within the context of a typical data modelling scenarios that involves either unlabelled **(unsupervised)** or labelled **(supervised)** data with the adopted strategy focusing on methods that have the potential for fulfilling the foregoing aspects - that is, methods capable of learning rules from data and utilising them to new cases. The chapter is organised as follows. Section 3.1 identifies the sources of data used in the analyses and it briefly outlines the collection methods and their amenability to answering the research question. From section 3.2 onwards the adopted modelling strategy encapsulates aspects of downscaling for labelled and unlabelled data - i.e., from key assumptions through Exploratory Data Analysis (EDA) to dimension reduction and predictive modelling are presented.

3.1 Data Sources and Study Area

3.1.1 Data Sources and Study Area for Downscaling An Air Quality Model

The measurements are hourly ozone observations in south-eastern USA over the summer period (June-August) of 2005 at stations maintained by the US Environmental Protection Agency (EPA). Data are available for 109 monitoring stations but some stations (in particular 15 stations) were outside or at the border of the grid cells range of the Air Quality Model in use (REAM). As we want to relate regional patterns of ozone to local observations, we discarded these 15 stations and carried out the analysis for the remaining 94 stations, shown in Figure 3.1. Since some of the stations had missing values (4.5% missing in all) we use linear interpolation to estimate the missing values. Ozone 24 hours ahead predictions are obtained from REAM model. REAM produces forecasts within 104 grid cells (which cover the south-eastern region in the US) with 70 km spatial resolution. Out of the 104 and grid cells 5 overlap with the sea, as shown in Figure 3.1. We do not consider these grid cells in our study. The data were highly skewed for some of the stations [see figure 3.2] and, as this could distort the relationships in the models build under distributional assumptions, the ozone data were converted to square root scale to remove the skewness. Further, prior to carrying out the analyses, the data were centred to avoid the non-stationary features that would prevent proper application of regression. More specifically, centring aimed at removing the diurnal cycles.

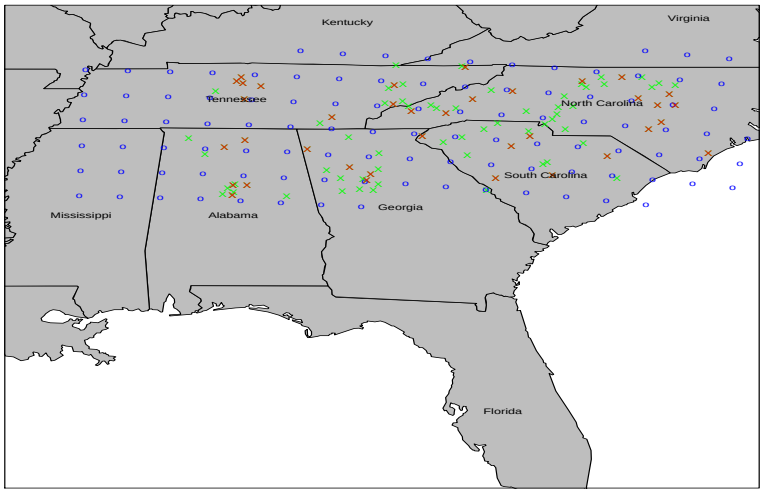

Figure 3.1: Ozone monitoring stations ($\times$) and REAM grid cells (circles).

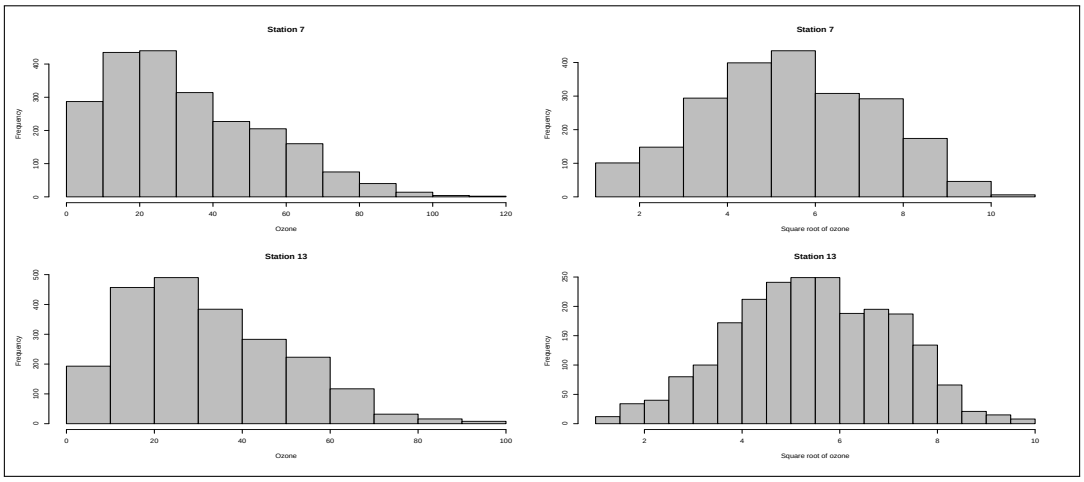

Figure 3.2: Histogram of ozone observations for stations 7 and 13. The plots on the left panel show the ozone observations on original scale and the plots on the right panel show the ozone observations on square root scale

The REAM Modeling System

REAM stand for the Regional chEmical trAnsport Model (Choi, Wang, Cunnold, T., Shim, Luo, Eldering, Bucsela and Gleason [2008]; Choi, Wang, Zeng, Cunnold, Yang, Martin, Chance, Thouret and E. [2008]; Wang et al. [2006, 2009]; Zeng et al. [2006]; Zhao et al. [2009, 2010]). It adopts the photochemical, dry deposition, and biogenic emission modules from the GEOS-CHEM model, see Bey et al. [2001] and references therein. It uses the same model setup by Choi, Wang, Zeng, Cunnold, Yang, Martin, Chance, Thouret and E. [2008] over North America. Anthropogenic and biogenic emission algorithms and inventories are adapted from the GEOS-CHEM model (Choi et al. [2005]; Choi, Wang, Cunnold, T., Shim, Luo, Eldering, Bucsela and Gleason [2008]). One exception is that the emissions of NO_x, CO, and $\geq$ C4 alkanes over the US are prepared by Sparse Matrix Operator Kernel Emissions (SMOKE) model [Houyoux et al., 2000] for 2005 projected from the VISTAS 2002 emission inventory.

REAM uses the National Center for Atmospheric Research/Penn State MM5 dynamical model to provide the meteorological fields using four-dimensional data assimilation based on the National Center for Environmental Prediction (NCEP) reanalysis and surface observations. The REAM model used in this study has 70 km horizontal resolution with 21 vertical layers in the troposphere. The five extra grids on each side of the REAM domain are for minimizing potential transport anomalies near the boundary. The 2005 summertime GEOS-CHEM global chemical transport model (version 7.2) simulations are used to specify initial and boundary conditions for trace gases for June-August 2005 time period. The regional simulations are carried out in the last two weeks of May for spin up, and used to determine the initial chemical condition in the troposphere for the June-August 2005 simulation.

3.1.2 Data Sources and Study Area for Downscaling An Ensemble of Air Quality Models

We use ground level ozone measurements obtained from the Air Base database which provides ozone concentrations in $\mu g/m^3$. We retain network stations that have sufficient amount of data so accordingly we discarded stations that has more that 100 missing values. There are three different classifications for the measurement networks: urban, suburban, and rural monitoring stations. Since sources and the buildings are mostly in urban areas, we would expect that downscaling is better illustrated for these stations, and hence we restrict our analysis to urban locations within France. There are a total of 116 urban stations (after discarding stations with more than 100 missing observations). For computational convenience we randomly select 35 station across France as shown in Figure 3.3. As summer is the most polluted season, our analyses were confined on summer months only (June to August 2001). An off-the-shelf ensemble containing 107 members generated by Damien and Mallet [2010] over all of Europe with a 0.5°horizontal resolution is used. The ensemble produces hourly ozone forecasts for the full year of 2001 over 3082 grid cells that covers all of the European region. The available data are significantly large which makes performing the analysis is computationally expensive and challenging. Hence we will restrict our analysis on France only as it contains many measurement stations. Therefore, we select the ensemble outputs with a total of 770 grid cells that cover all of France.

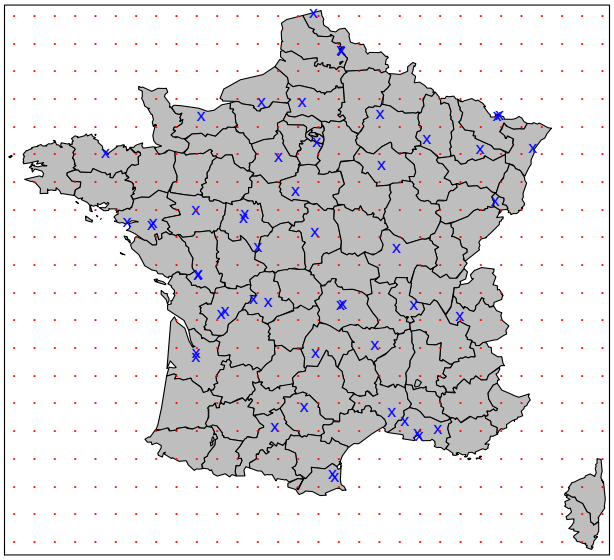

Figure 3.3: Ozone urban monitoring stations ($\times$) and Ensemble grid cells (dots).

The Polyphemus Modeling System:

The Polyphemus system is an ensemble forecasting system of air quality measurements that was development by Mallet and Sportisse [2006]. It is composed of three contents:

1. AtmoData library, which is a library of physical parameters that contains several parameterizations [Mallet and Sportisse, 2005].
2. The Polair3D chemistry transport model [Boutahar et al., 2004].
3. Programs that extract the input parameters from the AtmoData library.

The Polyphemus system was run by Damien and Mallet [2010] to generate an ensemble across Europe to forecast many atmospheric measurements such as, ozone and NO_2. The simulated models were created by using the following parameterization and input data options:

1. Physical Parameterization:
    - Land use cover.
    - Chemistry.
    - Cloud attenuation.
    - Critical relative humidity.
    - Vertical diffusion coefficient(K_z).
    - Deposition Velocity.
    - Coefficient Ra (for aerodynamic resistance in deposition velocities).
    - Vertical distribution emissions.
    - Photolysis rate.
2. Numerical Options:
    - Time step.
    - Vertical resolution.
    - First layer height.
    - Vertical wind diagnosis.
    - Minimal K_z.

- Minimal K_z in urban areas.
- Vertical application of minimal K_z.
- Exponent p to compute K_z.
- Boundary layer height.

For a detailed explanation of each parameter and how they were chosen to create the ensemble see Damien and Mallet [2010]. Ensembles created by using the Polyphemus system have an advantage of taking into account all sources of uncertainties at once: input data, physical parameterization and numerical options. This means that each member within the ensemble is a chemistry transport model on its own. An ensemble that contains 101 members was generated by modifying the stated physical parameterization and input data in the Polair3D air quality model. A detailed description on the ensemble generation process is stated in Damien and Mallet [2010]. The models are built for the year 2001 across Europe over (10.75°W, 22.75°E) ×(34.75°N, 57.75°N) at a 0.5°horizontal resolution. In addition six reference models (that were generated "by hand" as a combination of a pre-specified options) were included in the ensemble. The properties and the performance of the ensemble have been analyzed extensively in Damien and Mallet [2010].

3.2 Modelling Strategy

Linking up small and large scale variables involves relating the variables to one another, extracting the inferences that arise from those relationships and drawing statistically sound conclusions. The main purpose of downscaling is to correct the spatial-temporal mismatch among the variables while abiding by the fundamental theoretical foundations. This section describes the mechanics of some of the alternative methods to our own that were applied to the data. The sections highlights the methods suitability within the scope of our research question, objectives and their comparability to the proposed approach. The adopted downscaling strategies are essentially statistical and like all statistical techniques, there are essential assumptions that need to be taken into consideration prior to performing the downscaling. We shall be guided by the three main statistical downscaling assumptions identified by Giorgi et al. [2001] as follows

3.2.1 Key Statistical Downscaling Assumptions

1. **The chosen large scale predictors should be related to the local response variable realistically.** Furthermore, the predictors should be available at a spatial scale that covers the region of the downscaled response. Thus, the predictors must be selected such that they are realistically relevant to the climate response and spatially well

represented by the climate model. This indicates that the climate model needs to be verified that it can provide the variable at an adequate spatial domain. Whether it is to determine the variables responsible for the highest variation in data or the best predictors of a phenomenon under study, variable selection is always central to the process. Hammami et al. [2012] compared logistic regression - a combination of forward selection and backward elimination - with variants of the Least Absolute Shrinkage and Selection Operator (LASSO) based on the original ideas in Tibshirani [1996]. Not surprisingly, their findings prioritised the latter due to its computational advantages which, apparently arise from its underlying assumptions. Our work seeks to circumvent this issue.

2. **The selected transfer function should hold and stay valid for different climate conditions [Giorgi et al., 2001].** This means that the established relationship between the predictors and the predictand must remain valid at time periods outside the fitting period time range (when the model was established). For this reason, an independent set of observational data that are well apart from the original fitted period is needed to validate the model adequacy. This can be achieved by splitting the climate model outputs into two sets of present and future data. The downscaling equation is built using the present dataset. Then the model validity is tested by using the future dataset [Giorgi et al., 2001]. Due to inherent randomness in data, this approach is still susceptible to variability. Introducing spatial-temporal adaptive parameters can achieve more accurate and robust results. Laflamme et al. [2016] applied a CDF transformation function to local-level daily precipitation extremes, downscaling 58 locations in New England and calculating 25-year return levels from projected distribution of extreme precipitation local-level, they concluded that there had not been significant increases over the period. In comparing predicted with historical distributions they estimated uncertainty using three procedures - a parametric bootstrapping with mean corrected confidence intervals, a non-parametric bootstrapping with bias corrected and acceleration intervals and a Bayesian model. Our work follows in their footsteps but rather than deriving results from distributional differences, we introduce tuneable parameters that serve to adapt the models with respect to concept drift - i.e, changing assumptions and conditions.

3. **The selected set of predictors must be an accurate representation of the downscaled climate variable**. As mentioned above, variable selection methods (such as stepwise regression) might be used to eliminate essential predictors that could be beneficial in representing the future climate changes. As the techniques rely on the information in the data attributes, distributional assumptions remain fundamental to these models accuracy and reliability. Adopting a Bayesian approach that recursively computes prior and posterior probabilities in multi-sampled data potentially yields more accurate and robust outcomes. This idea derives from the original work by Jeong et al. [2013] who used a spatial interpolation approach, initialising the process by the initial observation site precipitation series, in order to establish a physical and/or statistical relationships between the simulated GCM grid point and precipitations and the observed local site precipitation series. Such relationships can be

established via multiple applications of dimensional-reduction methods.

3.2.2 Regression Analysis: The Basis for Statistical Downscaling

Regression analysis provides the primary insight into statistical downscaling in consideration of both labelled and unlabelled data characterised by variability as discussed above. Based on a simple illustration, given a response y and predictors $x_1, x_2, \cdots, x_p$, the parameters $\beta_0, \beta_1, \cdots, \beta_p$ in the basic linear regression model

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + \epsilon_i \tag{3.1}$$

are determined from available data - usually called training set. Computation of these parameters is based on the well-documented method of least squares [Neter et al., 1996] which minimises the sum of the squared deviations

$$Q = \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij} \right)^2 \tag{3.2}$$

Once the regression parameters are obtained, Equation 3.1 can be used to predict the response value for given values of the predictors. Our strategy proceeds on the basis of this simple understanding. In particular, we consider the unlabelled data case - in which the goal is to uncover inherent structures in the data matrix $\mathbf{X}$ as well as the labelled case whereby the data attributes in the data matrix $\mathbf{X}$ provide rules of associations with the label Y. Regression-based downscaling was accomplished by fitting a simple linear regression model for each station individually in Section 3.1.1, regressing hourly ozone observations on the grid cell that includes the station using the following model

$$O_t = \beta_0 + \beta_1 M_t + \epsilon_t \tag{3.3}$$

where the response variable, O_t, represent hourly ozone observations, M_t is the REAM model output of the grid cell that include the station, β_0 and β_1 are regression model parameters and are estimated by the method of least squares, and ϵ_t is an error vector with mean 0 and a constant variance σ^2. The model in Equation 3.3 is fitted to the historical data, and then used to predict hourly ozone observations. Forecast accuracy is measured by the root mean square error (RMSE) over several runs, the RMSEs are compared with results from the proposed PFC downscaling approach.

3.3 Dimensional-Reduction Methods

Recognising that spatial-temporal variability is an issue of concern and that statistical downscaling is a data-hungry approach, data quality becomes of paramount importance as it is the main source of variability. Our objectives in Section 1.2 are set to reflect the foregoing reality. Further, understanding the overall behaviour of the data at hand and gaining insight into the interactions of the data attributes will provide a basis for unsupervised learning while proper interpretations of the identified structures lead provide further insights, model comparability through training, validation and testing using multiple modelling techniques is the only way to confirm the results. Our modelling techniques are therefore chosen to reflect the foregoing characteristics. More specifically, we look at the dimensional reduction of data and predictive modelling in the Bayesian context. Finally, implementation to PFC models with multiple sampling from the data is adopted for the purpose of addressing variability. The material in this section are organised as follows. Section 3.3.1 presents the main ideas of Principal Component Analysis (PCA) and highlights its relevance to downscaling. It is followed, in Section 3.3.2 by the Principal Fitted Components (PFC) - a version of PCA that makes use of the response variable [Cook and Forzani, 2008]. Section 3.3.3 present the a methodology to performing PFC downscaling via a regularized covariance matrix. Section 3.3.4 presents the mechanics of the proposed PFC-based ensemble downscaling method followed by a Bayesian approach to compare air quality models performances.

3.3.1 Principal Component Analysis (PCA)

This section provides the basics for PCA application in downscaling and highlights the method's key concepts - matrix algebra, covariance, correlation matrices, eigenvalues, eigenvectors, normalisation and particularly how they tie in with the role of the method as a downscaling tool. The technique is applied for creating new variables which are a linear combination of the original variables [Anderson, 2003]. In particular, PCA can be applied to group our high-dimensional data, described above, together based on a smaller number of **super-variables** without loss of information. In other words, PCA seeks to transform a number of correlated variables into a smaller number of uncorrelated variables, called **P**rincipal **C**omponents. The technique uses the correlation among variables to develop a small set of components, which empirically summarise the correlations among them. Its main objective is to reduce data dimensionality while retaining most of the original variability in it. That is, it seeks to reduce the number of variables without losing important information which, in turn, helps detect naturally arising relationships among them. Principal components are extracted in succession, with the first component accounting for as much of the variability in the data as possible and each succeeding component accounting for as much of the remaining variability as possible.

Specifically, PCA is concerned with explaining the variance-covariance structure of a high dimensional random vector through a few linear combinations of the original component variables. It works by forming new variables such that the first variable accounts for the maximum variance in the original data with the second variable accounting for the maximum variation in the original data not accounted by the first variable while the p^{th} variable accounts for the maximum variation not accounted by the first $p-1$ variables. Thus, the first few PCs contains significant amount of information from the original data. PCA is a dimension reduction technique since one could use the first few PCs to represent the original data if the number of predictors is significantly large. The extracted components are inferred from the correlations among the number of variables, with each component being estimated as a weighted sum of the variables. If we let **X(t)**=$\mathbf{X}_1(t)$, $\mathbf{X}_2(t)$, $\cdots$, $\mathbf{X}_p(t)$ be an $n \times p$ centered matrix of predictors, where, p is the number of predictors, and $t = 1, 2, \cdots, n$ are observations, with the covariance matrix of predictors given as

$$\Sigma = \begin{bmatrix} [(X_1-\mu_1)(X_1-\mu_1)], & [(X_1-\mu_1)(X_2-\mu_2)], & \ldots, & [(X_1-\mu_1)(X_n-\mu_n)] \\ [(X_2-\mu_2)(X_1-\mu_1)], & [(X_2-\mu_2)(X_2-\mu_2)], & \ldots, & [(X_2-\mu_2)(X_n-\mu_n)] \\ [(X_3-\mu_3)(X_1-\mu_1)], & [(X_3-\mu_3)(X_2-\mu_2)], & \ldots, & [(X_3-\mu_3)(X_n-\mu_n)] \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ [(X_n-\mu_n)(X_1-\mu_1)], & [(X_n-\mu_n)(X_2-\mu_2)], & \ldots, & [(X_n-\mu_n)(X_n-\mu_n)] \end{bmatrix}$$

then we can think of linear combinations of the variables forming $C_1, C_2, \cdots, C_p$ principal components

$$\begin{aligned} C_1 &= \mathbf{a}_1^T\mathbf{X} = a_{11}X_1 + a_{12}X_2 + \cdots + a_{1p}X_p \\ C_2 &= \mathbf{a}_2^T\mathbf{X} = a_{21}X_1 + a_{22}X_2 + \cdots + a_{2p}X_p \\ C_3 &= \mathbf{a}_3^T\mathbf{X} = a_{31}X_1 + a_{32}X_2 + \cdots + a_{3p}X_p \\ &\cdots\cdots\cdots\cdots\cdots\cdots \\ C_p &= \mathbf{a}_p^T\mathbf{X} = a_{p1}X_1 + a_{p2}X_2 + \cdots + a_{pp}X_p \end{aligned} \tag{3.4}$$

such that $\mathbf{a}_1, \mathbf{a}_2, \cdots, \mathbf{a}_p$ are the eigenvectors that correspond to the eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p$ of the covariance matrix of predictors $\mathbf{\Sigma}$ = Var(**X**). The $\mathbf{a}'s$ are known as PC directions or PC loadings - i.e., the weights associated with each predictor. Note that from the p variables we can extract p components and, mathematically, this amounts to solving p equations with p unknowns. The variance in the correlation matrix is repackaged into p eigenvalues, with each eigenvalue representing the amount of variation that the component captures. Each extracted component is a linear combination of the p variables. The first component accounts for the largest possible amount of variance. The second component, formed from the variance remaining after that associated with the first component has been

extracted, accounts for the second largest amount of variance, etc. It can be shown that the total variance in any component k is $\mathrm{Var}(C_k) = \sum_{m=1}^{p} \lambda_m a_{mk}^2$ and so dividing the variance of a component with this total provides a good insight as to how much contribution to variability that component makes. The principal components are extracted with the restriction that they are orthogonal. Geometrically they may be viewed as dimensions in p-space, with each dimension perpendicular to each of the other dimensions. Intuitively, each component is a function of random data and so it retains an element of randomness for which assuming a distribution, it has a population variance defined as

$$\mathrm{Var}(C_i) = \sum_{m=1}^{p} \sum_{n=1}^{p} a_{im} a_{in} \sigma_{mn} = \mathbf{a}_i' \sum a_i \tag{3.5}$$

Further, two components, say, C_i and C_j vary together, hence can be said to have a population covariance defined as

$$\mathrm{Cov}(C_i, C_j) = \sum_{m=1}^{p} \sum_{n=1}^{p} a_{im} a_{in} \sigma_{mn} = \mathbf{a}_j' \sum a_j \tag{3.6}$$

The coefficients of the foregoing linear combinations form the vector

$$a_i' = \begin{bmatrix} a_{i1} & a_{i2} & a_{i3} & \dots & a_{i(p-1)} & a_{ip} \end{bmatrix}$$

For component one to explain the highest variation in the data, the coefficients $a_{11}, a_{12}, ..., a_{1p}$ must be selected such that its variance is maximised subject to the constraint that the sum of the squared coefficients is equal to one. That is,

$$\mathrm{Var}(C_1) = \sum_{m=1}^{p} \sum_{n=1}^{p} a_{1m} a_{1n} \sigma_{mn} = \mathbf{a}_1' \sum a_1 \quad \text{subject to} \quad a_1' a_1 = \sum_{j=1}^{p} a_{1j}^2 = 1 \tag{3.7}$$

The same applies to the second component, except that there is an additional stipulation that

$$\mathrm{Cov}(C_1, C_2) = \sum_{m=1}^{p} \sum_{n=1}^{p} a_{1m} a_{2n} \sigma_{mn} = \mathbf{a}_1' \sum a_2 = 0 \tag{3.8}$$

All subsequent components inherit this property, making them account for as much of the remaining variation as possible while remaining uncorrelated with other components - that is, for any k^{th} component, the constraints are

$$
\begin{aligned}
a_k' a_k &= \sum_{j=1}^{p} a_{kj}^2 = 1 \\
\mathrm{Cov}(C_1, C_k) &= \sum_{m=1}^{p}\sum_{n=1}^{p} a_{1m} a_{kn} \sigma_{mn} = \mathbf{a}_1' \sum a_k = 0 \\
\mathrm{Cov}(C_2, C_k) &= \sum_{m=1}^{p}\sum_{n=1}^{p} a_{2m} a_{kn} \sigma_{mn} = \mathbf{a}_2' \sum a_i = 0 \qquad (3.9) \\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
\mathrm{Cov}(C_{k-1}, C_k) &= \sum_{m=1}^{p}\sum_{n=1}^{p} a_{(k-1)m} a_{kn} \sigma_{mn} = \mathbf{a}_{k-1}' \sum a_k = 0
\end{aligned}
$$

Putting into Perspective

The extraction of principal components amounts to a **variance maximising** rotation of the original variable space. Thus, if there are $k < p$ variables, the k^{th} component is computed as a weighted sum of the variables as

$$C_k = \mathbf{a}_k^T \mathbf{X} = a_{k1} X_1 + a_{k2} X_2 + \cdots + a_{kp} X_p$$

The vectors $\mathbf{a}'s$ are chosen such that each of the determinants of **a** is 1; each of the principal components, C_k maximises $\mathrm{Var}(a_k X_i)$ and $\mathrm{Cov}(a_k X_i, a_r X_i) = 0, \forall r < k$. In other words, the principal components are extracted from the linear combinations of the original variables maximising the variance and have zero covariance with the previously extracted components. The rotation is called **variance maximising** because it seeks to maximise the variability of the **new variable**, while minimising the variance around it. To put this into perspective, consider our REAM model hourly data of 2203 hourly observations over 99 grid cells which, for convenience, we denote by $X_{ij}, i = 1, 2, 3, ..., 2203$ and $j = 1, 2, 3, ..., 99$. In this case, PCA will be seeking to reduce the 99 grid cells by generating new linear combinations of these cells that describe the variation in the predictions in a magnitude order. The same applies to the second part of the dataset consisting of actual ozone observations from 94 monitoring stations that scatter across the study region. Again, here, the objective will be to extract the variability from the data as described by the 94 stations.

Practical Interpretation of PCA

Once the variance maximising line has been found, there will still be some variability around it. Thus, extraction of components will continue by defining another line that maximises the remaining variability, and so on. By so doing, consecutive components are extracted. As each consecutive component is based on the maximisation of the variability

not captured by the preceding component, the resulting components are said to be independent. In other words, they are uncorrelated or orthogonal to one another. Since we seek to reduce the number of variables, the fundamental question is how many components we want to extract. Since the number of components is inversely related to the amount of variability - that is, as the number of variability goes up they account for less and less variability, common sense dictates that we stop when there is only very little **random** variability left. The nature of this decision is arbitrary. In practice, it is based on a number of guidelines which depend on both the theoretical foundations of the extraction of components and the underlying domain knowledge. PCA is applied in this study as one of the baseline techniques, so in its interpretation we shall strive to be as thrift as we possibly can - using the Kaiser and scree plot criteria as standard. Another important concept is that of "loadings". PCA amounts to transforming variables into linear combinations of an underlying set of hypothesized or unobserved components. The resulting components may be associated with 2 or more of the original variables. Loadings relate the specific association between factors and original variables. In particular, the concept of "loadings" refers to the correlation between the original variables and the factors, and the key to understanding the nature of a particular factor. Loadings derive from the magnitude of the eigenvalues associated with the individual variables. Squared factor loadings indicate what percentage of the variance in an original variable is explained by a component. Consequently, it is necessary to find the loadings, then solve for the factors, which will approximate the relationship between the original variables and underlying factors.

Downscaling an Air Quality Model by Principal Component Regression

Regressing ozone observations on the grid cell that contains the station may not be very efficient as the CTM may be misaligned and local ozone may be more closely related to regional conditions rather than the average over the grid cell. REAM forecasts ozone levels over $p = 99$ grid cells, and considering all 99 cells in the regression model raises the problems of "over-fitting" (such a model produces a good fitting performance but poor predictive performance on new data) and multicollinearity. A typical way to conquer this is to reduce data dimensionality - i.e., to reduce the number of predictors some of which might be autocorrelated. Principal Component Analysis (PCA) is one of the most popular dimensional-reduction techniques in use todate. As described in Section 3.3.1, the technique can be applied to the gird cells model output and then a few Principal Components that capture the highest variation in the original grid cells data can be used to explain the nature of variation in the data and potentially the components can be used to form new predictors for use in a regression model. Prior to applying PCA, the data are centred to enable the environment for forming components from the resulting covariance matrix and also to eliminate the influence of scale. For each station we regress hourly ozone observations on selected number of grid cell PCs. The model is defined as

$$O_t = \alpha_0 + \sum_{m=1}^{M} \alpha_m Z_m(t) + \epsilon_t \tag{3.10}$$

where, O_t is the model output, $M < p = 99$ is the number of PC scores in the model, $\alpha_0, \alpha_1, \cdots, \alpha_m$ are model parameters, $Z_1(t), Z_2(t), \cdots, Z_m(t)$ are PC scores, $t = 1, 2, \cdots, n$, and ϵ_t is an error vector with mean 0 and a constant variance σ^2. PCs can be used in regression to replace the original predictors. The use of PCs in regression overcomes the problem of predictors being correlated amongst each other. Furthermore, when the number of predictors is considerably large, few PCs can be used as regressors instead. Usually the first few PCs (the ones with the highest variance) are selected to replace the original predictors. However, the choice of how many PCs to consider is fairly subjective to the purpose of the analysis. Many methods have been developed to assist on deciding the number of PCs to consider, e.g. the scree test [Wilks, 2006]. As noted earlier, making use of the predictors alone and not making sufficient use of the response, is viewed as a weakness of PCA mainly because extracted components are not invariant or equivariant under full rank linear transformation of the predictors [Cook and Forzani, 2008]. Our work picks up on this point to deploy principal fitted components (PFC) as a downscaling technique. The next exposition presents the PFC approach.

3.3.2 Principal Fitted Component Analysis

As noted above, the extracted principal components are formed by transforming variables into linear combinations with each variable "weighted" by quantities with specific magnitudes and directions - referred to as "loadings." This way, each component can be interpreted as a predictor $\mathbf{X}$ of the problem under consideration Y although, clearly, the principal components are obtained from the predictors only without reference to the response variable. Cook and Forzani [2008] present a comprehensive discussion around issues around the use of principal components as predictors in a regression model. Firstly, extracted principle components become informative when the objective of the study is to identify why and how the predictors are related amongst each other (interdependence relationship). Secondly, it has also been argued that principal components are not invariant (remain unchanged under some transformation) or equivariant (changes in a convenient way under some transformations) under full rank linear transformations of the predictors. Cook [2007] introduced Principal Fitted Components (PFCs) for dimension reduction citing two major advantages over PCs when used as a dimension reduction in regression. Firstly, PFCs can be tailored to the value of the response if the value is known and, two, they are equivariant under full rank transformation of the predictors. PFCs are obtained by gaining a sufficient amount of information about the response Y from the predictors $\mathbf{X}$. PFC's main idea can be described as follows. Assume we have a dataset $(\mathbf{X}, \mathbf{Y})$ with n observations on p variables such that the target and predictor variables for the data matrix are, respectively, defined as follows

$$\mathbf{Y}_{n \times 1} = (y_1, y_2, ..., y_{n-1}, y_n)^T \quad \text{and} \quad \mathbf{X}_{n \times p} = (x_1, x_2, ..., x_{n-1}, y_n)^T \tag{3.11}$$

where $\mathbf{n} > \mathbf{p}$. Without loss of generality, we can assume that both $y_{i=1,2,\ldots,n}$ and $x_{i=1,2,\ldots,n}$ are normally distributed with zero means. Thus, if we let $X_{k\in p}$ denote a random vector distributed as $X|(Y = k)$ such that

$$\bar{\mu} = \mathbf{E}(X) \quad \text{and} \quad \mathbf{D} = \{\mu_k - \bar{\mu}|y = k \in \mathbf{S}_Y\} \tag{3.12}$$

where S_Y is the entire Y sample space and $D \in \mathbb{R}^{p \times d}$ is a semi-orthogonal matrix the columns of which form a basis for the d-dimensional subspace of the span, $\mathbf{D}$. The span is designed to ignore outlying cases which might either be random or their existence is well understood. Implementation of PFC is problem-data-specific. In the statistical package, **R**, for instance, the response variable Y representing the response vector of n observations, can take different bases - polynomial, categorical, fourier, piecewise continuous or piecewise discontinuous. Each of the basis function is vector-valued function of $Y \in \mathbb{R}$. The standard Gauss-Markov linear regression model, in this case is defined as $Y = \beta\mathbf{X} + \epsilon$ where $\beta \in \mathbb{R}^p$ and $X \sim \mathcal{N}(0,\, \sigma^2\mathbf{I})$ and our interest is in obtaining fitted parameters $\hat{\beta}$ as determined by the data. The PFC mechanics derive from those of the PCA as described above.

Putting into Perspective

The variability matrix $\sigma^2\mathbf{I}$ plays a crucial role in searching for the best fitted $\hat{\beta}s$. Note also that dimension reduction is a direct function of the eigenvectors $\mathbf{a}_1, \mathbf{a}_2, \cdots, \mathbf{a}_p$ that correspond to the eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p$ of this matrix. Apparently, dimension reduction cannot be achieved if $|\hat{\lambda}_i - \hat{\lambda}_j| \cong 0; \quad \forall i \neq j$. Now, from $\mathbf{X} \in \mathbb{R}^p$ we can envision a new function $\phi\,(\mathbf{X}) \in \mathbb{R}^d$, where $d \leq p$. Apparently, the new function carries as much regression information on the target variable as does the original data $\mathbf{X}$ and therefore it can replace the original variable as $Y|\mathbf{X} = \beta^T\phi(\mathbf{X}) + \epsilon$ without losing any regression information. Based on the foregoing reasoning, Cook [2007] show that for $\phi\,(\mathbf{X})$ to be a sufficient reduction, either of the three conditions in Table 3.1 must hold.

Condition #	Reduction Type	Condition-1	Relationship	Condition-2
1	**Inverse reduction:**	$X\vert\,[Y, \phi(X)]$	$\sim$	$X\vert\phi(X)$
2	**Forward reduction:**	$Y\vert X$	$\sim$	$Y\vert\phi(X)$
3	**Joint reduction:**	X	independent of	$Y\vert\phi(X)$

Table 3.1: Conditions for Sufficient Dimensional Reduction [Cook, 2007]

As noted above, reduction is based on the assumptions in Equations 3.11 and 3.12 - particularly that the former is a realisation of n independent copies of the random vector $(X; Y)^T$, where the predictor matrix is defined as

$$X = \mu_x + D\beta\,\{\xi\,(Y) - \mu_\xi\} \tag{3.13}$$

where $\mu_x \in \mathbb{R}^p$; $\quad$ D $\in \mathbb{R}^{p\times d}$ with rank d and $\beta \in \mathbb{R}^{d\times r}$ with rank d. Note that the the function ϕ mapping of $\mathbb{R}$ onto

$\mathbb{R}^r$, i.e., $\phi : \mathbb{R} \to \mathbb{R}^r$ is a known vector-valued function fulfilling the conditions in Table 3.1. That is

$$\mu_\xi = \mathbf{E}\left[\xi\left(Y\right)\right] : \quad \epsilon \sim \mathcal{N}(0,\, \Delta) \tag{3.14}$$

in which Y is independent of ϵ and with the central sub-space of Δ^{-1}. For downscaling purposes, our focus here is on model identifiability, i.e., the fundamental property Equation 3.13 must satisfy in order for downscaling to be valid. In other words we are looking to establish a theoretical justification for learning the true values of the models underlying parameters after obtaining an infinite number of observations from it. It is about model reliability which should ensure that varying the parameters will generate different probability distributions of the observable variables.

Typically, the model is identifiable only under certain technical restrictions, in which case the set of these requirements is called the identification conditions. The function in Equation 3.13 computes its own estimates of the model parameters in Equation 3.14 by imposing constraints for identifiability. Thus, recognising that the two mean parameters, in this case, are $\bar{x} = \dfrac{\sum_{i=1}^{n} x_i}{n}$ and $\bar{\xi} = \dfrac{\sum_{i=1}^{n} \xi\left(y_i\right)}{n}$, we can now define a non-negative function

$$\hat{\Gamma} = \frac{\left\{\sum_{i=1}^{n} \xi\left(y_i\right) - \bar{\xi}\right\}\left\{\xi\left(y_i\right) - \bar{\xi}\right\}^T}{n} \tag{3.15}$$

and, given user-input weight matrix Ψ, we can define the fitted function as

$$\left(\hat{D}, \hat{\beta}\right) = \arg \min_{\Omega \in \mathbb{R}^{p\times d}, \Theta \in \mathbb{R}^{d\times r}} \sum_{i=1}^{n} \left[x_i - \bar{x} - \Omega\Theta\left\{\xi\left(y_i\right) - \bar{\xi}\right\}\right]^T \hat{\Psi}\left[x_i - \bar{x} - \Omega\Theta\left\{\xi\left(y_i\right) - \bar{\xi}\right\}\right] \tag{3.16}$$

subject to $\Omega^T\Psi\Omega$ being diagonal and $\Theta\hat{\Gamma}\Theta^T = \mathbf{I}$. The sufficient reduction estimate $\hat{\mathbb{R}} : \mathbb{R}^p \to \mathbb{R}^d$ is defined as

$$\hat{G}\left(x\right) = \left(\hat{D}^T\hat{\Psi}\hat{D}^T\right)^{-1}\hat{D}^T\hat{\Psi}\left(x - \hat{x}\right) \tag{3.17}$$

Software-driven implementation of reduction estimates takes different parameters. In the software package **R**, for instance, the Fourier basis sets the vector $y_i - \bar{\xi}$ takes a trigonometrical form. Other forms include the polynomial basis which sequentially raises the response variable to the power $j = 1, 2, 3, ..., k-1, k$, and the slice basis which divides the response variable into "slices" of pre-defined width and ultimately applying an indicator variable to determine the number of observations in each slice.

