Modeling Aggregated Functional Data

Ronaldo Dias
UNICAMP, Campinas, Brazil – dias@ime.unicamp.br

In many areas of science, one aims to estimate latent subpopulation mean curves based only on observations of aggregated population curves. By aggregated curves we mean linear combination of functional data that cannot be observed individually. We assume that several aggregated curves with linearly independent coefficients are available, and each aggregated curve is an independent partial realization of a Gaussian process with mean modeled through a weighted linear combination of the disaggregated curves. The mean of the Gaussian process is modeled using B-splines basis expansion methods. We propose a semiparametric, valid covariance function that is modeled as the product of a nonparametric variance function by a correlation function. The variance function is described as the square of a function that is expanded using B-splines basis functions. This results in a nonstationary covariance function and includes constant variance models as special cases. Inference is performed following the Bayesian paradigm allowing experts? opinion, when available, to be accounted for. Moreover, it naturally provides the uncertainty associated with the parameters? estimates and fitted values. We analyze artificial datasets and discuss how to choose among the different covariance models. We focus on two different real examples: a calibration problem for NIR spectroscopy data and an analysis of distribution of energy among different types of consumers. In the latter example, our proposed covariance function captures interesting features of the data. Further analysis of different artificial datasets, as well as computer code and data is available as supplementary material online.

Key words: B-splines, nonparametric regression, chemomentrics