

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

William Casey^{1,*}, Leigh Metcalf², Shirshendu Chatterjee³, Heeralal Janwa⁴, Ernest Battifarano⁵, April Edwards¹, and Yaron Gvili⁶

¹United States Naval Academy

²Carnegie Mellon University

³City University of New York, Graduate Center and City College

⁴University of Puerto Rico, Rio Piedras

⁵New York University

⁶Independent Scholar

*corresponding wcasey@usna.edu

ABSTRACT

Many real-world problems feature nonlinear dynamic processes. Classical mathematical models may be adequate to describe a single dynamic process in isolation, but can be easily undermined by two natural and simple kinds of phenomenological variations: the emergence (or activation) of an additional dynamic process, and events that affect 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 to these critical variations.

We address the modeling issues with a novel mathematical framework that reenvision data as a mixture of multiple causal generating processes, each subject to possible parameter change points. The new viewpoint extends nonlinear 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. The new model space unifies a wider class of dynamics and is particularly effective at fitting multi-surge data and explaining key causal events related to surge origination. To demonstrate, we construct a mixture of logistic models termed the Adaptive Logistic Model (ALM), and then formulate appropriate nonlinear least squares optimization and regularization goals, and then apply ALM to data. To validate the approach, we return to COVID-19 forecasting (for case count), 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 (change-points) 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 ALM robustness through data experiments in hydrology, economics, cybersecurity, and social media.

1 Introduction

The infection dynamics of SARS-CoV-2 has proved challenging for mathematical modelers; nevertheless, 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 tools¹, such as the use of the classical compartmental SIR model to predict case loads. As the pandemic unfolded and more data became available, the limitations of classic models became more evident. Modelers identified a variety of novel factors including²⁻⁴: vaccination distribution^{5,6}, super-spreaders⁷, policy changes⁸, behavior changes⁹, economic impacts¹⁰, vaccination hesitancy, and misinformation or counterproductive narratives¹¹ are only some of the novel considerations that needed to be incorporated 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 events in⁷; multilayer network models overlays for epidemiology are considered in¹²; and self-exciting branching models are considered in¹. Those models discover similar phenomenological variations, and, like our method, are similarly motivated to address the shortcomings of standard models. Our approach, demonstrated with ALM, is to unify a wider class of models remixing classical ones, the new class can capture much more complex phenomena than an SIR model. Our contributions are 1) identifying phenomenological variations which challenge classical modeling, 2) developing a framework for a mixture model that addresses multi-causal dynamic processes, and 3)

connects inference, forecasting, causal events, change-point detection, and model adaptation with objective optimization within the framework.

Phenomenological variations: We argue that the inherent difficulties of COVID-19 modeling in the last few years stem from at least three kinds of adaptations; we refer to the list below as *phenomenological variations*. *First*, that multiple dynamic conditions can arise or be activated. COVID-19 case loads were the result of multiple simultaneously active infection processes caused by distinct viral variants. Viral variants were distinct enough to reinfect individuals, suggesting that the infection processes act (at least partially) independently. Additionally, new strains arise from mutation, introducing time delays between the infection processes as well as differing infection rates. Random events (e.g., the arrival of a new strain) that significantly alter the course of evolution are incorporated as change-points. Although these random events are initially unobserved, our procedure can estimate the change points and subsequent dynamics reasonably well and proficiently (meaning the new trend can be identified quickly after the change). Critically we argue that case load numbers are an aggregation or mixture of case load numbers arising from separate strains. Multi-surge models⁷ have also addressed the need for multiple independent dynamic processes. *Second*, that changing parameters of existing processes, policies, and behaviors have been aimed directly at modulating infection dynamics (e.g. *flatten the curve*). These effects may include vaccination distributions or behavioral changes in response to viral dynamics (i.e., isolating when infection rates increase). And *third*, and perhaps most difficult, is that novel nonlinear dynamic conditions can still emerge. Early efforts to model COVID-19 case loads, witnessed much speculation about novel nonlinear dynamics with various casual possibilities: silent-carriers, super-spreaders, behaviors such as super-spreader events. Our method is designed to address the first two issues with flexible model frameworks that can achieve model accuracy even with these phenomenological variations.

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 economics¹³ 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.

2 Background

To address the demands of modeling case load data, we have designed a method that considers mixtures of dynamic processes defined by ordinary differential equations. Our method is further structured to detect and self-adapt to changing parametric regimes. Our method draws upon mathematical aspects from several topical areas: mixture modeling, time series (temporal data) analysis, change-point detection, dynamical system inference and learning, nonlinear least squares, and infection models. To our knowledge (and surprise), no prior work has integrated these aspects into tools capable of addressing the demands of the applications examined in this work.

Mixture models have been explored in various mathematical and data science models that considered parameter estimation for mixture models. The Gaussian Mixture Model and associated Expectation Maximization (EM) algorithm¹⁴ considers mixture data to two Gaussian distributions. Radial Basis Boltzmann machine¹⁵ extends the theory to a larger number of Gaussian distributions. Mixture models have also been considered for temporal data^{16–21} however the aspect of identifying and adapting to change-points is unclear.

Time series analysis has a history of foundational work and many data-oriented methodologies²². Within this domain, Bayesian forecasting models²³ have been extensively explored, including approaches for change-point detection²⁴—especially in scenarios where the goal is to identify departures from a static distribution. Another major line of research emphasizes frequency-domain techniques and detrended time series models such as ARIMA²⁵. However, these traditional methods often struggle to accommodate changes in underlying trends²⁶ and can be computationally expensive²⁷—limitations that our model effectively addresses.

Numerous change-point detection methods have been proposed: Binary Segmentation²⁸, Bayesian models²⁴, and Total Variation Minimization²⁹. Other approaches include^{30–35}. Dynamic processes with changing parameters have also been considered with Non-Homogeneous Poisson Processes^{36,37}; however, these provide no immediate means to scale beyond one or two change-points. Another model along similar lines is overdispersion in Poisson processes³⁸ that has features of an adaptive model.

Although some approaches integrate time series analysis with change-point detection²⁴—closely resembling our own optimization strategy—to our knowledge, existing change-point methods have not addressed mixture models of dynamic equations, nor have they considered even the relatively simple case of growth or decay ODE models. Consequently, theoretical results in this area remain limited. To address these gaps, we provide the mathematical foundations of the ALM model in the Appendix, while reserving a deeper investigation of the model’s robustness for future work.

Dynamical inference models have included system identification³⁹ and, more recently, physics-inspired deep learning^{40–42}. Nonlinear least squares estimation problems have considered temporal data and more generally linear separability⁴³ and combination of nonlinear basis functions in the variable projection method⁴⁴, which we make use of in this work. These

techniques are inherently suited for mixture models, however, application to model growth and decay processes has been underdeveloped until now.

Our method was motivated by and developed to model multiple surges of SARS-CoV-2 case numbers with greater accuracy than compartmental SIR models. The SIR model is the classical mathematical model for epidemiology, while appropriate for a single pathogen affecting an isolated and simple population^{45,46}, it ultimately proved less useful for modeling SARS-CoV-2 spread within the human population. At the outset of COVID-19, the SIR model provided a reasonable approximation for case count; however, as the second wave of cases began, its validity was undermined by a widening discrepancy between model prediction and actual case count data. Unfortunately, as new variants emerged⁴⁷ that reinfected people who had been fully vaccinated for protection against earlier variants, the SIR model was shown to fail in these global circumstances⁴⁸. Also, at the beginning of the pandemic, an attempt was made to fit the logistic model to the cases⁴⁹. However, it was shown that in the absence of re-parameterization, it is a poor fit for the emerging surges. We believe our modification of the model as discussed below solves this problem effectively.