PFCs are obtained by performing PCA on the fitted sample covariance matrix $\hat{\Sigma}_{fit} = \frac{\hat{\mathbf{X}}^T\hat{\mathbf{X}}}{n}$ where $\hat{\mathbf{X}}$ is a matrix of fitted values of regressing $\mathbf{X}$ on a vector values function of the response y_i. Then, $\hat{\Phi}_1^T\mathbf{X}, \hat{\Phi}_2^T\mathbf{X}, \cdots, \hat{\Phi}_p^T\mathbf{X}$ are called **Principal Fitted Compnents** where $\hat{\Phi}_1, \hat{\Phi}_2, \cdots, \hat{\Phi}_p$ are the eigenvectors corresponding to eigenvalues $\hat{\lambda}_1^{fit}, \hat{\lambda}_2^{fit}, \cdots, \hat{\lambda}_p^{fit}$ of the covariance matrix $\hat{\Sigma}_{fit}$. Ordering the eigenvalues in descending order provides an insight into the amount of

variation in the original data each component accounts for. A few of these can be used in regression model instead of the original high dimensional predictors. Since PFCs are obtained with use of the response value we expect them to outperform PCs as regressors. A common practice is to choose the number of PFCs for use in regression modelling based on the magnitudes of their eigenvalues. Cook and Forzani [2008] proposed two approaches to making this choice - i.e., using the likelihood ratio test and Akaike's Information Criterion (AIC).

Downscaling an Air Quality Model by Principal Fitted Component Regression

We downscaled the REAM output and applied PFC analysis to reduce the dimension of the model outputs. We used polynomial basis function to compute the PFCs (having explored other basis functions, e.g. slice, when computing the PFCs and polynomial basis function seemed to give better results than the other methods). Then, we select few PFCs to use them as predictors in the regression model. For each station we regress hourly ozone observations on selected number of grid cell PFCs. The model is defined as

$$O_t = \delta_0 + \sum_{d=1}^{D} \delta_d P_d(t) + \epsilon_t \tag{3.18}$$

where, D is the number of PFC scores in the model, $D < p = 99$, $\delta_0, \delta_1, \cdots, \delta_d$ are model parameters, $P_1(t), P_2(t), \cdots, P_d(t)$ are PFC scores, $t = 1, 2, \cdots, n$, and ϵ_t is an error vector with mean 0 and a constant variance σ^2 $\left[\epsilon_t \sim \mathcal{N}\left(0, \sigma^2\right)\right]$. Todate, our published paper [Alkuwari et al., 2013] remains the only source for this type of application which makes its application for downscaling in this thesis fairly novel.

Our thresholding of the covariance matrix for PFC downscaling is based on the method proposed by Bickel and Levina [2008] with its main objective being to examine the effect of thresholding on the predictive ability of the PFC model. As the dimension of the data is significantly large, we would expect thresholding to add an improvement to the PFC model estimation, resulting in better prediction values. Further, we downscale the REAM model as described below using thresholded PFCs and we compare the results with the non-thresholded PFCs as well as with a thresholded PCA model. Details of the process are described below.

3.3.3 Downscaling via Regularized Covariance Data Matrix

If a p-variate population vector $\mathbf{X}$ has mean $\mu = (\mu_1, \mu_2, \cdots, \mu_p)$ then the covariance matrix of $\mathbf{X}$ is defined to be the square $p \times p$ symmetric matrix $\Sigma \equiv Cov(\mathbf{X}) = \mathbb{E}[(\mathbf{X} - \mu)(\mathbf{X} - \mu)^T]$, where the ij^{th} element in Σ is

$$\sigma_{ij} = \mathbb{E}[(X_i - \mu_i)(X_j - \mu_j)] \equiv Cov(X_i, X_j) \tag{3.19}$$

which is the covariance between X_i and $X_j, i \neq j$. In many multivariate statistical applications, understanding the eigenstrucuture (vectors and values) of the covariance matrix is fundamental as their magnitudes and dimensions determine the level of dimension reduction. Hence, it is essential that we use a good and well behaved estimator of the population covariance matrix so we can reach reliable results. The maximum likelihood estimator (MLE) is widely used to estimate the sample covariance. The MLE for a given sample covariance matrix, Σ_p, is defined as

$$\hat{\Sigma}_p = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{X}_i - \bar{\mathbf{X}})^T \tag{3.20}$$

Where $\mathbf{X}_1, \cdots, \mathbf{X}_n$ are i.i.d multivariate normal with mean $\mathbf{0}$ and covariance matrix Σ_p. The MLE estimator $\hat{\Sigma}_p$ is an unbiased estimator of the population covariance matrix Σ and under distributional assumptions it is considered to be a good representative of Σ as long as the dimension p is fixed and sample size n is large ($n \to \infty$). In other words, $\hat{\Sigma}_p$ tends to be unstable if p is large. Furthermore, the eigenvalues over disperse if p is large compared to n [Marcenko and Pastur [1967] and Johnstone [2001]] and the eigenvectors will not be consistent [Johnstone and Lu, 2004].

Covariance Matrix Thresholding

The estimate $\hat{\Sigma}_p$ converges to the population covariance Σ_p if p is fixed and n is significantly large in comparison to p. It is not uncommon to see cases where $p \geq n$ in which case we can expect to obtain unreliable estimates of Σ_p. Thus, typically, the resulting estimate $\hat{\Sigma}_p$ is not a good representative of Σ_p unless it is regularized in some way. We discuss and apply a regularizing technique based on covariance matrix thresholding method proposed by Bickel and Levina [2008]. The resulting thresholded matrix is simple with good theoretical properties and computationally inexpensive. The method adopted from Bickel and Levina [2008] is described as follows. For a square $p \times p$ matrix M, we define

$$\lambda_{max}(M) = \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p(M) = \lambda_{min}(M)$$

to be the eigenvalues of M and for any $s \leq \infty$ the thresholding operator is defined by:

$$T_s(M) = [m_{ij}\mathbf{1}\left(|m_{ij}| \geq s\right)] \tag{3.21}$$

Equation 3.21 is referred to as M *thesholded at* s. The thresholded matrix is invariant under permutations of the variable labels and is symmetric. Although the symmetry is preserved in the thresholded matrix, it does not necessarily preserve positive definiteness. To overcome the lack of positive definiteness in the thresholded matrix, the threshold value s must be chosen to satisfy the two following conditions [Bickel and Levina, 2008]

1. $\|T_s - T_0\| \leq \varepsilon$

2. $\lambda_{min}(M) > \varepsilon$

Bickel and Levina [2008] have proved theoretically that the thresholded matrix is consistent as long as $\frac{\log p}{n} \to 0$ which would guarantee positive definiteness with a probability converging to 1. Further, they show that the thresholded estimator is consistent in the operator norm over a pre-defined class of suitably sparse matrices which means that under pre-specified class of matrices, the thresholded estimator will converge to the matrix M if $\frac{\log p}{n} \to 0$. In the literature El Karoui [2008] presents arguments that convergence through the operator norm leads to convergence in the eigenvalues and eigenvectors. This makes the thresholded estimator particularly suitable for the application of PCA, implying that it would be a sensible choice to be used with PFCs as well. Detailed mechanics and theoretical proofs of the thresholded estimator properties can be found in Bickel and Levina [2008]. The choice of the threshold value s is rather arbitrary as long as it guarantees positive definiteness of the resulted estimator. However, Bickel and Levina [2008] proposed the cross-validation method below for selecting an appropriate threshold value.

1. Divide the sample randomly into two samples of sizes n_1 and n_2 respectively.
2. Choose $n_1 = n\left(1 - \frac{1}{\log n}\right)$ and $n_2 = \frac{n}{\log n}$.
3. Repeat the observations splitting N times.
4. Compute
$$\hat{R}(s) = \frac{1}{N}\sum_{\nu=1}^{N}\left\|T_s(\hat{\Sigma}_{1,\nu}) - \hat{\Sigma}_{2,\nu}\right\|_F^2 \tag{3.22}$$
Where $\|.\|_F$ is the Frobenius norm and $\hat{\Sigma}_{1,\nu}$ and $\hat{\Sigma}_{2,\nu}$ are the empirical covariance matrices of the n_1 and n_2 observations from the ν^{th} split.
5. Choose the threshold value s (for $s \geq \varepsilon_n \to 0, \varepsilon_n$ asymptotically equivalent to $\frac{n}{\log n}$) that minimizes $\hat{R}(s)$.

Downscaling by Principal Component Regression with a Thresholded Covariance Matrix

Since we are dealing with a high-dimensional dataset (99 grid cells), the resulted Σ might not be well behaved which makes it reasonable to carry out regularisation through thresholding as described below.

1. Compute the covariance matrix Σ of the gridded REAM outputs.
2. Threshold Σ using a selected threshold value and obtain the thresholded covariance matrix Σ_T.
3. Compute PC_T loadings, which are the PC loadings obtained using Σ_T.
4. For each station, regress the hourly ozone observations on selected number of grid cell PC_T scores using the

model

$$O_t = \alpha_0 + \sum_{m=1}^{M} \alpha_m Z_m(t) + \epsilon_t \tag{3.23}$$

where, M is the number of PC_T scores in the model, $M < p = 99$, $\alpha_0, \alpha_1, \cdots, \alpha_m$ are model parameters, $Z_1(t), Z_2(t), \cdots, Z_m(t)$ are PC$_T$ scores, $t = 1, 2, \cdots, n$, and ϵ_t is an error vector $\left[\epsilon_t \sim \mathcal{N}\left(0, \sigma^2\right)\right]$.

Downscaling by Principal Fitted Component Regression with a Thresholded Covariance Matrix

As explained earlier, PFCs make use of both the predictors and the response. Like in the PCA case, PFC loadings derive from the covariance matrix of the fitted values, $\hat{\Sigma}$, that results from the inverse regression of the predictors $\mathbf{X}$ on a function of the response value. PFC downscaling is therefore performed as outlined below

1. Compute $\hat{\Sigma}_{fit} = \frac{\hat{\mathbf{X}}^T\hat{\mathbf{X}}}{n}$ where $\hat{\mathbf{X}}$ is a matrix of fitted values of regressing $\mathbf{X}$ on the response y_i.
2. Threshold $\hat{\Sigma}_{fit}$ using a selected threshold value and obtains $\hat{\Sigma}_{fit,T}$.
3. Compute PFC$_T$ loadings, which are the PFC loadings obtained using $\hat{\Sigma}_{fit,T}$.
4. For each station regress hourly ozone observations on selected number of PFC$_T$ scores using the fitted model

$$O_t = \delta_0 + \sum_{d=1}^{D} \delta_d P_d(t) + \epsilon_t \tag{3.24}$$

Where, D is the number of PFC$_T$ scores in the model, $D < p = 99$, $\delta_0, \delta_1, \cdots, \delta_d$ are model parameters, $P_1(t), P_2(t), \cdots, P_d(t)$ are PFC$_T$ scores, $t = 1, 2, \cdots, n$, and ϵ_t is an error vector with mean 0 and a constant variance σ^2 $\left[\epsilon_t \sim \mathcal{N}\left(0, \sigma^2\right)\right]$.

3.3.4 PFC-based Ensemble Downscaling

Our contribution is bent on using this dimension reduction approach for downscaling. Our strategy is based on the foregoing discussions which provide a theoretical basis for expecting more reliable results. More specifically, by considering the response variable - as derived from actual measurements of observed periods and stations. It is reasonable to expect more reliable and effective downscaling results than in the case without a role for the response variable. Through comparative assessment, we also exhibit, in subsequent sections how downscaling ensembles using PFCs could produce more accurate predictions due to accounting for more information in the data attributes that cannot be captured by other downscaling methods. The strength of PFC ensemble downscaling relies on the following facts:

1. We are downscaling an ensemble of models, which means that we are taking into account the uncertainties of each ensemble member. In other words, we are maximising the potential of capturing variability.
2. We are applying a dimension reduction method to perform the downscaling. This means that we are maintaining all essential information about the models variability and discarding trivial information.
3. We are reducing the dimension with respect to the response. This indicates that we are considering the variability of the models (predictors) while preserving the information that lies within the response.

The stated facts imply that we can obtain forecasts that best represent the actual measurement at high spatial resolution. As, for air quality measurements, ensemble members produce gridded forecasts. This means that we have two levels of high dimensionality in air quality ensembles. The first level is the grid dimension in each model p and the second level is the number of models in the ensemble m. Therefore, the outputs of the ensemble is represented by $X_{m,t,p}$ - read as, the output of the ensemble member m at time t in grid cell p. So we have high dimensionality in both spatial and modular contexts. Our main objective, therefore, would be to be able to reduce the dimensions without causing major loss of spatial and modular information. One solution to this issue is to compute the mean gridded outputs across the ensemble members and carry out the PFC downscaling on the combined gridded outputs. Another, more sophisticated approach, is to perform a dimension reduction across both spatial and modular levels. We refer to this novel method as Double Dimension Reduction (DDR).

The approach has been used in other applications. For instance, Chiaromonte and Martinelli [2002] applied it to analyze global gene expression data. First, they reduced the dimension of gene expression using singular value decomposition (SVD) before performing further dimension reduction taking into account a response value derived from the Slice Inverse Regression (SIR) method. The method was also used by Li and Li [2004] to analyze microarrays of censored survival data. But to our knowledge the DDR approach has not been used elsewhere for downscaling air quality models. With DDR, we first reduce the dimension using a suitable dimension reduction technique (e.g PC or PFC) across space for each model in the ensemble to eliminate trivial spatial information. We select few leading spatial scores per model and reduce the dimension using an appropriate dimension reduction method across the models to eliminate modular redundancy. Next, we describe the mechanics of the DDR method based on different basis vector-valued functions of the response variable $Y \in \mathbb{R}$.

Statistical Ensemble Downscaling Using Double Dimension Reduction (DDR)

Let $X_{m,p|t}$ be the ensemble output of the model m on grid cell p at time t, where $m = 1, 2, \cdots, M$, (the total number of ensemble members); $p = 1, 2, \cdots, g$ (the number of grid cells in the forecasting domain) and time $t = 1, 2, \cdots, n$. The ensemble forecasts can be represented as a set of M output matrices, each of dimension $n \times p$. Our main purpose

is to reduce the current dimension of the X from $M \times n \times p$ to a lower dimensional set of forecasts while maintaining sufficient spatial and modular information. In other words, we have to reduce the dimension twice in order to eliminate both spatial and modular redundancy. The procedure runs through two steps as outlined below

1. **First Stage:** Eliminating Spatial Redundancy

   This step performs dimension reduction for each ensemble member individually in order to eliminate spatial redundancy within each model. Therefore, for each model the spatial dimension is reduced from p to k ($k < p$) using a suitable dimension reduction method. The resulted outputs will have a dimension $n \times K$, where $K = M \times k$.

2. **Second Stage:** Eliminating Model Redundancy

   After extracting sufficient spatial information through reduced space in each model, further dimension reduction is performed in order to eliminate modular redundancy. The justification for performing this step is that air quality ensembles are built based on several pre-specified initial parameters. Hence, while the difference in the initial conditions helps in assessing the uncertainties, not all the information provided by the parameterization are significant. Therefore, we select an appropriate dimension reduction technique to reduce the modular dimension from K to d, where $d < K$ and so the dimension of the resulting ensemble outputs now becomes $n \times d$.

Once the new lower dimensional ensemble output is obtained, we use regression methods to carry out the statistical downscaling. Figure 3.4 graphically illustrates the DDR method. We will perform the DDR using PCs and PFCs to downscale an air quality ensemble and compare the performance of PC-DDR *vs.* PFC-DDR.

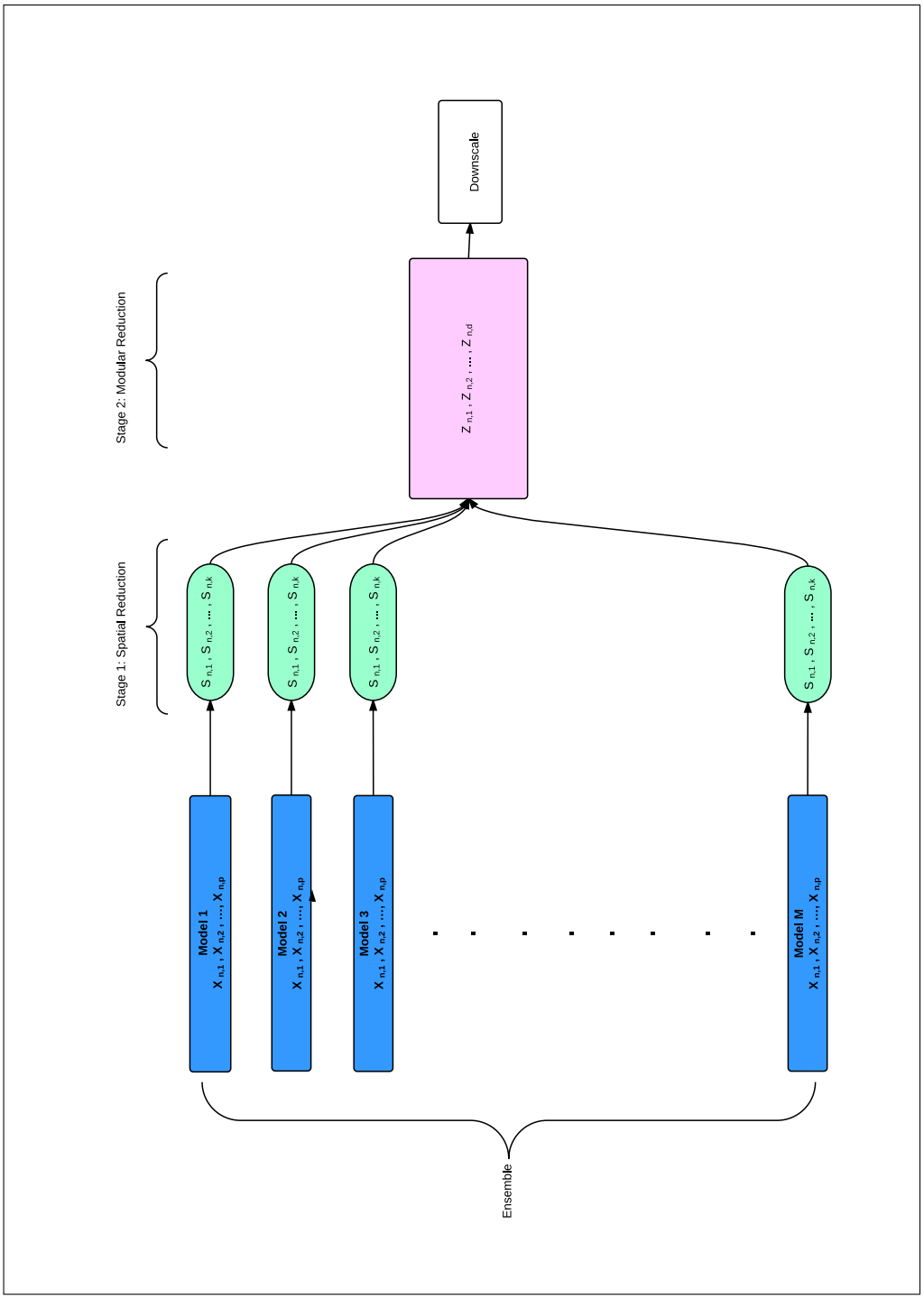

Figure 3.4: An illustrative plot of the DDR method

Statistical Downscaling of an Air Quality Ensemble using PCs

Dimension reduction using PCs helps in eliminating the dimension redundancy while maintaining the important structural variability in the original variables. Prior to performing the analysis we center the data as PCs should be obtained based on a covariance matrix formed from centered data. Let $\mathbf{X}_{1,p|t}, \mathbf{X}_{2,p|t}, \cdots, \mathbf{X}_{M,p|t}$ be the set of ensemble outputs of models $1, 2, \cdots, M$. Where $p = 1, 2, \cdots, g$ are the grid cells and $t = 1, 2, \cdots, n$ is the time points, hence each $\mathbf{X}$ has dimension $n \times p$. The DDR using PCs is performed as follows

1. Apply PCA on the outputs of each ensemble member separately.
2. Obtain the leading PC_1 loadings (the index 1 is used to identify the values computed in the first reduction stage) for each ensemble member and compute the corresponding PC_1-scores accordingly. This results in a lower dimensional new set of ensemble outputs per model which retains the maximum structural variability from the original forecasts and eliminates trivial spatial information.
3. Set $\mathbf{A}$ to be a matrix that has dimension $n \times k$ of all the PC_1-scores of all ensemble members obtained in step 2.
4. Perform PCA to reduce the dimension of $\mathbf{A}$.
5. Obtain the first d leading PC_2 loadings (the index 2 identifies the values computed at the second reduction stage) and compute the corresponding PC_2-scores. This results in a lower dimensional new set of combined ensemble outputs which retains the maximum structural variability from $\mathbf{A}$ and eliminates redundant modular information.
6. Perform the downscaling using multiple regression based on the model:

$$O_t = \alpha_0 + \sum_{d=1}^{D} \alpha_d Z_d(t) + \epsilon_t \tag{3.25}$$

Where O_t is the actual measurement of interest, D is the number of PCs_2 scores in the model, $\alpha_0, \alpha_1, \cdots, \alpha_d$ are model parameters, $Z_1(t), Z_2(t), \cdots, Z_d(t)$ are PC_2-scores, $t = 1, 2, \cdots, n$, and ϵ_t is an error vector $\left[\epsilon_t \sim \mathcal{N}\left(0, \sigma^2\right)\right]$.

Statistical Downscaling of an Air Quality Ensemble using PFCs

We noted earlier that, when it comes to dimension reduction, PFCs have an upper hand over PCs, as the mechanics of the former are tailored to function with respect to the response variable while those of the former are not. This indicates that PFCs are more suitable when dimension reduction is required in regression since the reduced data will capture the interdependence between the response and the predictor variables. Similar to PCs, the data needs to be centred before carrying out with the PFC analysis. Let $\mathbf{X}_{1,p|t}, \mathbf{X}_{2,p|t}, \cdots, \mathbf{X}_{M,p|t}$ be the set of ensemble outputs of models $1, 2, \cdots, M$ where $p = 1, 2, \cdots, g$ are the grid cells and $t = 1, 2, \cdots, n$ are the time points. Each of the

resulting data matrix **X** has dimension $n \times p$. The DDR using PFCs is performed as outlined below

1. Apply PFC analysis (after choosing the appropriate basis function) on the outputs of each ensemble member separately. Parameter tuning is possible here - including experimenting with various basis functions.
2. Obtain the leading PFC_1 loadings (the index 1 is used to identify values computed at the first reduction stage) for each ensemble member and compute the corresponding PFC_1-scores accordingly. This results in a lower dimensional new set of ensemble outputs that captures the spatial structural variability from the original forecasts and discards redundant spatial information while taking the actual measurement (the response) into account.
3. Set **B** to be a matrix that has dimension $n \times k$ of the all PFC_1-scores obtained in step 2.
4. Perform PFC analysis to reduce the dimension of **B**.
5. Obtain the first d leading PFC_2 loadings (the index 2 is used to identify the values computed in the second reduction stage) and compute the corresponding PFC_2-scores. Thus, the selected PFC_2-scores are a new set of combined and lower dimensional ensemble forecasts that retains both spatial and modular structural variability while taking actual measurement (the response) into account.
6. Preform the downscaling using multiple regression where the model is given as

$$O_t = \delta_0 + \sum_{d=1}^{D} \delta_d P_d(t) + \epsilon_t \tag{3.26}$$

Where O_t is the actual measurement of interest, D is the number of PFCs_2 scores in the model, $\delta_0, \delta_1, \cdots, \delta_d$ are model parameters, $P_1(t), P_2(t), \cdots, P_d(t)$ are PFC_2-scores, $t = 1, 2, \cdots, n$, and ϵ_t is an error vector $\left[\epsilon_t \sim \mathcal{N}\left(0, \sigma^2\right)\right]$.

3.4 Comparability and Assessment of Downscaling Models

As stated in Section 1.4, the thesis presents a novel downscaling method to enhance air pollutants predictions. Thus, in addition to proving its robustness - accuracy and consistency, its superiority is to be validated through a comparative analysis of its performance with that of alternative downscaling techniques in use. In the next exposition we present brief discussions and mechanics of some of these methods. It particularly highlights some of the key issues relating to model performance and outlines technical justifications deriving from potential gaps that ought to be filled within those methods.

Model performance plays a central role in inferential statistics and the quest for attaining robust - accurate and consistent results has never been higher than it is in the modern era of Big Data. Having closely examined the downscaling models applied in this work, a comparative analysis is carried out based on a number of model features, not least performance. This section outlines the approach adopted in comparing downscaling models against observations. It

adopts Derwent et al. [2010] position that for any model to influence policy-makers, it must be able to reproduce real-world behaviour. To fulfil this condition, the proposed model will be compared to alternative models' performance on the same data. In addition to that, the proposed model will be used in predicting air quality at unobserved stations.

3.4.1 Some Fundamental Questions

While comparison of models plays a crucial role in predicting air quality, there are fundamental questions which we are going to try and address within the scope of our research question and objectives. In particular, how should the comparison be carried out? How should inconsistencies be addressed - for instance, what conclusions should be drawn in cases of partial agreement - i.e., good performance in one model setting against poor performance elsewhere? How do we assess model accuracy and reliability? How do we account for the effect of data variability and what volume of the environment do monitoring site observations represent? Comparing models with observations requires that we provide a framework for comparison. Typical air quality model applications rely on historical data as well as monitored data streams from networks - hence some authors refer to the process as "history matching" [Derwent et al., 2010]. Air quality scientists have generally found "history matching" quite challenging - see, for instance, Oreskes et al. [1994] and Beck [2002]. Like in many other modelling applications, finding the relevant data attributes and parameters for model optimisation has been the main issue of concern. Specifically in environmental science, gaining access to sufficiently accurate historic emissions and meteorological data has never always been straightforward.

3.4.2 A Bayesian Approach to Comparability of Models

In many applications, including environmental science, it is common practice to use part of the data for model calibration and tuning and use the calibrated model, with no further adjustment, on test data. One of the challenges environmental scientists have faced over the years is that often these calibration and tuning steps have been carried out elsewhere and the data may no longer be accessible. As a result, many model inter-comparisons would omit the first stage and go straight to the second - see, for instance, Van Loon et al. [2007*a*]. The adopted practice in the environmental scientific community has been to reach an agreement between the modelling groups concerning pollution episodes or time periods (from a month to a year) that are to be studied and from that harmonise historic emissions data and collect observations from network databases. Any discrepancies would usually be resolved by re-running the models. Another issues that has intrigued environmental scientists is the performance evaluation of of Air Quality Models. In particular, they emphasise the differences between model evaluation and verification or validation [Oreskes et al., 1994] - with verification being attributed to model "truthfulness", implying that the model is "reliable" and can therefore form a basis for policy and decision making. On the other side, validation would imply that the model pre-

dictions are consistent with the data and that it is an accurate representation of the physical reality.

PFC predictions are notionally based on a maximum of $n \times p$ data where $p = 99$ are the number of extracted components readings. By discretising the actual and the predictions we can infer the potential number of categories of ozone concentration levels is $j = 2$ and the overall misclassification error can be computed as the sum of the weighted probabilities of observing ozone data belonging to one of the ozone categories given that we are not in that class - which is analogous to observing high ozone levels at a global level which are not representative of regional conditions. The error formula is

$$\psi^{\text{model}} = \sum_{j=1}^{k} \sum_{i=1}^{n} = p\left(c_j = y_i\right) p\left(x_i \in c_j | y_i \notin c_j\right) \propto \frac{\pi_j f_j\left(x\right)}{\sum_{j=1}^{k} \pi_j f_j\left(x\right)} \tag{3.27}$$

where π_j denotes the prior likelihood for belonging to the j^{th} category of ozone concentrations and $f_j\left(x\right)$ is the distribution density of the data in that category. Data and model-dependent variation is common problem in predictive modelling which we tackle by a combination of techniques including using ROC curves [Egan, 1975] to select an optimal model. A graphical illustration of a binary problem scenario and ROC plot is given in Figure 3.5. According to this technique, a classifier is optimal only if it yields results in the top left corner of the plot. Its rationale derives from the following conditions and scenarios. Assume that a predictive model yields four possible outcomes true positive (n_{tp}), false positive (n_{fp}), true negative (n_{tn}) and false negative (n_{fn}), the ROC accuracy and error are

$$\textbf{Accuracy} = p\left(c_j|x\right) = \frac{n_{tp} + n_{tn}}{n_{tp} + n_{tn} + n_{fp} + n_{fn}} \Leftrightarrow 1 - p\left(c_j|x\right) = \textbf{Misclassification Error} \tag{3.28}$$

Equation 3.28 calculates the probability of predicting an observation to belong to its true class. In a two-class scenario where $c_k = \{y_1, y_2\}$, this probability is equivalent to $p\left(c_k|x\right) = p\left(y_k \in c_k\right) = \sum_{j=1}^{k} p\left(y_i\right) \int p\left(x|y_i\right) dx$ where the integral is over both classes. Krzanowski and Hand [2009] demonstrate various ways in which ROC curves can be used as performance measures by focusing on *inter-alia* statistical tests for ROC curves and their summary statistics.

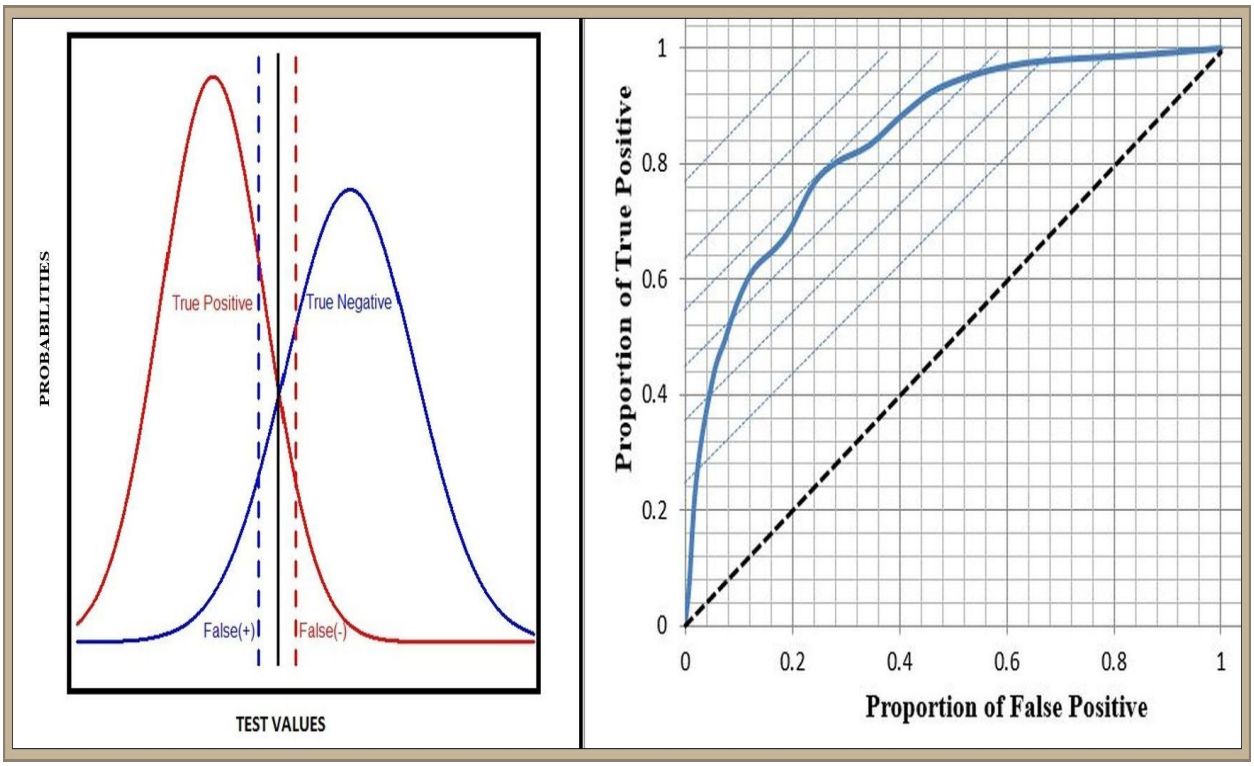

Figure 3.5: An hypothetical binary problem scenario (LHS) and the corresponding optimising ROC plot

By shifting the diagonal line into the north-western direction; fixing the south-western and moving the north-eastern tip anti-clockwise or fixing the latter and moving the former clock-wise you create iso-performance lines [Provost and Fawcett, 2001] which form the assessment base. The diagonal line can be viewed as the strategy of randomly guessing an unobserved station has **high** or **low** ozone levels as an increased **sensitivity** decreases the likelihood of **false negatives** while increasing **specificity** decreases chances of **false positives.** The consequences are quite analogous to **Type-I** and **Type-II** errors in hypothesis testing. By learning rules from the data the classifier makes intelligent guesses and, according to this technique, a classifier is optimal only if it yields results in the top left corner of the plot. One of most attractive properties of ROC curves is their insensitivity to changes in class distribution.

The downscaling analogy to Equations 3.27, 3.28 and the plots in Figure 3.5 can be conceptualised by discretising the PFC downscaling model predictions which, without loss of generality, we can assume a binary scenario representing the global and regional conditions as illustrated by the bi-modal distribution in 3.5. This scenario describes the proportions of true positives, true negatives, false positives and false negatives with the solid vertical line representing the **Bayesian** minimum error. Various approaches for minimising this error have been proposed (Mwitondi et al. [2002] and Freund and Schapire [1997]) but a commonly acceptable practice is to vary the allocation rule to meet specific requirements of an application. Separating the two modes requires maximising the allocation rule in Equation 3.27.

The allocation rule in 3.27 can be implemented via numerous domain-partitioning models, raising, *inter-alia*, model

performance and selection issues. Hence, to assess PFC performance, we test the model on new, previously unseen data and we compare its performance with available alternative predictive techniques. ROC curves are fitted based on the information obtained from repeated searches using different multiple model parameters and recording the attained performances associated with multiple iso-performance (tangent) lines as described below.

Algorithm 1 Multiple Samples/Models Comparability Algorithm

procedure SAMPLE, FIT, ASSESS, REPEAT
  Top Level 1: Set Training and Testing Sample Sizes
  Initialise $i = 1,\ \kappa$, **sample size** s **and** $\Theta_\kappa = c\,(.)\,[$ **a** $\kappa -$ **long parameters holding vector**$]$
  For $\quad i \leq \kappa \quad Do$
  While $\quad s \leq n \quad Do$
  Sampling $\leftarrow$ Obtain Samples $\quad \mathcal{S}_{s=1,2,3,\ldots,S\leq n} \in X$
  Fit initially large, over-fitting models $\quad \hat{\mathcal{L}}_{g\geq 2}$
  Update $\Theta_{s+1} := \Theta_s \left\{\pi_s, \psi^{\text{model}}, f_s\,(.)\right\} \leftarrow$ **Extract key parameters (*priors*, densities, *posteriors*, cost) from** $\quad \hat{\mathcal{L}}_s$
  Cross-validate - plot the specificity and sensitivity for $\hat{\mathcal{L}}_s$
  Select η_s **best-performing models, where** $\eta_s \subseteq g\left\{\hat{L}_{\eta_s}\right\}$ **based on Cross-validation results**
  Select Model that Maximises *posteriors* $\quad \hat{\mathcal{L}}_{s.g}$
  End While
  Update $\Theta_{s.(\kappa+1)} := \Theta_s.\kappa \left\{\pi_{s.\kappa}, \psi^{\text{model}}, f_{s.\kappa}\,(.)\right\} \leftarrow$ **Extract key parameters (*priors*, densities, *posteriors*, cost) from** $\quad \hat{\mathcal{L}}_\kappa$
  Top Level 2: Selecting Optimal from Many Competing Models
  Select $\eta_{s.\kappa}$ **best-performing models, where** $\eta_{s.\kappa} \subseteq g\left\{\hat{L}_{\eta_{s.\kappa}}\right\}$ **based on Cross-validation Results**
  Select Model that Maximises *posteriors* $\quad \hat{\mathcal{L}}_{s.\kappa.g}$
  End For
end procedure

The algorithm fits initially large, over-fitting models and based on cross-validation based assessment, drops worst-performing models. The optimal balance between accuracy and reliability is decided based on performance maximisation for each model obtained by updating *priors* by *posteriors*. Selection criteria may vary, this work searches for the iso-performance (tangent) line yielding the best balance. That is, assuming constant loss, the algorithm searches for the optimal location using iso-performance lines and extract the consensus level between the models. Model improvement parameters are tested for significance with respect to accuracy and reliability and the algorithm stops if a set criterion is achieved. Otherwise the model parameters are repeatedly computed across samples and iterations.

The algorithm summarises the proposed statistical framework for dealing with variability and model performance assessment. It addresses variability through repeated samples and multiple learning models on the data in Section 3.1 as well as demonstrating how rules can be learnt from training data and applied to new, previously unseen, test data. The second issue is addressed via model assessment techniques which address both accuracy and reliability. The final part of the algorithm seeks to obtain optimal models based on the information obtained from repeated searches using transformed plots. Multiple models are plotted and from them we can generate multiple tangent lines, iso-performance [Provost and Fawcett, 2001], optimising or otherwise. Extracted parameters from the plots can be used as inputs in

repeated training. The relationships between different slopes corresponding to the different model versions may guide us choose a slope or slopes in different points which we can use to adapt the model architecture and so on. The algorithm also contains the Bayesian-like iteration for improving both the graphics and the model. It is also possible to explore the impact of covariates on the ROC curves by examining the way they inter-cross. Chapter 4 presents our exploratory, comparative and final results alongside discussions pertaining to our research question and objectives.

Chapter 4: Results

This chapter presents the overall results from the analyses carried out throughout this work. It is designed to reflect the research question and the objectives hence its structure. Understanding the distributional behaviour of the data used in the analyses was considered fundamental to the planned types of analyses. For instance, discretising the actual ozone levels to produce categories creates notional *prior* and *posterior* probability of memberships to these categories. The number of categories is not decided arbitrarily and one way to determine it is to look at the distributional patterns based on different smoothing parameters, say. The chapter presents results from four main routes - an overall Exploratory Data Analysis (EDA) and Downscaling; Downscaling REAM Model Ground Level Ozone Outputs, Downscaling an Air Quality Model with Regularized Covariance Data Matrix and Downscaling an Ensemble Air Quality Models.

