Research Statement

Jesse Wheeler

Dynamic systems are the subject of many ecological and financial studies. Recent advances in the collection and storing of data have lead to an abundance of measurements taken on dynamic systems that may evolve over time and space. Such datasets often exhibit spatial-temporal dependence between observations that render many traditional statistical models useless. A modern approach to modeling these data is the use of state-space models, which involve the proposal mathematical equations that describe how unobservable states of the system evolve over time and how measurements of the system variables are obtained. The use of state-space models is appealing because researchers may include mechanisms in the model that reflect our current scientific understanding of the system; this enables researchers to estimate the effect a potential intervention may have on a system, provides a framework to perform statistical testing of our current understanding against alternative explanations, and make inference on unobservable variables that may be scientifically interesting.

Despite the utility of these models, statistical inference of a dynamic system using state-space models remains a challenging task. The primary goal of my current research is to expand the current state-of-the-art for state-space modeling. In particular, I develop methodology, theory, and software for likelihood based inference of state-space models with the *plug-and-play* property, which is that one only needs the ability to simulate from the model in order to perform inference. While these methods have a variety of applications, most published articles using this class of algorithms are found in the field of Epidemiology, including my own applied research.

While my current research is focused on state-space modeling of dynamic systems, I am interested in many other statistical topics and am open to new research topics and collaborations. Some themes common to each of my research interests are transparency, reproducibility and simplicity. I believe that many of the challenges faced by various scientific disciplines are a direct result of the poor practice of these three principles. I also believe that statisticians must play a more active role in promoting and executing these principles. As such, one of my primary research goals is to positively impact the scientific community by writing papers and developing software that encourage and enable other researchers to incorporate these principles in their own work. An example of how my current work relates to these principles is my contribution to likelihood based inference for auto-regressive, moving-average time series models [1]. In this work I show how current software implementations of the maximum likelihood estimation routine for ARMA models may result in sub-optimal parameter estimates and other misleading results; I then describe an approach to obtain better parameter estimates for these models, and provide a detailed description of the implemented algorithm in order to facilitate reproducibility and transparency.

1 State-space modeling in higher dimensions

The complex, non-linear nature of many dynamic systems of scientific interest make the state-space modeling approach a challenging task. Contemporary approaches that have been successfully used for non-linear, low-dimensional systems are based on sequential Monte Carlo (SMC) techniques. As the dimensionality of the system increases, however, the approximation error of SMC increases exponentially. This necessitates the development of algorithms that can be used to perform inference on high-dimensional non-linear state-space models. Recent research by my research group have enabled some progress in this area [2].

One simplifying assumption that enables likelihood based inference for high-dimensional, non-linear state-space models is that of approximate independence between spatial units. I am currently

developing an algorithm called Block Panel Iterated Filter (BPIF) that will extend the current state-of-the-art for this class of models under this assumption [3].

The algorithms described in [2] and [3] enable the estimation of both system wide (or shared) parameters and unit-specific parameters. My future work in this area includes the natural extension of [2] and [3] to include the ability of estimating random effects for these classes of models.

2 State-space modeling of infectious diseases

The ongoing global COVID-19 pandemic has sparked interest in the statistical modeling of infectious diseases. State-space models are well-suited for epidemiological systems as researchers have long used partially observed dynamical models to describe the development of an infectious disease outbreak. The methodological, theoretical, and computing developments mentioned in Sections 1 and 3 provide epidemiologists with a new set of tools for modeling infectious disease outbreaks. In [4], I provide practical advice on how to use these tools and avoid common mistakes that are made in infectious disease modeling.

3 Software

The development of theory and methods for likelihood based inference of state-space models requires development of new software. This section briefly describes some of the open-source software packages to which I have contributed.

- `arima2`: This package provides useful functions for fitting auto-regressive, integrated, moving-average (ARIMA) models in `R`. The most important function of this package is the `arima` function, which fits ARIMA models using a multiple restart algorithm that results in models with higher likelihoods than the `arima` function of the `stats` package. I am the creator and primary contributor of this package, which is used and described in [1].

- `spatPomp`: This package provides a framework for developing spatiotemporal partially observed Markov process (SpatPOMP) models for data analysis. An implementation of the method described in [2] is found in this package. My contributions to this package include the implementation of new package features, as well as some bug fixes.

- `pomp`: This widely used package provides a framework for developing partially observed Markov process (POMP) models; the `spatPomp` package described previously is an extension of this package. My contributions to this package are bug fixes.

References

[1] J. Wheeler and E. L. Ionides, "Likelihood based inference of ARMA models," *The R Journal*, Working Paper, 2023.

[2] E. L. Ionides, N. Ning, and J. Wheeler, "An iterated block particle filter for inference on coupled dynamic systems with shared and unit-specific parameters," *Statistica Sinica*, pp. pre–published online, 2022.

[3] J. Wheeler and E. L. Ionides, "Block panel iterated filtering for unit-specific parameters," *TBD*, Working Paper.

[4] J. Wheeler, A. Rosengart, J. Zhuoxun, K. Tan, N. Treutle, and E. L. Ionides, "Informing policy via dynamic models: Cholera in Haiti," *The Annals of Applied Statistics*, Submitted, 2023.