Our method is next demonstrated on the logistic nonlinear ordinary differential equation, which can model both growth and decay processes characterized by two parameters: capacity and rate.

3 Materials and methods

We begin with the logistic equation, a nonlinear ordinary differential equation (ODE) whose solution is the core nonlinear function we use to demonstrate our methodology. Next we consider the problem of parameter estimation. We expand the model using a class of affine combinations over the model class, enabling successful modeling of multiple nonlinear surges of the same mathematical nature. We readdress the estimation problem and provide a regularized objective that, while preventing overfitting, can be designed to detect change-points. We detail the numerical methods enabling fast convergence for model estimation; we discuss the adaptive capability of the model and illustrate its connection to change-point detection and how the model adapts to novel changes in data.

3.1 Nonlinear dynamics: the logistic model

The origin of the logistic model dates to the late 18th and early 19th century. At that time, mathematical and philosophical discussions on the stability of growth were common; one such economic example posed by Malthus is whether abundance should be used to further growth or invested in enhancements toward a utopian society. Mathematical models describing growth available at that time included *arithmetic* and *geometric* growth now respectively termed linear and exponential [Geometric models include a wide class of solutions to difference equations with constant coefficients such as the Fibonacci recurrence $F(n) = F(n-1) + F(n-2)$, whose closed bounded solution: $F(n) = \frac{\phi^n - (1-\phi)^n}{\sqrt{5}}$ combines two geometric growth modes]. These models generally struggled to capture both accelerating growth and any reasonable counterbalancing dynamics to yield an equilibrium. Pierrer-François Verhulst working with Alphonse Quetelet adapted the exponential growth model to include a term with sufficient resistance to limit growth, thus capable of obtaining an equilibrium. During the course of a three-paper sequence in 1838-1847, the logistic model emerged⁵⁰.

Today the logistic model is widely known in its ordinary differential equation (ODE) form, letting $X(t)$ representing mass at time t :

$$\frac{dX}{dt} = \underbrace{rX(t)}_{\text{growth}} \underbrace{(1 - X(t)/K)}_{\text{resistance}}, \quad (\text{M1})$$

and owing to its elegance and expressivity, its applications are numerous and diverse. It is widely applied in many fields: biology, ecology, physics, chemistry, material sciences, electrical engineering, and machine learning where the sigmoid function is an important neural network activation function.

Logistic growth **M1** is a first-order nonlinear ordinary differential equation, having two parameters: capacity K , and rate r . With $r < 0$ the dynamics can also express capacity-limited decay. Solutions for **M1** with initial conditions (t_0, X_0) exist and describe capacity-limited growth or decay under reasonable assumptions for X_0 and K listed in Table 1.

The equation **M1** with initial conditions (IC) of X_0 at time 0 has a closed-form solution:

$$X(t) = \begin{cases} \frac{X_0 K}{(K - X_0)e^{-r(t-t_0)} + X_0} & \text{for } t \geq t_0 \\ 0 & \text{otherwise when } t < t_0. \end{cases} \quad (\text{S1})$$

For $t > t_0$ the solution coincides with the sigmoid function (or the hyperbolic tangent function), a smooth infinitely differentiable function whose domain is the real line:

$$f(t; r, K, t_0, X_0) = \frac{X_0 K}{(K - X_0)e^{-r(t-t_0)} + X_0}. \quad (\text{S2})$$

Slightly different from considering governing dynamics (equation M1) and initial conditions, is to consider a class of solution curves (equation S2) and one observed point on the solution curve. For the logistic model, this alternate framing uniquely determines the same solution, and is particularly useful if the initial conditions are not observed – such as the case of parameter estimation from data.

behavior	r	K	$\lim_{t \rightarrow \infty} X(t)$
capacity-limited growth	$r > 0$	$K > X_0$	K
constant	$r = 0$ or $K = X_0$		X_0
capacity-limited decay	$r < 0$	$X_0 > K$	K

Table 1. Behavior characterization for the solution S1. The logistic equation is capable of expressing capacity-limited growth, and capacity-limited decay, as well as a constant (nonchanging) dynamics. The constraints and asymptotic limits are shown for each case.

The solution S1 and sigmoid function S2 match identically for $t > t_0$, and are capable of expressing either growth or decay, as well as the desired properties, sought by Verhulst, initially explosive then increasingly resistive resulting in frozen equilibrium at capacity.

Additional properties of the mathematical model and derived formulae for the logistic equation can be found in the Appendix. We summarize the main derivations needed.

First, the inflection point of the function $X(t)$ in equation S1, is the point at which growth is momentarily matched by resistance to growth. It is given by:

$$\tau = \frac{\log\left(\frac{K-X_0}{X_0}\right)}{r}. \quad (\text{S3})$$

Second, using the inflection point τ , we can develop formulae to relate the amount of change in the equation S2 to a time interval symmetric about τ . For equation S2, the time when any fraction (say q) of total change (K) occurs can be calculated by solving $X(t) = qK$ for t , the crossing time is given by:

$$t_q = \tau - \frac{\log\left(\frac{q}{1-q}\right)}{r}. \quad (\text{S4})$$

These formulae are useful for interpreting dynamics. For example, when parameters are fixed they allow us to determine a point in time t_q when a fraction q of the total dynamic change for $X(t)$ will be observed as historical. Noting the symmetry in q around $1/2$, we can derive a formula that determines double sided time-bounds for which a given amount of dynamic change occurs (see appendix). Additionally, since the solution for S1 and equation S2 only differ for $t < t_0$, the formula can also be used to measure the difference between an ODE solution with initial conditions and our alternate framing. Further explanations are given in the appendix.

Summary: Not every nonlinear model admits a closed-form solution, but the logistic equation yields helpful and elegant formulae when considering prediction and adaptive aspects of the full model. Still, the logistic function is expressive enough for a host of interesting capacity/rate limited dynamics. For nonlinear ODEs without analytic solutions, numerical techniques can augment approaches along these lines, albeit much more computationally intense augmentations of methodology are required. We continue the methodology assuming the logistic formula as our nonlinear model.

3.2 Logistic model estimation

Given data for a growth/decay process governed by the logistic equation M1, one may ask which curve (S2) fits the data best.

Letting $D = \langle (t_i, d_i) \rangle_{i=1}^N$ be data, the squared residual can be defined by:

$$R(r, K, t_0, X_0 | D) = \sum_{i=1}^N (f(t_i; r, K, t_0, X_0) - d_i)^2.$$

Minimizing $R(r, K, t_0, X_0 | D)$ can be carried out using numerical methods, a technique known as nonlinear least squares regression.

In this setting, the variables t_0, X_0 can be interpreted as any point of the curve. Letting $\theta = \langle r, k, t_0, X_0 \rangle$,

Model estimation is a search extending over the admissible domain $\mathcal{X}$ to determine a minimizer:

$$\hat{\theta} = \arg \min_{\theta \in \mathcal{X}} R(\theta | D).$$

More generally, the model estimation procedure can be further refined and tailored if one knows the error distribution.