4.1 Exploratory Data Analysis (EDA), Simulation and Downscaling

Results for downscaling an Air Quality Model are obtained via simple simulations to help gain further understanding of the features of PCs and PFCs in the context of our research question and objectives. The simulation follows Cook [2007] in which the random variables generated are seasonal, which mimics our situation where a diurnal cycle is present (and removed). Thus, we generate an $n-$ sized and centred normal random variable Y with mean 0 and variance σ_Y^2. We add a seasonal pattern to the generated random variable and, for simplicity, a seasonality component of repetitive cycles of 10 values: $1, 2, 3, \cdots, 10$ is added. We further generate an $n \times p$ matrix, defined by the inverse model in Equation 4.1

$$\mathbf{X} = \boldsymbol{\Gamma y} + \boldsymbol{\sigma\varepsilon} \tag{4.1}$$

where $\boldsymbol{\Gamma}$=$\boldsymbol{(1, 0, \cdots, 0)^T}$ and $\boldsymbol{\sigma > 0}$, and ε is a standard normal random variable. With 100 predictors, we perform the simulation with sample sizes $\boldsymbol{n = 200,} \quad \boldsymbol{500} \quad$ and $\quad \boldsymbol{1000}$ based on the forward regression model

$$\boldsymbol{Y = \alpha_0 + \alpha^T x + \sigma_{Y|X}\varepsilon}$$

where $\mathbf{x}$ is the observed value of $\mathbf{X}$, $\boldsymbol{\sigma_{Y|X}}$ is constant, and ε is a standard normal random variable. Finally, we apply three different approaches to model the simulated data. For a straightforward comparison, amongst all methods and

as in Cook [2007], we restrict ourselves to the dimension of the reduced space, i.e. the number of PCs and PFCs to be included in the regression model is 1, i,e., $\boldsymbol{d} = \mathbf{1}$. We fit the PC model with one PC and the PFC model with one PFC. For each simulated dataset we use the first 80% as a fitting (training) period and the remaining 20% as a validation period. These proportions are not rigid and, given the data size, they can be apportioned differently without significantly affecting results. Table 1 summarizes the results based on 100 replications and, as it can be seen here, the PFC model seems to show a better predictive ability than the other models at all sample sizes.

Sample Size	linear regression	PC regression	PFC regression
200	1.02	0.79	0.67
500	0.78	0.76	0.71
1000	0.74	0.73	0.71

Table 4.1: Simulation Results: RMSEs averaged over 100 replications for linear, PC, and PFC regression. The PC model was fitted with one PC and the PFC model was fitted with one polynomial PFC.

4.2 Downscaling REAM Model Ground Level Ozone Outputs

The geographical loadings of the corresponding PC or PFC can provide great insights into the global and local understanding. These insights come via interpretations of spatial EOFs and F-EOFs plots, basically, display the locations at which PCs and PFCs contribute more strongly or weakly. The EOFs and the F-EOFs are associated with eigenspaces of dimension one. Figure 4.1 shows the leading EOFs. The EOF distributions reflect the general variation of ozone, the value of which is higher over emission regions and over land than over ocean. EOF1 illustrates, mainly, the ozone gradient decreasing towards the eastern coastline while EOF4 shows the ozone gradient decreasing towards the southern coastline, reflecting generally much lower ozone concentrations over the ocean than on land. On the other hand, EOF2 and EOF3 are associated with regional ozone distribution patterns, in which the variations are lower in Alabama and northern Georgia, and Mississippi and Tennessee, respectively. These spatial patterns are likely driven by meteorological systems that transport low ozone air masses to these regions.

Figures 4.2 and 4.3 show the first F-EOFs of REAM outputs corresponding to eight selected stations over the period 6-25 June 2005. Notice that stations 35 and 60 are generally associated with the west to east gradient of low ozone in the eastern coastline, which is somewhat similar to the EOF1 distribution in Figure 4.1. The F-EOFs of Stations 75 and 5 have a mixture of EOF1 and EOF4 distributions, both showing lower ozone variation in the eastern and southern coastlines. The F-EOF of Station 51 has some resemblance to that of EOF2, showing lower ozone variations in Mississippi and Tennessee respectively. Station 66 exhibits an F-EOF that is similar to that of EOF3 - that is, showing low variation in Alabama and northern Georgia, although this feature extends further to the eastern part of South Carolina. The F-EOF of Stations 96 and 83 are more complicated, none of which is a clear extension of the 4 EOFs.

The uniqueness of the first F-EOFs implies that the PFC method is able to more efficiently reduce the dimension of the problem and capture the regional distribution pattern specifically relevant to the site of interest.

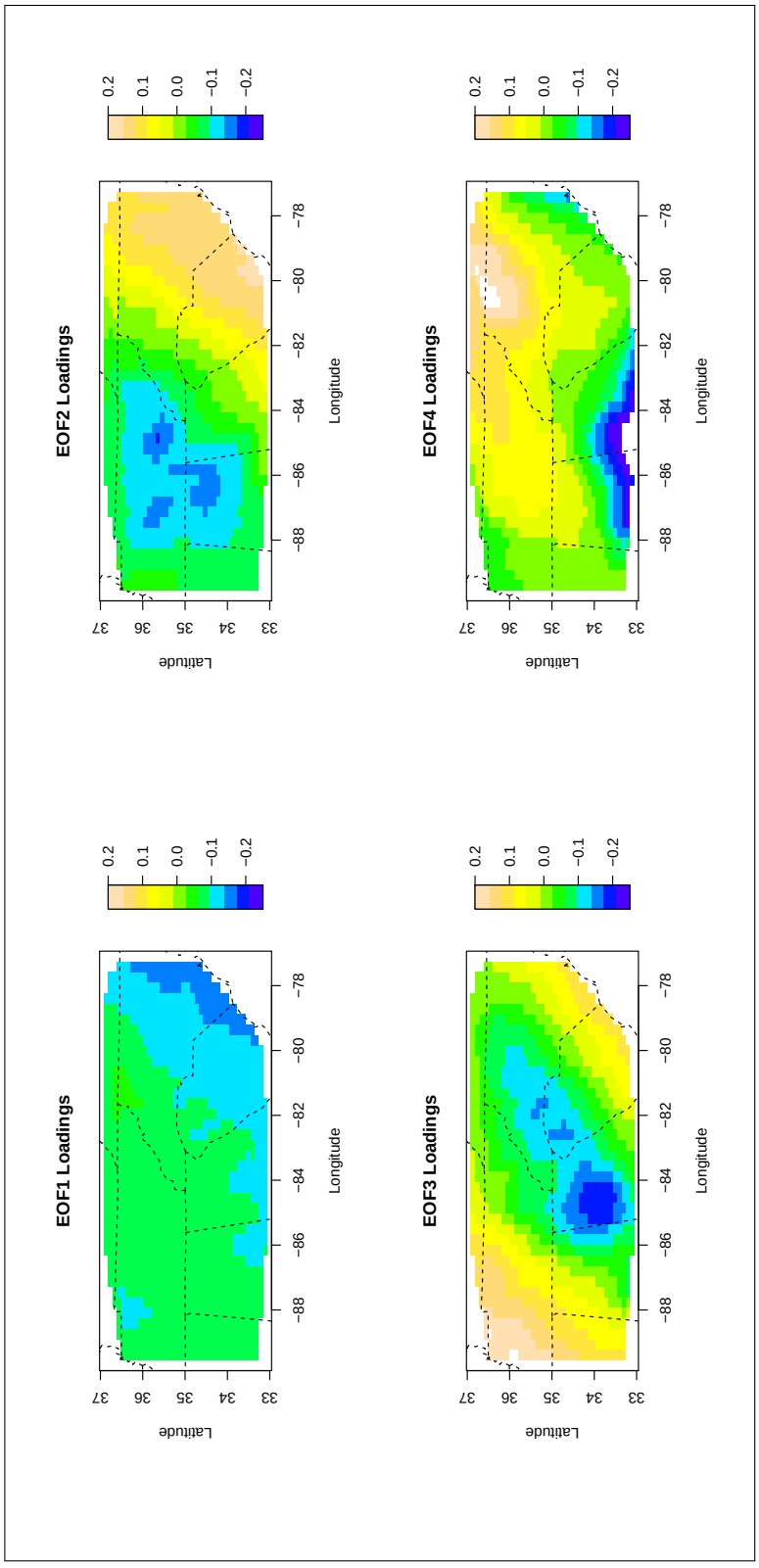

Figure 4.1: The first four leading EOFs of the gridded REAM output from 6 June to 25 June 2005.

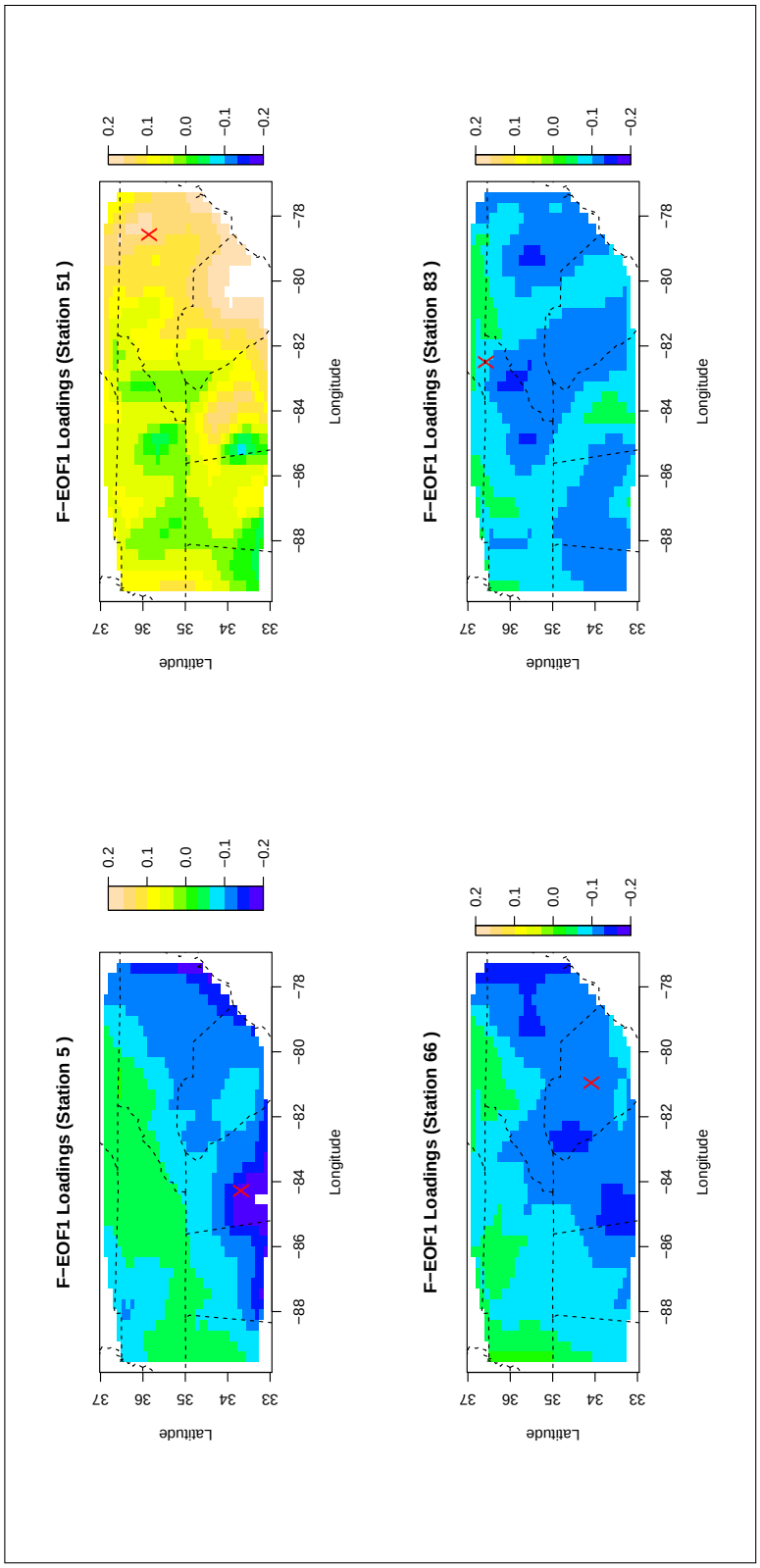

Figure 4.2: The first F-EOFs (polynomial basis function with degree 1) of the REAM outputs for four stations, estimated over 6-25 June 2005. The location of the station is marked by '×'.

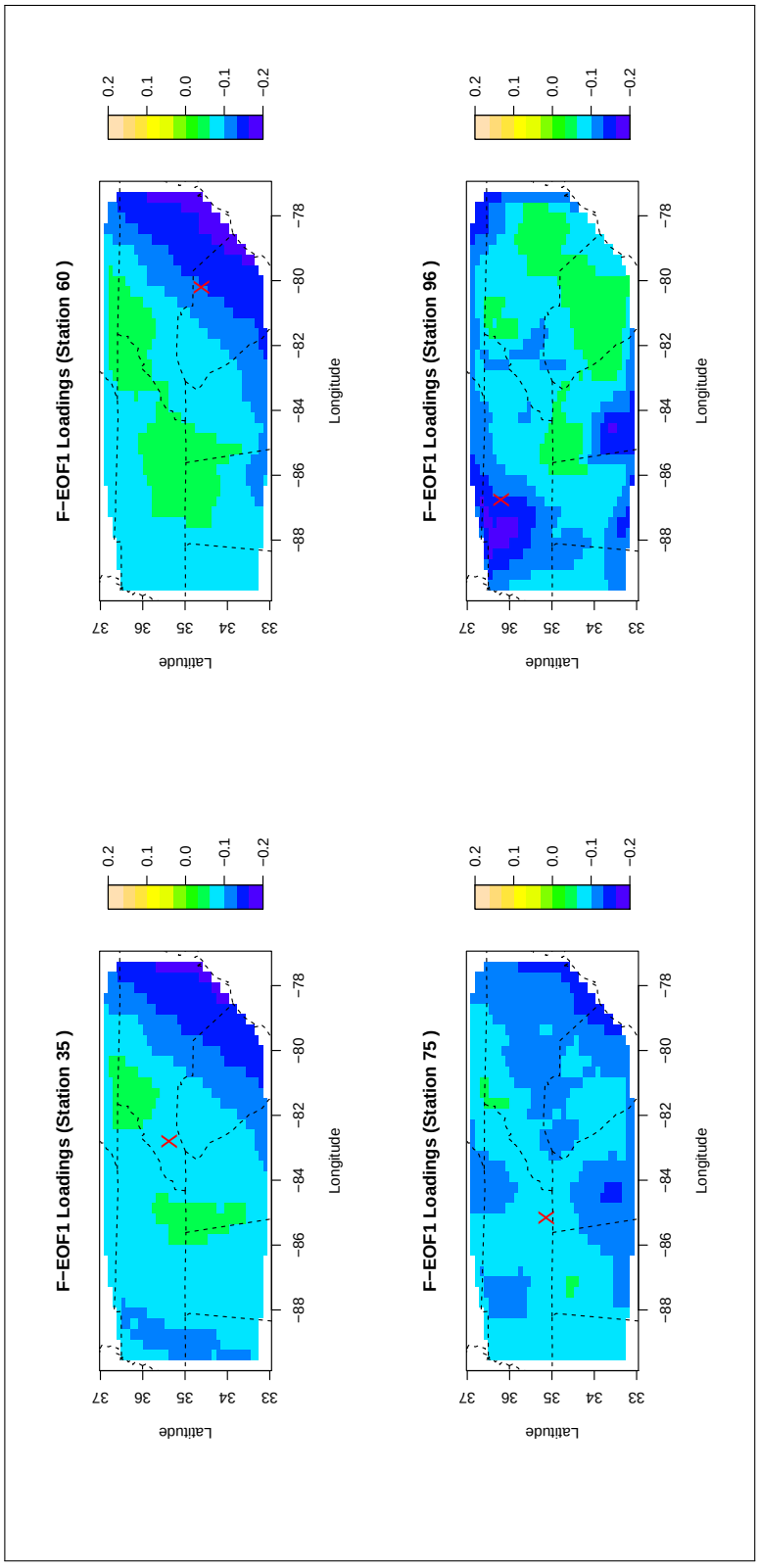

Figure 4.3: The first F-EOFs (polynomial basis function with degree 1) of the REAM outputs for four stations, estimated over 6-25 June 2005. The location of the station is marked by '×'.

4.2.1 PC and PFC Performance on Simulations

According to the simulation results in section 4.1, PFCs performs comparatively better when the sample size is small compared to the dimension of predictors. This is the situation for the ozone data at hand, as the sample size (fitting period) is relatively small compared to the number of grid cells. Using the three approaches discussed in the earlier sections, we use the period from 6 to 25 June as a training period and the period from 26 to 30 June as validation period. Although the data were converted to the square root scale to adjust the skewness, the plots and tables presented in this section show the results after converting them back to the original scale. The number of PCs selected in the regression model was determined after performing the standard and well-documented leave one out cross validation (see, for instance, Mertens et al. [1995]). Thus, we performed cross-validation for each stations individually and limited the number of PCs in the regression model to a maximum of 20 to avoid over-fitting - a phenomenon that typically arises when the fitting (training) and validation errors start deviating apart, with the former going down while the latter is rising. The cross-validation results show that each station should be fitted with a different number of PCs. For example, for some stations using only one PC in the regression model seems to be enough, while for other stations all 20 PCs should be used as predictors to obtain significant results. Hence, according to the cross validation results a PC regression model with different number of PCs have been fitted for each station. Figure 4.4 shows a summary of the number stations (i.e regression models) versus the number of PCs needed to fit the regression model. The plot shows that for most of the stations in the study area using only one PC in the regression model seems to be significant.

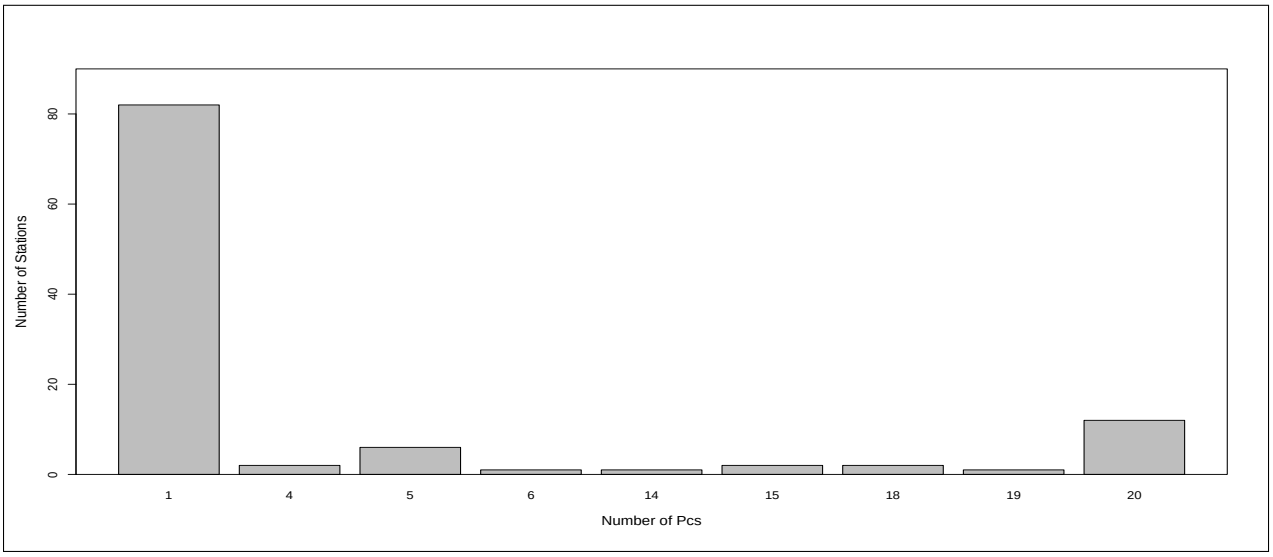

Figure 4.4: The bar chart shows a summary of the number of stations versus the number of PCs needed to fit the regression model for the station. The number of PCs were determined using the PRESS cross validation method by Mertens et al. [1995]

We fit the PFC model using one PFC (PFCs were obtained using a polynomial basis function of degree one). Table 4.2 shows the RMSEs for some stations within the study region and the average RMSE. The average RMSE indicate that overall, PFC regression outperformed both PC and simple regression methods. The PFC model has significantly improved the predictive ability relative to the REAM model (the ozone prediction error has been reduced by 52% relative to the REAM model predictions). Using PFCs improved ozone predictions by approximately 10% relative to the simple regression model, but (on average) it showed a 3% improvement relative to the PC model. Although RMSE values indicate that the PFC model has better predictive ability than the other approaches, they show that PFCs did not perform best in 36 stations compared to the linear and PC regression, which approximately accounts for 38% of the stations in the study region (these stations are marked in red 'x' in Figure 3.1). In 18 out of these 36 stations the PC model seems to outperform the other methods. Simple regression appears to perform worst on average, which reinforces our view that regional variations ought to be taken into account.

Regression Model	All Stations	Station 29	Station 84	Station 107
REAM	18.98	15.10	26.95	27.40
Linear	10.01	11.57	13.33	10.08
PC	9.35	11.31	10.27	9.89
PFC	9.13	9.86	10.36	8.91
Linear with AR(2) errors	10.66	12.77	12.33	9.17
PC with AR(2) errors	9.61	11.17	10.27	9.53
PFC with AR(2) errors	9.39	10.14	10.36	8.76

Table 4.2: RMSEs: training period is 6-25 June, validation period is 26-30 June. The PC models were fitted for: station 29 with 18 PCs, station 84 with 5 PCs, and station 107 with 20 PCs. The PFC models were fitted with 1 PFC (polynomial basis function with degree one)

Figures 4.5 and 4.6 display ozone observations and the corresponding REAM outputs, simple regression forecasts, PC regression forecasts, and PFC regression forecasts over the period from 26 to 30 June for some selected stations. The plots indicate that for the selected prediction period, the PFC model produced forecasts that are very close to the actual ozone concentrations at most times of the day.

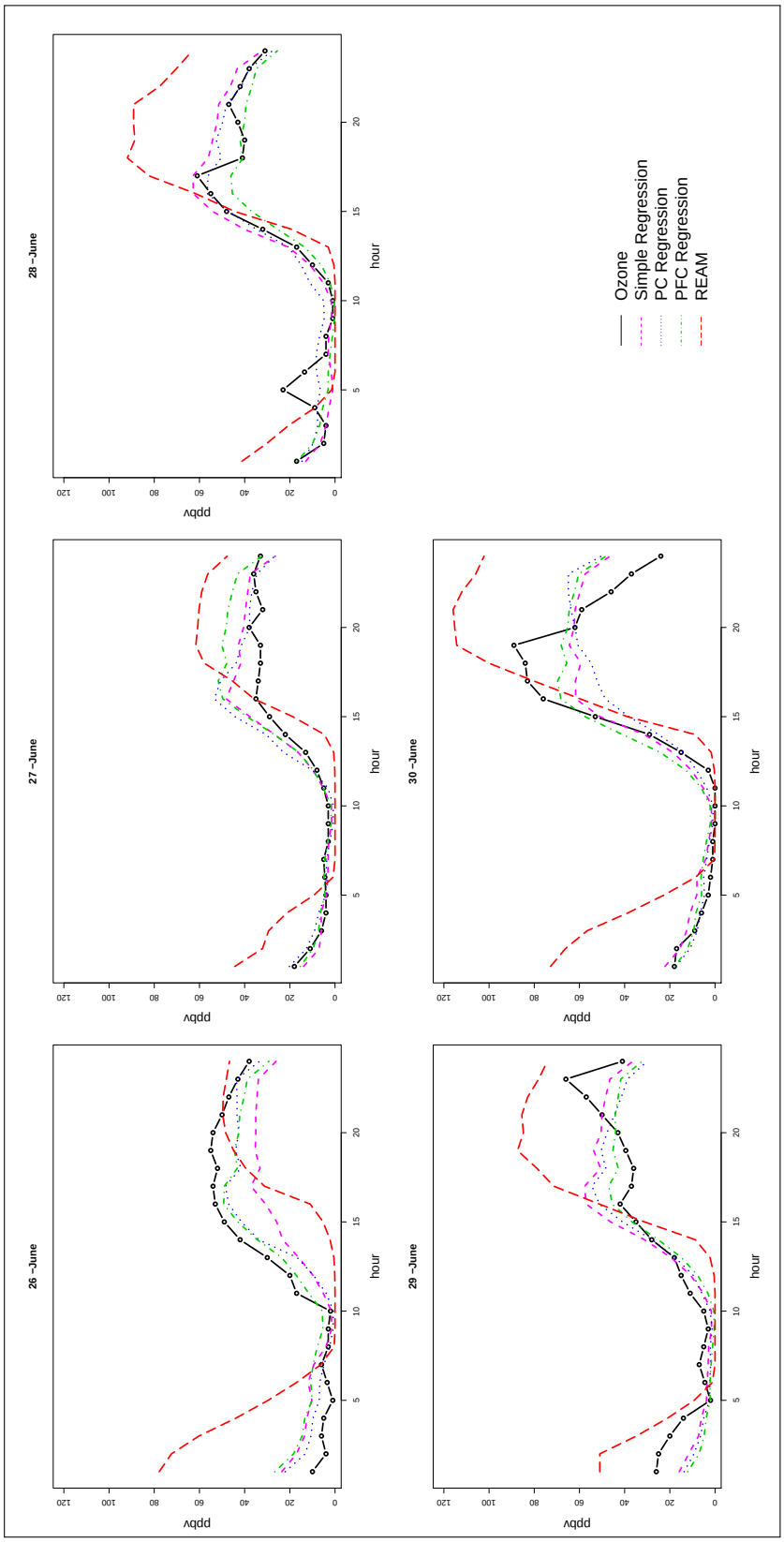

Figure 4.5: Prediction plots for station 107 (26-30 June). Observations (black line), REAM outputs (dashed red line), linear regression predictions (dashed pink line), PC predictions (20 PCs, dotted blue line), and PFC predictions (polynomial basis function with degree one, dashed green line).

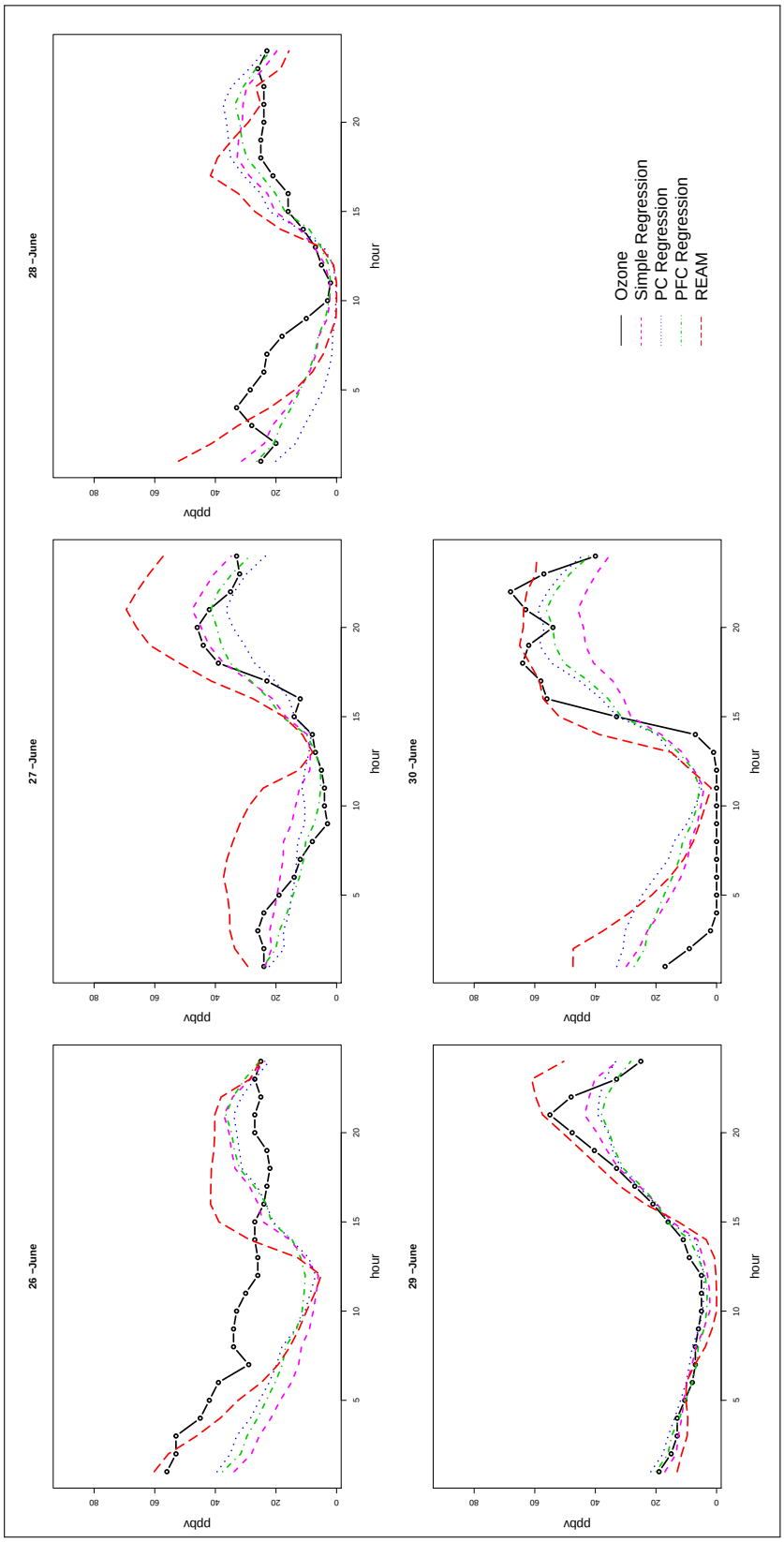

Figure 4.6: Prediction plots for station 29 (26-30 June). Observations (black line), REAM outputs (dashed red line), linear regression predictions (dashed pink line), PC predictions (20 PCs, dotted blue line), and PFC predictions (polynomial basis function with degree one, dashed green line).

The nature of our data suggests that there might be a great possibility that the model errors could be correlated. One way of verifying this, is to plot the autocorrelation function (ACF) and the partial autocorrelation function (PACF) of the model residuals. The ACF and the PACF plots (not shown here) for the residuals of the simple regression, PC regression, and PFC regression indicated that the models errors were autocorrelated and so they needed to be modelled as an autoregressive model (AR). Note that the foregoing errors structure is actually due to the different short-term behaviours in time of observations and REAM outputs, and was already modelled in Guillas et al. [2008]. The ACF and PACF plots suggest that modelling the residuals using a second order autoregressive model AR(2) seems to be a sensible choice overall. We refitted the regression models presented in this paper with an AR(2) model for the error. Table 4.2 shows the RMSEs averaged over the stations within the study region and for three randomly selected stations. We would expect that the RMSE values to be smaller for the models with and AR errors. However, the results show that modelling the errors with an AR(2) models did not improve the models predictions in general.

To investigate the change in prediction errors when using a longer fitting period, we repeated the analysis using the period from 6 June to 10 July as a fitting (training) period and used the proceeding five days as validation period. Table 4.3 shows the RMSE values averaged over all stations in the study region and for selected stations. The results indicate that the PFC model improved the predictive performance by 45% compared to the REAM model. Moreover, using PFCs improved ozone predictions by 2% relative to the simple regression model and by 4% compared to the PC model. On average, the PC model seems to perform the worst in this case.

It is important to note that although predictions errors seem to be smaller when using a longer fitting period, the PFC model does not seem to have a significant prediction improvement compared to the other downscaling methods. This coincides with the simulation results when we used a relatively large sample size.

Regression Model	All Stations	Station 29	Station 84	Station 107
REAM	15.84	11.46	17.61	18.07
Linear	8.87	9.66	7.18	8.06
PC	9.04	11.56	6.45	7.82
PFC	8.68	9.27	6.54	7.82

Table 4.3: RMSEs: training period is 6 June to 10 July, validation period is 11 to 15 July. The PC models were fitted for: station 29 with 18 PCs, station 84 with 5 PCs, and station 107 with 20 PCs. The PFC models were fitted with 1 PFC (polynomial basis function with degree one)

4.2.2 Further Validation of Performance

To further verify the consistency of our findings, we repeated the analysis and model fitting for different fitting and validation periods. We selected a fitting period of size $\boldsymbol{n}$ days and predict for the next $\boldsymbol{k}$ days. Then, we move the $\boldsymbol{n}$ day fitting period one day ahead and predict for the next $\boldsymbol{k}$ days and so on. The data are available from 2 June 2005 to 31 August 2005. Within 2 June to 26 August, we select a fitting period of $\boldsymbol{n = 20}$ consecutive days (i.e. 480 data

points) and we allocate the following $\boldsymbol{k} = \mathbf{5}$ days (i.e. 120 data points) to the validation period. Then we compute the RMSE for each set of predictions. This computation is repeated 60 times, as we move the fitting period by one day a head and proceed with the prediction for the corresponding validation period.

Table 4.4 shows the RMSEs (averaged over the 60 runs) for stations 29, 80, and 107. For station 29 the PC model was fitted with 18 PCs and for stations 80 and 107 the PC models were fitted with 20 PCs. The PFC models were fitted with 1 PFC, which was computed using a polynomial basis function with degree one. The table also shows the average RMSE over all stations and over all 60 runs. PFCs outperform other methods in terms of predictive ability. However, PFCs did not perform well for 38 stations, which is approximately 40% of the stations in the study region. These results coincide with the results we obtained from Table 4.2. We conclude that overall, PFCs show a significant improvement over PCs as our analysis rely on the 60 chosen fitting and validation periods.

Regression Model	All Stations	Station 29	Station 80	Station 107
Linear	9.33	9.50	8.79	10.31
PC	9.32	9.96	8.84	10.94
PFC	9.03	9.23	7.44	9.52

Table 4.4: The RMSE value for some selected stations in the study region. We selected a fitting period of 20 days (i.e. 480 data points) and we used the following 5 days (i.e. 120 data points) as a validation period. We chose 60 different fitting and validation periods for each station. The RMSE values are averaged based on 60 runs

To assess the intrinsic uncertainties in the PC and PFC regression approaches we employed the Jackknife process as outlined in Efron and Tibshirani [1994]. We remove one day (i.e. 24 data points) out of the fitting period (6-25 June) and carry out the estimation and the prediction of the validation period (26-30 June). This procedure is repeated removing one day at a time over the fitting period. Moreover, we add each model predictor (PC or PFC score) progressively capping the number of predictors in the models to a maximum of 10, to both avoid over-fitting and provide computational convenience. The computation of the PFCs was based a polynomial basis function with degree 10 as the number of PFCs in a model should not exceed the size of the basis function. Computed RMSEs for each set of predictions are exhibited in Table 4.5 - which are basically the Jackknife RMSEs averaged over all 94 stations for the PC and the PFC models. It can be seen that the PFCs have better predictive ability than PCs when considering one PC and one PFC as predictors in the regression model. Furthermore, it was also established that increasing the number of predictors in the PC and PFC models does not improve the predictive ability of the model and that adding more predictors adds more noise for both types of models. This final observation implies over-fitting as the models start interpreting noise as meaningful data. These results conclude that having only one PFC in the regression does not only yield better predictive performance, but also indicates that PFCs outperform PCs in reducing data dimensionality.

No. of Predictors	PC Model	PFC Model
1	9.20	9.14
2	9.24	9.34
3	9.34	9.45
4	9.48	9.50
5	9.45	9.56
6	9.49	9.63
7	9.39	9.70
8	9.43	9.73
9	9.44	9.76
10	9.42	9.82

Table 4.5: Jackknife RMSE for the PC and PFC regressions. The values are averaged over all 94 stations of the study region. The training period was 6 June to 25 June and the validation period was from 26 June to 30 June. The PFC model was computed based on a polynomial basis with degree 10

In summary, the foregoing section presented Principal Fitted Components (PFCs) to downscale an air quality model for ozone over the southeastern U.S. The analyses were carried out for each site separately and the results were compared to two downscaling approaches - linear regression and principal components (PC) regression. The PFC regression outperformed the other methods in terms of predictive ability in most stations within the study region. However, as illustrated above, the PFC method did not work better in roughly one third of the stations. This might be because there are limited grid cells covering the locations of those stations, which might be resolved by enlarging the domain of the model, as some of stations in the study area are located at the border of the grid cells domain. This is common problem and it is analogous to near-overlapping clusters in cluster analysis.

We considered the autocorrelated nature of the data by fitting an AR(2) model for errors and the results did not show any improvement in the models predictive ability. This may be because the autoregressive structure of the errors was not strong, hence the models without AR(2) errors might already have captured the essential features of the relationship between REAM and the observations - hence, the AR modeling step simply adds noise instead of reducing uncertainty. We repeated the analyses for different fitting and validation periods and the results coincided with the illustrative period we initially chose. We examined the uncertainties in the PC and the PFC models by applying the Jackknife method and the results confirmed that PFCs outperformed PCs as a dimension reduction technique. These results are consistent with the simulation results reported in Cook [2007].

As PFCs estimates are based on the sample covariance matrix, which been shown to not being a good estimator of the population covariance matrix [Dempster, 1969], one approach that could be used to obtain a better estimate of the covariance matrix is thresholding [Bickel and Levina, 2004]. One of the main advantages of thresholding is that it is computationally inexpensive and so, we extend our work to thresholding the covariance matrix used in the estimation of PFCs. It would be interesting to examine the effect of thresholding on the predictive ability of the PFC regression model. In the next exposition we investigate the impact of thresholding the covariance matrix on the forecasting

performance of the PC and PFC models.

4.3 Downscaling an Air Quality Model with Regularized Covariance Matrix

This section presents downscaling of air quality model based on Regularized Covariance Data Matrix. As implied above, the approach seeks to examine the effect of thresholding on the predictive ability of the PFC regression model, in particular and, potentially, enhance its predictive power. Similar to the previous section, it consists the following components - simulation illustrations and downscaling REAM Ozone Outputs using PFCs.

4.3.1 Simulation Illustration

The simulation performing in this sub-section is fairly similar to the one in 4.1. Its main objective is to examine the effect of thresholding the covariance matrix had on the downscaling results. More specifically, through simulation we try to create an ideal situation of which thresholding and downscaling are applied to a data set that satisfies their theoretical assumptions. Similar to section 4.1, our simulations follow Cook [2007] and to imitate the realistic features of climate data we simulate seasonal random variables with a diurnal cycle.

The simulations are performed as follows. First, we generate an $\boldsymbol{n}$ sized normally distributed variable $\boldsymbol{Y}$ with mean 0 and variance $\boldsymbol{\sigma_y^2}$ and, for simplicity, we add a repetitive cycle of 10 values: $\mathbf{1, 2, \ldots, 10}$ to include a seasonal pattern to the data. We then generate an $\boldsymbol{n \times p}$ matrix $\mathbf{X}$ according to the inverse function 4.1, fixing the number of predictors to be $\boldsymbol{p = 100}$. The simulations are performed on three sample sizes, $\boldsymbol{n = 100,\ \ 150,\ \ 200,\ \ 500,}$ and $\mathbf{1000.}$ As in the previous example, on each simulated dataset we use the first 80% of the data as fitting (training) period and the remaining as a validation period.

