Modeling and Inference of Mixed Dynamics and Detection of Causal Emergent Features

W. Casey, L. Metcalf, S. Chatterjee, H. Janwa, E. Battifarano, A. Edwards, and Y. Gvili

September 30, 2024

Abstract

Many real-world problems feature non-linear dynamic processes. Classical mathematical models may adequately describe a single dynamic process in isolation,

but are easily obstructed by two natural and simple kinds of phenomenological variations. First, the emergence (or activation) of an additional dynamic process, and second, changes affecting the parameters of an active process.

COVID-19 data offers an important case study expressing these phenomenological variations that deeply challenge the classical SIR epidemiological model, and call for novel mathematical methods to detect and adapt projections around these critical obstructions. We address the modeling issues with a novel mathematical framework that re-envisions data as a mixture of multiple causal generating processes, each subject to possible parameter change-points. The new viewpoint extends non-linear classical models in a manner that overcomes many of these types of phenomenological variations, and enables a highly adaptive modeling closely linked to causal events. To demonstrate, we construct a mixture of logistic models we term Adaptive Logistic Model (ALM), forming appropriate non linear least squares optimization and regularization goals.

To validate the approach, we return to COVID-19 case count forecasting, and compare ALM directly to other forecasting methods. ALM forecast accuracy is competitive with all leading forecast methods, but its greatest utility may be in how it detects changing dynamics (changepoints) and retains far fewer but more interpretable parameters relating naturally to cause and intervening change.

The method can be applied more generally as it adapts well to the multi-generative nature of many time series data problems. We demonstrate examples in hydrology, economics, cyber-security, and social media.

The infection dynamics of SARS-CoV-2 has proved challenging for mathematical modelers; nonetheless, mathematical models that offer reliable, repeatable predictions are essential for informed policy that can reduce human harm. The pandemic has highlighted multiple limitations of standard mathematical modeling tools [15], such as the classical compartmental SIR model for predicting case loads. As the pandemic unfolded and more data become available, a variety of novel factors started to reshape more standard models: [37, 31, 58], vaccination distribution [21, 22], super-spreaders [23], policy shifts [20], behavior [66], economic impacts[44], vaccination hesitancy and misinformation or counterproductive narratives [48] are but some heretofore novel considerations applied to improve infection case load prediction models.

Pandemic modeling has also motivated novel advances and enhancements of standard mathematical models. The standard SIR models have been enhanced to account for multi-surge requirements in [23], multi-layer network models overlays for epidemiology are considered in [60], and self-exciting branching model are considered in [15]. Those models discover similar phenomenological variations, and like our method are similarly motivated to address the shortcomings of standard models. Our contributions are 1) identifying phenomenological variations which challenge classical modeling, 2) developing a framework for mixture model addressing multiple causal dynamic processes, and 3) Connecting inference, forecasting, causal events, change-point and model adaptation with objective optimization within the framework.

Phenomenological variations: We argue, the inherent difficulties of COVID-19 modeling in the last few years appears to stem from at least three kinds of adaptations; we refer to the list below as phenomenological variations. Firstly, that multiple dynamic conditions can arise or be activated. COVID-19 case loads were the result from multiple simultaneously active infection processes corresponding to distinct viral variants. Viral variants were distinct enough to re-infect individuals, suggesting that the infection processes act (at least partially) independently, critically we argue that case load numbers are an aggregation or mixture of case load numbers arising from each strain. Additionally, new strains arise from mutation, introducing time delays between the infection processes as well as differing infection rates. Multi-surge models [23] have also addressed the need for multiple independent dynamic processes. Secondly, that changing parameters of existing processes, polices and behaviors have been aimed directly at modulating infection dynamics (e.g. *flatten the curve*). These effects may include vaccination distributions, or behavioral shifts in responses to dynamics (i.e., isolating when infection rates increase). And *thirdly*, and perhaps most difficult, is that novel non-linear dynamic conditions can still emerge. Early efforts to model COVID-19 case loads, witnessed much speculation about novel non-linear dynamics with various casual possibilities: silentcarriers, super-spreaders, behaviors such as super-spreader events. Our method is designed to address the first two issues and compensate the phenomenological departure they cause.

The inherent difficulties above are neither isolated to infection dynamics, nor less appreciated in other scientific fields. We sense that there are many applications. In particular, the application of virology to economics [59] and other social sciences are ready examples. Other examples from earth science and biology include hydrology, geology, climate change. We illustrate our method on problems from diverse disciplines, which also importantly require policy tools.