3.3 The affine model

To provide some modeling resolution for COVID-19's multiple waves, we use a linear sum of these nonlinear components. We term this additive model composed of logistic components the *Adaptive Logistic Mixture Model* or **ALM Model**. Let $\mathcal{M}$ be the class of all curves from equation S2. Letting $\theta_i = \langle r_i, K_i, t_{0i}, X_{0i} \rangle$, we may account for various time conditions of the causal effect for a set of independent surges indexed by i . A discussion of the admissible set of parameters θ can be found in the mathematical model portion of the supplemental materials. We use $\mathcal{P}$ here.

Let

$$\mathcal{M}^N = \left\{ \sum_{i=1}^N f(t; \theta_i) \mid \theta_i \in \mathcal{P} \right\}, \quad (\text{M1})$$

and let:

$$\mathcal{M}^+ = \bigcup_{J=1}^{\infty} \mathcal{M}^J,$$

that is, the set of all finite sums over members of $\mathcal{M}$.

3.3.1 Nonlinear least squares with regularization

We generalize a procedure for fitting data by residual formulae R_N applied to curves in the class $\mathcal{M}^N$. Next, we describe a regularization term C that acts to select the number of components required, penalizing the over-utilization of additional logistic components. Together, R_N and C provide a residual function R_* for the class $\mathcal{M}^+$, thereby defining the model estimation objective for the additive model.

Letting $D = \langle (t_i, d_i) \rangle_{i=1}^n$ be data, the squared residual displacement from the data to a curve of $\mathcal{M}^N$ specified by $\langle \theta_1, \dots, \theta_N \rangle$ can be defined as:

$$R_N(\theta_1, \dots, \theta_N | D) = \sum_{j=1}^n \left(\left(\sum_{i=1}^N f(t_j | \theta_i) \right) - d_j \right)^2. \quad (\text{R1})$$

A nonlinear least squares procedure can determine efficiently local numerical minimizers on $\mathcal{M}^N$ given data. Let $\Theta = \langle \theta_1, \theta_2, \dots, \theta_N \rangle$, ranging over admissible value set $\mathcal{P}^N$, then:

$$\hat{\Theta}(N) = \arg \min_{\Theta \in \mathcal{P}^N} R_N(\Theta | D). \quad (\text{R2})$$

Letting $\mathcal{P}^+ = \bigcup_{j=1}^{\infty} \mathcal{P}^j$, a minimizer of $\mathcal{M}^+$ can be defined with an appropriately chosen cost function. For $\Theta \in \mathcal{P}^+$ let $N(\Theta)$ measure the number of parameter indices. Finally, the regularized residual is:

$$R_*(\Theta | D) = R_{N(\Theta)}(\Theta | D) + C(N(\Theta)), \quad (\text{R3})$$

where C is a monotonically increasing cost function, is designed to penalize the use of additional parameters for curve estimation. The minimizer is defined by:

$$\hat{\Theta} = \arg \min_{\Theta \in \mathcal{P}^+} R_*(\Theta | D). \quad (\text{R4})$$

3.4 Change prediction and auto-adaptive modeling

Here we describe one of the model's key features: its ability to recognize when a new dynamical regime is emerging and to adapt quickly with new parameters. A new parametric regime or a change-point imputing significant changes to existing parameters can be detected in our model in a variety of ways including: statistical testing, hypothesis testing, and optimized dynamical programming.

During the COVID-19 epidemic, novel strains of the SARS-CoV-2 virus emerged to reinfect individuals who had recovered from earlier strains, thereby generating multiple waves (or surges, for example, see Figure 1). Nonetheless, the effect on cumulative case counts appears to be driven by the aggregation of two independent growth components each with its own parameters. For capacity/rate dynamic evolution such as the other examples below, and as described by the affine model with N surges (equation M1), the model, owing to its regularization term, is equipped to determine when an additional growth component should be included. This can happen when

$$\frac{|R_N(\hat{\Theta}(N) | D) - R_{N+1}(\hat{\Theta}(N+1) | D)|}{C(N+1) - C(N)} < 1. \quad (\text{D1})$$

Said differently, additional parameters, stepping from a model in $\mathcal{M}^N$ to a model in $\mathcal{M}^{N+1}$, is worth it when the reduction in residual outweighs the cost. As such, these events occur as jump processes in $N(\Theta)$ with respect to equation R3.

More generally, we can adopt the view that *distinct causes* can arise at any point, resulting in the addition of new dynamic components that are governed by the same dynamic laws.

Change Detection: We propose two methods to enable the model to self-identify the emergence of novel dynamic components and consider the decision criteria of equation D1 to trigger the use of additional parameters. Also important for policy and prediction is the time required to detect change, as well as the time required for parameters to stabilize after the change is detected. First, by monitoring the skewness for residual fit, the models can determine the plausibility of evidence of a recent change. Letting $r_k = (\sum_{i=1}^N f(t_k|\theta_i)) - d_k$. A skewness test can be applied to r_k , under the assumption that $\langle r_k \rangle_k$ follow a normal distribution with mean $\bar{r}$ and standard deviation σ for data $r[:m]$:

$$\mu_3(m) = \frac{\sum_{k=1}^m (r_k - \bar{r})^3}{(m-1)\sigma^3}. \quad (D2)$$

A threshold can be set once $|\mu_3(m)| > \tau$, and a wider search for extended parameters can be triggered.

Second, by direct model hypothesis testing, letting $D[..k]$ be the observed data up to and including time step k , we extend equation R4 to account for a number of observations. Let

$$\hat{\Theta}_k = \arg \min_{\Theta \in \mathcal{P}^+} R_*(\Theta|D[..k]), \quad (D3)$$

and let $\mathcal{J}_k = N(\hat{\Theta}_k)$ describe the number of parameters best describing data at time k .

At time $k-1$, assume $\mathcal{J}_{k-1} = J$, when data sample for time k is observed, perform modeling in both classes, $\mathcal{M}^J$ and $\mathcal{M}^{J+1}$, then by directly applying the decision criteria of equation D1, the model can self-evaluate if more parameters are required. In this way, the model is constantly considering hypothesis testing in two classes of models, $\mathcal{M}^J$ the current model, and $\mathcal{M}^{J+1}$ which can improve model fits under the assumption that a novel dynamic component has recently emerged.

Additionally, there are variations. Hypothesis testing can be expanded to consider many classes, or more efficiently, one can use the skewness test to trigger hypothesis testing.

The ability of the model to detect change given recent observations, is consistent with other change-point methods, such as online Bayesian change-point detection²⁴; however, note that our method requires no prior distributions for parameters. The ability to self tune fills a gap noted in⁴⁹, where the need for new parameters was recognized. Our method detects change and determines new parameters.

3.5 Numerical algorithms for multi-causal model

The nonlinear least squares problem is most commonly addressed and solved 'locally' with the Levenberg-Marquardt method^{51,52}. Given an initial setting for parameters, a search in parameter space, guided by the objective of decreasing the residual fit, will determine a minimizer (at least locally in the vicinity of the initial parameters). Quicker convergence is offered by various enhancing search strategies: the Trust Region Reflective (TRR) algorithm⁵³ offers improved handling of over/under shooting parameters bounds, and the Variable Projection method⁴⁴ method decomposes the search into two sub-searches: parameters affecting nonlinear kernels and parameters which combine nonlinear kernels into an affine model. Our code, available as open source (see appendix), leverages each of these methods via the Levenberg-Marquardt solvers `curve_fit` within the Python modular library `scipy.optimize`⁵⁴, and the Variable Projection library within the `VarPro` Rust language library⁴⁴.