Prior to carrying out the downscaling using PCs and PFCs, we need to threshold the covariance matrices that are used to obtain the PCs and the PFCs based on the simulated data. To determine the threshold value we rely on the selection method originally proposed by Bickel and Levina [2008], but we make a slight modification to the threshold selection method to adapt it to the seasonal nature of our data. As proposed by Bickel and Levina [2008], the thresholding selection method relies on randomly dividing the data into two samples of sizes $\boldsymbol{n_1 = n\left(1 - \frac{1}{\log n}\right)}$ and $\boldsymbol{n_2 = \frac{n}{\log n}}$ which is based on single observations where the order is absent.

However, since our data are daily time points with seasonal patterns (*e.g* hourly, seasonally, monthly, $\cdots$, *etc*), instead of randomly splitting the data according to single data points, we randomly divide them in terms of days to maintain the seasonal patterns within each day. Since the sample size $\boldsymbol{n = d \times k}$, where $\boldsymbol{d}$ is the number of days in the sample and $\boldsymbol{k}$ is the seasonal reoccurrence of the observations in each day (e.g number of hours), we replace $\boldsymbol{n_1}$ and $\boldsymbol{n_2}$ by

$\boldsymbol{d_1}$ and $\boldsymbol{d_2}$, where $\boldsymbol{d_1}$ is the number of days in the first sample and $\boldsymbol{d_2}$ is the number of the days in the second sample. The modification is performed as follows

$$\boldsymbol{n_2 = d_2 \times k = \frac{n}{\log n} = \frac{d \times k}{\log(d \times k)} = \frac{d \times k}{\log d + \log k}}$$
$$\boldsymbol{\Rightarrow d_2 = \frac{d}{\log d + \log k}} \quad \text{and, accordingly} \quad \boldsymbol{d_1 = d - d_2}$$

The foregoing modification ensures that not only do we maintain the similar number of observations in each sample according to the original method by Bickel and Levina [2008] but also we maintain the seasonal pattern within each sample. The threshold value is then chosen to be the value $\boldsymbol{s}$ that minimizes $\boldsymbol{R(s)}$ in equation 3.22.

Based on 50 runs, and after examining several threshold values for different sample sizes, the chosen threshold value along with their corresponding sample size are shown in Table 4.6. The main idea is to threshold the covariance matrices using the selected threshold values and carry out the PC and PFC downscaling as explained above.

For straightforward comparison to the simulation in 4.1, we downscale using PCs and PFCs with a reduced dimension space that equals 1 - that is, we fit the PC and PFC models with one predictor. For the PFC model we use a first degree polynomial basis function. Table 4.6 summarizes the results based on 100 replications. The table shows that, in general, thresholding seems to improve the predictive ability of both PCs and PFCs compared to the results in table 4.1. Furthermore, the simulation results show that downscaling based on a thresholded matrix performs comparatively better when the sample size is small compared to the dimension of the predictors. This coincides with the main objective of thresholding, which is to improve the estimation of the covariance matrix if $\boldsymbol{p}$ is large compared to $\boldsymbol{n}$, therefore the results of a covariance matrix based statistical techniques could be improved too.

Sample Size	Without Thresholding		With Thresholding			
	PC	PFC	Threshold Value	PC	Threshold Value	PFC
100	0.86	0.78	0.35	0.80	0.05	0.74
150	0.89	0.68	0.35	0.68	0.05	0.65
200	0.87	0.65	0.30	0.69	0.25	0.64
500	0.77	0.67	0.50	0.70	0.05	0.67
1000	0.75	0.69	0.40	0.70	0.05	0.68

Table 4.6: Simulation Results: RMSEs averaged over 100 replications for thresholded and non-thresholded PC and PFC regression. The PC model was fitted with one PC and the PFC model was fitted with one PFC using a polynomial basis function.

Table 4.7 shows the prediction improvement percentage comparing the RMSEs prior and after thresholding. In general, thresholding seems to improve the predictive ability of both PC and PFC models. Moreover, PCs have bigger improvement percentages compared to PFCs - results which indicate that thresholding had a greater impact on PCs. For both models, thresholding seems to have better effect when size is relatively small, where the PC model has im-

proved by 23.6% and the PFC model showing an improved performance of about 4.41%. Although the results of Table 4.7 show improvement in the predictive ability, there is no significant difference between PCs and PFCs. It is worth mentioning that the simulation results revealed an insignificant difference between the performance of PCs and PFCs (PFCs had a trivial lower RMSE value). Next section takes a closer look at real ozone concentrations.

Sample Size	PC regression	PFC regression
100	7%	4.4%
150	23.60%	4.41%
200	20.69%	1.54%
500	9.09%	0%
1000	6.67%	0%

Table 4.7: RMSE improvement percentages prior and after thresholding the covariance matrix

4.3.2 Downscaling the REAM Model Ozone Outputs

This section considers applications to real data set of ozone concentrations in the South-Eastern United States. Again, the period from 6 to 25 June is used for training the model while the period from 26 to 30 June is adopted for model validation. For the sake of straightforward comparison to the previous downscaling results in section 4.2, we maintain similar number of PCs and PFCs. The fitted model was fitted with one PFC, obtained using a polynomial basis function of degree one. Table 4.8 presents the RMSEs for few selected stations and the averaged RMSE amongst all stations in the study region. The threshold value was selected according to the modified Bickel and Levina [2008] method to be 0.20 for PCs and 0.15 for PFCs. On average, thresholding does not seem to improve the downscaling results. Even within the selected stations, the RMSE values do not significantly differ from the values in table 4.2 where the regular covariance matrix was used to obtain the PCs and PFCs.

Regression Model	All Stations	Station 29	Station 84	Station 107
PC	9.34	11.22	10.37	9.71
PFC	9.14	9.84	11.02	9.03

Table 4.8: RMSEs: training period is 6-25 June, validation period is 26-30 June. The PC models were fitted for: station 29 with 18 PCs, station 84 with 5 PCs, and station 107 with 20 PCs. The PFC models were fitted with 1 PFC (polynomial basis function with degree one). The threshold values are 0.20 and 0.15 for PCs and PFCs respectively.

To further verify the consistency of our findings, we repeated the analysis and model fitting for different fitting and validation periods. We selected a fitting period of size $\boldsymbol{n}$ days and predict for the next $\boldsymbol{k}$ days. Then, we varied the $\boldsymbol{n}$ day fitting period one day ahead and predicted for the next $\boldsymbol{k}$ days and so on, working with data covering the period from 2 June 2005 to 31 August 2005. Within 2 June to 26 August, we select a fitting period of $\boldsymbol{n = 20}$ consecutive days (i.e. 480 data points) and we allocate the following $\boldsymbol{k = 5}$ days (i.e. 120 data points) to the validation period

and we compute the RMSE for each set of predictions. This computation is repeated 60 times, as we move the fitting period by one day a head and proceed with the prediction for the corresponding validation period.
Table 4.9 shows the RMSEs (averaged over the 60 runs) for stations 29, 80, and 107. The results appear to coincide with the results of table 4.8. The table indicates that even for different fitting periods, thresholding does not seem to add an improvement to the predictive performance of the models. There is a minor improvement of the thresholded PCs and PFCs in station 29 where the RMSEs decreased by 5% for PCs and by 1%. However, this improvement does not seem to be significant.

Regression Model	All Stations	Station 29	Station 80	Station 107
PC	9.32	9.94	8.71	10.94
PFC	9.05	9.21	7.46	9.52

Table 4.9: The RMSE value for some selected stations in the study region. We selected a fitting period of 20 days (i.e. 480 data points) and we used the following 5 days (i.e. 120 data points) as a validation period. We chose 60 different fitting and validation periods for each station. The RMSE values are averaged based on 60 runs. The threshold value is 0.20 for PCs and 0.15 for PFCs. The threshold value is 0.20 for PCs and 0.15 for PFCs.

The plotted EOFs and F-EOFs provide an exploratory analysis of the effect of thresholding on the spatial patterns. Figures 4.7 and 4.8 show the first four EOFs obtained from the thresholded covariance matrix (at a threshold value of 0.20) versus not thresholded PCs. The plots do not show any dramatic change between the EOFs that were obtained from a regular covariance matrix and the ones obtained from a thresholded covariance matrix. A minor change in the spatial patterns can be seen in the first EOF around Georgia and around North Carolina in the third EOF. However, these alterations in the spatial patterns are trivial. Figures 4.9 and 4.10 show the first F-EOFs for some selected stations in the study region. The plots represent the F-EOFs that were obtained from a first degree polynomial basis function and a covariance matrix that was thresholded at a value of 0.15. The plots show that thresholding appears to have an effect on the spatial patterns on stations 51 and 96. It does indeed tend to sparse the eigenstructure everywhere except where the stations are located.

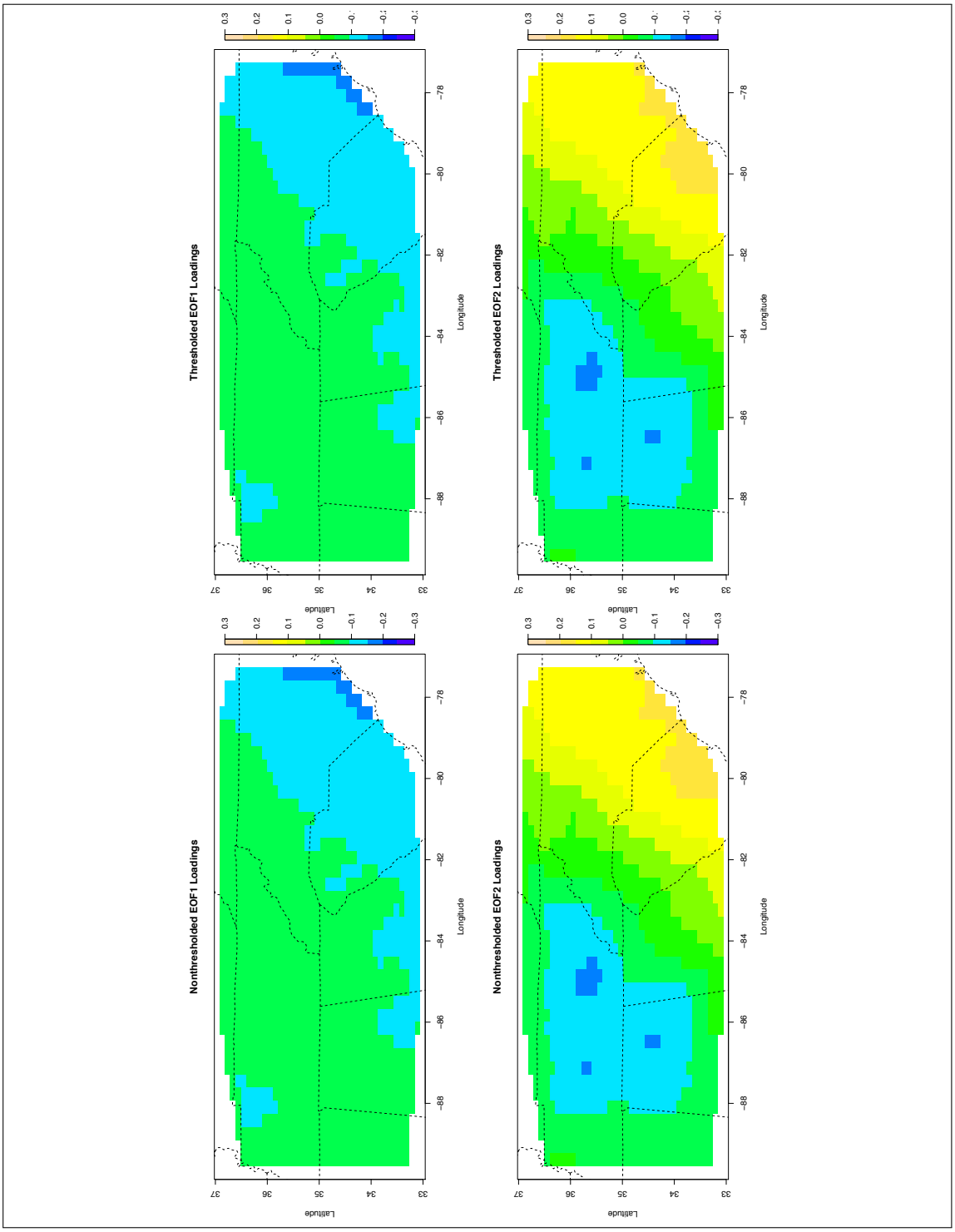

Figure 4.7: The first and second EOFs (thresholded at 0.20) of the gridded REAM output from 6 June to 25 June 2005.

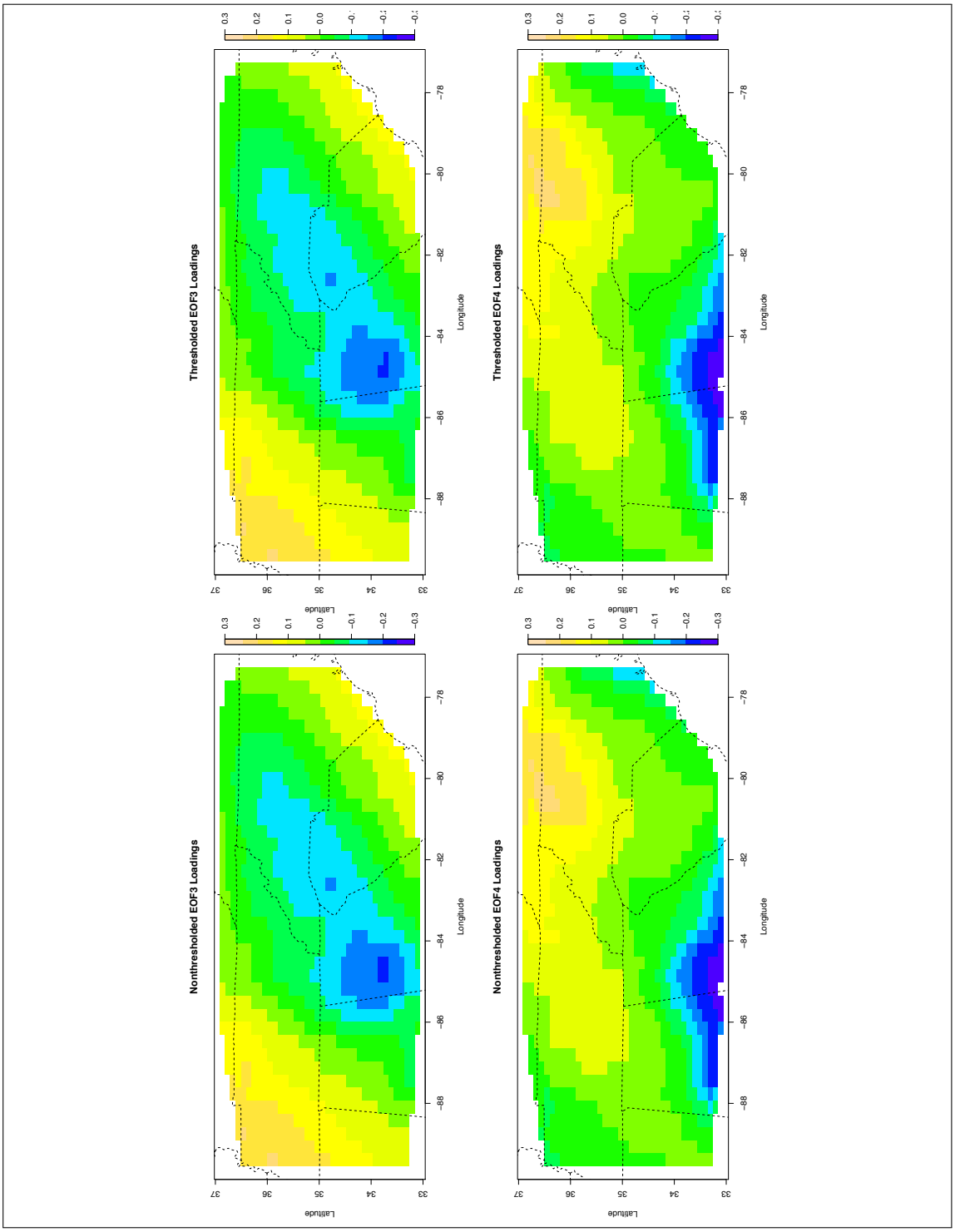

Figure 4.8: The third and forth EOFs (thresholded at 0.20) of the gridded REAM output from 6 June to 25 June 2005.

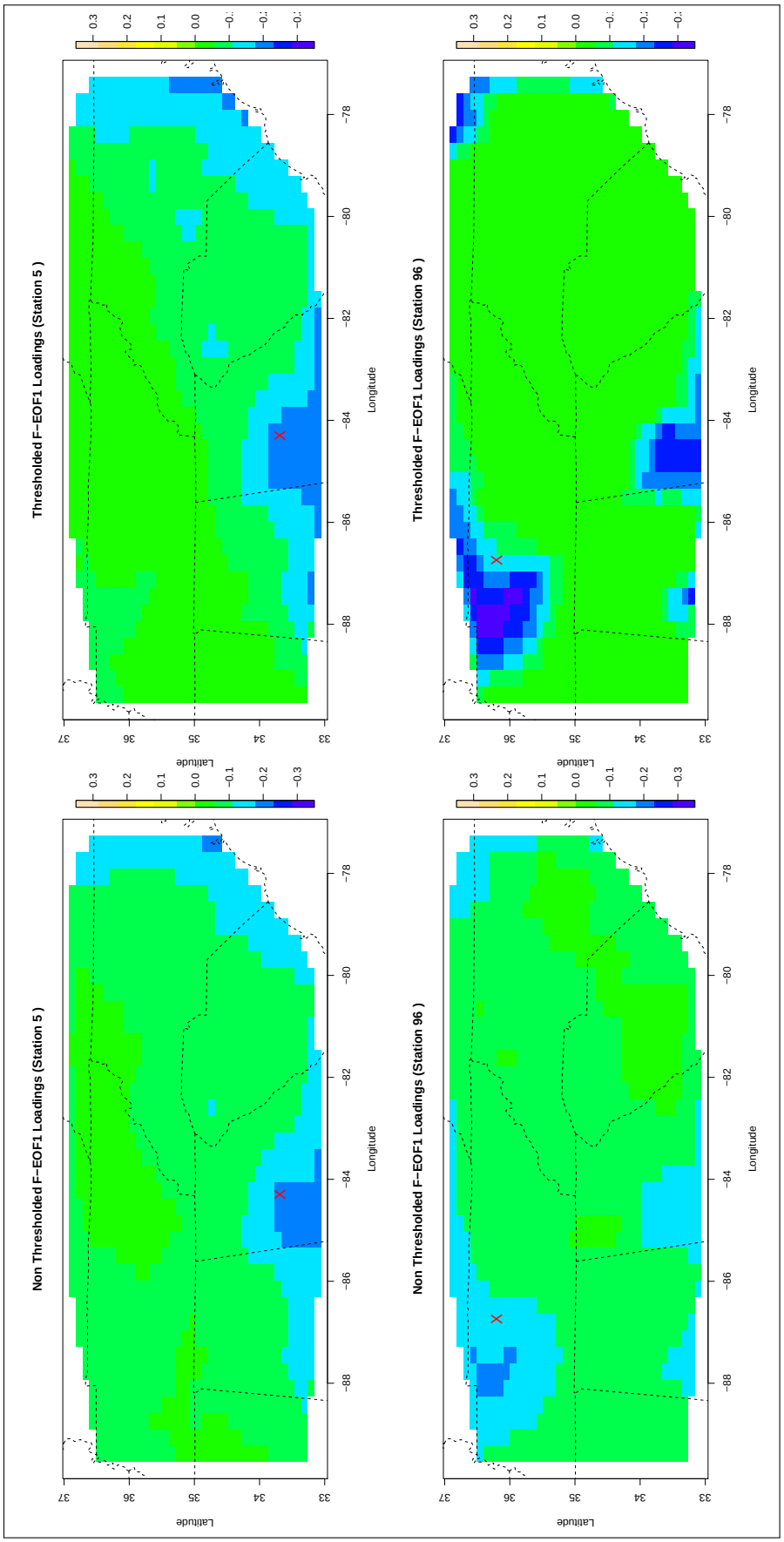

Figure 4.9: The first F-EOFs (polynomial basis function with degree 1 and thresholded at 0.15) of the REAM outputs for selected stations, estimated over 6-25 June 2005. The location of the station is marked by '×'.

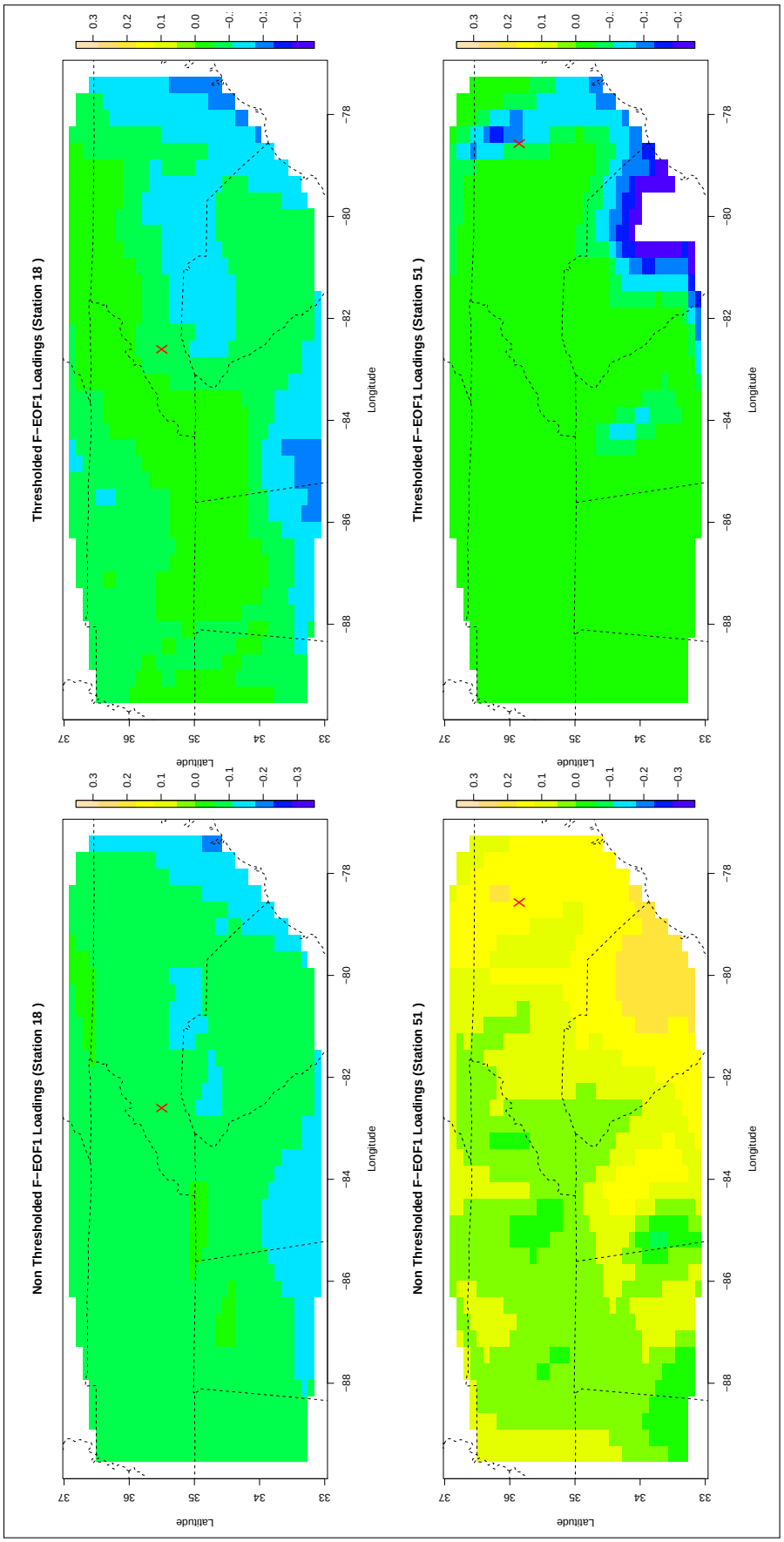

Figure 4.10: The first F-EOFs (polynomial basis function with degree 1 and thresholded at 0.15) of the REAM outputs for selected stations, estimated over 6-25 June 2005. The location of the station is marked by '×'.

Despite the change in some of the spatial patterns for some stations, generally, thresholding does not improve the downscaling results. One reason that can justify this, is that thresholding is most useful when the sample size is small compared to the dimension of the data, which is not the case for the results in table 4.8. That is, there are 480 observations in the fitting period versus a dimension of 99 grid cells. Thus, we repeated the analysis with a reduced fitting period size from 16 to 25 of June. The results are shown in table 4.10. The threshold value was chosen according to the modified Bickel and Levina [2008] selection method to be 0.05 and 0.50 for PCs and PFCs respectively. Even for a smaller sample size, the results do not coincide with the simulation results. Here too, thresholding does not seem to add any improvement to the predictive abilities of the PC and PFC models on average. In fact, in this case, thresholding appears to worsen the performance of the PC model.

Regression Models	Without Thresholding				With Thresholding			
	All	Station 29	Station 84	Station 107	All	Station 29	Station 84	Station 107
PC	11.23	13.41	12.64	11.61	11.26	15.11	12.66	11.81
PFC	10.82	12.01	12.36	11.49	10.68	11.39	11.91	10.95

Table 4.10: RMSEs: training period is 16-25 June, validation period is 26-30 June. The PC models were fitted for: station 29 with 18 PCs, station 84 with 5 PCs, and station 107 with 20 PCs. The PFC models were fitted with 1 PFC (polynomial basis function with degree one). The threshold values are 0.05 and 0.50 for PCs and PFCs respectively.

For some stations, thresholding slightly improves the predictive ability of the PFCs model but, overall, the RMSE values indicate that the predictive performance of the PC model was better prior thresholding. Table 4.11 summarizes the results of similar analysis based on 60 replications using different fitting and validation periods. The table shows that thresholding does not appear to improve the predictive performance of both PC and PFC model even when the dimension of the data is small compared to the sample size

Regression Model	Without Thresholding				With Thresholding			
	All	Station 29	Station 84	Station 107	All	Station 29	Station 84	Station 107
PC	10.23	10.90	9.38	9.88	10.09	12.40	10.59	10.56
PFC	9.81	10.10	7.99	7.95	9.77	9.63	9.06	9.07

Table 4.11: The RMSE value for some selected stations in the study region. We selected a fitting period of 10 days (i.e. 240 data points) and we used the following 5 days (i.e. 120 data points) as a validation period. The RMSE values are averaged based on 60 runs. The threshold value is 0.20 for PCs and 0.15 for PFCs. The threshold value is 0.05 for PCs and 0.50 for PFCs.

The prediction plots for stations 29, 84, and 107 are shown in figures 4.11, 4.12, and 4.13. The plots indicate in general that the PC model (whether the PCs were thresholded or not) is not able to outperform the PFC model. Moreover, the PFC model outperforms the PC model regardless of the size of the fitting period. In addition, the prediction plots show that the thresholded PFC predictions appears to outperforms the non-thresholded predictions especially in the

afternoon.

Generally, for this particular data, thresholding does not improve the downscaling results. We searched exhaustively over the covariance matrices that were used to compute the PCs and PFCs to gain an understanding of the lack of improvement when applying thresholding. For the data in hand the covariance matrices seem to be already sparse with a significant amount of matrix entries that are close to zero. Hence, for this dataset, thresholding might not enhance the estimation of a matrix since the matrix appears to be already regularized and does not seem to need further improvement.

In summary, this section downscaled an air quality model using PFCs with a regularized covariance matrix. To regularize the covariance matrix we employed a thresholding method that was proposed by Bickel and Levina [2008]. The rationale for selecting this specific technique was its many practical virtues. The results obtained from downscaling using thresholded PFCs were compared with non-thresholded PFCs as well as with thresholded PCs. For consistency, we used the same ozone data that were used in the previous section. Prior to applying thresholding to the ozone data, we performed a simple simulation and the results showed that thresholding helped improve the predictive ability of both PC and PFC models. Compared to the non thresholded results, thresholding seems to have greater impact on PCs more than PFCs. Moreover, as expected, thresholding appears to show significant improvement when the sample size $\boldsymbol{n}$ is small compared to the dimension $\boldsymbol{p}$**.** We applied thresholding when downscaling the REAM model outputs using PFCs and the results we obtained did not coincide with the simulation results.

Thresholding did not add any improvement to the predictive ability of both, PC and PFC models. Furthermore, thresholding appeared to worsen the performance of the PC models. Since thresholding is supposed to be more beneficial when the sample size is small relative to the dimension, we repeated the analysis using a shorter fitting period. For a shorter fitting period, thresholding appeared to add a minor improvement to the PFC model predictions while, on the other hand, the predictions of the PC model appeared to get worse.

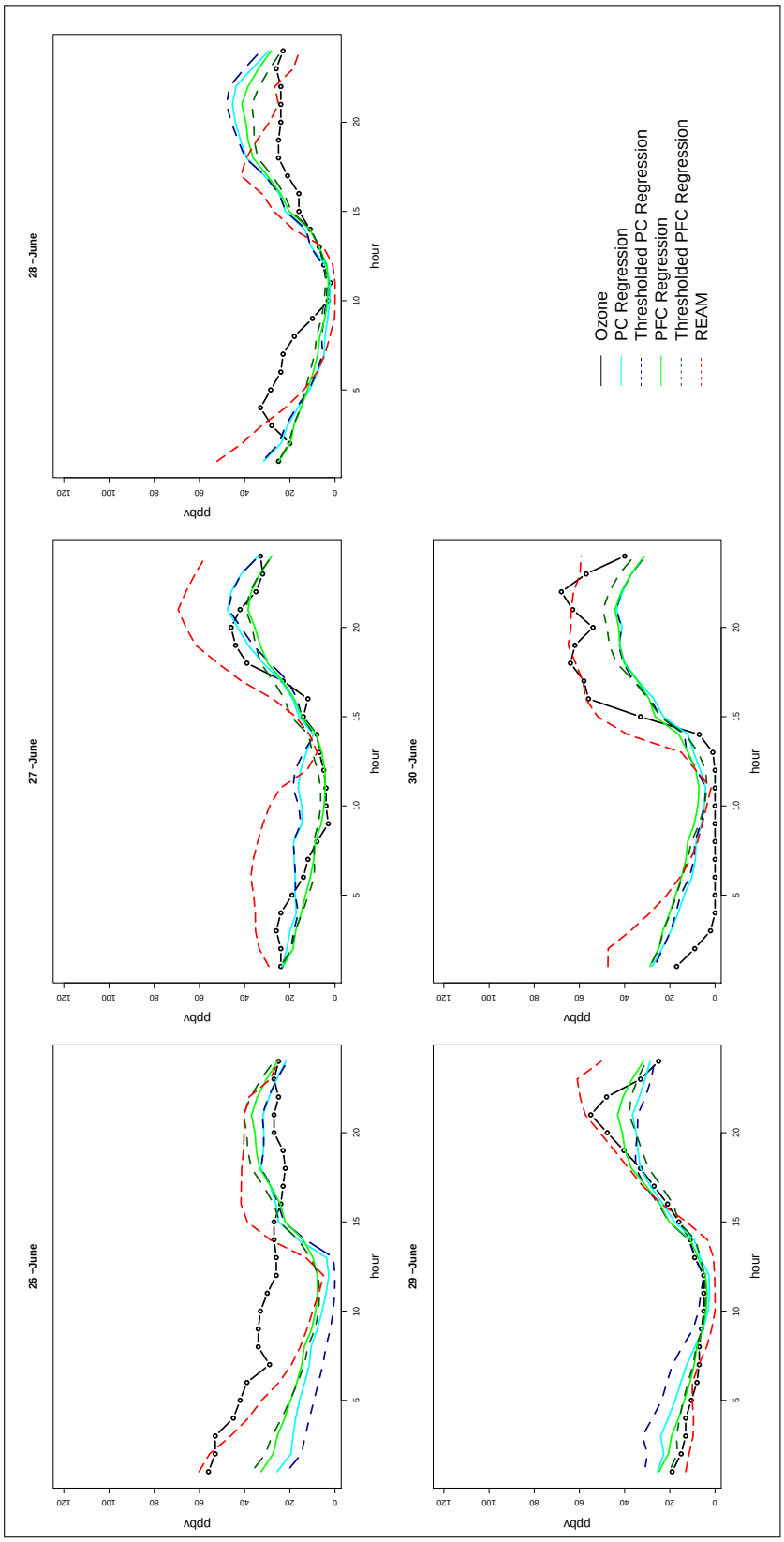

Figure 4.11: Prediction plots for station 29 (26-30 June). Observations (black line), REAM outputs (dashed red line), PC predictions (cyan line), Thresholded PC predictions (dashed dark blue line), PFC predictions (polynomial basis function with degree one, green line), thresholded PFC predictions (polynomial basis function with degree one, dashed dark green line).

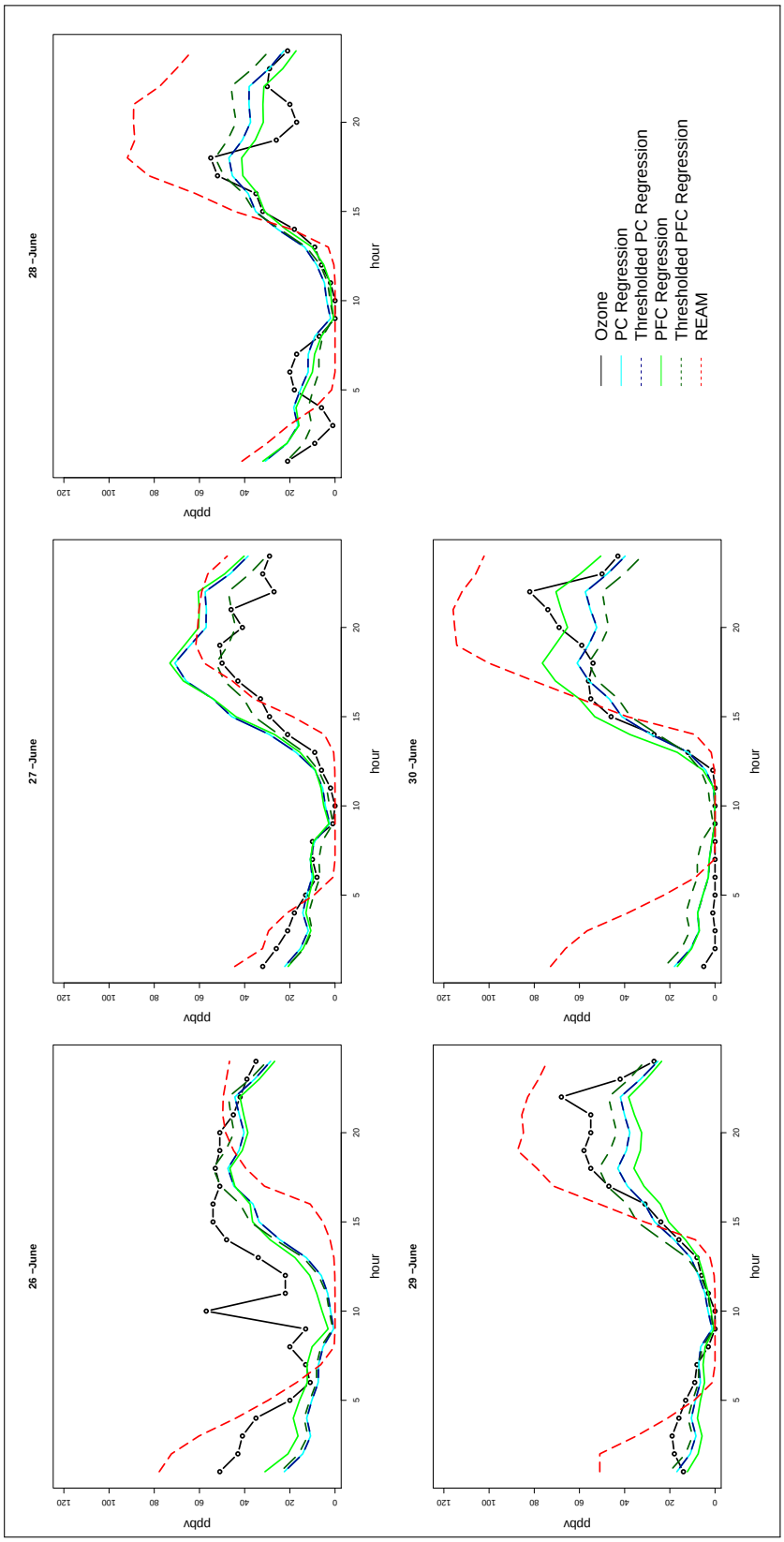

Figure 4.12: Prediction plots for station 84 (26-30 June). Observations (black line), REAM outputs (dashed red line), PC predictions (cyan line), Thresholded PC predictions (dashed dark blue line), PFC predictions (polynomial basis function with degree one, green line), thresholded PFC predictions (polynomial basis function with degree one, dashed dark green line).

