**SIAM 2019 Abstract**

*A dynamical systems approach to transforming disparate timescales in data driven equation-free modeling of disease dynamics*

Mollie M. Van Gordon and Joshua L. Proctor

Effective prediction of disease dynamics often requires knowledge of both endogenous and exogenous factors. Incorporating diverse types of information, however, introduces covariates at different observational timescales. Despite advances in time dependency and discretization in data driven equation-free modeling, dissimilarity in time intervals has remained a challenge. We present a method that addresses this limitation, judiciously unifying timescales of input data. To demonstrate, we develop a malaria transmission model. Relevant remote sensing data includes weekly land cover data alongside sub-daily precipitation and temperature. Plasmodium incidence is reported at a weekly to monthly frequency. Our approach rigorously connects these timescales using principles from non-autonomous dynamical systems and a recently developed method for sparse identification of nonlinear dynamics (SINDy). We employ this method in multiple layers: identifying the disease transmission model as well as optimizing the timescale transformation itself. This layered SINDy approach identifies the most appropriate functional form for converting inputs to a unified timescale. Incorporation of environmental and epidemiological data initially at different timescales expands the options for feature selection and increases the model’s predictive power. Applications of this approach are broadly relevant to disease modeling as well as any data driven equation-free modeling incorporating diverse input data.

Key questions:

* How can we identify useful metrics or transformations that best aggregate a fast timescale variable (such as rainfall) to a slower timestep for incorporation into model identification?
* How can we do that aggregation such that it can be optimally incorporated into the feature library of the long timestep system?

Iterative steps:

* Demonstrate identification of a simple memory kernel in the fast timescale variable.
  + Include past timesteps of fast variable as linear, non-interacting features in the slow timescale model identification.
  + Limitations: additive only; no interaction terms.
* Demonstrate simple layered SINDy approach.
  + Build a set of possible aggregation features of the fast variable without interactions with other variables.
  + Conduct model identification with only these features. High accuracies are not expected; coefficient values are not expected to be useful. We are looking for a measure of importance for different possible aggregation functions.
  + Limitations: initial aggegation feature search does not include interactions with long timestep variables; user is responsible for generating fast timescale feature library.
  + [Minimum viable product]
* Options for further exploration:
  + Compare with other dimensionality reduction techniques for exogenous fast timescale variable.
  + Explore possibilities for including interactions between fast and slow variables in the fast timescale aggregation feature search.
  + Explore memory kernel in the initial aggregation step.

Technical complications:

* SINDy is sensitive to coefficient values, normalization, transformation, lambda space, etc.
* Current implementations of SINDy are very slow.

Possible approaches:

* Theory forward: write skeleton code that follows the theoretical design. Troubleshooting for robustness and tractability post hoc.
* Implementation forward: explore and build robust tests and implementations that ensure valid results. Optimize/refactor code for computational tractability. Subsequent theoretical implementations delayed but easier.
* [Timeline dependent on approach]