Our code, leveraging the efficiencies of the Levenberg-Marquardt solvers, can infer model parameters rather quickly. We implement the nonlinear least squares curve-fitting inductively. Starting with a single logistic, we determine ϕ_1 , then building inductively we determine ϕ_n by holding all prior estimated parameters (i.e., $\phi_1, \phi_2 \dots \phi_{n-1}$) fixed, and apply the solver to minimize R_n defined in equation R1. The routine can be stopped when the residual improvements fall below a specified threshold.

Running on a dual-core MacBook Pro (Version Ventura 13.4.1) with 32G of memory, our code implementing reflection solvers from `scipy.optimize`, can determine the parameters for a two surge model given 500 data points in an average 2.5 seconds. When we increase the data set size to 1500 points, curve-fitting can be achieved in an average 3.5 seconds. Our code which implements the `VarPro` method is even faster (see nonlinear least squares in the appendix)

4 Results

We first illustrate the outcome of nonlinear least squares regression for ALM on COVID-19 data in Figure 1. We conclude from its residual, that ALM models effectively COVID-19 case load in ways that other approaches, such as compartmental

SIR models, struggle with. Further, we compare ALM to more than 30 statistical forecasting methods (additional materials on COVID-19 forecasting are found in the appendix) by directly comparing Least Squared Error (LSE) of forecasting to data in a retro-diction study. There we find that ALM with little configuration, is already a notable competitor among forecasting methods, even with 21 or fewer parameters to develop an entire baseline model for over three years of forecasting. An additional feature of ALM includes the elucidation of inflection points. These aspects enhance the explainability of the model and can potentially enhance policymakers reasoning about diverse causes including emergence of novel strains as well as interventions such as policy shifts. ALM achieves outstanding model fit and explainable summary to include surge inflections that feature multiple growth phenomena as time localized surges.

Next, we consider, more broadly, applications of ALM by examining model related problems in various science fields where capacity/rate parameterized growth/decay components are episodically active due to potentially hidden causes. Further, we discuss ALM's potential enhancing use in those problem domains. We discuss change prediction and what insights change prediction can offer for policymakers, in particular by detecting emergent and novel dynamics as they unfold, timely reasoning and decision-making can be applied. The logistic model, and its historical resolution of capacity-limited growth, has been a powerful general model for dynamics in many areas of science. We believe that this data-driven extension, called ALM, may also help capture essential features in a wide range of scientific problems that feature confounding aggregations of multiple components of capacity-limited growth and decay, including biology, epidemiology, earth science, hydrology, economics, cybersecurity, and social media. We conclude this results section with a discussion of model limitations.

4.1 Biology and epidemiology

We apply the ALM model for prediction of case loads due to SARS-CoV-2 during the COVID-19 pandemic. Better prediction of case loads is important for planing and policy-making and can save lives. In Figure 1, ALM is applied to cases from Los Angeles County. Additionally, to compare with state of the art statistical methods, we provide a comprehensive comparison study of methods for COVID-19 case count prediction tasks. The study limits the scope to the state of California over more than three years of time and compares only to methods that offered more than 80 predictions in Covid-Forecast hub. See the appendix for comparison and results.

4.2 General applications

Here, we consider and illustrate the use of the additive logistic model ALM to various scientific subjects.

Earth science and hydrology, such as water levels in reservoirs (for example Lake Mead), have extrinsic and intervention causal components which modulate capacity and rate for water levels as a function of time. Causes include drought, changes in demand, and policy⁵⁵. Figure 2a illustrates ALM on Lake Mead water levels.

Economics, such as the Consumer Price Index (CPI), has been studied since 1913 and is used in determining inflation⁵⁶. Figure 2b illustrates the applications of ALM to CPI. The number, published monthly, is essentially the rate of change of the data plotted in 2b. Additionally, the possible use of Narrative Economics¹³, features an overlap of applications of economics with social media. Analogously to the famed butterfly effect in weather prediction, it would be interesting and perhaps useful to quantify how talking about the economy changes the economy.

Cybersecurity, vulnerabilities are important information regarding code bugs that hackers can exploit. As such, they represent an informational commodity for cyber operations and are tracked by the US National Institute of Standards and Technology (NIST) within their National Vulnerability Database (NVD). By considering the frequency and date of incident report associated to each vulnerability, a type of epidemiology can be developed to track cybersecurity. Accordingly, the ALM model also performs well within this domain as shown in Figure 2c.

Social media. The rise of social media has enabled novel modes of information sharing, and raised new concerns of viral miss-information. We apply the ALM model to counts of news articles re-posted per day that mentioned the COVID-19 pandemic, we present ALM modeling in Figure 2d.

4.2.1 Predicting change and auto-adaptive models

The ALM model responds to and auto-tunes new parameters to emergent dynamic components (surges) found in data from LA county COVID-19 cases for various time periods (i.e., using $D[.k]$ to represent the first k samples of data as would be observable on day k). We illustrate how the model in classes $\mathcal{M}^2$ and $\mathcal{M}^3$ and adapts to minimize residual shortly after observing data from the third wave of cases. In Figure 3 the considerations used by the model to change and re-trigger estimation with additional parameters are visible. In Figure 3a we observe that at around 2021-08-12 the fit for a two surge model begins to struggle, in 3b the residual distributions for $r[.k]$ are visualized for weekly updated data (i.e., for k equal to 2021-12-03, 2021-12-10, 2021-12-17, and 2021-12-24). In contrast, model estimation with three surges (additional parameters) are found to fit better in Figure 3c, reduce residuals and restore symmetry in residual distributions (see the violin plots of Figure 3c displaying the residual for $\mathcal{M}^3$ for same dates).