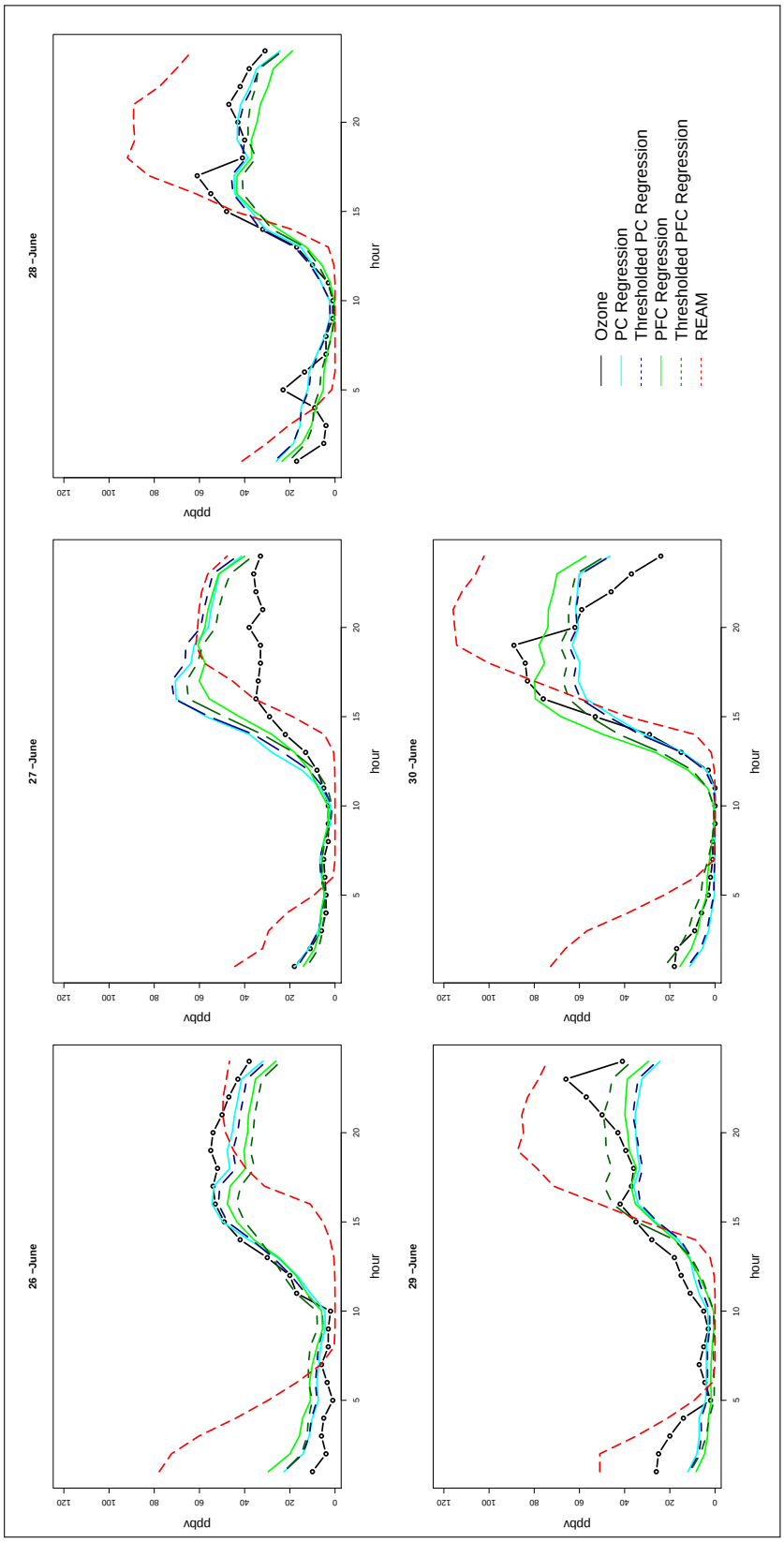

Figure 4.13: Prediction plots for station 107 (26-30 June). Observations (black line), REAM outputs (dashed red line), PC predictions (cyan line), Thresholded PC predictions (dashed dark blue line), PFC predictions (polynomial basis function with degree one, green line), thresholded PFC predictions (polynomial basis function with degree one, dashed dark green line).

4.4 Downscaling an Ensemble of Air Quality Models

This section focuses on downscaling an ensemble air quality models and, like the previous sections, it consists of a couple of components - notably, simulation aimed at providing insights into the performance of ensemble downscaling using the DDR method as described in Section 3.3.4 and downscaling ground level ozone. As we are focusing on the application of PFC, we adapt the simulation conducted by Cook [2007]. However, we modify it to accommodate an ensemble of models instead of a single model. An ensemble is originally built based on a reference model that in application is thought to best predicts the actual measurements. Ensembles are created by altering some initial conditions in the reference model.

4.4.1 Simulation Illustrations

As stated above, ensemble members are supposed to maintain the information provided by the reference model and must have better performance compared to single reference models as an ensemble accounts for the uncertainties and may include a better model than the default model. Consequently, our simulation need to take into consideration two characteristics when generating the data that serves the purpose of ensemble downscaling using PFCs - that is, firstly, imitate the mechanisms of how an ensemble is created and, secondly, generated ensemble members have to be suitable for the use of PFCs, i.e. the generated data should be tailored with respect to the response value. Furthermore, since our applications are on atmospheric models, we need to mimic the gridded structure of a geophysical model in our application. Thus, assuming that the best model that predicts the actual measurement is defined as

$$\mathbf{X_1}, \mathbf{X_2}, \cdots, \mathbf{X}_p$$

where each $\mathbf{X}_i$ is a vector that represents a geophysical model output at a grid cell i and has size n. Thus, $\mathbf{X}$ have $n \times p$ dimension. For simplicity we generated each $\mathbf{X}$ to be a normally distributed variable with mean 0 and standard deviation 1. Then we generate the response value (actual measurement) y according to the following forward linear model $y = \mathbf{X}W + \epsilon$, where ϵ is a standard normal error term variable and $\mathbf{W}$ is a user-specified weights vector of length p. We are trying to mimic a gridded model where each X represents the model output at a specific grid cell. In application we would expect that the measurement of interest is best predicted at a grid cell where the measurement monitoring station is located. Furthermore, the measurement of interest also might depend on the outputs of neighbouring grid cells while the outputs of far away cells would have a weak or no impact on the measurement value. Hence, we selected the weights such that optimum weight 1 was given to one variable (representing the grid cell where the station is located) and then the weights gradually decrease to 0 (as the grid cell gets further). Hence we specify the weights vector $\mathbf{W}$ as follows

- $\boldsymbol{W_1 = 1}$.
- $\boldsymbol{0.4 \leq W_2, \cdots, W_{10} \leq 0.60}$
- $\boldsymbol{0.1 \leq W_{11}, \cdots, W_{30} \leq 0.35}$
- $\boldsymbol{0.05 \leq W_{31}, \cdots, W_{40} \leq 0.09}$
- $\boldsymbol{W_{41}, \cdots, W_p = 0}$

For each set of weights, we randomly select a value within the specified range and assign this value randomly to the weight value $\mathbf{W_i}$. We then generate a reference model $\boldsymbol{M_{ref}}$ to be vector function of the best model as follow

$$\boldsymbol{M_{ref} = f_1(\mathbf{X_1}), f_2(\mathbf{X_2}), \cdots, f_p(\mathbf{X}_p)}$$

In practice, the functions $\boldsymbol{f(.)}$ are unknown, but represent the knowledge of geophysics modellers. The reference model $\boldsymbol{M_{ref}}$ is created by experts to mimic the best model. Since, in real life the best model is not known, $\boldsymbol{M_{ref}}$ is created as substitute of the best model depending on the specification by climate experts on what represents the best model in practice. For simplicity, we use power functions ($\boldsymbol{f(x) = x^q}$) to create $\boldsymbol{M_{ref}}$, as in practice the function $\boldsymbol{f(.)}$ is nonlinear. After generating $\boldsymbol{M_{ref}}$ we simulate the ensemble members.

In application, ensemble models are built using the reference model by adding slight alterations. But, generally, we would expect that an ensemble would outperform a single reference model because it accounts for many input related uncertainties. A good ensemble will have the majority of its members to show good predictive ability of response variable, with few models that do not perform very well. While a poor ensemble would have the majority of its members to show a poor predictive performance. As the purpose of this study is the examine the use of PFCs in downscaling ensembles based models, we need to adapt the inverse simulation model of Cook [2007]. Now, consider the following inverse models

$$\mathbf{X} = \boldsymbol{\Gamma y + \sigma \varepsilon} \tag{4.2}$$

$$\mathbf{X}_M = \boldsymbol{\Gamma M_{ref} + \sigma \varepsilon} \tag{4.3}$$

where $\mathbf{X}$ is a model that depends on the response value $\boldsymbol{y}$ (similar to the simulation inverse model in Cook [2007], $\mathbf{X}_M$ depends on the reference model $\boldsymbol{\Gamma}$=$\boldsymbol{(1, 0, \cdots, 0)^T}$, $\boldsymbol{\epsilon}$ is a standard normal error term and $\boldsymbol{\sigma > 0}$. The best ensemble (almost impossible to have in practice) would have all its members generated using the model in Equation 4.2. In equations 4.2 and 4.3 the parameter $\boldsymbol{\sigma}$ specifies how the generated data varies from the response value in equation 4.2 and from the reference model in equation 4.3. A bad ensemble would have all its members generated using Equation 4.3. In reality an ensemble would include a mixture of both good and poor models, depending on having more good models than poor models to get satisfactory prediction results. We generate 4 sets of ensembles to apply the DDR

method as follows

- The best ensemble: where all its members is created using equation 4.2.
- Good ensemble: where the majority of its members is created using equation 4.2.
- Poor ensemble: where the majority of its members is created using equation 4.3.
- Bad ensemble: where all its members is created using equation 4.3.

We generate 50 ensemble members and fix the dimension in each model to p =100. We perform the simulation with four sample sizes 100, 200, 500, and 1000. In each case, we use the last 50 observations as a validation period. For computational convenience, as well as straightforward comparison, we only choose one PC_1 per member in the first reduction stage. Then, we choose one PC_2 in the second reduction stage to carry out the PC downscaling. Similarly, for the PFC downscaling we select one PFC_1 reduction stage and the first PFC_2 in the second reduction stage.
PFCs are computed using a polynomial basis function with degree one. Table 4.12, summarizes the simulation results based on 100 runs for the four sets of ensembles (best, good, poor, and bad) and the PC and PFC downscaling results of the reference model. For the purpose of comparison we repeat the simulation where we specify the weights $\mathbf{W}$ to be equal to $\mathbf{\Gamma}=(\mathbf{1},\mathbf{0},\cdots,\mathbf{0})^{\boldsymbol{T}}$. The results are shown in table 4.13. We repeated the simulation using $\boldsymbol{\sigma} = \mathbf{5}$ to examine the effect of increased variation on the predictive performance. The results are shown in table 4.14. Figure 4.14 summarizes the simulation results. In general, the simulation results indicate that ensemble downscaling using the DDR method shows a better predictive ability than single model downscaling. Furthermore, the RMSE value tend to decrease when the sample size n increases. This is expected since we are using regression models even if the model is not a good fit. In addition, when the sample size is relatively large (n=1000) there does not seem to be a significant difference between the performance of the PC and PFC models.
Figure 4.14 shows that in all simulation trials, ensemble downscaling using the DDR method appears to improve the predictive ability when applied to the best ensemble and the good ensemble. This indicates that DDR downscaling seems to performs best when applied to a stable more consistent ensemble with a good predictive ability. On the other hand, the DDR method appears to perform worst when applied to the poor and bad ensemble compared to the other sets of ensembles. The PFC model appears to outperform the PC model in terms of predictive performance. However, when downscaling the single reference model PFC seems to perform worst when the sample size is relatively small. To summarize, downscaling using the DDR method appears to work best when the ensemble members have a good satisfactory predictive ability. Moreover, **PFC-DDR appears to outperform PC-DDR regardless of whether the ensemble has a good or poor predictive ability.**

Sample Size	Best Ensemble		Good Ensemble		Poor Ensemble		Bad Ensemble		One Reference Model	
	PC	PFC	PC	PFC	PC	PFC	PC	PFC	PC	PFC
100	0.72	0.70	0.73	0.69	1.71	1.61	1.53	1.50	1.64	1.67
200	0.38	0.36	0.41	0.40	1.56	1.45	1.49	1.50	1.29	1.45
500	0.19	0.18	0.20	0.19	1.63	1.42	1.63	1.67	1.33	1.32
1000	0.13	0.13	0.14	0.14	1.71	1.69	1.76	1.78	1.49	1.47

Table 4.12: Simulation Results: RMSEs averaged over 100 replications for best, good, poor, and bad ensembles and the reference model. The PC model was fitted with one PC and the PFC model was fitted with one polynomial PFC. $\mathbf{W}$ is pre-specified and $\boldsymbol{\sigma} = \mathbf{1}$

Sample Size	Best Ensemble		Good Ensemble		Poor Ensemble		Bad Ensemble		One Reference Model	
	PC	PFC	PC	PFC	PC	PFC	PC	PFC	PC	PFC
100	0.73	0.66	0.76	0.70	1.34	1.29	1.21	1.11	1.64	1.67
200	0.42	0.39	0.45	0.44	1.20	1.18	1.27	1.30	1.29	1.45
500	0.21	0.20	0.23	0.22	1.21	1.17	1.54	1.34	1.33	1.32
1000	0.15	0.15	0.16	0.16	1.13	1.10	1.31	1.29	1.49	1.47

Table 4.13: Simulation Results: RMSEs averaged over 100 replications for best, good, poor, and bad ensembles and the reference model. The PC model was fitted with one PC and the PFC model was fitted with one polynomial PFC. $\mathbf{W} = \mathbf{\Gamma}$, and $\boldsymbol{\sigma} = \mathbf{1}$

Sample Size	Best Ensemble		Good Ensemble		Poor Ensemble		Bad Ensemble		One Reference Model	
	PC	PFC	PC	PFC	PC	PFC	PC	PFC	PC	PFC
100	1.17	1.02	1.64	1.28	1.82	1.77	1.85	1.82	1.40	1.38
200	0.86	0.81	0.98	0.85	1.80	1.64	1.59	1.56	1.47	1.45
500	0.61	0.52	0.54	0.45	1.41	1.25	1.66	1.68	1.35	1.34
1000	0.35	0.34	0.39	0.35	1.76	1.76	1.90	1.89	1.35	1.36

Table 4.14: Simulation Results: RMSEs averaged over 100 replications for best, good, poor, and bad ensembles and the reference model. The PC model was fitted with one PC and the PFC model was fitted with one polynomial PFC. $\mathbf{W}$ is pre-specified, and $\boldsymbol{\sigma} = \mathbf{5}$

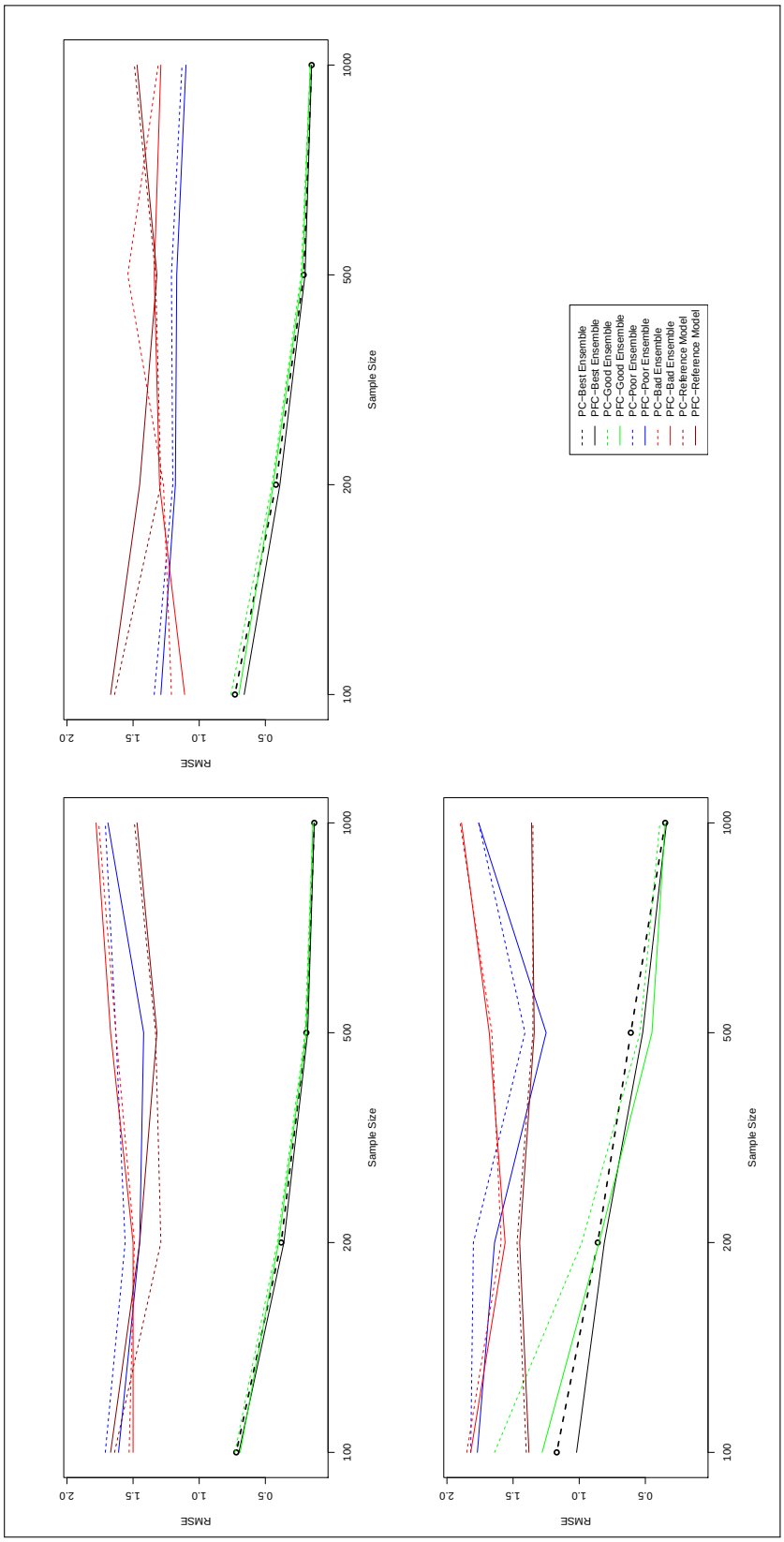

Figure 4.14: The top left plot shows that simulation results when the weighs $\mathbf{W}$ are specified and $\boldsymbol{\sigma} = \mathbf{1}$. The top right plot shows the simulation results when $\mathbf{W} = \boldsymbol{\Gamma}$ and $\boldsymbol{\sigma} = \mathbf{1}$. The bottom left plot shows that simulation results when $\mathbf{W}$ are specified and $\boldsymbol{\sigma} = \mathbf{5}$. The best ensemble is colour coded in black, the good ensemble is colour coded in green, the poor ensemble is colour coded in blue, the bad ensemble is colour coded in red, and the reference model is colour coded in dark red. In each case the dashed line represents the PC model and the solid line represent s the PFC model

4.4.2 Downscaling Ground Level Ozone

In this section, we use the Polyphemus ensemble, which produces ozone forecasts across France to illustrate our proposed methodology. The simulation results in section 4.4.1 indicate that the proposed DDR technique would perform better if ensemble members have a relatively good predictive performance. The Polyphemus ensemble have 107 members some of which would have an inadequate performance, possibly due to the fact that was built using a significantly large number of parameters, which could be a source to many performance deficiencies. Hence, we only consider models that show a relatively satisfactory performance.

There are many selection methods that can be employed in our case. One approach would be to discard the models with the highest RMSE, however, the drawback of this method is that the selection would be carried out after knowing all the facts, i.e. all observations are known. Another approach is to filter out the models randomly. However, this might lead to a loss of some good models within the ensemble. An alternative approach is to use a testing period with a significant number of observations, then we can measure how often a model computed the closest concentration to the observation. We adapt this approach as follows

1. We choose a testing period with a relatively sufficient number of observations.
2. For each observation, we give one point to the model that best predicts the observations.
3. We loop over all observations, summing up the assigned points of each model.
4. We filter out the models with the lowest number of points.

As noted above, the Polyphemus ensemble has 107 members of which six are reference models. These models are built by hand, and not generated automatically. The developers of the Polyphemus system have selected the parameters of theses reference models without any perturbation in the input field [Damien and Mallet, 2010]. Therefore, it makes sense to retain these six models. When applying our proposed selection method, we retain 46 models. Thus, we have an ensemble that has a total of 52 members including the six reference models. Thus, we perform our analysis over France using 52 members of the Polyphemus ensemble and each member produces ozone concentrations forecasts over 770 grid cells that cover all of France. We use measurements of 35 urban stations that scatter across France and perform the downscaling for the summer period of 2001. Our analyses commence performing the first reduction stage of the DDR method, which is to eliminate spatial redundancy of each of the 52 members. For each model we apply both PC and PFC analysis. To compute the PFCs we use a third degree polynomial basis function.

Prior to performing the second reduction step, we plotted the EOFs and F-EOFs for each model to have an insight on the spatial contribution of PCs and PFCs in each model. As the number of PCs and PFCs are relatively large we narrow them by presenting the first EOF and the first F-EOF plot for some selected models, after all the first EOFs

account for the maximum variation of the data. Figures 4.15 to 4.18 show the first EOF versus the first F-EOF for stations 9 and some selected ensemble members. The general variation of ozone appears to be higher over land than over ocean on the West. Furthermore, the distribution appears to be relatively higher over the southeastern coastline and the Mediterranean sea, where it is known to be close to emission regions. The plots show that F-EOFs appear to catch more information around the location at which the station is located. Furthermore, F-EOFs distribution appears to reflect the variation of ozone in neighbouring locations, which indicates that it captures more regional spatial information than EOFs. The plots for station 9 (figures 4.15 to 4.18) show that the F-EOFs reflect the distribution of ozone concentrations of the south-eastern coastline of France and over the Mediterranean. Moreover, the ozone gradient seems to decrease toward the eastern coastline. On the other hand, the EOFs appear to be associated with regional ozone patterns at which the variations are lower at the northwestern coastline, which are far from the stations location. However, for models 37 and 45, EOFs show a relative resemblance to ozone variations in the F-EOFs.

To examine the consistency of our findings, we plotted the first EOFs and F-EOFs for station 27, which is located at a more central location than station 9. The plots are for some selected ensemble members and are shown in figures 4.19 to 4.22. For this station, ozone variation distribution appears to be relatively similar for both EOFs and F-EOFs. The noticeable difference is that F-EOFs show higher concentration at the stations location than the EOFs. To summarize, the graphs suggest that the PFC method seems to be able to capture regional information that are relevant to the location of interest as well as reducing the dimension efficiently.

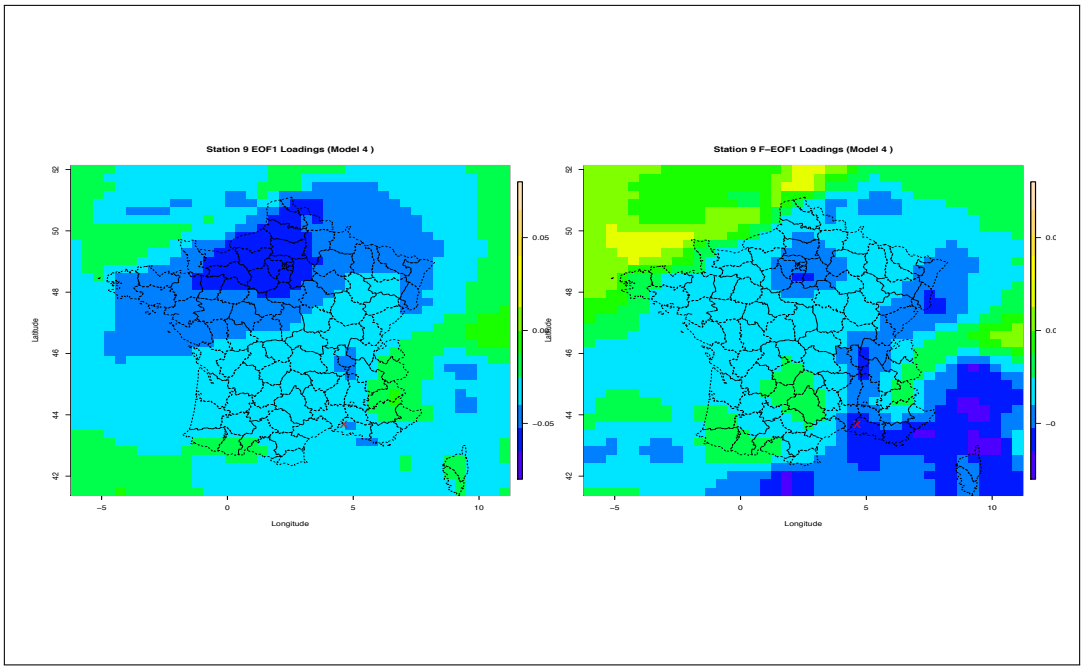

Figure 4.15: First EOF and F-EOF plot for ensemble member 4 of station 9. The F-EOF is obtained using a third degree polynomial basis function

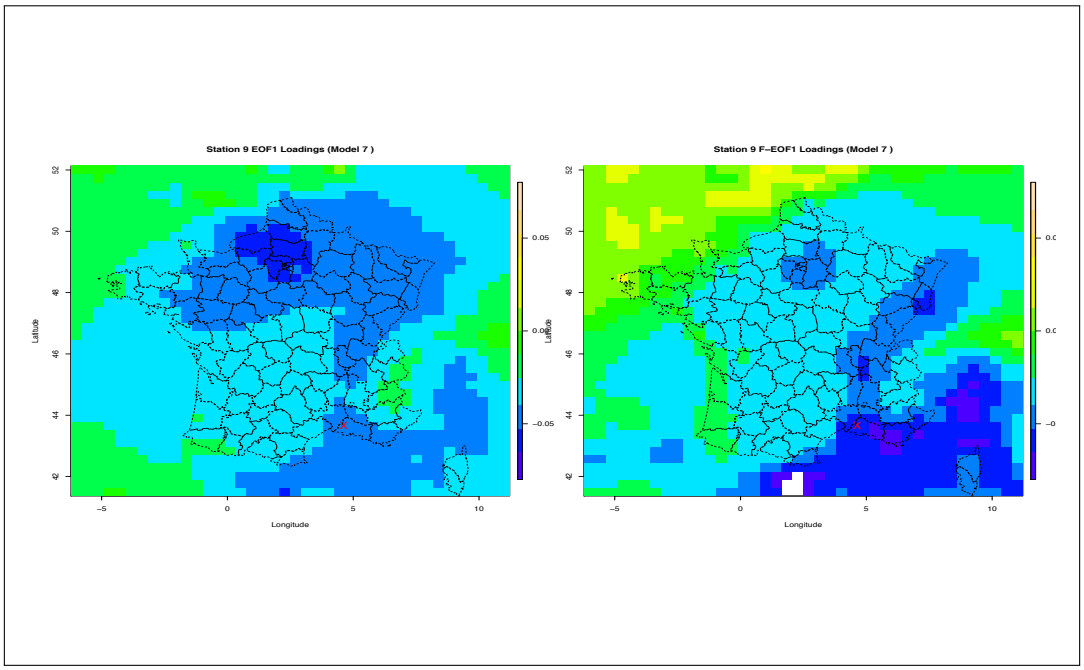

Figure 4.16: First EOF and F-EOF plot for ensemble member 7 of station 9. The F-EOF is obtained using a third degree polynomial basis function

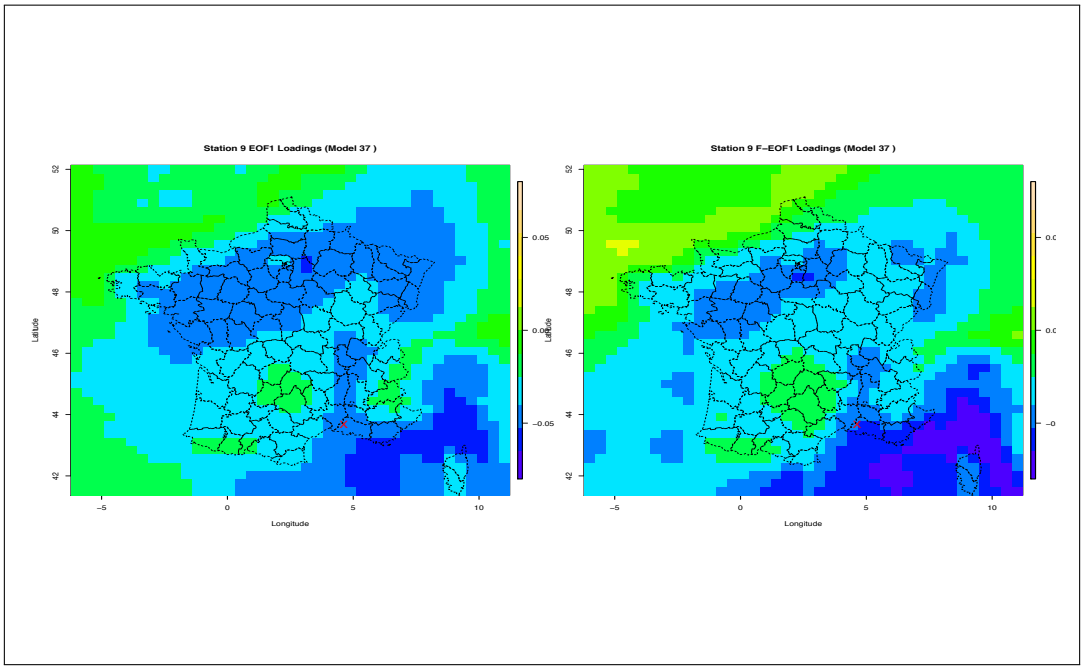

Figure 4.17: First EOF and F-EOF plot for ensemble member 37 of station 9. The F-EOF is obtained using a third degree polynomial basis function

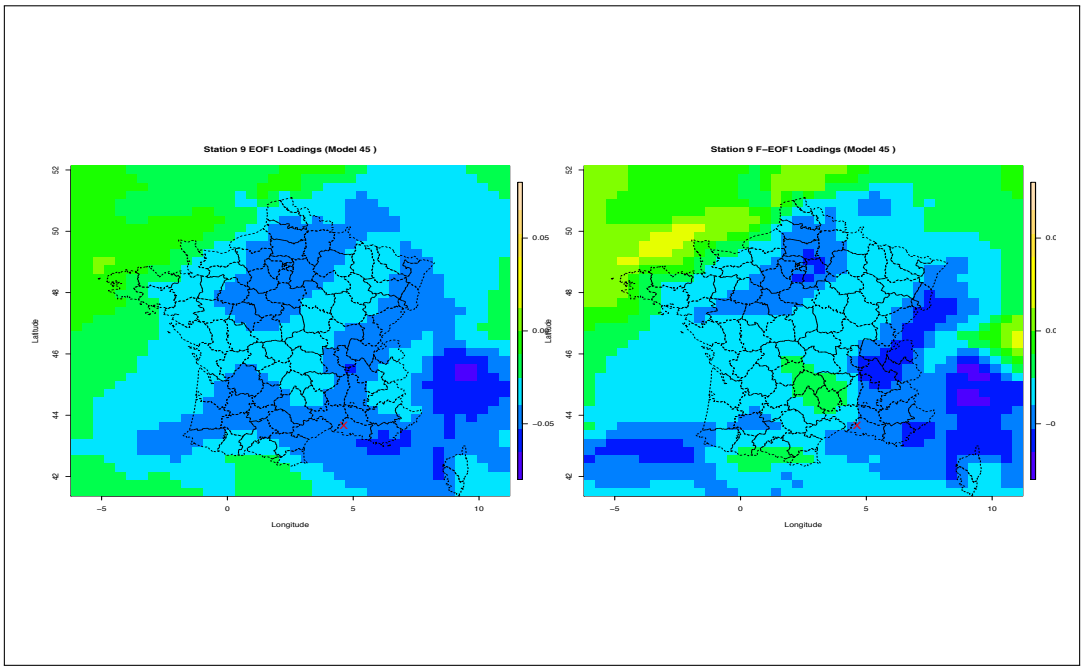

Figure 4.18: First EOF and F-EOF plot for ensemble member 45 of station 9. The F-EOF is obtained using a third degree polynomial basis function

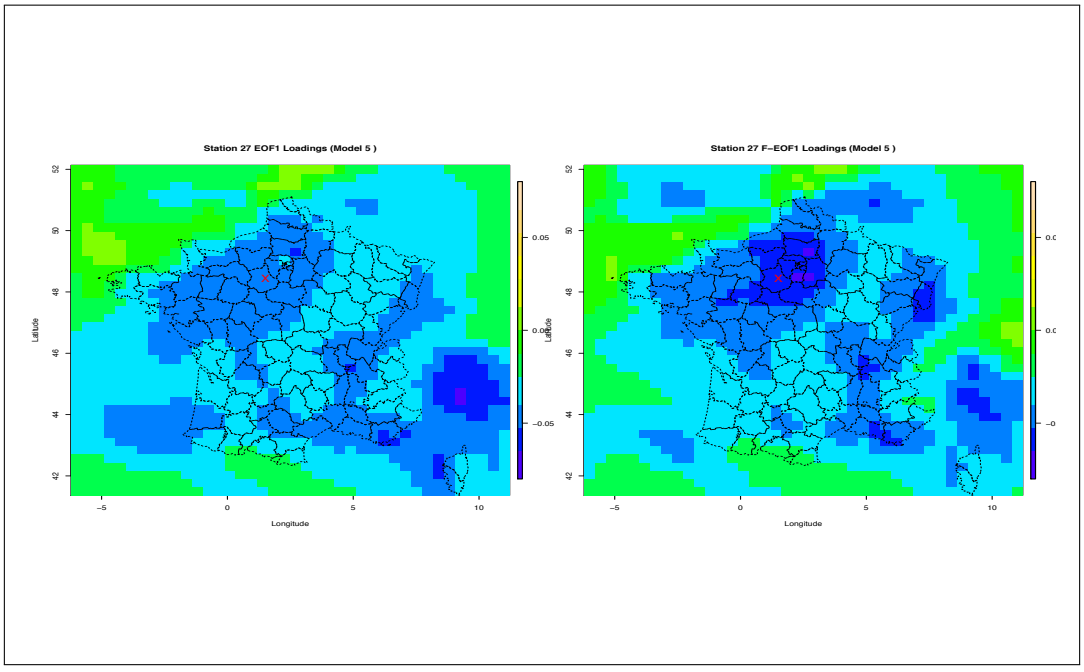

Figure 4.19: First EOF and F-EOF plot for ensemble member 5 of station 27. The F-EOF is obtained using a third degree polynomial basis function

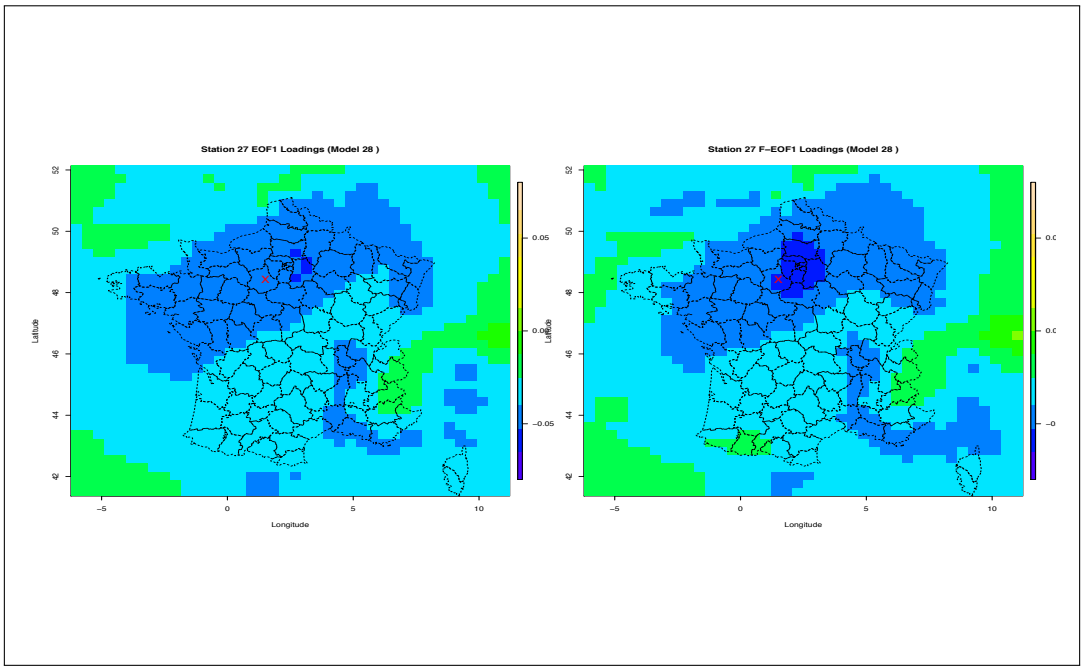

Figure 4.20: First EOF and F-EOF plot for ensemble member 28 of station 27. The F-EOF is obtained using a third degree polynomial basis function

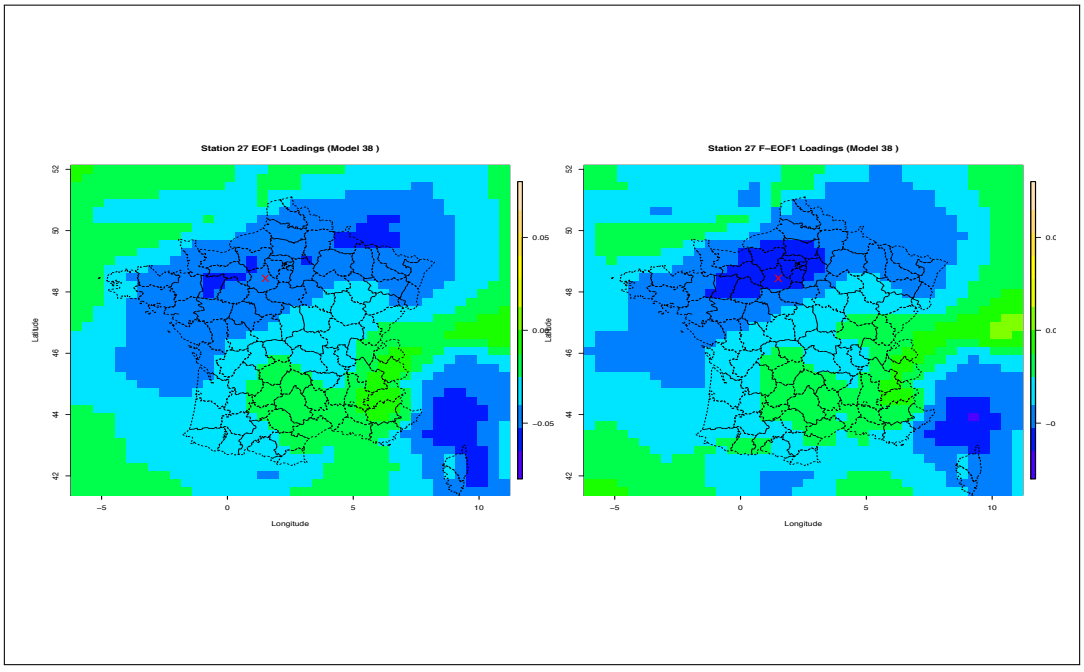

Figure 4.21: First EOF and F-EOF plot for ensemble member 38 of station 27. The F-EOF is obtained using a third degree polynomial basis function

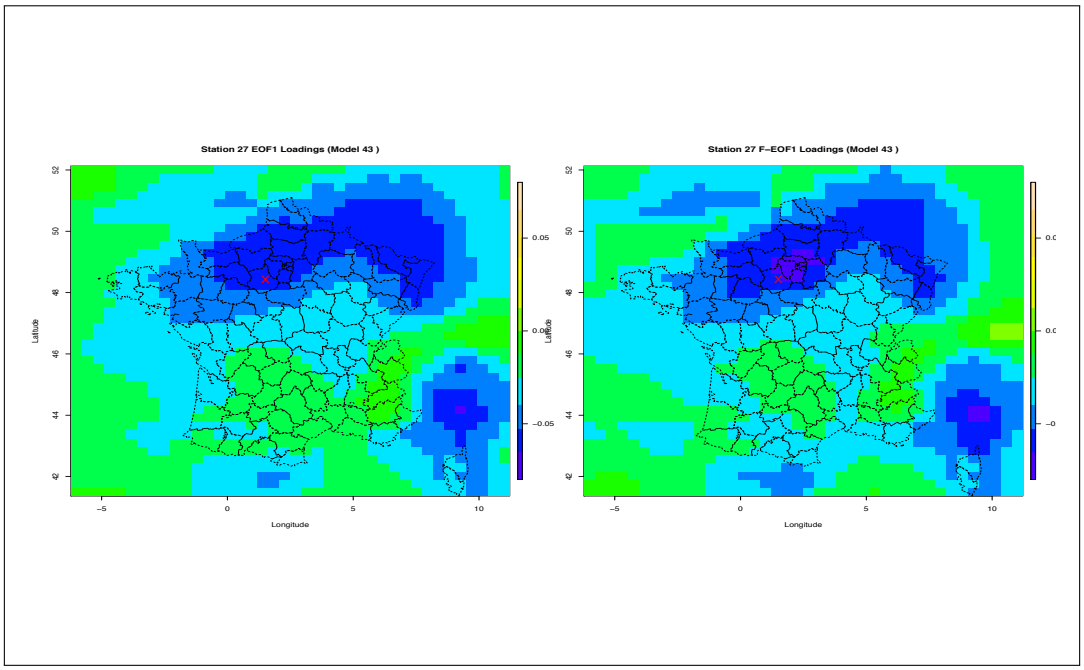

Figure 4.22: First EOF and F-EOF plot for represents the RMSE member 43 of station 27. The F-EOF is obtained using a third degree polynomial basis function

4.4.3 Identifying the Best PCs and PFCs

While the first reduction stage provided us with essential spatial information, further reduction across the models is needed to carry out the downscaling. At this stage we use the PFC_1 − scores obtained at the first stage to perform further dimension reduction across the models to downscale the ensemble. Since we performed the first reduction step using a polynomial basis function with degree three, we are restricted to use a maximum of three scores per model [Cook, 2007]. Initially we maintain the first PC score and the first PFC score for each model to reduce the dimension across the models. That is we reduce the dimension of a total of 52 scores (one for each model).

