

Weighted Supervised Functional Principal Components Analysis

Zewen Zhang¹, Chunzheng Cao^{1,*} and Shuren Cao¹

¹ School of Mathematics and Statistics, Nanjing University of Information Science and Technology, Nanjing 210044, China

Abstract. In functional linear regression, a supervised version of functional principal components analysis (FPCA) can automatically estimate the leading functional principal components (FPCs), which not only represent the major source of variation of the functional predictor but also are simultaneously correlated with the response. However, the existing supervised FPCA (sFPCA) is only applicable to single modal functional data. In this paper, we propose a weighted version of supervised FPCA (w-sFPCA) by considering the adaptive weighting of multi-modal functional predictors. The new w-sFPCA not only assigns corresponding weights to each modal of functional predictors, but also automatically estimates the leading FPCs associated with response variables, representing the main sources of variation in functional predictors. The method is demonstrated to have a better prediction accuracy than the conventional sFPCA method by using one real application on meteorological data and two carefully designed simulation studies.

AMS subject classifications: 62H12, 62G05

Key words: Functional data analysis, Functional regression, Supervised learning, Classification.

1 Introduction

Functional data become widely used in various fields such as chemometrics, climatology, economy, image analysis, linguistics, meteorology and other areas. As a consequence, there is a recent interest in methods dealing with functional data that include functional principal component analysis and functional linear models [1–3]. In the study of meteorological problems, functional statistical downscaling method is a new tool to catch the statistical relationship between some regional predictable factors and the large-scale circulation data from Global climate model (GCM) [4, 5].

*Corresponding author. Email addresses: caochunzheng@nuist.edu.cn (C. Cao)

©2023 by the author(s). Licensee Global Science Press. This is an open access article distributed under the terms of the Creative Commons Attribution (CC BY) License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

We study the following functional linear model [6]

$$Y = \beta_0 + \int \beta(t)[X(t) - \mu(t)]dt + \varepsilon, \quad (1.1)$$

where Y is a regional scalar response, β_0 is the intercept, $\mu(t)$ is the mean function, and $X(t)$ stands for the functional large-scale predict factors. Model (1.1) has been used to predict climate change scenarios in the future by linking a scalar response with an integral form of a functional predictor [7].

Due to the presence of response variables, functional predictor and response variables should have correlations [8,9]. Functional partial least-squares regression is a important method to solve this problem [10]. [11] proposed supervised sparse functional principal component, which incorporate supervision information to recover the true functional principal components (FPCs). An supervised functional principal components analysis (sFPCA) has been proposed [12], which is able to borrow the information from the response variable Y to estimate the leading FPCs. The estimated FPCs have a better performance in functional regression work. Supervised FPCA can be applied to two-dimensional functional data [13] as well.

However, a common limitation of the above methods is that when the functional predictors have different levels, the information from all these levels are totally separated from each other [14]. For example, when predicting precipitation, many studies may consider adding one or several humidity factors instead of considering the atmospheric circulation as a predictor [15]. They may adding the specific humidity (hus) of 850hPa or 500hPa only, when the factor has different observations on 17 pressure level [16, 17]. This may lead to the loss of information about this preparatory factor. For another example, for predictors which are observed at only one pressure level, it may have simulated data from different global climate models. Existing studies often use only a single global climate model [18, 19]. There is great uncertainty in the use of climate models for future scenario projections, so statistical downscaling simulations and projections using a single climate model can no longer meet the needs of climate impact assessment. Ensemble methods are beginning to be applied by meteorologists to statistical downscaling to reduce the uncertainty of the projections, thereby improving the level of regional climate prediction [20–22].

In this paper, we proposed a weighted supervised functional principle components analysis (w-sFPCA) method. Our target is to study different levels of predictors at the same time. We use weight parameters to find the relationships between these levels to obtain better prediction performance. The novelty of the paper is twofold. Firstly, we propose a framework to calculate weights to consider two or more types of functional predictors and utilize the scalar response variable, to obtain better prediction performance. Secondly, our estimation algorithm is based on eigenvalue decomposition which is much easier to implement.

The rest of the paper is organized as follows. Details of our method is described in Section 2. Two carefully designed simulation studies are used to evaluate the finite