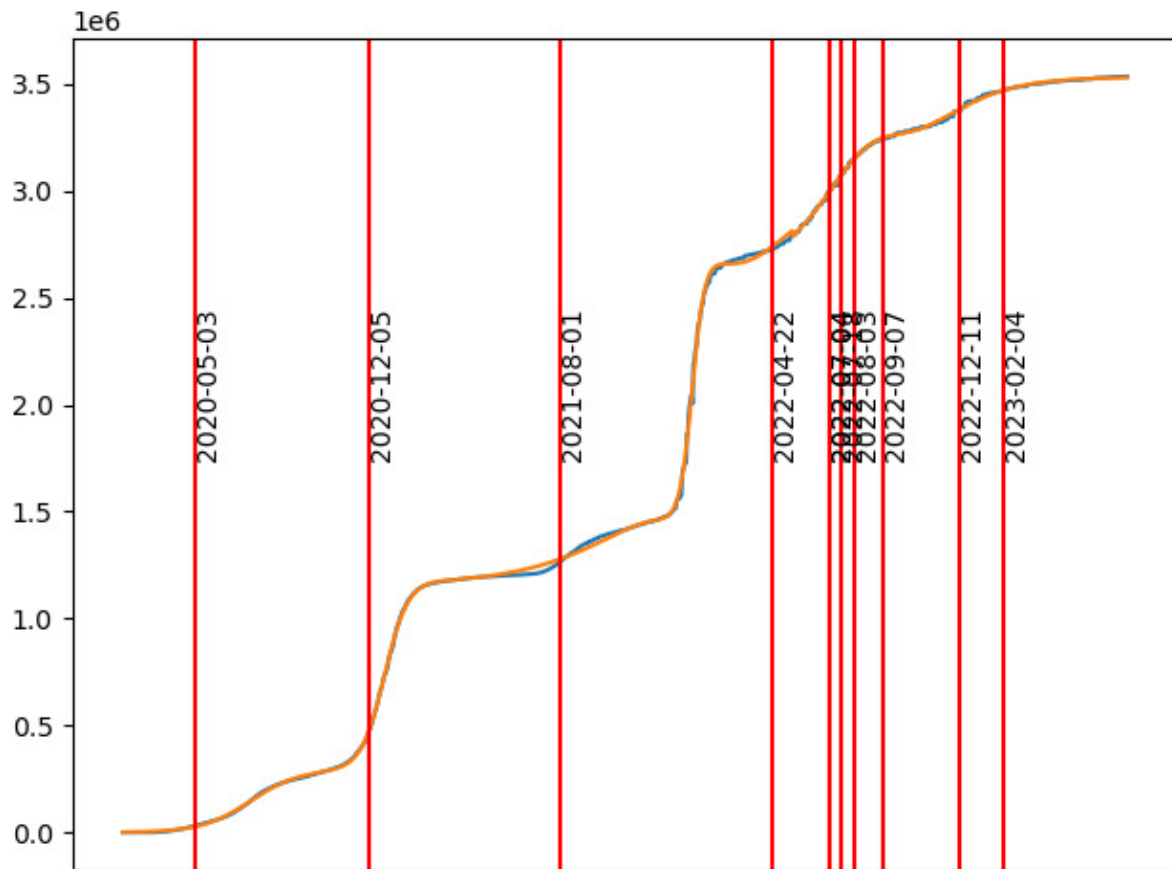


Figure 1. COVID-19 case (cumulative) count dynamics in Los Angeles County 2020-02-02 through 2023-07-23

illustrates the basic modeling problem. Cumulative case counts (blue) arise from several waves of growth. Our Model, called ALM, considers affine mixtures of multiple nonlinear dynamic models. By extending classical models in this way, data can drive a nonlinear regularized least squares optimization to achieve better models (orange) whose parameters indicate important changes points (red vertical lines). For COVID-19 case loads, the model determines change-points consistent with surge events. Importantly, ALM provides a means to model multi-surge growth in ways that compartmental models are not suited for.

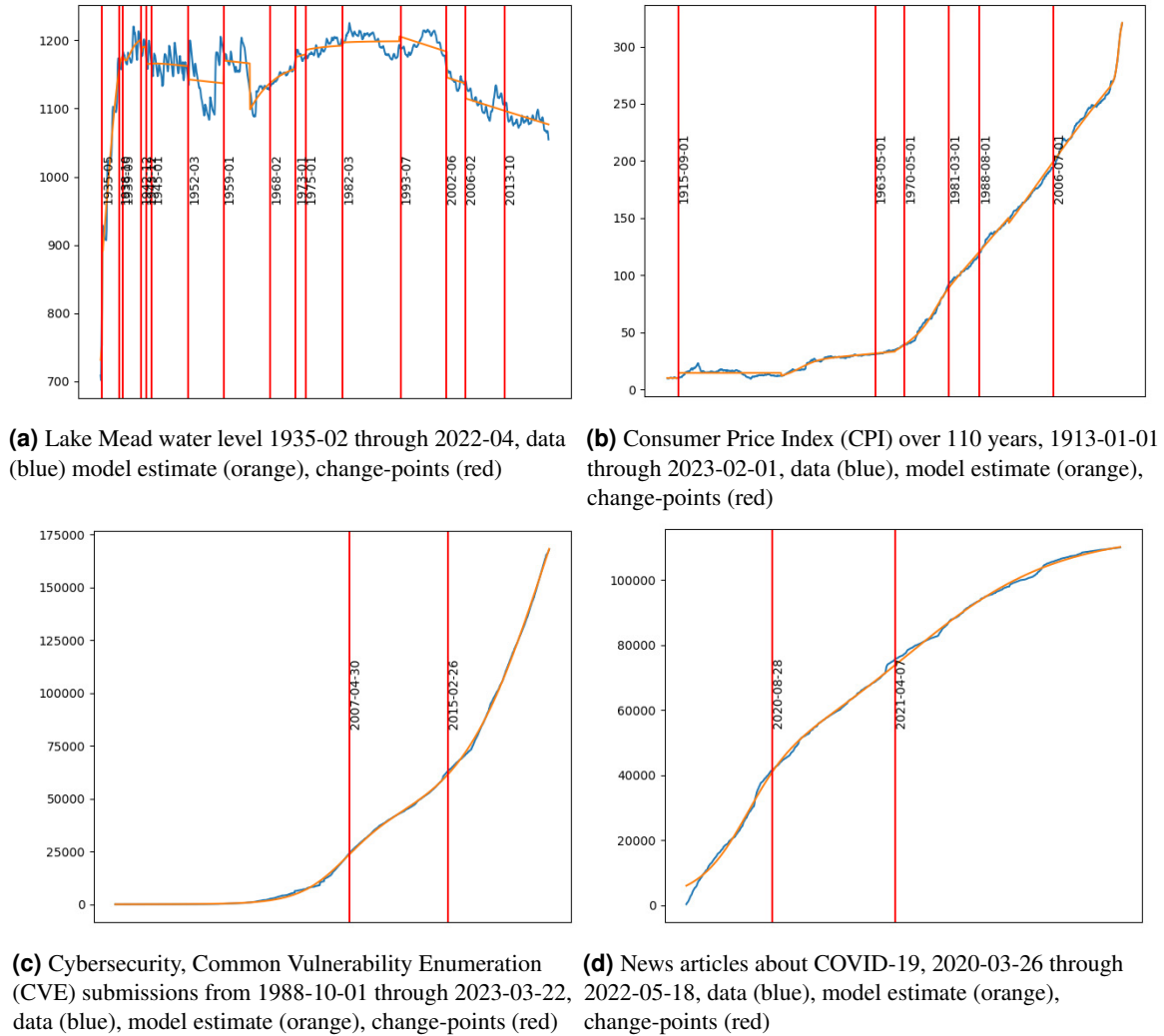


Figure 2. Applications of the ALM model to various dynamic processes that are challenging to model with a single isolated classical model. By extending the hypothesis space to a mixture of multiple repeated models, the classical model repertoire can be used as a basis to assemble well fit models that better reflect data and identify changing dynamic conditions. The ALM model which considers a mixture of logistic growth and decay modes can generally fit a wide range of problems involving rate and capacity-limited growth or decay.