To perform the downscaling, we used the period from 1 June to 31 July as a fitting period and validate our method by using the period from 1 to 10 August as a validation period. The downscaling is performed on all the selected 35 stations in the study region. Table 4.15 shows the RMSE values averaged over all selected stations and for different PC and PFC models. This is to help in identifying the best PC and PFC regression models that can produce the best downscaling results for our data. To avoid over-fitting we restrict the maximum number of predictors to 10 (through cross-validation). To clarify why some cells in table are empty, according to Cook [2007], the number of predictors should not exceed the power of the basis function, this is why we do restrict the model fitting to a maximum of the power used in the polynomial function (for detailed theoretical justification see Cook [2007]).

The table suggests that a PC model with one PC score as predictor seems to be the best fit as it holds the smaller averaged RMSE value compared to the other PC models. Moreover, the table shows that a PFC mode with one PFC score that was obtained using a second degree polynomial basis function, appears to outperform other PFCs models. Hence, we limit our analysis by adapting these two downscaling models: a PC model with one PC score as a predictor, and a PFC model with one PFC score (computed using a second degree polynomial basis function) as a predictor.

Table 4.16 shows that averaged RMSE values over all 35 stations and for some selected stations. In addition, we selected one of the six reference models in the ensemble and used to perform the downscaling. The bar plot in figure 4.23 summarizes the results of table 4.16. The results in the left side panel of table 4.16 are similar to the analysis performed earlier, where we downscaled a single air quality model. The right side panel of table 4.16 shows the downscaling results of our proposed DDR techniques. The results for downscaling the reference model are in total agreement with the results obtained then - that is, PFCs outperform PCs. On average the prediction error improved by approximately 4% when downscaling using PFCs. Improvement percentage is relatively small, however, the table shows that averaged RMSEs yield relatively large improvements within each station individually.

Comparing ensemble downscaling to single model downscaling, both table 4.16 and the bar plot indicate that generally downscaling an ensemble of models show better predictive performance than downscaling a single air quality model. This could be because ensembles accommodate many input parameters and parameter values, which might help in accounting many uncertainties. The table shows that all RMSE values when downscaling an ensemble are smaller than their corresponding values when downscaling a single reference model. Furthermore, PC-DDR downscaling show a 5% improvement compared to PC downscaling, and PFC-DDR downscaling show a 4% improvement compared to PFC downscaling. This implies that taking the actual measurement into account could improve the downscaling results. For each station individually the PFC-DDR method still showed a better performance that the PC-DDR method.

No. of Predictors	PC	POLY1	POLY2	POLY3	POLY4	POLY5	POLY6	POLY7	POLY8	POLY9	POLY10
1	19.694	19.240	19.229	19.232	19.232	19.233	19.233	19.233	19.233	19.233	19.233
2	19.891		20.884	20.585	20.502	20.457	20.513	20.426	20.467	20.442	20.456
3	19.863			20.555	20.431	20.322	20.335	20.320	20.296	20.320	20.341
4	19.881				20.251	19.871	19.950	19.785	19.841	19.762	19.893
5	19.983					19.942	19.861	19.738	19.770	19.750	19.777
6	20.353						19.686	19.812	19.751	19.801	19.814
7	20.564							19.749	19.743	19.915	19.982
8	20.447								19.621	19.646	19.671
9	20.102									19.446	19.652
10	20.598										19.579

Table 4.15: RMSE values averaged over the 35 stations in the study area. The rows represent the number of predictors used to fit the regression model. The first column shows the averaged RMSE values of PC regression models with 1, 2, $\cdots$, and, 10 PC scores. The preceding columns show the averaged RMSE values of PFC regression models. The column label "POLY1" means that the PFCs were computed using a polynomial basis function with degrees one, the column label "POLY2" means that the PFCs were computed using a polynomial basis function with degrees two, and so on.

Regression Model	Single Reference Model Downscaling				Ensemble DDR Downscaling			
	All	Station 3	Station 15	Station 121	All	Station 3	Station 15	Station 21
PC	20.65	22.99	22.85	18.11	19.70	21.49	21.86	17.06
PFC	20.01	22.30	22.04	17.04	19.23	20.74	21.16	16.95

Table 4.16: Downscaling results: RMSE values averaged over all stations in the study region and for some selected stations. The PC model is fitted with on PC and PFC model is fitted with one PFC. The PFC is computed using a second degree polynomial function. The fitting period is from 1 June to 31 July 2001, and the validation period is from 1 to 10 August 2001.

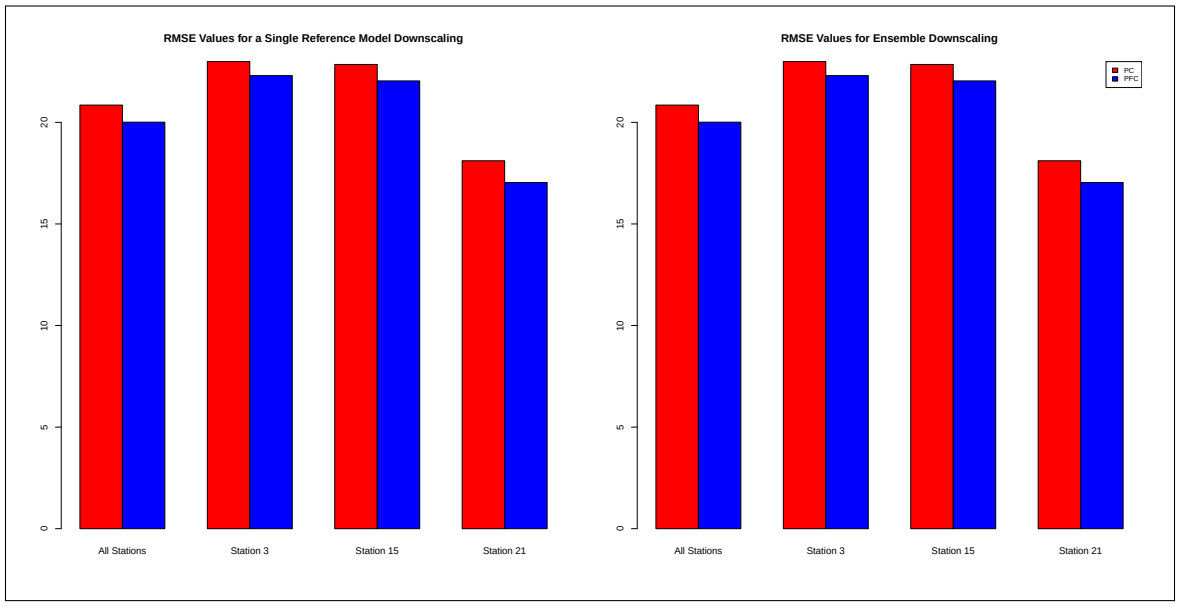

Figure 4.23: A graphical summary of the results of table 4.16

To verify the consistency of our findings and validate out results we repeated out analysis for 10 different samples. We randomly select a fitting period of similar size as the period used in table 4.15, i.e. 61 one consecutive days, and use to perform the DDR method and the appropriate final regression models to perform the downscaling. Then we used the preceding 10 days to validate our results. Figures 4.24 to 4.28 show graphical summaries of our findings for each one of the randomly chosen samples. On the left hand side panel of each plot are box plots, providing insightful information that is needed here such as the minimum RMSE value and not emphasize on the variation or distributional information. The right hand side panel shows line plots of the averaged RMSE values for different regression models with different number of predictors. Basically these line plots represent RMSE tables similar to table 4.15. That is why some of the blue lines cut off after a certain number of predictors.

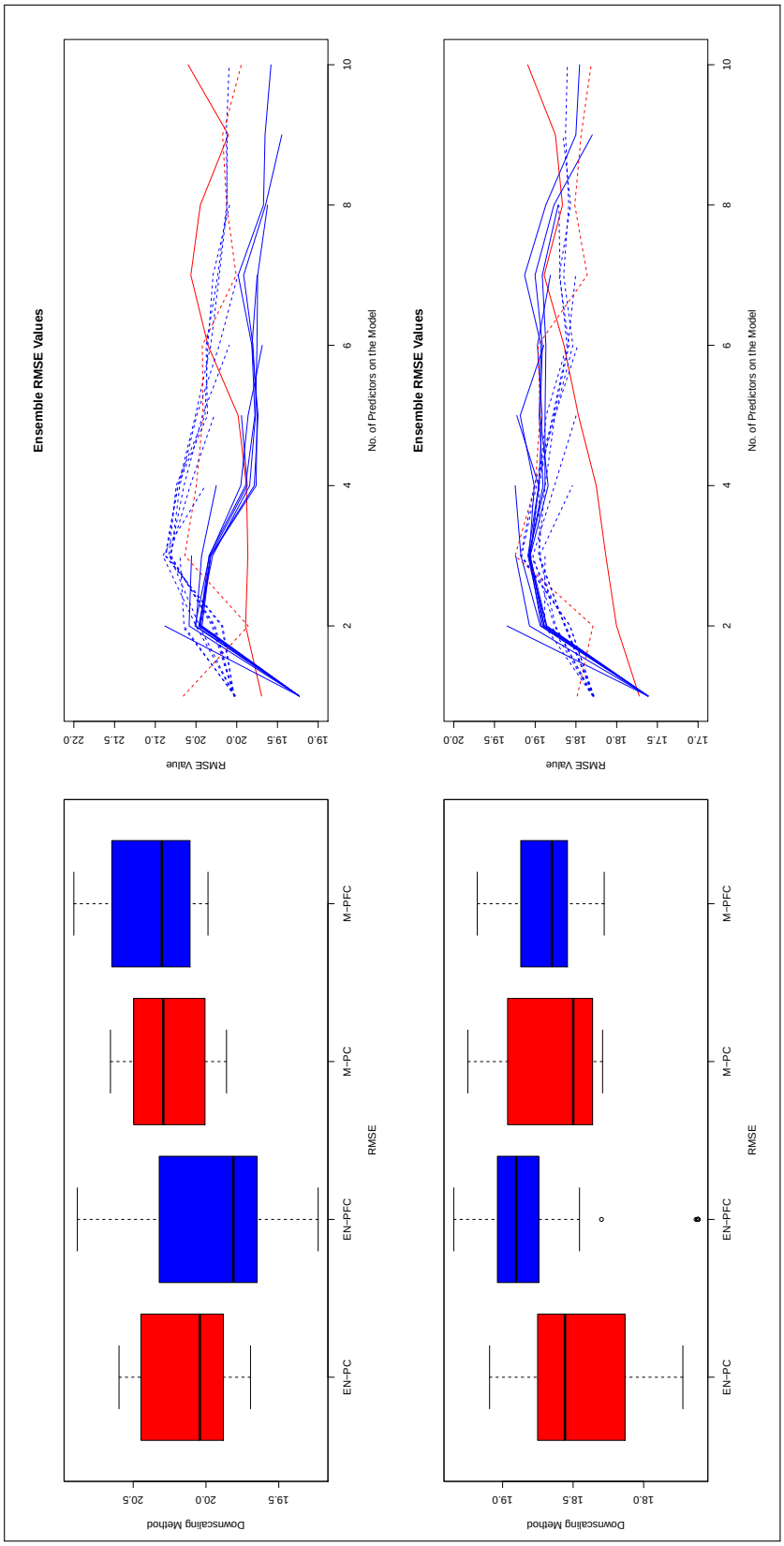

Figure 4.24: RMSE plots for different periods. Each horizontal panel represents the results of the sample period. The plots show the RMSE values for PC and PFC downscaling for the ensemble (DDR method) and for one reference model. The results are show for different models with different number of predictors. The colour code blue represents the results of a PFC method. The colour code red represents the results of a PC model. In the line plot, the dashed line represents the RMSE values of a single reference model. and the solid line represents the RMSE values for ensemble (DDR method). The RMSE values are averaged over all 35 stations.

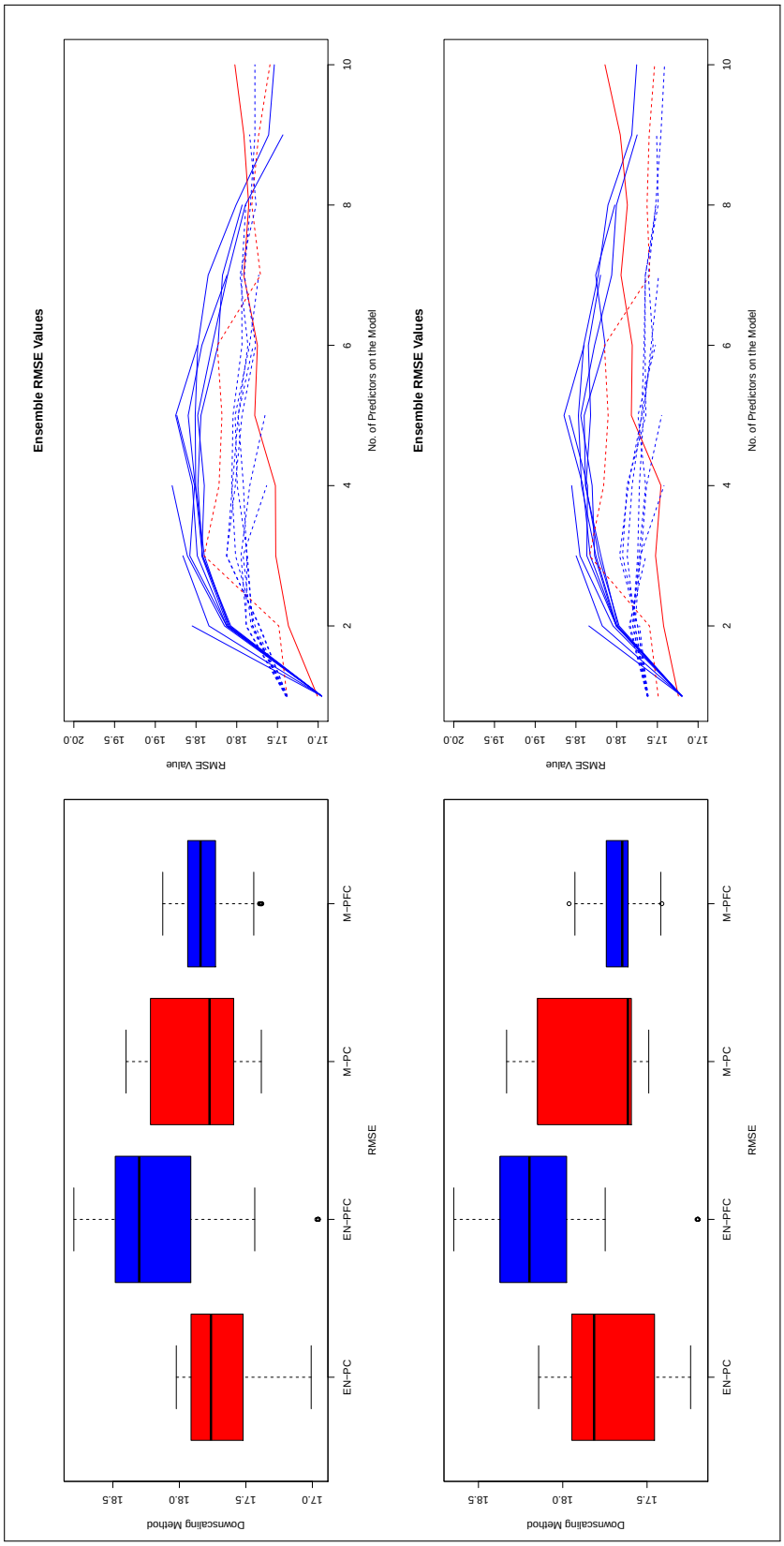

Figure 4.25: RMSE plots for different periods. Each horizontal panel represents the results of the sample period. The plots show the RMSE values for PC and PFC downscaling for the ensemble (DDR method) and for one reference model. The results are show for different models with different number of predictors. The colour code blue represents the results of a PFC method. The colour code red represents the results of a PC model. In the line plot, the dashed line represents the RMSE values of a single reference model. and the solid line represents the RMSE values for ensemble (DDR method). The RMSE values are averaged over all 35 stations.

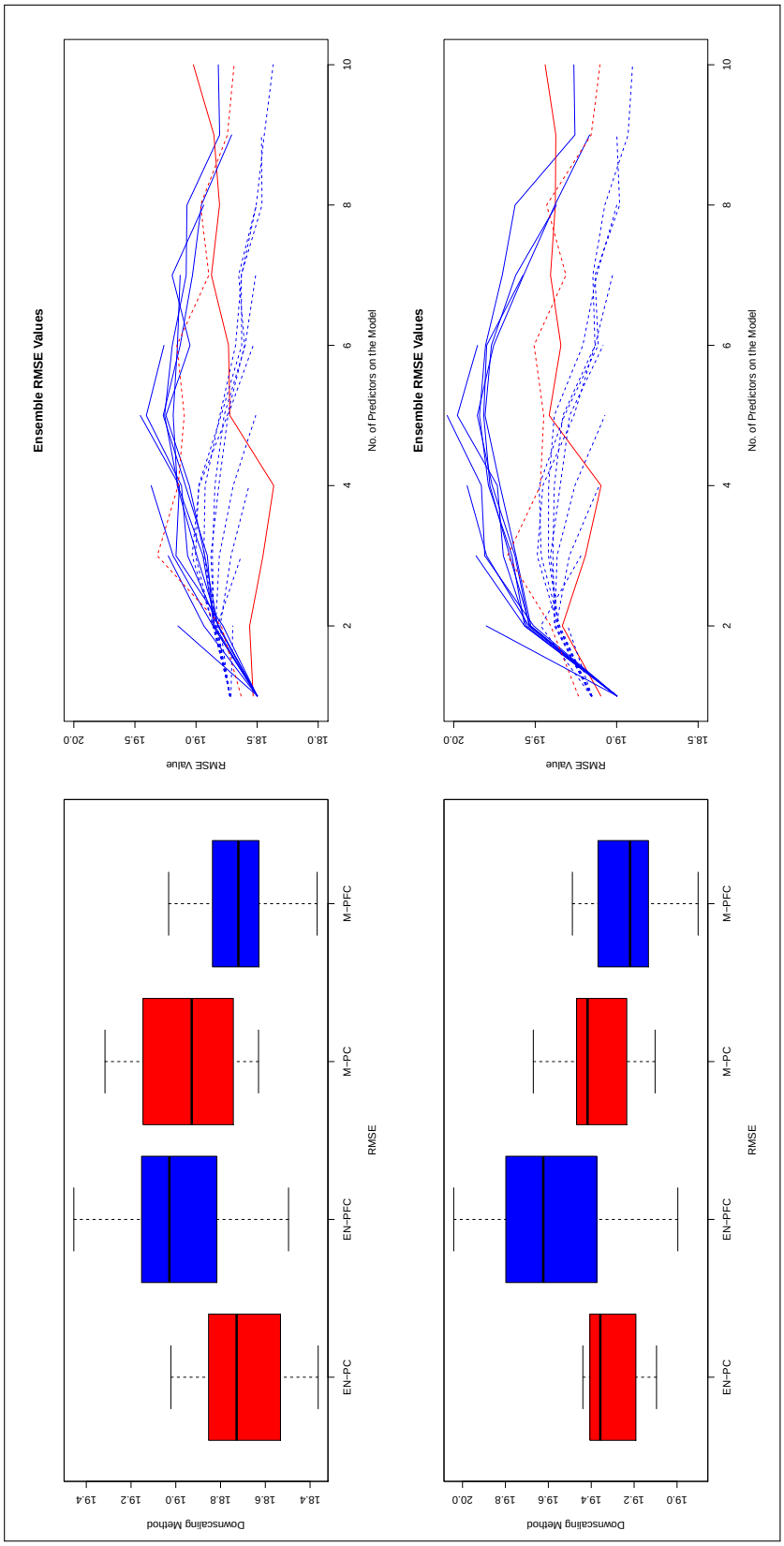

Figure 4.26: RMSE plots for different periods. Each horizontal panel represents the results of the sample period. The plots show the RMSE values for PC and PFC downscaling for the ensemble (DDR method) and for one reference model. The results are show for different models with different number of predictors. The colour code blue represents the results of a PFC method. The colour code red represents the results of a PC model. In the line plot, the dashed line represents the RMSE values of a single reference model. and the solid line represents the RMSE values for ensemble (DDR method). The RMSE values are averaged over all 35 stations.

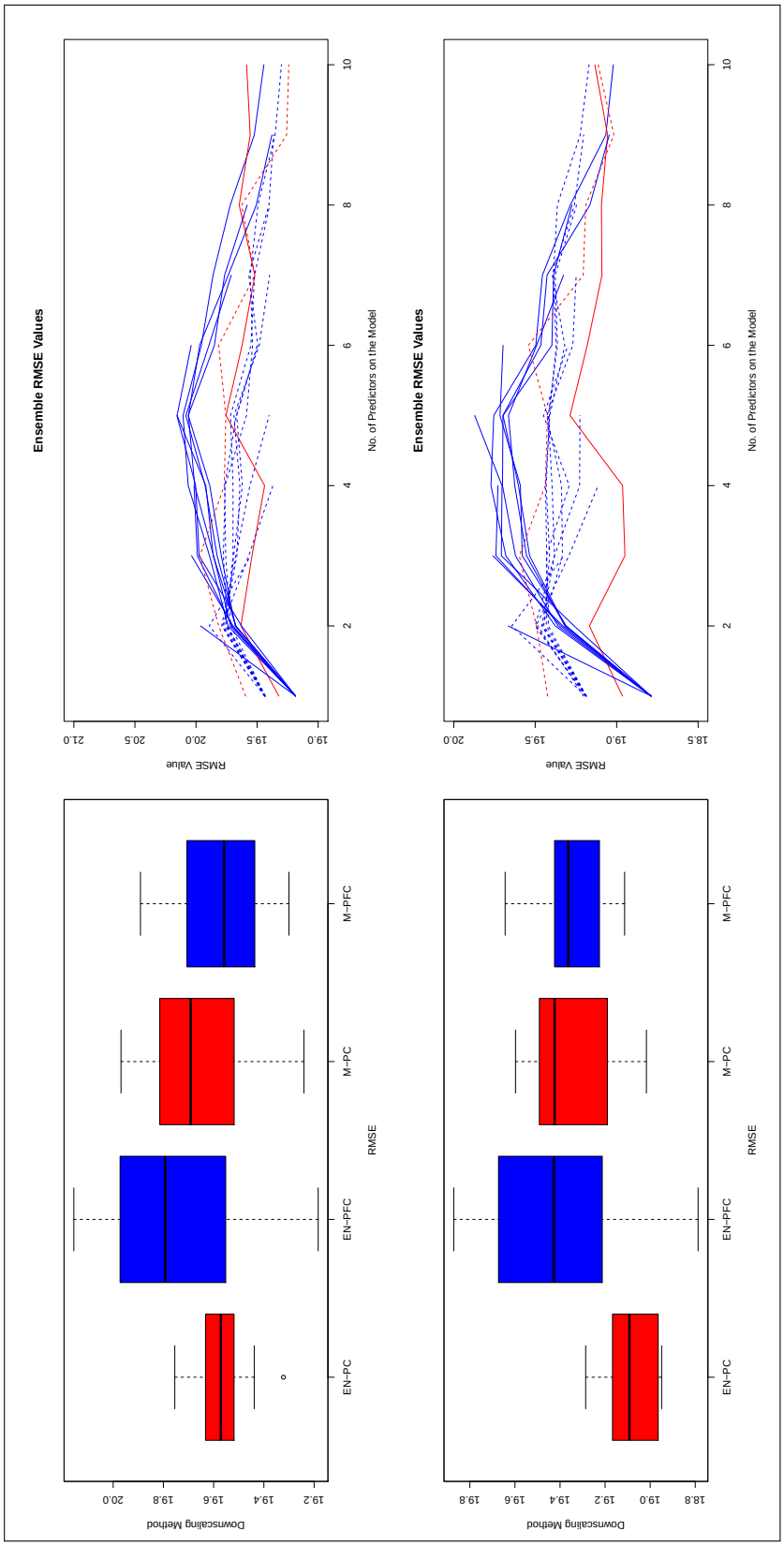

Figure 4.27: RMSE plots for different periods. Each horizontal panel represents the results of the sample period. The plots show the RMSE values for PC and PFC downscaling for the ensemble (DDR method) and for one reference model. The results are show for different models with different number of predictors. The colour code blue represents the results of a PFC method. The colour code red represents the results of a PC model. In the line plot, the dashed line represents the RMSE values of a single reference model. and the solid line represents the RMSE values for ensemble (DDR method). The RMSE values are averaged over all 35 stations.

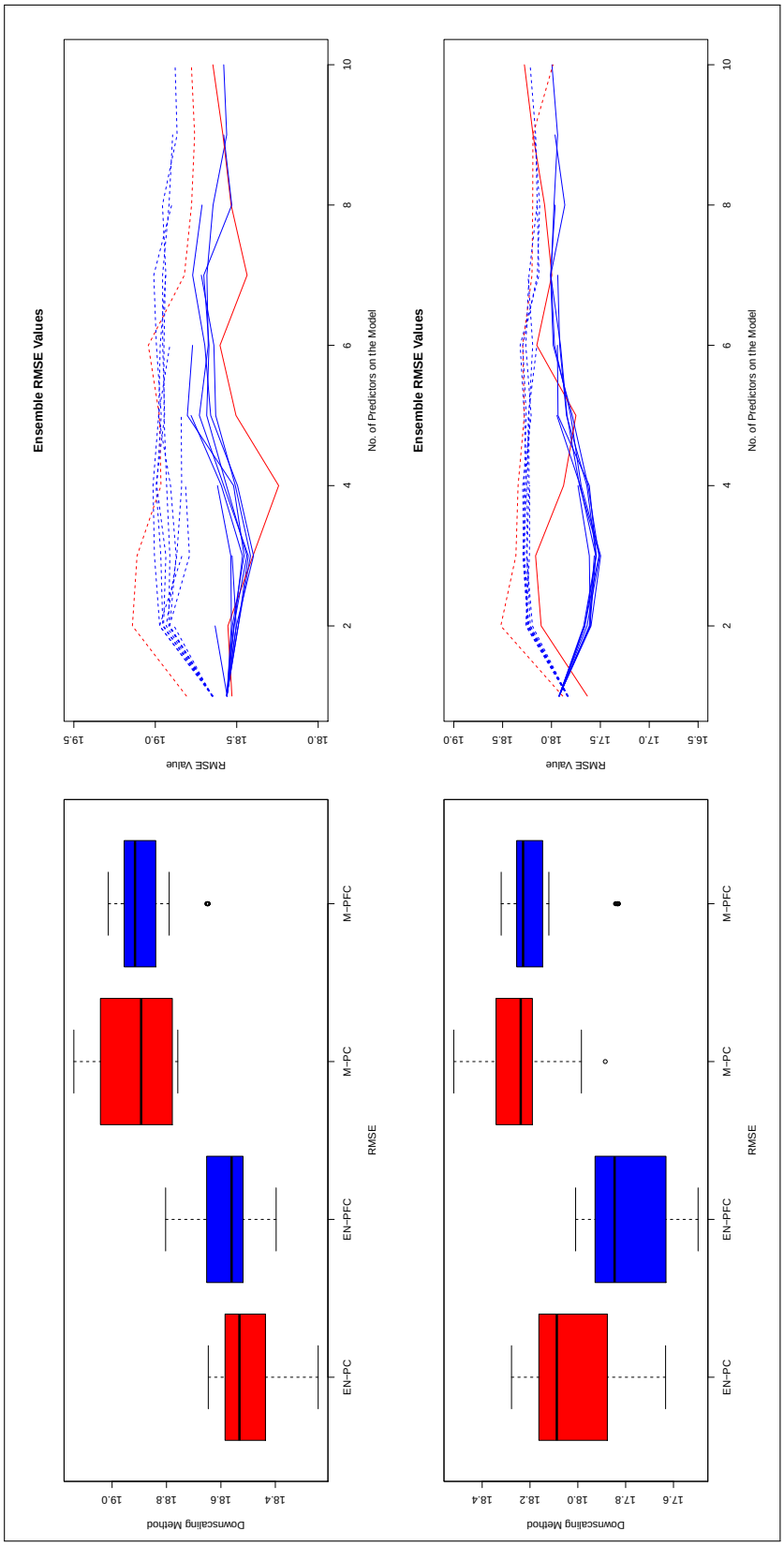

Figure 4.28: RMSE plots for different periods. Each horizontal panel represents the results of the sample period. The plots show the RMSE values for PC and PFC downscaling for the ensemble (DDR method) and for one reference model. The results are show for different models with different number of predictors. The colour code blue represents the results of a PFC method. The colour code red represents the results of a PC model. In the line plot, the dashed line represents the RMSE values of a single reference model. and the solid line represents the RMSE values for ensemble (DDR method). The RMSE values are averaged over all 35 stations.

The plots indicate that downscaling using PFCs outperform PCs. More specifically, ensemble downscaling using PFC-DDR method seem to show better predictive ability than other downscaling methods (except for the sample in the top panel of figure 4.26). In addition, a PFC based method (whether it is a DDR method or a single model method) appears to be superior to PCs as a regression dimension reduction technique.

We repeated our analysis using two scores per model in the second reductive step. That is, for each model we selected the first two scores to carry out the modular reduction, yielding a total of 104 (2×52) variables to be used in the second reduction step. Table 4.17 shows the averaged RMSEs for different regression models and different number of predictors. In this table we use the period from 1 June to 31 July as a fitting period and the period from 1 to 10 August as a validation period. The table shows that the PC predictions improves as the number of PCs in the regression model increase, on the other hand, PFCs show better predictive performance when the number of predictors in the model is relatively small. Moreover, the PC-DDR models appear to outperform the PFC-DDR method only when nine PC scores are included in the model.

For additional insights, we repeated the analysis for 10 different samples and the results are summarized in figures 4.29 to 4.33. The plots show that PC-DDR model have better predictive performance compared to PFC-DDR model only when the regression model contains six or more predictors. This means that PFC-DDR method appears to be a better dimension regression reduction method than PC-DDR. It is worth mentioning that in all cases either using PCs or PFCs, downscaling an ensemble using DDR method seem to yield more accurate forecasts than downscaling a single air quality models (in all plots the solid line, which refers to the ensemble, is usually below the dashed line, which refers to the single model).

No. of Predictors	PC	POLY1	POLY2	POLY3	POLY4	POLY5	POLY6	POLY7	POLY8	POLY9	POLY10
1	19.564	19.195	19.126	19.130	19.128	19.129	19.130	19.131	19.131	19.130	19.129
2	19.353		20.324	20.167	20.162	20.072	20.010	19.981	19.946	19.946	19.942
3	20.097			20.643	20.844	20.763	20.817	20.749	20.729	20.724	20.728
4	19.896				20.334	20.139	20.202	20.135	20.196	20.189	20.230
5	20.051					19.987	20.047	20.231	20.206	20.137	20.122
6	20.004						19.997	19.949	20.139	20.004	20.026
7	19.322							19.922	19.980	20.096	20.144
8	18.596								19.675	19.792	19.881
9	18.021									19.680	19.571
10	18.284										19.380

Table 4.17: RMSE values averaged over the 35 stations in the study area, where two score per model were used in the second reduction stage. The rows represent the number of predictors used to fit the regression model. The first column shows the averaged RMSE values of PC regression models with 1, 2, $\cdots$, and, 10 PC scores. The preceding columns show the averaged RMSE values of PFC regression models. The column label "POLY1" means that the PFCs were computed using a polynomial basis function with degrees one, the column label "POLY2" means that the PFCs were computed using a polynomial basis function with degrees two, and so on.

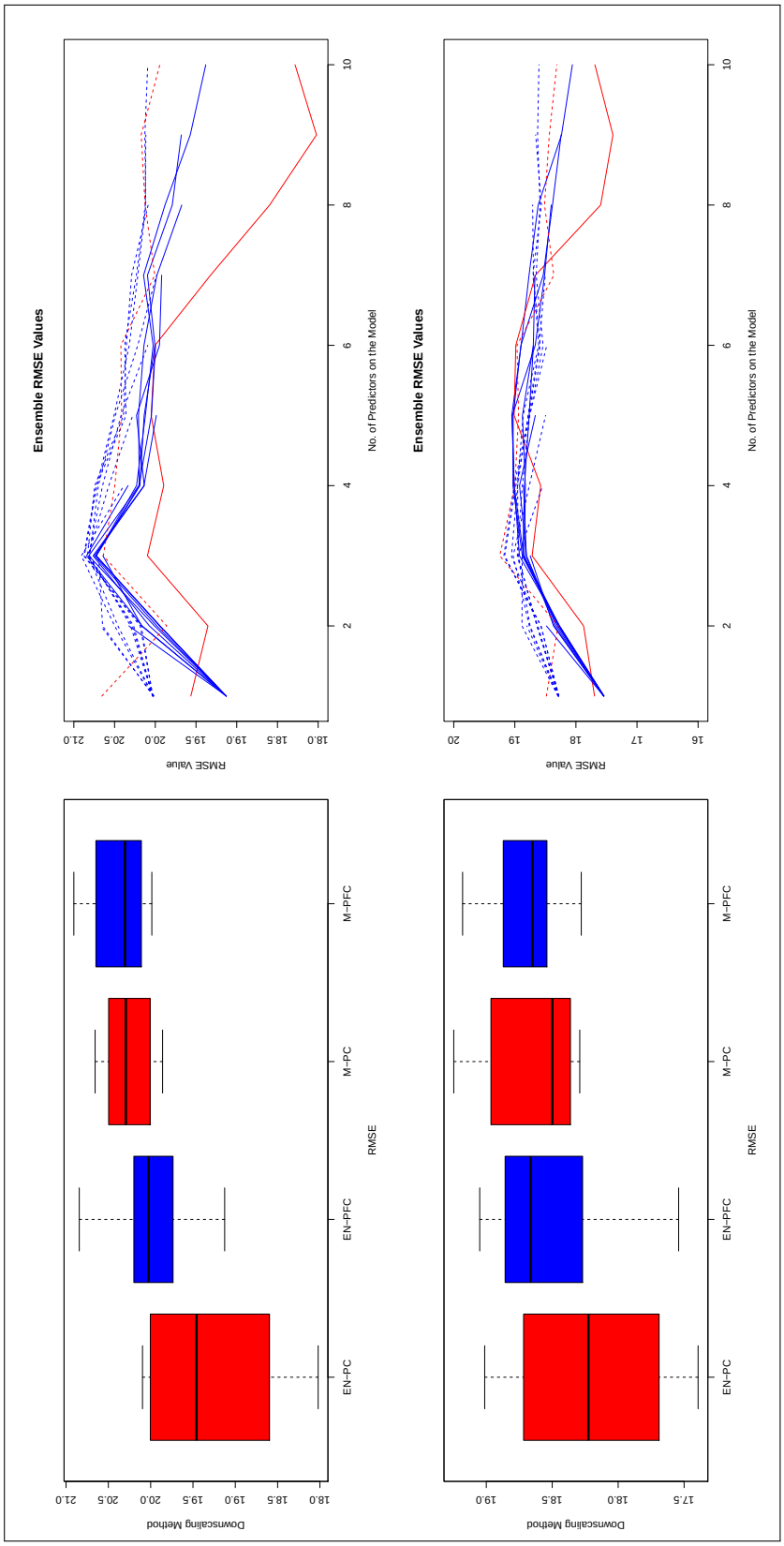

Figure 4.29: RMSE plots for different periods where two scores per model were used in the second reduction stage. Each horizontal panel represents the results of the sample period. The plots show the RMSE values for PC and PFC downscaling for the ensemble (DDR method) and for one reference model. The results are show for different models with different number of predictors. The colour code blue represents the results of a PFC method. The colour code red represents the results of a PC model. In the line plot, the dashed line represents the RMSE values of a single reference model. and the solid line represents the RMSE values for ensemble (DDR method). The RMSE values are averaged over all 35 stations.

