Ecological Prediction with High-Frequency "Big Data" Covariates.

Christopher K. Wikle [a]

[a] 146 Middlebush Hall
Department of Statistics
University of Missouri
Columbia, MO 65211
wiklec@missouri.edu

Abstract: Time-frequency analysis has become a fundamental component of many scientific inquiries. Due to the improvements in technology, the amount of high-frequency signals that are collected for ecological and other scientific processes is increasing at a dramatic rate. Incorporating such information into traditional models is complicated by the inherent differences in temporal scales between the response and the predictors. Salient features of high-dimensional time-dependent outcomes and/or predictors may be difficult to discern through scientific or statistical examination in the time domain. Such features often become more pronounced and possibly more interpretable when considered from a time-frequency perspective. Critically, such time-frequency based representations can be considered analogous to spatial image processes, which can be effectively represented by common reduced-rank methods to deal with the inherent dependence between time-frequency "pixels." When combined with efficient variable selection approaches, such representations can improve prediction, classification, and interpretation of spatial and temporal responses on different scales of resolution than the high-frequency covariates.

In order to facilitate the use of these data in ecological prediction and inference, we present a class of nonlinear multivariate time-frequency functional models that can identify important features of each signal as well as the interaction of signals corresponding to response variables. The proposed methods utilize various methods to estimate time-frequency "images", rank reduction, and stochastic search variable selection to effectively reduce the dimensionality and to identify important time-frequency (and, hence, time-domain) features. The methods are demonstrated through various ecological and environmental examples, such as predicting phenotypic selection from insect communication signals, and predicting spawning success of shovelnose sturgeon on the Lower Missouri River from high-frequency data storage tag information.