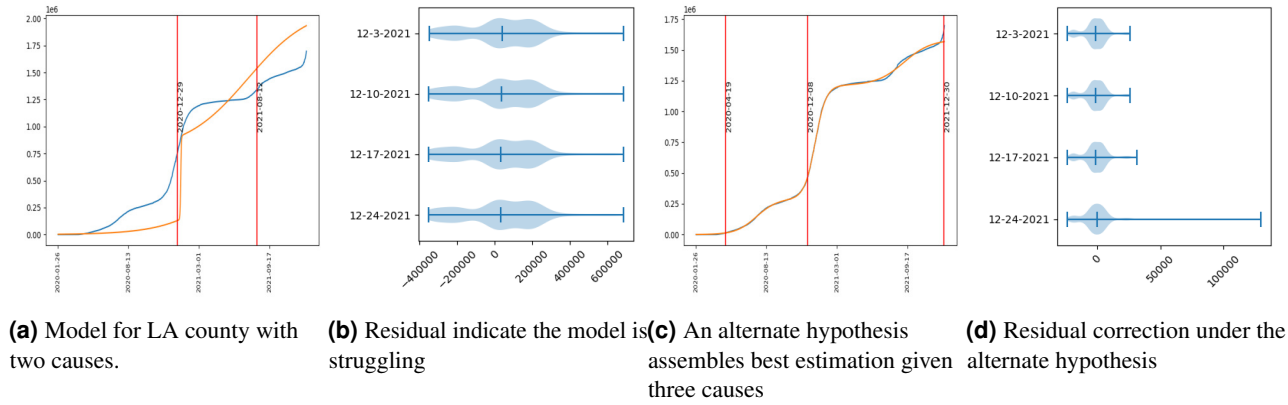


Figure 3. Forecasting with the ALM model: One advantage of this approach is that the model is able to self-evaluate its residual fit on a running basis. Should the residual, assumed to be Normal, accumulate sufficient skewness, the model can form a model selection test to include a hypothesis that a change-point has recently occurred. By comparing the current model (e.g., in (a) having two causes for growth), to an alternate model which posits an additional generating process (e.g., in (b) a third causes is introduced). Change to the model is accepted if the model improvements (correction for skewness or to improve the residual) outweigh the cost for adding an additional component imposed by the regularization term. This can be efficiently automated and performed by the model on a recurring basis. New dynamic regimes can be identified shortly after evidence mounts that data diverges from the current best forecast model.

Accordingly, the loss of symmetry for two surge models can be seen in the distributions of Figure 3b. Further, the loss of symmetry is detectable with a skewness measure. See equation D2. Also, residual reduction can be detected within the outlined hypothesis testing and decision criteria of equation D1. Note that the decisions to switch from a model in $\mathcal{M}^2$ to a model in $\mathcal{M}^3$ can be made fairly early on, as early as 2021-08-12 as the beginning of the third wave is experienced.

4.2.2 Explainable models and interpretations

Fitted models are amenable to interpretation. First, our software can decompose the components of growth. Note that the inflection point is highlighted as red vertical lines in Figures 1, 2a, 2b, 2c, and 2d to represent a specific time centering the dynamics of a single additive component. Additionally, the inflection point is where the growth rate for this component is optimal. Using equation S4, the inflection point can be used to locate a time when a percentage of change (due to the parameters of the given component) has occurred. Our software locates the positions of these inflection points and relevant bands around them and finds the associated capacity and rate parameters. where any portion of its effect is observed, as well as component parameters. This information is intended to enable policy formation and assessment tasks, where explainable AI algorithms are preferred to black-box methods that struggle to reveal what features of data are critical to the optimization results.

4.2.3 Limitations

While the additive logistic model (ALM) performs well on various data sets that feature temporally localized growth or decay modes, its effectiveness may be limited by both noise and the possibility of temporally overlapping growth/decay epochs that confound model estimation.

The design of the cost function found in equation R3 is an important and interesting problem associated with change-point detection and identification. Here, we argue that many problems are amenable to a design approach where domain experts, knowledgeable of what constitutes change, could develop appropriate cost functions that capture change and avoid overfitting hazards. For simpler cases, such as low noise and separated nonlinear phenomena as the examples visualized above, supervised learning approaches where an expert demonstrates a few examples of emergent dynamics, the system could potentially search for cost-thresholds that yield appropriate model responses in regards to equation R3, such as a jump in the number of model parameters, to adapt to the emergent feature. The general problem is clearly challenging. For more complex cases, such as greater noise-to-signal ratio in data and non-isolated nonlinear phenomena, it is likely that the same approach may break down. Exploration of these limitations is deferred to future work. A similar approach is suggested regarding the skewness test of equation D2. Likewise, for simple cases such as the data sets modeled above, a design approach can be pursued. Additionally, in problems where stronger assumptions concerning the residual distribution can be made (e.g., Gaussian), skewness testing may be more interpretable and prescriptive than our first problem of designing cost functions that induce model adaptability.

5 Conclusion

The additive logistic model is a straightforward method to fit data that has several surges. Owing to the flexibility and elegance of the logistic function at its core, ALM can be seen as a flexible extension and is shown to be able to model causally episodic growth and decay components in a variety of problems. ALM features efficient construct-ability, simple theory, and interpretability and appears useful as a data-driven model approach for general application. It also points to a general approach for surrogate model construction by introducing other nonlinear component functions. We have demonstrated the utility of this approach and observed how it yields straightforward models to problem domains that lack classical models that can perform as well. Further, our approach requires no outside information in relation to the information about changes and novel surges, but can potentially self-identify change-points and tune novel parameters when novel surges arise.

The model can also be used in more fields beyond those cited here. The model can be used not only to predict change-points, but to aid in considering the severity of events. A major goal in volcanology is to predict how severe the next eruption will be^{57,58}. For rainfall, most research doesn't report the time when the trend changes⁵⁹, something the model is designed to accomplish. Another field that looks at trends and changes in the trends is the study of invasive species⁶⁰. Climate change is also interested in not just the change in the data, but the magnitude of the change⁶¹. In future work, we plan to consider multidimensional adaptive models (spatial-temporal) in this context. In the broader field of ecology, estimating the amount of change as well as the time of change is a difficult problem⁶². In some disciplines, the causal events inducing changing dynamics are either not directly observable or (for example, in cybersecurity) are strategically withheld from the modeler, such as a hidden population. For such problems, our framework that links observable change to causal event inference could prove useful to evaluate forensically causal conditions or interventions, and thereby offer enhanced analysis modalities to a variety of pressing problems.

Data availability

The Datasets used for model development are publicly available from the following detailed sources [For a complete list of sources, including appended materials, see: <https://github.com/austincasey/humpty/tree/main/data>]. COVID-19 case loads were collected daily (up to mid-2023) on a per-US-county basis and shared via the usafacts.org website⁶³, this was filtered by county and used as input with ALM to form Figures: 1, 3a, 3b, 3c, and 3d. The US Bureau of Reclamation measures Lake Mead monthly to determine how much water was lost or gained and shares that information with the public⁵⁵. This data was used with our ALM model to generate Figure 2a. The Consumer Price Index (CPI) is tracked by the St. Louis Federal Reserve Bank and is published monthly⁵⁶. This data was used with ALM to produce Figure 2b. Data on cybersecurity vulnerabilities are collected by Mitre Corporation and made available to the public⁶⁴. This data is used with ALM to form Figure 2c. COVID-19 social media data was collected from the GDELT⁶⁵ project. These data were filtered by location (Las Vegas) and used with ALM to form Figure 2d. All source code developed for this research is publicly available at <https://github.com/austincasey/humpty> along with copies of filtered data sets from above. For any additional questions, the corresponding authors can assist upon reasonable request.