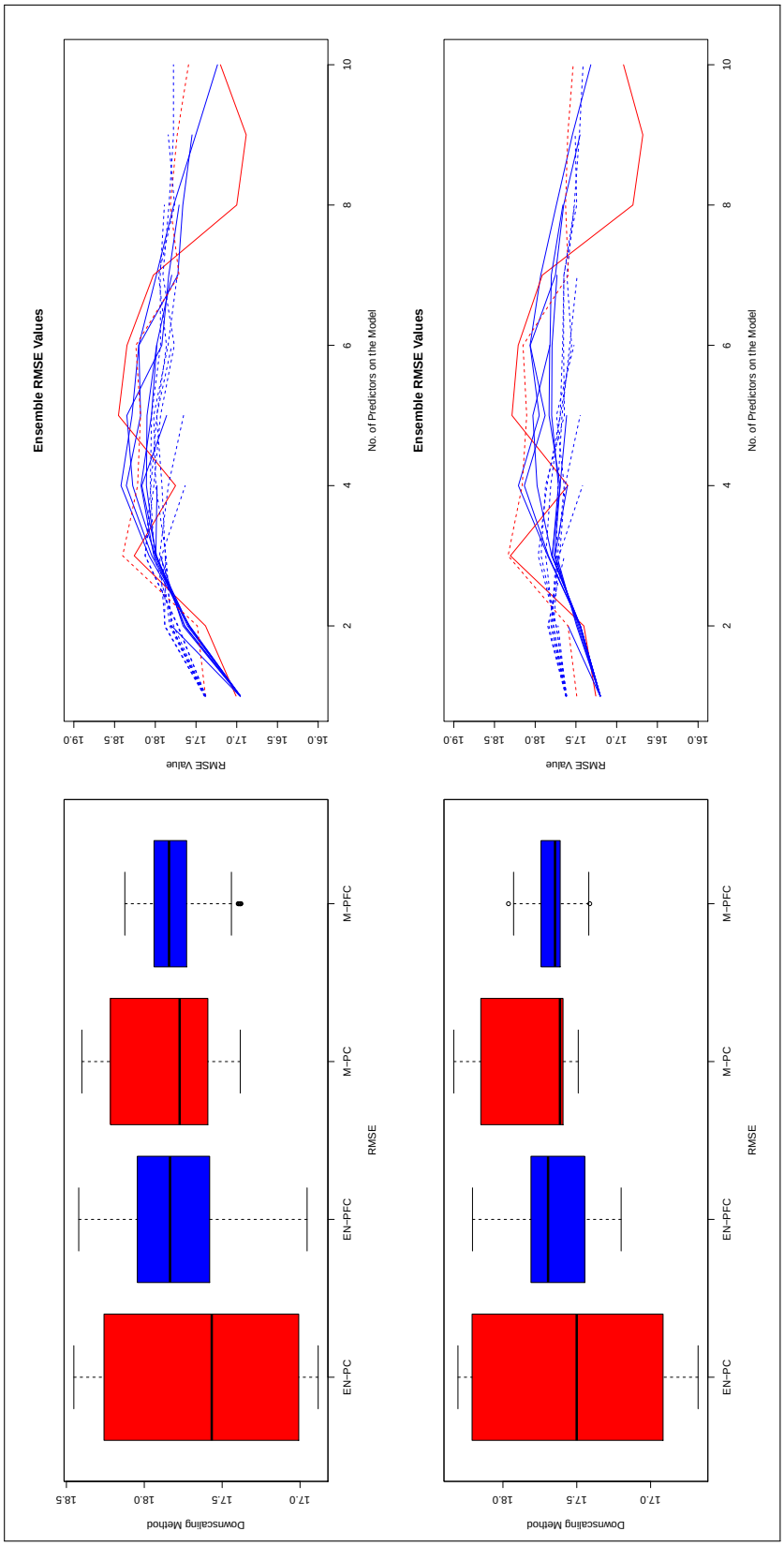

Figure 4.30: RMSE plots for different periods where two scores per model were used in the second reduction stage. Each horizontal panel represents the results of the sample period. The plots show the RMSE values for PC and PFC downscaling for the ensemble (DDR method) and for one reference model. The results are show for different models with different number of predictors. The colour code blue represents the results of a PFC method. The colour code red represents the results of a PC model. In the line plot, the dashed line represents the RMSE values of a single reference model. and the solid line represents the RMSE values for ensemble (DDR method). The RMSE values are averaged over all 35 stations.

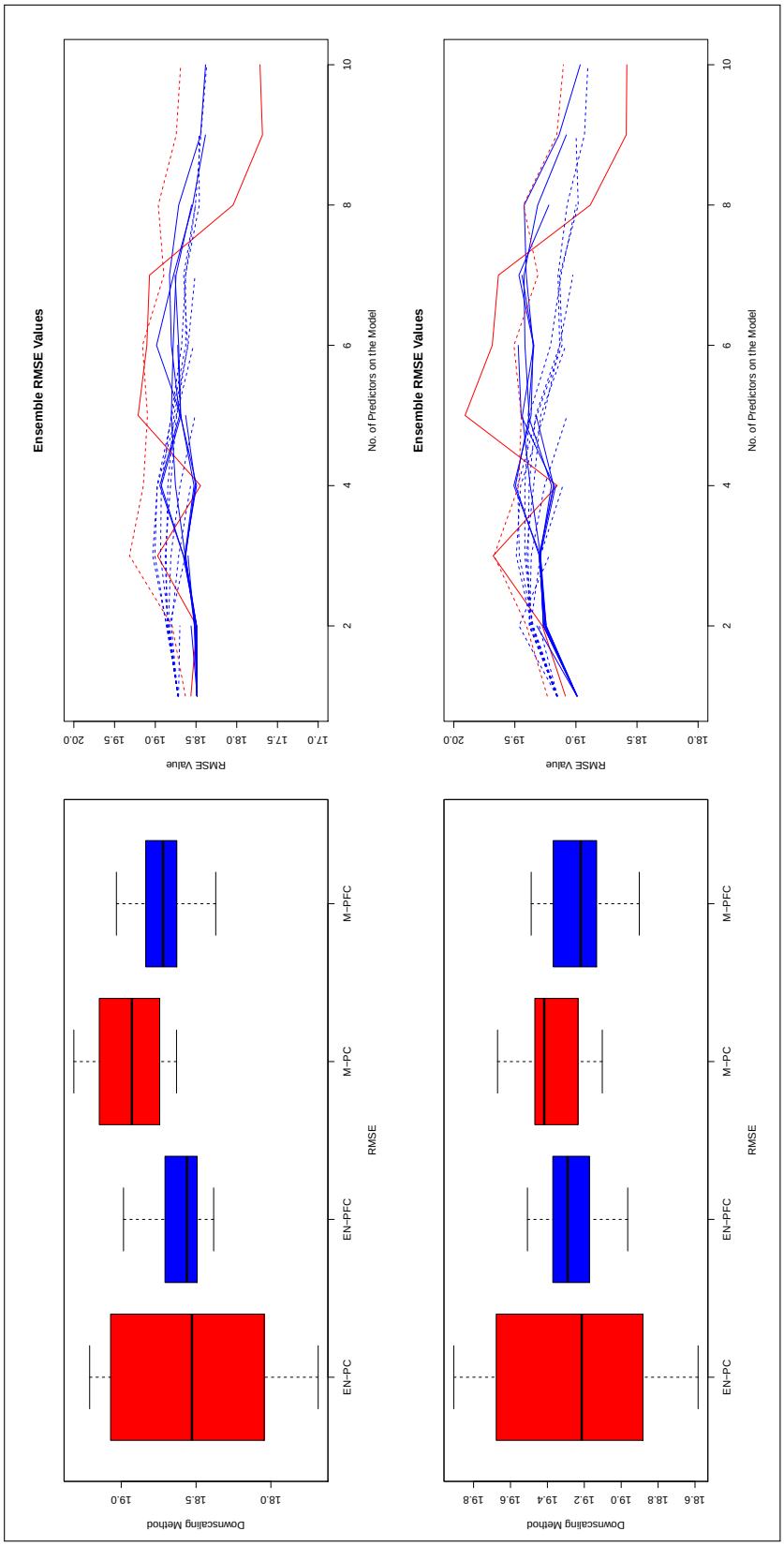

Figure 4.31: RMSE plots for different periods where two scores per model were used in the second reduction stage. Each horizontal panel represents the results of the sample period. The plots show the RMSE values for PC and PFC downscaling for the ensemble (DDR method) and for one reference model. The results are show for different models with different number of predictors. The colour code blue represents the results of a PFC method. The colour code red represents the results of a PC model. In the line plot, the dashed line represents the RMSE values of a single reference model. and the solid line represents the RMSE values for ensemble (DDR method). The RMSE values are averaged over all 35 stations.

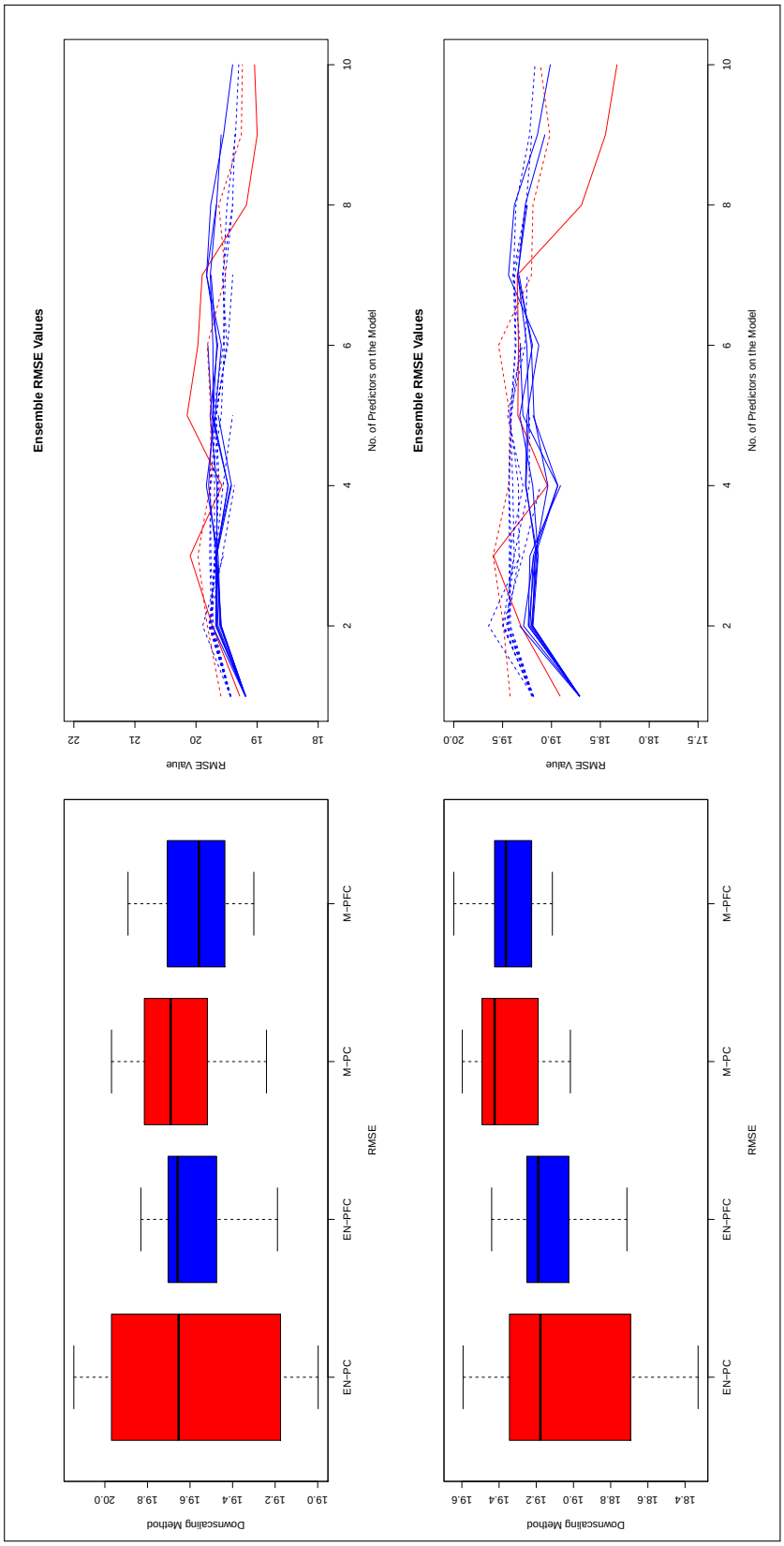

Figure 4.32: RMSE plots for different periods where two scores per model were used in the second reduction stage. Each horizontal panel represents the results of the sample period. The plots show the RMSE values for PC and PFC downscaling for the ensemble (DDR method) and for one reference model. The results are show for different models with different number of predictors. The colour code blue represents the results of a PFC method. The colour code red represents the results of a PC model. In the line plot, the dashed line represents the RMSE values of a single reference model. and the solid line represents the RMSE values for ensemble (DDR method). The RMSE values are 240 data points.

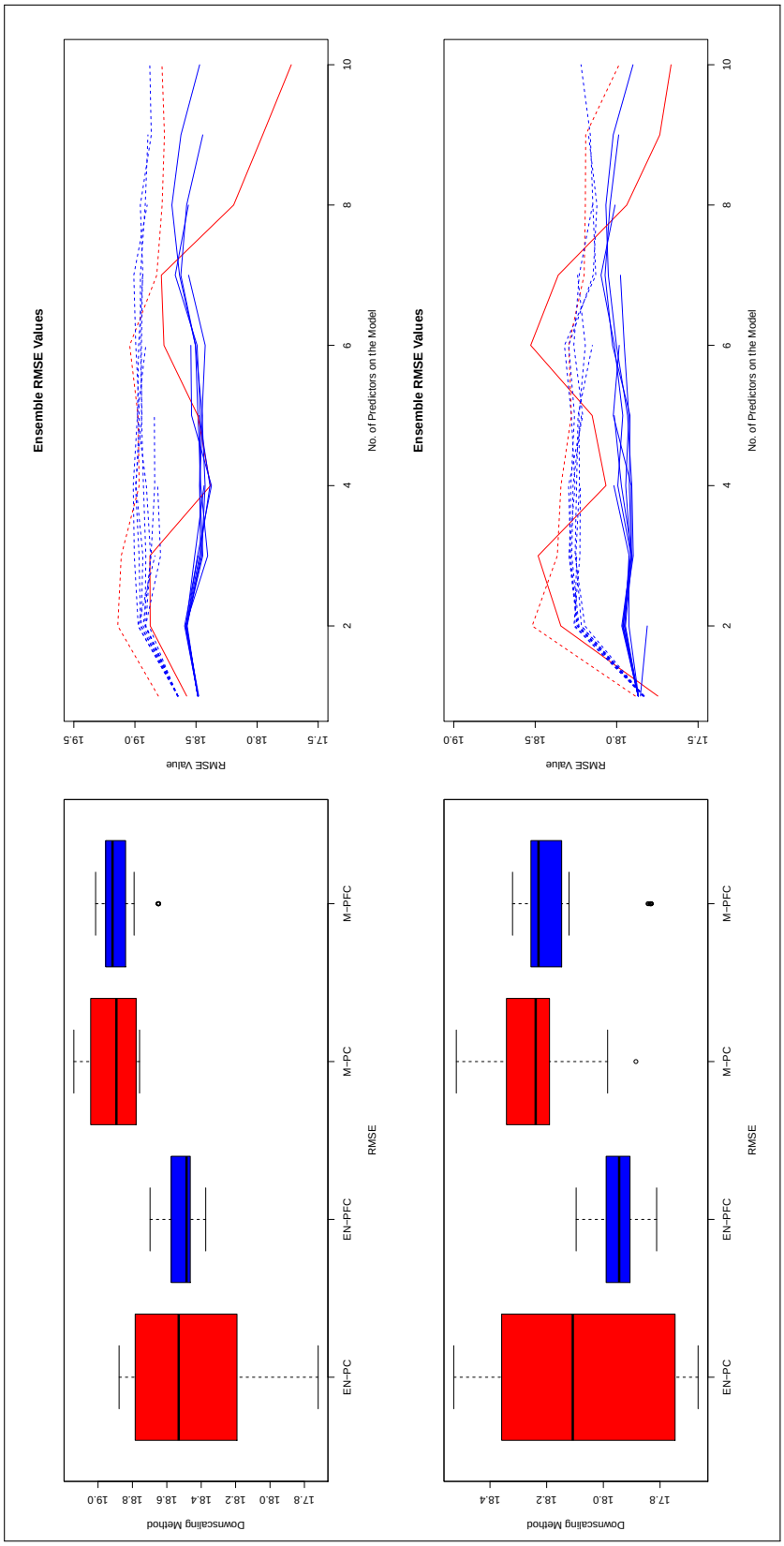

Figure 4.33: RMSE plots for different periods where two scores per model were used in the second reduction stage. Each horizontal panel represents the results of the sample period. The plots show the RMSE values for PC and PFC downscaling for the ensemble (DDR method) and for one reference model. The results are show for different models with different number of predictors. The colour code blue represents the results of a PFC method. The colour code red represents the results of a PC model. In the line plot, the dashed line represents the RMSE values of a single reference model. and the solid line represents the RMSE values for ensemble (DDR method). The RMSE values are averaged over all 35 stations

In summary, the section downscaled an air quality ensemble using a new proposed method called the double dimension reduction technique (DDR). The proposed technique is a two step reduction procedure that serves two purposes. First, to reduce the dimension of ensemble outputs while maintaining sufficient spatial information. The second purpose is to combine the ensemble forecasts, thus, it would be appropriate for the of downscaling. In the first step we filter out redundant spatial information of each ensemble member while in the second stage we use the spatially reduced variables that were obtain in the first step to perform further reduction across the models of an ensemble, yielding in elimination of redundant models.

We performed a simulation study to get an insight on the efficiency of our proposed method. The simulation results indicate that ensemble downscaling outperformed single model downscaling in general which is expected because ensembles account for many uncertainties which a single model cannot capture. Furthermore, the simulation results showed that DDR downscaling would perform best when an ensemble have a good forecasting ability. That is, the DDR technique seems to work best when all or the majority of ensemble members have a satisfactory performance in predicting the measure of interest. We applied the DDR method to downscale ground level ozone outputs over France using the Polyphemus ensemble.

In general, the obtained results coincide with the simulation results when considering only one score per member in the second reductive stage. The results indicate that PFC-DDR downscaling outperform PC-DDR downscaling. We repeated the analysis for different fitting and validations periods and results coincides with the initially chosen period. We repeated the analysis using two scores per ensemble member in the second reductive step. The results show that PC-DDR method seems to outperform PFC-DDR only when the we consider six or more PC scores in the regression model. Although this means better prediction results, this does not serve the purpose of dimension reduction. Generally, either we use PC or PFC in the DDR method, ensemble downscaling seem to have more accurate forecasts than when downscaling a single reference model. The findings in this chapter - sections 4.1 through 4.4 cover several aspects of predictive modelling by highlighting the importance of not only model training and validation, but also the scope for embedding expert knowledge in the models. These characteristics are crucial in model performance - accuracy and reliability. In the next exposition, we outline performance assessment of the PFC downscaling method using alternative methods for performance comparability as suggested in Section 3.4. These results are designed to validate the proposed method.

4.5 Performance Assessment Using Alternative Predictive Methods

This section provides a comparative assessment of PFC model predictions using two predictive modelling techniques - Support Vector Machines (SVM) and Decision Trees (DT) models. To carry out the analysis we use the US ozone data presented in section 3.1.1. For simplicity, we discretise the target variable based on its distributional behaviour. The left and right hand side panels of Figure 4.34 present actual and forecast densities respectively. Both densities exhibit a pronounced bi-modal pattern on the basis of which we discretise the vector using the mean-based rule illustrated in Figure 3.5.

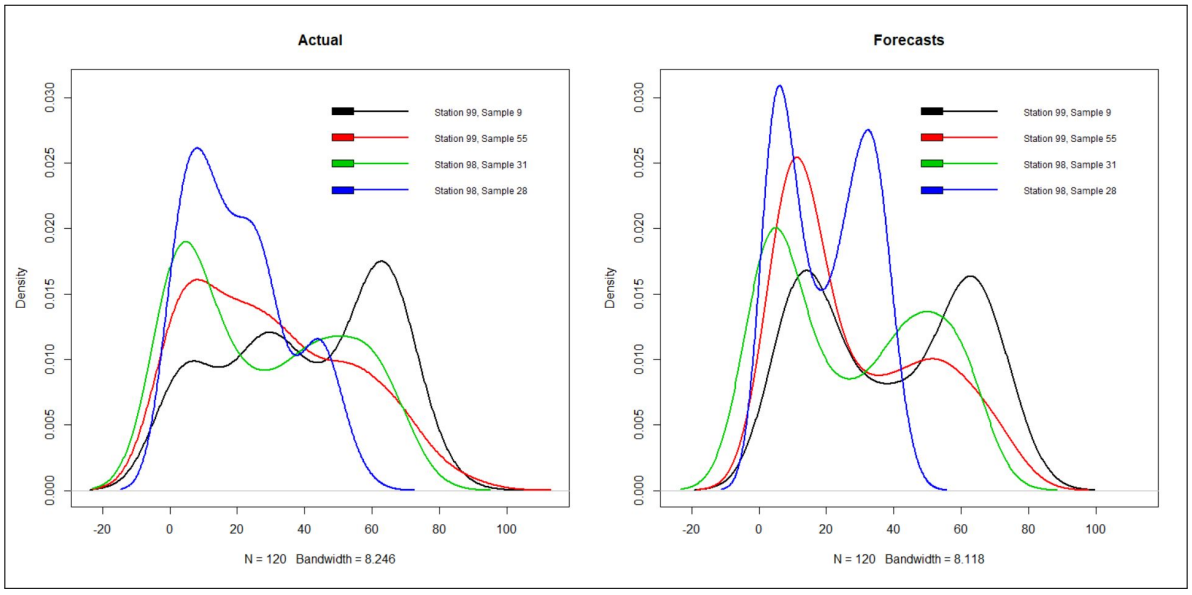

Figure 4.34: Plot of actual (LHS) and forecast (RHS) densities for selected samples

The bimodality of the actual and forecast ozones is evident in the class proportions shown in Table 4.18. The proportions, obtained from discretisation, correspond to the *priors* and *posteriors* in **Algorithm 1** and, depending on the nature and stage of modelling, either of these can be used as input parameters to a predictive model.

Station	Sample	Priors		Posteriors	
		Class One	Class Two	Class One	Class Two
Station 99	Sample 9	0.5166667	0.4833333	0.5	0.5
Station 99	Sample 55	0.4583333	0.5416667	0.4	0.6
Station 98	Sample 31	0.475	0.525	0.4916667	0.5083333
Station 98	Sample 28	0.4666667	0.5333333	0.5	0.5

Table 4.18: Estimated Class Proportions - *priors* and *posteriors*

Comparatives data are sampled from an $\boldsymbol{n} \times \boldsymbol{p}$ data matrix used to do the downscaling in air quality model. The

rows, $\boldsymbol{n = 2203,}$ represent hours of day from 05:00hrs to 23:00hrs on $\mathbf{01}^{st}$ June while columns 1 to 109 are the actual ground level ozone concentrations for 109 monitoring stations scattered across South-Eastern United States and columns 110 to 213 are the Air Quality Model output (REAM output) for 104 grid cells covering the same South-Eastern region. More specifically, we computed two averages across columns one for the ground level ozone concentrations and another for REAM outputs. These were then discretised to form a binary class variable based on the class weights rule in Table 4.18. We then sampled training data from the full data matrix, creating a new matrix $\boldsymbol{X^* \subset X}$ with the discretised class label. The algorithm was trained on this sample and tested on a new unlabelled sample of notionally unmonitored stations which we treat as unobserved and, without loss of generality, we carry out performance comparative analyses modelled on algorithm 1 and Table 4.18. Algorithm 1 provides scope for conducting various performance comparative analyses. Plotting fitted against observed values as the four panels in Figure 4.35 shows, is one way of assessing performance via data visualisation. The four panels represent samples from four different stations. As the locations of these stations are known, any set can be adopted as a regional set from which one or more local stations may be isolated for testing by simply masking actual (observed) values.

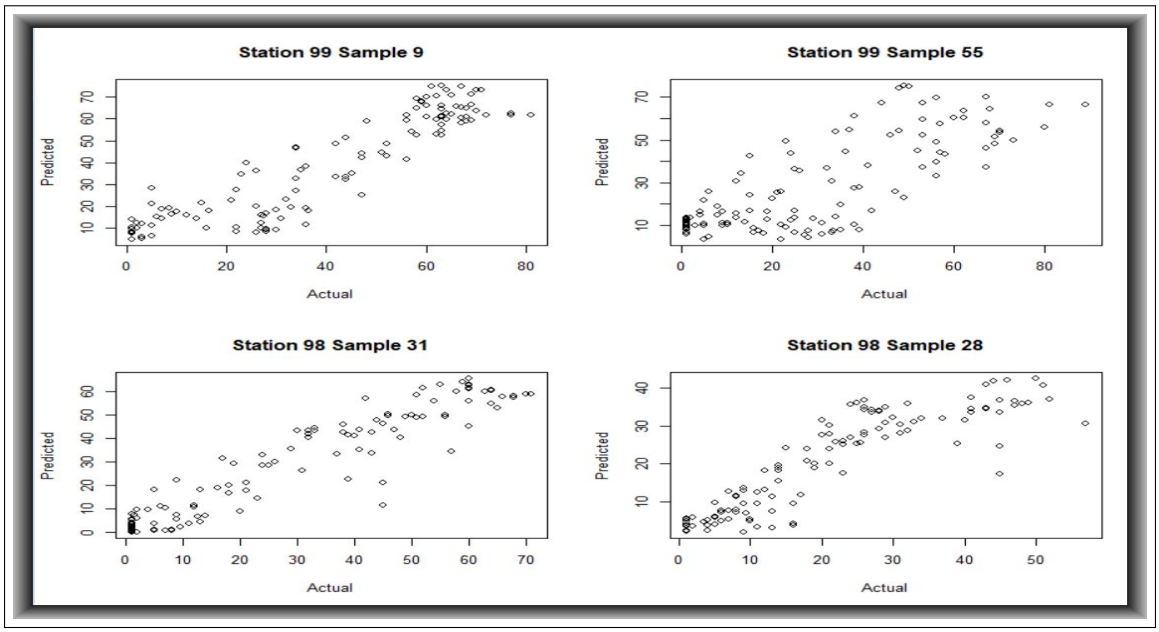

Figure 4.35: Patterns of four stations' PFC forecasts versus observed ozone levels from selected samples

The choice of the test stations can be carried out automatically by setting specific parameters in the algorithm, it can also be via visual inspection. For instance, good predictions such as Station 98 Sample 28 and weakly fitted Station 99 Sample 55 provide good candidates for testing. On the other hand, generated fitted parameters can be retained for use in further analyses - for instance, the confusion matrix will provide information on regression or misclassification

errors; plotting training versus cross-validation errors provide information on model over-fitting, if any.

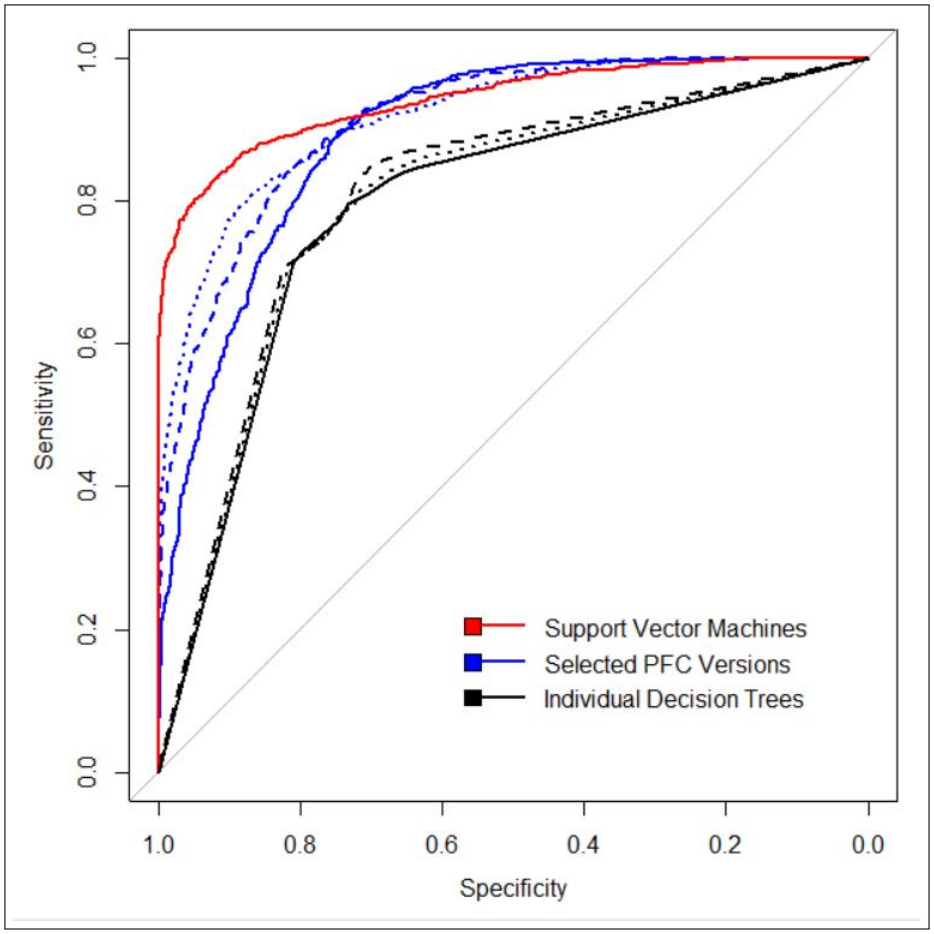

Figure 4.36: ROC curves for SVM, selected PFCs and Decision Trees

One typical performance assessment uses ROC curves for which we follow our setting in Section 3.4.2 in which PFC predictions are derived from an $\boldsymbol{n} \times \boldsymbol{p}$ matrix. We appraise the power of PFCs by adapting downscaling to a standard predictive problem setting the stations and REAM outputs as predictors of ozone levels. Averaging across stations provides a notional regional parameter on the basis of which ozone levels of any of the stations can be predicted and/or tested. Hence, we discretised the average columns following the rule in Figure 3.5 to create two classes, as suggested by the distributional behaviour of the target variable in Figure 4.34. The binarisation creates two categories of ozone - that is, **high** and **low** ozone levels lying above and below the mean respectively. We then trained Support Vector Machines and Decision Trees models on the ozone/REAM output data and tested the models on samples from the same lot not used in training. The ROC curve plots [Egan, 1975] in Figure 4.36 represent the performance assessment of the three methods and the dominance of Support Vector Machines (SVM) is obvious. They were fitted based on the information obtained from repeated searches using different multiple model parameters and recording

the attained performances associated with multiple iso-performance (tangent) lines. Since sensitivity and specificity quantify avoidance of false negatives and false positives respectively, alarm could be triggered for marginally high ozone levels in which case, the PFCs appear to be making good attempts versus SVM. Completely outperformed are decision trees.

It is important to note that while a classifier is usually said to be optimal if it yields results in the north-western corner of the plot, classifier superiority must always be decided by taking variation into consideration which is precisely what Algorithm 1 seeks to achieve. By setting κ large, averaging of ROC curves can yield good, reliable results through cross-validation or bagging techniques. Multiple runs on κ provide scope for measuring the margins by which curves vary and by so doing attain sensible averaging measures. Hence, we iteratively assessed multiple models on the basis of iso-performance lines and, in this case, we assessed PFC performance through comparative analyses versus alternatives modelling techniques - in this case, SVM and Decision Trees.

Chapter 5: Conclusion

This thesis seeks to develop a suitable approach towards improving air quality models forecasts. More specifically, it employs statistical downscaling to improve the predictions of air quality models based on two important issues relating to downscaling air quality model. Firstly, air quality model forecasts tend to have low spatial resolutions and, secondly, the gridded model outputs tend to have relatively large spatial dimensions. It was this issue of high dimensionality that motivated us to focus on performing the downscaling using a dimension reduction methods on the rationale that the proposed dimension reduction method would serve its purpose of reducing dimension while maintaining the influential spatial and regional information provided by the air quality model. Hence our focus was mainly on statistical downscaling using PFC as a newcomer to the air quality downscaling portfolio of techniques. Practical applications of PFCs in statistical dimension reduction, as motivated by Cook [2007], provide scope for reducing, typically, high-dimensional data without ignoring the response variable. This feature makes PFC a sensible choice for downscaling when using regression as well as classification methods. A significant amount of work in the thesis therefore went down to adapting the use of PFCs in air quality downscaling by investigating a wide range of techniques the evolution of which would have helped develop interest in the use of PFCs in downscaling. The work is presented in various scenarios that fall under a similar research umbrella - downscaling of air quality models.

While the work consolidates its knowledge base by investigating into the aforementioned downscaling techniques, the focal point was to develop a novel approach to downscaling air quality model using PFCs which we consider as our first step. The next step was to extend work on PFC downscaling to tackle the problem of inconsistently large covariance matrices as they are crucial to obtaining PFC's key parameters and this is followed by a generalization of the application of PFCs in downscaling by creating a downscaling method to downscale an ensemble of air quality models. Strongly relating to this is an appraisal of the PFC as downscaling technique, which we accomplish via a comparative analysis with other regression and classification-based techniques as detailed in Section 3.4.2. The following is the summary of the work accomplished including our contribution to knowledge.

5.1 Summary of Thesis Results, Contributions, and Limitations

As noted above, the thesis sought to develop, apply and promote the application of PFCs in downscaling air quality models. Further, it aimed at assessing the predictive performance and comparability of selected downscaling models, specifically focusing on downscaling ground level ozone and predicting its concentration levels - the seriousness of ground level ozone is well-documented. Hence, the thesis sets off, in Chapter 1, Section 1.2, by providing clear problem definition and objectives around this scenario as well as providing the work motivation and contribution to knowledge. Current practices are detailed in Chapter 2. Technical details and theoretical underpinnings for the development and deployment of PFCs for downscaling appear in Section 3.2 under Chapter 3 which also provides a description of data sources. Contribution to knowledge is via answering the research question in Section 1.2 and goes through PFC downscaling, model comparability and assessment. The material in the subsections between sections 3.2 and 3.4 is all devoted to answering this question. As noted earlier, its main purpose is to provide fundamental insights as to whether the scientific community is capable of narrowing down the global environmental phenomena via regional discordance and scales - i.e., by blending dynamical and statistical models into robust non-linear methods for a better understanding of regional and global climate data. This section summarises the results and objectives' level of attainment.

5.1.1 Contribution Components

The thesis developed a downscaling method that employs PFCs to reduce the spatial dimension of an air quality model outputs and then proceeded with the downscaling using the spatially reduced variables. Further, the proposed method was compared not only to a wide range of downscaling methods - notably simple regression and PCA regression, but a comparative analysis was extended to other predictive methods as detailed in Section 3.4. These comparisons - based both on real and simulated data provided great support for our proposed approach as they typically exhibited a better performance of PFC than most competing models. Comparisons with predictive models such as Support Vector Machines (SVM) sought to highlight its potential to addressing Bayesian oriented approaches in learning and mapping regional to local features or vice versa. While PFC did not prove superior over SVM its double-role as dimensional-reduction and predictive method still gives it an upper hand, in a downscaling domain.

Component I

In terms of performance, simulation results indicated that PFCs outperform other downscaling methods and provide great predictive ability. PFCs were applied to downscale ozone outputs over the South-Eastern region of the United States and ozone outputs that were simulated by the REAM model were used. Ozone measurements from 94 monitoring stations within the study region made it possible to carry out analyses for each station separately which saw PFC regression outperforming other downscaling methods at the majority of stations in the study area. There were cases in which PFCs did not work well, and these could be attributed to the limited number of grid cells in the study region and there was evidence to suggest that these stations were located at the border of the grid cell domain. These findings are typical in data clustering applications which are quite similar to the downscaling problem being tackled in this thesis. The solution was, therefore, to cross-validate them by repeating our analysis for different fitting and validation periods and the overall findings, obtained based on the mechanics of Algorithm 1, confirmed that PFC downscaling yields more accurate forecasts than the other methods. Results from uncertainty testing by the Jackknife technique were consistent with Cook [2007] - confirming that PFCs outperform PCs as a regression dimension reduction.

One might argue that our method lacks the information that can be obtained by using a spatially and temporally varying coefficient in the regression model such as the downscaler used in Berrocal et al. [2011]. This type of weighted downscalers have the advantage of borrowing strength for coefficients in neighbouring locations and across space, while in our case we obtain a different coefficient for each location (station) individually. Although our technique does not borrow strength from neighbouring grid cell, we believe that our method could be more adaptive to spatial properties that might exist in the region, such as teleconnections and anisotropy. Another important avenue to improve the forecasting ability of the PFC model would be to consider additional factors that affect ozone levels (e.g. temperature). One may think of the common EOFs [Benestad, 2002*b*] as a potential tool for that goal. Initial investigation was done using Temperature as a predictor in the downscaling process in addition to REAM.

Although we would expect the predictive ability of the model to improve, the results showed that adding temperature as a predictor in the downscaling process did not improve the models predictive performance. This might be because REAM already seems to capture well the impact of temperature on ozone. Finally, the uncertainties in the model itself, due to numerical errors or unknown parameterizations of chemistry and transport of ozone and its precursors, ought to also trickle down to the location of interest through downscaling. This is a challenging task that most requires a Bayesian framework to reflect prior scientific knowledge. Consequently, as part of assessing the performance of PFCs, we have set a Bayesian scenario in Section 3.4.2 with corresponding results presented in Section 4.5.