Acknowledgments

We wish to thank the RxCovea group. The research of SC has been supported partially by NSF grant DMS 2154564. The research of Heeralal Janwa was supported by the NASA grant 80NSSC22M0248 and NASA-EPSCoR grant 80NSST24M0107 and also by the PR NASA Grant Consortium and its Director, Dr. Gerardo Morell. We also thank the anonymous reviewers for the helpful comments on the initial version that helped us improve the manuscript.

Copyright 2024 Carnegie Mellon University, Ernest Battifarano, Heeralal Janwa, Shirshendu Chatterjee and Will Casey

NO WARRANTY. THIS CARNEGIE MELLON UNIVERSITY AND SOFTWARE ENGINEERING INSTITUTE MATERIAL IS FURNISHED ON AN "AS-IS" BASIS. CARNEGIE MELLON UNIVERSITY MAKES NO WARRANTIES OF ANY KIND, EITHER EXPRESSED OR IMPLIED, AS TO ANY MATTER INCLUDING, BUT NOT LIMITED TO, WARRANTY OF FITNESS FOR PURPOSE OR MERCHANTABILITY, EXCLUSIVITY, OR RESULTS OBTAINED FROM USE OF THE MATERIAL. CARNEGIE MELLON UNIVERSITY DOES NOT MAKE ANY WARRANTY OF ANY KIND WITH RESPECT TO FREEDOM FROM PATENT, TRADEMARK, OR COPYRIGHT INFRINGEMENT.

DM24-0612

Author contributions statement

W.C. contributed model development, implementation, visualization, conducted forecasting comparisons with other methods, and prepared images in the appended material. L.M. contributed model development, data development, testing validation, implementation, visualization, and prepared images in the main manuscript. W.C., L.M., S.C., H.J., and E.B. contributed to

methodological and algorithm development and manuscript preparation. A.E. and Y.G. assisted with early technical discussion and manuscript review.

References

1. Bertozzi, A. L., Franco, E., Mohler, G., Short, M. B. & Sledge, D. The challenges of modeling and forecasting the spread of COVID-19. *Proc. Natl. Acad. Sci.* **117**, 16732–16738 (2020).
2. Khan, M. A. & Atangana, A. Mathematical modeling and analysis of COVID-19: A study of new variant omicron. *Phys. A: Stat. Mech. its Appl.* **599**, 127452 (2022).
3. Gonzalez-Parra, G., Martínez-Rodríguez, D. & Villanueva-Micó, R. J. Impact of a new sars-cov-2 variant on the population: A mathematical modeling approach. *Math. Comput. Appl.* **26**, 25 (2021).
4. Saban, M., Myers, V. & Wilf-Miron, R. Changes in infectivity, severity and vaccine effectiveness against delta COVID-19 variant ten months into the vaccination program: The israeli case. *Prev. Medicine* **154**, 106890 (2022).
5. Caldwell, J. M. *et al.* Vaccines and variants: Modelling insights into emerging issues in COVID-19 epidemiology. *Paediatr. Respir. Rev.* **39**, 32–39 (2021).
6. Chen, M. K., Zhuo, Y., de la Fuente, M., Rohla, R. & Long, E. F. Causal estimation of stay-at-home orders on sars-cov-2 transmission. *arXiv preprint arXiv:2005.05469* (2020).
7. Chen, Y.-C., Lu, P.-E., Chang, C.-S. & Liu, T.-H. A time-dependent SIR model for COVID-19 with undetectable infected persons. *IEEE transactions on network science engineering* **7**, 3279–3294 (2020).
8. Caldwell, J. M. *et al.* Understanding COVID-19 dynamics and the effects of interventions in the philippines: A mathematical modelling study. *The Lancet Reg. Heal. Pac.* **14** (2021).
9. Yuan, P. *et al.* Efficacy of a “stay-at-home” policy on sars-cov-2 transmission in toronto, canada: a mathematical modelling study. *Can. Med. Assoc. Open Access J.* **10**, E367–E378 (2022).
10. Mbogo, R. W. & Orwa, T. O. Sars-cov-2 outbreak and control in Kenya-mathematical model analysis. *Infect. Dis. Model.* **6**, 370–380 (2021).
11. Nannyonga, B. K. *et al.* Infodemic: How an epidemic of misinformation could lead to a high number of the novel corona virus disease cases in Uganda. *Preprints* (2020).
12. Sood, M. *et al.* Spreading processes with mutations over multi-layer networks. *arXiv preprint arXiv:2210.05051* (2022).
13. Shiller, R. J. *Narrative economics: How stories go viral and drive major economic events* (Princeton University Press, 2020).
14. Xuan, G., Zhang, W. & Chai, P. Em algorithms of gaussian mixture model and hidden markov model. In *Proceedings 2001 international conference on image processing (Cat. No. 01CH37205)*, vol. 1, 145–148 (IEEE, 2001).
15. Jang, J.-S. & Sun, C.-T. Functional equivalence between radial basis function networks and fuzzy inference systems. *IEEE transactions on Neural Networks* **4**, 156–159 (1993).
16. McDowell, I. C. *et al.* Clustering gene expression time series data using an infinite gaussian process mixture model. *PLoS computational biology* **14**, e1005896– (2018).
17. Albert, P. S. A two-state markov mixture model for a time series of epileptic seizure counts. *Biometrics* 1371–1381 (1991).
18. Eirola, E. & Lendasse, A. Gaussian mixture models for time series modelling, forecasting, and interpolation. In *Advances in Intelligent Data Analysis XII: 12th International Symposium, IDA 2013, London, UK, October 17-19, 2013. Proceedings 12*, 162–173 (Springer, 2013).
19. Kalliovirta, L., Meitz, M. & Saikkonen, P. A gaussian mixture autoregressive model for univariate time series. *J. Time Ser. Analysis* **36**, 247–266 (2015).
20. Xiong, Y. & Yeung, D.-Y. Mixtures of arma models for model-based time series clustering. In *2002 IEEE International Conference on Data Mining, 2002. Proceedings.*, 717–720 (IEEE, 2002).
21. Zhang, Z., Li, W. K. & Yuen, K. C. On a mixture garch time-series model. *J. Time Ser. Analysis* **27**, 577–597, DOI: <https://doi.org/10.1111/j.1467-9892.2006.00467.x> (2006). <https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1467-9892.2006.00467.x>.
22. Prado, R., Ferreira, M. A. R. & West, M. *Time Series: Modeling, Computation, and Inference*. Chapman & Hall/CRC Texts in Statistical Science (Chapman & Hall/CRC, Boca Raton, FL, 2021), 2 edn.

23. West, M. & Harrison, J. *Bayesian forecasting and dynamic models* (Springer, 1997).
24. Adams, R. P. & MacKay, D. J. Bayesian online changepoint detection. *arXiv preprint arXiv:0710.3742* (2007).
25. Hamilton, J. D. *Time Series Analysis* (Princeton University Press, Princeton, NJ, 2006 - 2017).
26. Newbold, P. Arima model building and the time series analysis approach to forecasting. *J. forecasting* **2**, 23–35 (1983).
27. Nickerson, D. M. & Madsen, B. C. Nonlinear regression and arima models for precipitation chemistry in east central florida from 1978 to 1997. *Environ. Pollut.* **135**, 371–379 (2005).
28. Lee, S.-S. Non-linear prediction problems for ornstein-uhlenbeck process. *Nagoya Math. J.* **91**, 173–184 (1983).
29. Osher, S., Solé, A. & Vese, L. Image decomposition and restoration using total variation minimization and the h. *Multiscale Model. & Simul.* **1**, 349–370 (2003).
30. Nikiforov, I. & Basseville, M. Detection of abrupt changes: theory and application. *Englewood Cliffs, NJ: PTR Prentice Hall* (1993).
31. Kim, K., Park, J. H., Lee, M. & Song, J. W. Unsupervised change point detection and trend prediction for financial time-series using a new cusum-based approach. *IEEE Access* **10**, 34690–34705 (2022).
32. Hushchyn, M. & Ustyuzhanin, A. Generalization of change-point detection in time series data based on direct density ratio estimation. *J. Comput. Sci.* **53**, 101385 (2021).
33. Namono, B., Starr, A., Emmanouilidis, C. & Cristobal, R. C. Online change detection techniques in time series: An overview. In *2019 IEEE international conference on prognostics and health management (ICPHM)*, 1–10 (IEEE, 2019).
34. Pastor-Barriuso, R., Guallar, E. & Coresh, J. Transition models for change-point estimation in logistic regression. *Stat. medicine* **22**, 1141–1162 (2003).
35. Pettitt, A. N. A non-parametric approach to the change-point problem. *J. Royal Stat. Soc. Ser. C (Applied Stat.)* **28**, 126–135 (1979).
36. Achcar, J. A., Coelho-Barros, E. A. & de Souza, R. M. Use of non-homogeneous poisson process (nhpp) in presence of change-points to analyze drought periods: a case study in brazil. *Environ. Ecol. Stat.* **23**, 405–419 (2016).
37. Achcar, J. A. & de Oliveira, R. P. Climate change: use of non-homogeneous poisson processes for climate data in presence of a change-point. *Environ. Model. & Assess.* 1–14 (2022).
38. Dean, C. & Lawless, J. F. Tests for detecting overdispersion in poisson regression models. *J. Am. Stat. Assoc.* **84**, 467–472 (1989).
39. Åström, K. J. & Eykhoff, P. System identification—a survey. *Automatica* **7**, 123–162 (1971).
40. Sun, L., Gao, H., Pan, S. & Wang, J.-X. Surrogate modeling for fluid flows based on physics-constrained deep learning without simulation data. *Comput. Methods Appl. Mech. Eng.* **361**, 112732 (2020).
41. BR Nogueira, I., V. Santana, V., Ribeiro, A. M. & Rodrigues, A. E. Using scientific machine learning to develop universal differential equation for multicomponent adsorption separation systems. *The Can. J. Chem. Eng.* **100**, 2279–2290 (2022).
42. Pestourie, R., Mroueh, Y., Rackauckas, C., Das, P. & Johnson, S. G. Physics-enhanced deep surrogates for pdes. *Springer* (2023).
43. Golub, G. & Pereyra, V. Separable nonlinear least squares: the variable projection method and its applications. *Inverse problems* **19**, R1 (2003).
44. O’Leary, D. P. & Rust, B. W. Variable projection for nonlinear least squares problems. *Comput. Optim. Appl.* **54**, 579–593 (2013).
45. Takahashi, A., Spreadbury, J. & Scotti, J. Modeling the spread of tuberculosis in a closed population. *Laporan Hasil Penelitian* (2010).
46. Britton, T. *et al. Stochastic epidemic models with inference*, vol. 2255 (Springer, 2019).
47. Sars-cov-2 variants of concern as of 10 august 2023. <https://www.ecdc.europa.eu/en/covid-19/variants-concern> (2023).
48. Moein, S. *et al.* Inefficiency of sir models in forecasting covid-19 epidemic: a case study of isfahan. *Sci. reports* **11**, 4725 (2021).
49. Abusam, A., Abusam, R. & Al-Anzi, B. Adequacy of logistic models for describing the dynamics of covid-19 pandemic. *Infect. Dis. Model.* **5**, 536–542 (2020).

50. Cramer, J. S. The early origins of the logit model. *Stud. Hist. Philos. Sci. Part C: Stud. Hist. Philos. Biol. Biomed. Sci.* **35**, 613–626 (2004).
51. Levenberg, K. A method for the solution of certain non-linear problems in least squares. *Q. applied mathematics* **2**, 164–168 (1944).
52. Marquardt, D. W. An algorithm for least-squares estimation of nonlinear parameters. *J. society for Ind. Appl. Math.* **11**, 431–441 (1963).
53. Li, Y. Centering, trust region, reflective techniques for nonlinear minimization subject to bounds. Tech. Rep., Cornell University (1993).
54. Virtanen, P. *et al.* Scipy 1.0: fundamental algorithms for scientific computing in python. *Nat. methods* **17**, 261–272 (2020).
55. Lake mead at hoover dam, end of month elevation (feet). <https://www.usbr.gov/lc/region/g4000/hourly/mead-elv.html> (2023). Accessed: 2023-06-28.
56. FRED economic data. <https://fred.stlouisfed.org/categories/9> (2023). Accessed: 2023-06-28.
57. Bebbington, M. S. Long-term forecasting of volcanic explosivity. *Geophys. J. Int.* **197**, 1500–1515, DOI: [10.1093/gji/ggu078](https://doi.org/10.1093/gji/ggu078) (2014). <https://academic.oup.com/gji/article-pdf/197/3/1500/27640885/ggu078.pdf>.
58. Marzocchi, W. & Bebbington, M. S. Probabilistic eruption forecasting at short and long time scales. *Bull. volcanology* **74**, 1777–1805 (2012).
59. Patakamuri, S. K., Muthiah, K. & Sridhar, V. Long-term homogeneity, trend, and change-point analysis of rainfall in the arid district of ananthapuramu, andhra pradesh state, india. *Water* **12**, 211 (2020).
60. DeBoer, J. A., Anderson, A. M. & Casper, A. F. Multi-trophic response to invasive silver carp (*hypophthalmichthys molitrix*) in a large floodplain river. *Freshw. Biol.* **63**, 597–611 (2018).
61. Bloomfield, P. & Nychka, D. Climate spectra and detecting climate change. *Clim. Chang.* **21**, 275–287 (1992).
62. Beckage, B., Joseph, L., Belisle, P., Wolfson, D. B. & Platt, W. J. Bayesian change-point analyses in ecology. *New Phytol.* **174**, 456–467 (2007).
63. Us COVID-19 cases and deaths by state. <https://usafacts.org/visualizations/coronavirus-covid-19-spread-map/>. Accessed: 2023-06-28.
64. Cve. <https://cve.org> (2024). Accessed: 2024-02-29.
65. Blog, T. G. P. Now live updating & expanded: A new dataset for exploring the coronavirus narrative in global online news.
66. Cramer, E. Y. *et al.* The united states covid-19 forecast hub dataset. *Sci. Data* DOI: [10.1101/2021.11.04.21265886](https://doi.org/10.1101/2021.11.04.21265886) (2022).
67. Jahja, M., Chin, A. & Tibshirani, R. J. Real-time estimation of covid-19 infections: Deconvolution and sensor fusion. *Stat. Sci.* **37**, 207–228 (2022).
68. Bai, J. Estimating multiple breaks one at a time. *Econom. theory* **13**, 315–352 (1997).
69. Fryzlewicz, P. Wild binary segmentation for multiple change-point detection. *The Annals Stat.* **42**, 2243 – 2281, DOI: [10.1214/14-AOS1245](https://doi.org/10.1214/14-AOS1245) (2014).