Component II

PFC downscaling was extended to consider another dimension reduction related issue which is the estimation of the covariance matrix. We argued that the covariance matrix is not a good estimate for the actual population matrix if the dimension of the variables is large compared to the size of the sample [Dempster, 1969]. For this reason we were motivated to improve our results by applying matrix regularization prior to computing the PFCs. We adapted a matrix thresholding technique that was proposed by Bickel and Levina [2004] to regularize the covariance matrices that are used to compute PFCs and then proceed with the downscaling using the thresholded PFCs. We performed a simulation study to examine the effect of thresholding the covariance matrix on both PC and PFC downscaling and results showed that thresholding the covariance matrix improved the PC and PFC downscaling performance while thresholded PFCs yielded more accurate predictions than those attained by thresholded PCs. Furthermore, thresholding appeared to significantly improve the forecasting ability of both downscaling models when the sample size was relatively small compared to the dimension. These sample-dependent findings are interesting in that they can be pursued in future applications by adapting the mechanics of Algorithm 1 with new datasets, different $\boldsymbol{\kappa}$ and varying sample sizes.
Results from downscaling of REAM ozone outputs using thresholded PFCs were not consistent with simulation results. In fact, thresholding appeared to marginally worsen the predictive ability of the models - this feature was observed via repeated analyses using different fitting and validation periods sampled from $\boldsymbol{X}$**.** This should not be surprising as, in practice, some techniques might fail to give good estimators as we require when applied to real life data which could be due to the nature of the dataset in hand. That is precisely why EDA is such a fundamental stage in any serious modelling exercises and it is on this reasoning that we set it as part of our very first objective. In this case, lack of improvement in downscaling results could be explained by the fact that for the ozone data at hand, the resulting covariance matrices were already quite sparse and they did not need further regularizing. Another possible reason was that the spatio-temporal characteristic of the ozone data exhibited a systematic ordering that referred to the time point each value was measured at. Further, due to the spatial nature, the ozone data do not have an organized structural order on the plane and so, although we can index the distance between the variables using some kind of metric labelling, we cannot guarantee a clear ordered pattern. Thus because of this ordered/unordered feature of the data in hand, it is worthwhile to apply other regularizing methods that can handle such data. Other regulating approaches such as banding or tapering could be used when the data are ordered, however, they still do not take into account the unordered nature of the data. More sophisticated approaches may be needed for this case.

Component III

Generally, we proposed a new downscaling method that is based on a two-step dimension reduction procedure called double dimension reduction (DDR). The DDR method starts by first eliminating spatial redundancy, and then goes on to use the spatially reduced variables to perform further reduction across ensemble members. The thesis conducted a simulation study to verify the efficiency of the proposed downscaling technique and the findings were compared to the single model downscaling method described in Chapter 3. The simulation results indicated that downscaling an ensemble of models using the DDR method is successful when the majority of ensemble members have an adequate forecasting ability - which is not surprising. Furthermore, downscaling an ensemble using the DDR method yielded better predictive ability than the single model and, most importantly, PFC-DDR downscaling yielded more accurate forecasts that PC-DDR downscaling. We used DDR downscaling to downscale ozone outputs simulated by the Polyphemus ensemble and we performed downscaling over France in the summer of 2005 with convincing results. The implementation procedure involved downscaling ozone predictions of 52 ensemble members and using one measurements for 35 stations scattered over urban areas of the French region. We illustrated our proposed methodology on each measurement network individually and also downscaled a single reference model to compare the results to ensemble downscaling. We used one PFC-score per model in the second reductive step, the downscaling results coincided with the simulation results. That is PFC- DDR downscaling yielded the best predictive ability compared to the other approaches. On the other hand, increasing the number of variables to be used in the second reduction stage, yielded results which were slightly divergent from the simulation results - that is, the PFC-DDR did not show an adequate predictive ability compared to PC-DDR. One possible reason could be because there might still be models within the ensemble that do not predict ozone very well, and by including two scores per model in the second stage we might have increases the inadequate contribution of these models in our downscaling, yielding an insignificant downscaling performance. Nevertheless, in this case PC-DDR only outperformed PFC-DDR when a relatively large number of predictors were used to fit the regression model - indicating that PFCs remains a superior tool.

Finally, comparative assessment of the predictive power of PFC downscaling was conducted by measuring up the method against SVM and DT models. The main idea here was to draw it up against typically predictive modelling techniques without downscaling features. PFC downscaling outperforming of sampled trees and close performance with SVM provided further grounds to assert that our findings have great scope for extension into predictive modelling.

5.2 Potential Future Directions for Research

Like all research projects, it is expected that this work will lead to new research directions. More specifically, we do expect some of its results to trigger new interesting research questions and establish a basis for future work. One natural path towards motivating that goal is to use PFCs in other climate studies to address other research problems other than downscaling. Potentials examples of PFC applications in climate studies are briefly presented below.

5.2.1 Bayesian Approach for Ensemble Downscaling and Calibration:

Many ensemble calibration techniques have been introduced in literature. The gridded nature of ensemble members motivates the calibration of ensemble forecasts because observations for every forecast grid point are available. In particular, techniques such as Bayesian Model Averaging (BMA) [Raftery et al., 2005] that takes each ensemble member into account have gained great deal of importance. The BMA method has been used to combine ensemble models outputs by assigning weights to each model based on their probabilistic information. BMA and PFCs could potentially be used to downscale an ensemble with BMA combining gridded ensemble models outputs and applied to each grid cell output separately in an ensemble of $\boldsymbol{m}$ models, say, as follows

$$\begin{aligned} \boldsymbol{M_1} &= \boldsymbol{x_{n,1}, x_{n,2}, \cdots, x_{n,p}} \\ \boldsymbol{M_2} &= \boldsymbol{x_{n,1}, x_{n,2}, \cdots, x_{n,p}} \\ &\vdots \\ \boldsymbol{M_m} &= \boldsymbol{x_{n,1}, x_{n,2}, \cdots, x_{n,p}} \end{aligned}$$

Where $\boldsymbol{n}$ is the sample size, $\boldsymbol{p}$ the number of grid cells. Combining outputs of each grid cell separately based on BMA yields a set of combined forecasts $\boldsymbol{\mathcal{M} = y_{n,1}, y_{n,2}, \cdots, y_{n,p}}$. Then we can apply PFCs to reduce the dimension of the new combined forecasts and proceed with the downscaling and forecast calibration. A similar approach was used to develop and assess the classifiers used in the comparative analyses of Section 3.4 where the $\boldsymbol{m}$ models correspond to the $\boldsymbol{n \times p}$ matrix representing hours of day from 05:00hrs to 23:00hrs on $\boldsymbol{01^{st}}$ June from the South-Eastern region of the United States with class labels being discretised averages across columns.

5.2.2 Statistical Emulation of an Air Quality Model:

Emulation of a full air quality model directly is not an easy task to accomplish due to their high-dimensionality. Hence, emulation of a full CTM model needs to take into account the input parameters $\boldsymbol{P_t}$ and a state vector of the targeted output $\boldsymbol{y_t}$. Then the CTM updates the state vector $\boldsymbol{y_t}$ at time $\boldsymbol{t}$ to time $\boldsymbol{y_{t+1}}$ at step $\boldsymbol{t+1}$ using the input $\boldsymbol{P_t}$ as follows

$$\boldsymbol{y_{t+1}} = \boldsymbol{\mathcal{M}}(\boldsymbol{y_t}, \boldsymbol{P_t}) \tag{5.1}$$

Hence, in order to emulate the full model we need to reduce the dimension of the parameters and state vector - that is, reduce the dimension of $\boldsymbol{P}$, model inputs $\boldsymbol{x_t}$ and outputs $\boldsymbol{x_{t+1}}$ using a suitable dimension reduction method - which may reasonably be PFCs. As fitting PFCs requires specification of a response variable, a sensible choice for the response variable would be the state vector of the output variable obtained from a previous training period. For example if the purpose of the study requires the emulation of ozone concentrations, we can use its state vector variable as a predictor in the inverse regression to compute the PFCs and select the leading PFC modes of the input, output and parameters before performing the emulation. Evaluation of this method might be through comparison of PFCs and PCs emulations or a comparative analysis involving alternative predictive methods as illustrated in Section 3.4.

The proposed method in this work derived from Cook and Li [2009] and should provide scope for developing new ideas in tackling likelihood-based solutions to dimensional reduction, in particular. There is also great potential for extending applications to predictive modelling - especially as the target variable is not restricted to take only certain values - making it open to both regression and classification problems. One of the main downsides noted by Cook and Li [2009] - i.e., the requirement that $var(X|Y)$ should be constant is a typical distributional assumption can addressed by adopting a Bayesian-based non-parametric approach implemented by algorithm 1. Many problems in real life generate numerical predictors for which this proposed method, via the algorithm, and as such it provides a good source of interdisciplinary comparisons across applications which creates scope for extensions into non-climate applications.

Bibliography

Alkuwari, F. A., Guillas, S. and Wang, Y. [2013], 'Statistical downscaling of an air quality model using fitted empirical orthogonal functions', *Journal of Atmospheric Environment* **81**, 1–10.

Anderson, T. W. [2003], *An Introduction to Multivariate Statistical Analysis*, Wiley.

Beck, B. [2002], 'Model evaluation and performance. encyclopaedia of envirometrics', *Science* **3**, 1275–1279.

Benestad, R. [2001], 'A comparison between two empirical downscaling strategies', *International Journal of Climatology* **21**(13), 1645–1668.
URL: *http://dx.doi.org/10.1002/joc.703*

Benestad, R. [2002*a*], 'Empirically downscaled multimodel ensemble temperature and precipitation scenarios for Norway', *Journal of Climate* **15**(21), 3008–3027.
URL: *http://dx.doi.org/10.1175/1520-0442(2002)015¡3008:EDMETA¿2.0.CO;2*

Benestad, R. [2002*b*], 'Empirically downscaled temperature scenarios for northern Europe based on a multi-model ensemble', *Climate Research* **21**(21), 105–125.

Benestad, R., Hanssen-Bauer, I. and Chen, D. [2008], *Empirical - Statistical Downscaling*, World Scientific Publishing.

Berrocal, V., Gelfand, A. and Holland, D. [2009], 'A spatio-temporal downscaler for output from numerical models', *Journal of Agricultural, Biological, and Environmental Statistics* **15**(2), 176–197.

Berrocal, V., Gelfand, A. and Holland, D. [2010], 'A bivariate space-time downscaler under space and time misalignment', *The Annals of Applied Statistics* **4**(4), 1942–1975.

Berrocal, V. J., Gelfand, A. E. and Holland, D. M. [2011], 'Space-time data fusion under error in computer model output: An application to modeling air quality', *Biometrics* .

Bey, I., Jacob, D. J., Yantosca, R., Logan, J., Field, B., Fiore, A., Li, Q., Liu, H., Mickley, L. and Schultz, M. [2001],

'Global modeling of tropospheric chemistry with assimilated meteorology: Model description and evaluation', *Journal of Geophysical Research* **106**.

Bickel, P. J. and Levina, E. [2004], 'Some theory for fisher's linear discriminant function, 'naive bayes', and some alternatives when there are many more variables than observations', *Bernoulli* **10**(6), 989–1010.
URL: *http://dx.doi.org/10.3150/bj/1106314847*

Bickel, P. J. and Levina, E. [2008], 'Covariance regularization by thresholding', *The Annals of Statistics* **36**(6), 2577–2604.

Boutahar, J., Lacour, S., Mallet, V., Quelo, D., Roustan, Y. and Sportisse, B. [2004], 'Development and validation of a fully modular platform for numerical modelling of air pollution: Polair', *International Journal of Environment and Pollution* **22**(1), 17–28.
URL: *http://inderscience.metapress.com/content/N39E0W3UVJ51DDLB*

Chandler, R. E. [2013], 'Exploiting strength, discounting weakness: combining information from multiple climate simulators', *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* **371**(1991).

Chen, S.-T., Yu, P.-S. and Tang, Y.-H. [2010], 'Statistical downscaling of daily precipitation using support vector machines and multivariate analysis', *Journal of Hydrology* **385**(1), 13–22.

Chiaromonte, F. and Martinelli, J. [2002], 'Dimension reduction strategies for analyzing global gene expression data with a response', *Mathematical Biosciences* **176**(1), 123 – 144.
URL: *http://www.sciencedirect.com/science/article/pii/S0025556401001067*

Choi, Y., Wang, Y., Cunnold, D., T., Z., Shim, C., Luo, M., Eldering, A., Bucsela, E. and Gleason, J. [2008], 'Spring to summer northward migration of high O_3 over the western North Atlantic', *Geophysical Research Letters* **35**.

Choi, Y., Wang, Y., Zeng, T., Cunnold, D., Yang, E., Martin, R., Chance, K., Thouret, V. and E., E. [2008], 'Springtime transitions of NO_2, CO, and O_3 over North America: Model evaluation and analysis', *Journal of Geophysical Research* **113**.

Choi, Y., Wang, Y., Zeng, T., Martin, R., Kurosu, T. and K., C. [2005], 'Evidence of lightning NO_x and convective transport of pollutants in satellite observations over North America', *Geophysical Research Letters* **32**.

Coelho, C. A. S., Stephenson, D. B., Doblas-Reyes, F. J., Balmaseda, M., Guetter, A. and van Oldenborgh, G. J. [2006], 'A Bayesian approach for multi-model downscaling: Seasonal forecasting of regional rainfall and river flows in South America', *Meteorological Applications* **13**(1), 73–82.
URL: *http://dx.doi.org/10.1017/S1350482705002045*

Cook, R. D. [2007], 'Fisher lecture: Dimension reduction in regression', *Statistical Science* **22**(1), 1–26.

Cook, R. and Forzani, L. [2008], 'Principal fitted components for dimension reduction in regression', *Statistical Science* **23**(4), 485–501.

Cook, R. and Li, L. [2009], 'Dimension reduction in regressions with exponential family predictors', *Journal of Computational and Graphical Statistics* **18**(3), 774–791.

Damien, G. and Mallet, V. [2010], 'Automatic generation of large ensembles for air quality forecasting using the Polyphemus system', *Geoscientific Model Development* **3**(1), 69–85.

Delle Monache, L. and Stull, R. B. [2003], 'An ensemble air-quality forecast over western Europe during an ozone episode', *Atmospheric Environment* **37**(25), 3469–3474.
URL: *http://www.sciencedirect.com/science/article/pii/S1352231003004758*

Dempster, A. [1969], *Elements of continuous multivariate analysis*, Addison-Wesley series in behavioral sciences. Quantitative methods.

Derwent, D., Fraser, A., Abbott, J., Jenkin, M., Willis, P. and Murrells, T. [2010], 'Evaluating the performance of air quality models', *www.defra.gov.uk* **3**, 1–77.

Duan, Q., Ajami, N., Gao, X. and Sorooshian, S. [2007], 'Multi-model ensemble hydrologic prediction using bayesian model averaging', *Advances in Water Resources* **30**(5), 1371 – 1386.
URL: *http://www.sciencedirect.com/science/article/pii/S030917080600220X*

Eder, B., Kang, D., Mathur, R., Pleim, J., Yu, F., Otte, T. and Pouliot, G. [2009], 'A performance evaluation of the national air quality forecast capability for the summer of 2007', *Atmospheric Environment* **43**(14), 2312 – 2320.

Efron, B. and Tibshirani, R. J. [1994], *An Introduction to the Bootstrap*, Chapman and Hall.

Egan, J. P. [1975], 'Signal detection theory and roc analysis', *Academic Press* .

El Karoui, N. [2007], 'Tracy–widom limit for the largest eigenvalue of a large class of complex sample covariance matrices', *The Annals of Probability* **35**(2), 663–714.
URL: *http://dx.doi.org/10.1214/009117906000000917*

El Karoui, N. [2008], 'Operator norm consistent estimation of large-dimensional sparse covariance matrices', *The Annals of Statistics* **36**(6), 2717–2756.
URL: *http://dx.doi.org/10.1214/07-AOS559*

EPA, U. [2006], Air quality criteria for ozone and related photochemical oxidants (2006 final), Technical report, Washington, DC.

Etzion, D. and Aragon-Correa, J. [2016], 'Big data, management, and sustainability: Strategic opportunities ahead', *Organization and Environment* **29**(2), 147–155.
URL: *http://epubs.surrey.ac.uk/810620/*

Ezzati, M., Cohen, A. J., Anderson, H. R., Ostro, B., Pandey, K., Krzyzanowski, M., Kuenzli, N., Gutschmidt, K., Pope, C. A., Romieu, I., Samet, J. M. and Smith, K. [2004], *Comparative Quantification of Health Risks: Global and Regional Burden of Disease Attributable to Selected Major Risk Factors*, Vol. 1 of *Comparative Quantification of Health Risks: Global and Regional Burden of Disease Attributable to Selected Major Risk Factors*, World Health Organization.
URL: *https://books.google.co.uk/books?id=NzL7MAAACAAJ*

Feddersen, H. and Andersen, U. [2005], 'A method for statistical downscaling of seasonal ensemble predictions', *Tellus A* **57**(3), 398–408.
URL: *http://dx.doi.org/10.1111/j.1600-0870.2005.00102.x*

Freund, Y. and Schapire, R. [1997], 'A decision-theoretic generalization of on-line learning and an application to boosting', *Journal of Computer and System Sciences* **55**, 119–139.

Frey-Buness, F., Heimann, D. and Sausen, R. [1995], 'A statistical-dynamical downscaling procedure for global climate simulations', *Theoretical and Applied Climatology* **50**(3-4), 117–131.
URL: *http://dx.doi.org/10.1007/BF00866111*

Furrer, R. and Bengtsson, T. [2007], 'Estimation of high-dimensional prior and posterior covariance matrices in kalman filter variants', *Journal of Multivariate Analysis* **98**, 227–255.

Galmarini, S., Bianconi, R., Addis, R., Andronopoulos, S., Astrup, P., Bartzis, J., Bellasio, R., Buckley, R., Champion, H., Chino, M., D'Amours, R., Davakis, E., Eleveld, H., Glaab, H., Manning, A., Mikkelsen, T., Pechinger, U., Polreich, E., Prodanova, M., Slaper, H., Syrakov, D., Terada, H. and der Auwera, L. V. [2004], 'Ensemble dispersion forecasting—part ii: application and evaluation', *Atmospheric Environment* **38**(28), 4619 – 4632.
URL: *http://www.sciencedirect.com/science/article/pii/S1352231004004960*

Giorgi, F., Christensen, J., Hulme, M., von Storch, H., Whetton, P., Jones, R., Mearns, L., Fu, C., Arritt, R., Bates, B., Benestad, R., Boer, G., Buishand, A., Castro, M., Chen, D., Cramer, W., Crane, R., Crossly, J., Dehn, M., Dethloff, K., Dippner, J., Emori, S., Francisco, R., Fyfe, J., Gerstengarbe, F., Gutowski, W., Gyalistras, D., Hanssen-Bauer, I., Hantel, M., Hassell, D., Heimann, D., Jack, C., Jacobeit, J., Kato, H., Katz, R., Kauker, F., Knutson, T., Lal, M., Landsea, C., Laprise, R., Leung, L., Lynch, A., May, W., McGregor, J., Miller, N., Murphy, J., Ribalaygua, J.,

Rinke, A., Rummukainen, M., Semazzi, F., Walsh, K., Werner, P., Widmann, M., Wilby, R., Wild, M. and Xue, Y. [2001], *Regional Climate Information- Evaluation and Projections*.

Guillas, S., Bao, J., Choi, Y. and Wang, Y. [2008], 'Statistical correction and downscaling of chemical transport model ozone forecasts over Atlanta', *Atmospheric Environment* **42**(6), 1338–1348.

Hammami, D., Lee, T. S., Ouarda, T. and Lee, J. [2012], 'Predictor selection for downscaling gcm data with lasso', *Journal of Geophysical Research: Atmospheres* **117**(D17).

Hanna, S. R., Chang, J. and Fernau, M. E. [1998], 'Monte carlo estimates of uncertainties in predictions by a photochemical grid model (uam-iv) due to uncertainties in input variables', *Atmospheric Environment* **32**(21), 3619 – 3628.
URL: *http://www.sciencedirect.com/science/article/pii/S1352231097004196*

Hanna, S. R., Lu, Z., Frey, H. C., Wheeler, N., Vukovich, J., Arunachalam, S., Fernau, M. and Hansen, D. A. [2001], 'Uncertainties in predicted ozone concentrations due to input uncertainties for the uam-v photochemical grid model applied to the july 1995 {OTAG} domain', *Atmospheric Environment* **35**(5), 891 – 903.
URL: *http://www.sciencedirect.com/science/article/pii/S1352231000003678*

Hay, L. and Clark, M. [2003], 'Use of statistically and dynamically downscaled atmospheric model output for hydrologic simulations in three mountainous basins in the western united states', *Journal of Hydrology* **282**(1), 56 – 75.
URL: *http://www.sciencedirect.com/science/article/pii/S002216940300252X*

Hessami, M., Gachon, P., Ouarda, T. and A., S.-H. [2008], 'Automated regression-based statistical downscaling tool', *Environmental Modelling and Software* **23**(6), 813–834.

Houyoux, M., Vukovich, J. and Brandmeyer, J. [2000], *Sparse Matrix Kernal Emission Modeling System: SMOKE User Manual. MCNC-North Carolina Supercomputing Center.*
URL: *http://www.smoke-model.org*

Huang, J., Liu, N., Pourahmadi, M. and Liu, L. [2006], 'Covariance matrix selection and estimation via penalized normal likelihood', *Biometrika* **93**(1), 85–98.

Jacob, D. [2004], *Introduction to Atmospheric Chemistry*, Princeton University Press.
URL: *http://www.jstor.org/stable/j.ctt7t8hg*

Jenkins, G. S. and Barron, E. J. [1997], 'Global climate model and coupled regional climate model simulations over the eastern united states: Genesis and regcm2 simulations', *Global and Planetary Change* **15**(1–2), 3–32.
URL: *http://www.sciencedirect.com/science/article/pii/S0921818196000112*

Jeong, D., St-Hilaire, A., Ouardab, T. and Gachond, P. [2013], 'A multi-site statistical downscaling model for daily precipitation using global scale gcm precipitation outputs', *International Journal of Climatology* **33**, 24312447.

Johnson, O. [2008], 'Theoretical properties of Cook's PFCs dimension reduction algorithm for linear regression', *Electronic Journal of Statistics* **2**, 807–828.

Johnstone, I. [2001], 'On the distribution of the largest eigenvalue in principal components analysis', *The Annals of Statistics* **29**(2), 295–327.

Johnstone, I. and Lu, A. Y. [2004], 'On consistency and sparsity for principal components analysis in high', *Unpublished manuscript* .

Kidson, J. and Thompson, C. [1998], 'A comparison of statistical and model-based downscaling techniques for estimating local climate variations', *Journal of Climate* **11**(4), 735–753.

Kim, J. W., T., C. J., Baker, N. L., Wilks, D. S. and Gates, W. L. [1984], 'The statistical problem of climate inversion: Determination of the relationship between local and large-scale climate', *American Meteorological Society* **112**(10), 2069–2077.

Klien, W. H. [1963], 'Specification of precipitation from the 700-millibar circulation', *Monthly Weather Review* **91**(10), 527–536.
URL: *http://dx.doi.org/10.1175/1520-0493(1963)091¡0527:SOPFTC¿2.3.CO;2*

Krishnamurti, T. N., Kishtawal, C. M., Zhang, Z., LaRow, T., Bachiochi, D., Williford, E., Gadgil, S. and Surendran, S. [2000], 'Multimodel ensemble forecasts for weather and seasonal climate', *Journal of Climate* **13**(23), 4196–4216.

Kryzhov, V. [2012], 'Downscaling of the global seasonal forecasts of hydromet center of Russia for north Eurasia', *Russian Meteorology and Hydrology* **37**(5), 291–297.
URL: *http://dx.doi.org/10.3103/S1068373912050019*

Krzanowski, W. J. and Hand, D. J. [2009], 'Roc curves for continuous data', *Chapman and Hall* .

Krzyzanowski, M. and Cohen, A. [2008], 'Update of who air quality guidelines', *Air Quality, Atmosphere & Health* **1**(1), 7–13.
URL: *https://doi.org/10.1007/s11869-008-0008-9*

Laflamme, E., Linder, E. and Pan, Y. [2016], 'Statistical downscaling of regional climate model output to achieve projections of precipitation extremes', *Weather and Climate Extremes* **12**, 15–23.

Lee, S. M., Princevac, M., Mitsutomi, S. and Cassmassi, J. [2009], 'MM5 simulations for air quality modeling: An application to a coastal area with complex terrain', *Atmospheric Environment* **43**(2), 447 – 457.

Li, L. and Li, H. [2004], 'Dimension reduction methods for microarrays with application to censored survival data', *Bioinformatics* **20**(18), 3406–3412.
URL: *http://bioinformatics.oxfordjournals.org/content/20/18/3406.abstract*

Lorenz, E. N. [1956], Empirical orthogonal functions and statistical weather prediction, Technical Report 1, M.I.T. Statistical Forecasting Project.

Mallet, V. and Sportisse, B. [2005], Data processing and parameterization in atmospheric chemistry and physics: the AtmoData library, Technical report.

Mallet, V. and Sportisse, B. [2006], 'Uncertainty in a chemistry-transport model due to physical parameterization and numerical approximations: An ensemble approach applied to ozone modeling', *Journal of Geophysical Research* **111**(D01302).

Marcenko, V. and Pastur, L. [1967], 'Distribution of eigenvalues for some sets of random matrices', *Mathematics of the USSR-Sbornik* **1**(4), 507–536.

McKeen, S., Chung, S. H., Wilczak, J., Grell, G., Djalalova, I., Peckham, S., Gong, W., Bouchet, V., Moffet, R., Tang, Y., Carmichael, G. R., Mathur, R. and Yu, S. [2007], 'Evaluation of several pm2.5 forecast models using data collected during the icartt/neaqs 2004 field study', *Journal of Geophysical Research: Atmospheres* **112**(D10), n/a–n/a.
URL: *http://dx.doi.org/10.1029/2006JD007608*

Mertens, B., Fearn, T. and Thompson, M. [1995], 'The efficient cross-validation of principal components applied to principal component regression', *Statistics and Computing* **5**, 227–235. 10.1007/BF00142664.

Min, Y. M., Kryjov, V. N. and Oh, J. [2011], 'Probabilistic interpretation of regression-based downscaled seasonal ensemble predictions with the estimation of uncertainty', *Journal of Geophysical Research: Atmospheres* **116**(D8), n/a–n/a.
URL: *http://dx.doi.org/10.1029/2010JD015284*

Ministry of Development Planning and Statistics (Qatar) [n.d.], 'Environmental Statistics Annual Report', www.mdps.gov.qa, Year = 2013.

Monteiro, A., Ribeiro, I., Tchepel, O., Carvalho, A., Martins, H., Sá, E., Ferreira, J., Martins, V., Galmarini, S., Miranda, A. and Borrego, C. [2013], 'Ensemble techniques to improve air quality assessment: Focus on o3 and pm', *Environmental Modeling and Assessment* **18**(3), 249–257.
URL: *http://dx.doi.org/10.1007/s10666-012-9344-0*

Murphy, J. [1999], 'An evaluation of statistical and dynamical techniques for downscaling local climate', *Journal of*

Climate **12**(8), 2256–2284.
URL: *http://dx.doi.org/10.1175/1520-0442(1999)012¡2256:AEOSAD¿2.0.CO;2*

Mwitondi, K. S., Taylor, C. C. and Kent, J. T. [2002], 'Using boosting in classification', *Proceedings of the Leeds Annual Statistical Research (LASR) Conference; Leeds University Press* pp. 125–128.

Neter, J., Kutner, M., Nachtsheim, C. and Wasserman, W. [1996], *Applied Linear Statistical Models*, Mc Graw Hill.

Ngan, F., Byun, D., Kim, H., Lee, D., Rappenglück, B. and Pour-Biazar, A. [2012], 'Performance assessment of retrospective meteorological inputs for use in air quality modeling during texaqs 2006', *Atmospheric Environment* **54**(0), 86 – 96.

Oreskes, N., Shrader-Frechette, K. and Belitz, K. [1994], 'Verification, validation, and confirmation of numerical models in the earth sciences', *Science* **263**, 641–646.

Ostro, D. [2004], *Outdoor Air Pollution: Assessing the Environmental Burden of Disease at National and Local Levels*, Environmental burden of disease series, World Health Organization, Protection of the Human Environment.
URL: *https://books.google.co.uk/books?id=dIpBtwAACAAJ*

Pagowski, M., Grell, G. A., McKeen, S. A., Dévényi, D., Wilczak, J. M., Bouchet, V., Gong, W., McHenry, J., Peckham, S., McQueen, J., Moffet, R. and Tang, Y. [2005], 'A simple method to improve ensemble-based ozone forecasts', *Geophysical Research Letters* **32**(7), n/a–n/a.
URL: *http://dx.doi.org/10.1029/2004GL022305*

Paul, D. [2007], 'Asymptotics of the leading sample eigenvalues for a spiked covariance model', *Statistica Sinica* **17**(4), 1617–1642.

Provost, F. and Fawcett, T. [2001], 'Robust classification for imprecise environments', *Proceedings of Machine Learning: 25th Anniversary of Machine Learning,* **42**(3), 203–231.

Racsko, P., Szeidl, L. and Semenov, M. [1991], 'A serial approach to local stochastic weather models', *Ecological Modelling* **57**(1–2), 27–41.
URL: *http://www.sciencedirect.com/science/article/pii/0304380091900534*

Raftery, A. E., Gneiting, T., Balabdaoui, F. and Polakowski, M. [2005], 'Using bayesian model averaging to calibrate forecast ensembles', *Monthly Weather Review* **133**(5), 1155–1174.
URL: *http://dx.doi.org/10.1175/MWR2906.1*

Rothman, A., Levina, E. and Zhu, J. [2009], 'Generalized thresholding of large covariance matrices', *The Annals of Statistics* (485).

Rotman, D. A., Atherton, C. S., Bergmann, D. J., Cameron-Smith, P. J., Chuang, C. C., Connell, P. S., Dignon, J. E., Franz, A., Grant, K. E., Kinnison, D. E., Molenkamp, C. R., Proctor, D. D. and Tannahill, J. R. [2004], 'Impact, the llnl 3-d global atmospheric chemical transport model for the combined troposphere and stratosphere: Model description and analysis of ozone and other trace gases', *Journal of Geophysical Research: Atmospheres* **109**(D4).
URL: *http://dx.doi.org/10.1029/2002JD003155*

Sahu, S., Gelfand, A. and Holland, D. [2007], 'High resolution space-time ozone modeling for assessing trends', *Journal of the American Statistical Association* **102**(480), 1221–1234.

Schmidt, H. [2002], 'Sensitivity studies with the adjoint of a chemistry transport model for the boundary layer', *Air Pollution Modelling and Simulation* pp. 400–410.

Sloughter, J. M., Gneiting, T. and E., R. A. [2012], 'Probabilistic wind vector forecasting using ensembles and bayesian model averaging', *Monthly Weather Review* **141**(6), 2107–2119.
URL: *http://dx.doi.org/10.1175/MWR-D-12-00002.1*

Sloughter, J. M., Gneiting, T. and Raftery, A. E. [2010], 'Probabilistic wind speed forecasting using ensembles and bayesian model averaging', *Journal of the American Statistical Association* **105**(489), 25–35.
URL: *http://dx.doi.org/10.1198/jasa.2009.ap08615*

Smith, M. and Kohn, R. [2002], 'Parsimonious covariance matrix estimation for longitudinal data', *Journal of the American Statistical Association* **97**(460), 1141–1153.

Solazzo, E., Bianconi, R., Vautard, R., Appel, W. K., Moran, M., Hogref, C., Bessagnet, B., Brandt, J., Christensen, J., Chemel, C., Coll, I., Van der Gon, H. D.and Ferreira, J., Forkel, R., Francis, X. V., Grell, G., Grossi, P., Hansen, A. B. and Galmarini, S. [2012], 'Model evaluation and ensemble modelling of surface-level ozone in europe and north america in the context of aqmeii', *International Journal of Climatology* **53**, 60–74.

States, U. [2007], *The Plain English guide to the Clean Air Act*, U.S. Environmental Protection Agency, Office of Air Quality Planning and Standards Washington, DC.

The Official Journal of the European Union [2009], 'THE EUROPEAN PARLIAMENT AND THE COUNCIL OF THE EURO PEAN UNION'. Regulation (EC) No 1005/2009 of The European Parliament and of the Council of 16 September 2009 on substances that deplete the ozone layer.

Tibshirani, R. [1996], 'Regression shrinkage and selection via the lasso', *Journal of the Royal Statistical Society. Series B (Methodological)* pp. 267–288.

Van Loon, M., Vautard, R., Schaap, M., Bergstrom, R., Bessagnet, B., Brandt, J., Builtjes, P., Christensen, J., Cuvelier, C., Graff, A., Jonson, J. E., Krol, M., Langner, J., Roberts, P., Rouil, L., Stern, R., Tarrason, L., Thunis, P., Vignati,

E., White, L. and Wind, P. [2007*a*], 'Evaluation of long-term ozone simulations from seven regional air quality models and their ensemble', *Atmospheric Environment* **41**, 2083–2097.

Van Loon, M., Vautard, R., Schaap, M., Bergström, R., Bessagnet, B., Brandt, J., Builtjes, P., Christensen, J., Cuvelier, C., Graff, A., Jonson, J., Krol, M., Langner, J., Roberts, P., Rouil, L., Stern, R., Tarrasón, L., Thunis, P., Vignati, E., White, L. and Wind, P. [2007*b*], 'Evaluation of long-term ozone simulations from seven regional air quality models and their ensemble', *Atmospheric Environment* **41**(10), 2083 – 2097.
URL: *http://www.sciencedirect.com/science/article/pii/S1352231006011046*

Vautard, R., Schaap, M., Bergström, R., Bessagnet, B., Brandt, J., Builtjes, P., Christensen, J., Cuvelier, C., Foltescu, V., Graff, A., Kerschbaumer, A., Krol, M., Roberts, P., Rouïl, L., Stern, R., Tarrason, L., Thunis, P., Vignati, E. and Wind, P. [2009], 'Skill and uncertainty of a regional air quality model ensemble', *Atmospheric Environment* **43**(31), 4822 – 4832. Urban Air Quality Selected Papers from the 6th International Conference on Urban Air Quality.
URL: *http://www.sciencedirect.com/science/article/pii/S1352231008009072*

Vrac, M., Drobinski, P., Merlo, A., Herrmann, M., Lavaysse, C., Li, L. and Somot, S. [2012], 'Dynamical and statistical downscaling of the french mediterranean climate: uncertainty assessment', *Natural Hazards and Earth System Sciences* **12**(9), 2769.

Wang, J., Wang, C., Lai, C., Chang, C., Liu, Y., Zhang, Y., Liu, S. and Shao, M. [2008], 'Characterization of ozone precursors in the pearl river delta by time series observation of non-methane hydrocarbons', *Atmospheric Environment* **42**(25), 6233 – 6246.
URL: *http://www.sciencedirect.com/science/article/pii/S1352231008000848*

Wang, X. L., Swail, V. R. and Cox, A. [2010], 'Dynamical versus statistical downscaling methods for ocean wave heights', *International Journal of Climatology* **30**(3), 317–332.

Wang, Y., Choi, Y., Zeng, T., Ridley, B., Blake, N., Blake, D. and Flocke, F. [2006], 'Late-spring Increase of Trans-Pacific Pollution Transport in the Upper Troposphere', *Geophysical Research Letters* **33**(1), L01811.

Wang, Y., Hao, J., McElroy, M. B., Munger, J. W., Ma, H., Chen, D. and Nielsen, C. P. [2009], 'Ozone air quality during the 2008 Beijing Olympics: effectiveness of emission restrictions', *Atmospheric Chemistry and Physics* **9**, 5237–5251.

Weaver, C. P., Cooter, E., Gilliam, R., Gilliland, A., Grambsch, A., Grano, D., Hemming, B., Hunt, S. W., Nolte, C., Winner, D. A., Liang, X.-Z., Zhu, J., Caughey, M., Kunkel, K., Lin, J.-T., Tao, Z., Williams, A., Wuebbles, D. J., Adams, P. J., Dawson, J. P., Amar, P., He, S., Avise, J., Chen, J., Cohen, R. C., Goldstein, A. H., Harley, R. A.,

Steiner, A. L., Tonse, S., Guenther, A., Lamarque, J.-F., Wiedinmyer, C., Gustafson, W. I., Leung, L. R., Hogrefe, C., Huang, H.-C., Jacob, D. J., Mickley, L. J., Wu, S., Kinney, P. L., Lamb, B., Larkin, N. K., McKenzie, D., Liao, K.-J., Manomaiphiboon, K., Russell, A. G., Tagaris, E., Lynn, B. H., Mass, C., Salathé, E., O'neill, S. M., Pandis, S. N., Racherla, P. N., Rosenzweig, C. and Woo, J.-H. [2009], 'A preliminary synthesis of modeled climate change impacts on U.S. regional ozone concentrations', *Bulletin of the American Meteorological Society* **90**(12), 1843–1863.

URL: *http://dx.doi.org/10.1175/2009BAMS2568.1*

Wilby, R., Charles, S., Zorita, E., Timbal, B., Whetton, P. and Mearns, L. [2004], 'Guidelines for use of climate scenarios developed from statistical downscaling methods'.

Wilby, R. L., Wigley, T. M. L., Conway, D., Jones, P. D., Hewitson, B. C., Main, J. and Wilks, D. S. [1998], 'Statistical downscaling of general circulation model output: A comparison of methods', *Water Resources Research* **34**(11), 2995–3008.

URL: *http://dx.doi.org/10.1029/98WR02577*

Wilks, D. S. [1999], 'Interannual variability and extreme-value characteristics of several stochastic daily precipitation models', *Agricultural and Forest Meteorology* **93**(3), 153–169.

URL: *http://www.sciencedirect.com/science/article/pii/S0168192398001257*

Wilks, D. S. [2006], *Statistical Methods in the Atmospheric Sciences*, 2 edn, Academic Press.

Wilks, D. S. and Wilby, R. L. [1999], 'The weather generation game: a review of stochastic weather models', *Progress in Physical Geography* **23**(3), 329–357.

URL: *http://ppg.sagepub.com/content/23/3/329.abstract*

Wong, F., Carter, C. and Kohn, R. [2003], Efficient estimation of covariance selection models, Technical report, SAMSI.

Wu, W. B. and Pourahmadi, M. [2003], 'Nonparametric estimation of large covariance matrices of longitudinal data', *Biometrika* **90**(4), 682–693.

URL: *http://biomet.oxfordjournals.org/content/90/4/831.abstract*

Zeng, T., Wang, Y., Chance, K., Blake, N., Blake, D. and Ridley, B. [2006], 'Halogen-driven low altitude O_3 and hydrocarbon losses in spring at northern high latitudes', *Journal of Geophysical Research.* **111**(D17313), 55–557.

Zhao, C., Wang, Y., Yang, Q., Fu, R., Cunnold, D. and Choi, Y. [2010], 'Impact of east Asian summer monsoon on air quality over China: The view from space', *Journal of Geophysical Research.* **115**(D09301).

Zhao, C., Wang, Y. and Zeng, T. [2009], 'East China plains: A basin of ozone pollution', *Environmental Science and Technology* **43**, 1911–1915.

Zhou, B. and Du, J. [2010], 'Fog prediction from a multimodel mesoscale ensemble prediction system', *Weather and Forecasting* **25**(1), 303–322.

URL: *http://dx.doi.org/10.1175/2009WAF2222289.1*