

## TWO FILTERING METHODS OF FORECASTING LINEAR AND NONLINEAR DYNAMICS OF INTENSIVE LONGITUDINAL DATA

MICHAEL D. HUNTER 

PENNSYLVANIA STATE UNIVERSITY

HAYA FATIMAH AND MARINA A. BORNOVALOVA

UNIVERSITY OF SOUTH FLORIDA

With the advent of new data collection technologies, intensive longitudinal data (ILD) are collected more frequently than ever. Along with the increased prevalence of ILD, more methods are being developed to analyze these data. However, relatively few methods have yet been applied for making long- or even short-term predictions from ILD in behavioral settings. Applications of forecasting methods to behavioral ILD are still scant. We first establish a general framework for modeling ILD and then extend that frame to two previously existing forecasting methods: these methods are Kalman prediction and ensemble prediction. After implementing Kalman and ensemble forecasts in free and open-source software, we apply these methods to daily drug and alcohol use data. In doing so, we create a simple, but nonlinear dynamical system model of daily drug and alcohol use and illustrate important differences between the forecasting methods. We further compare the Kalman and ensemble forecasting methods to several simpler forecasts of daily drug and alcohol use. Ensemble forecasts may be more appropriate than Kalman forecasts for nonlinear dynamical systems models, but further forecasting evaluation methods must be put into practice.

Key words: dynamical systems, forecasting, time series, Kalman filtering, intensive longitudinal data, drug and alcohol use.

Prediction of human behavior is famously difficult, but psychology is entering an era in which enough data are collected on a sufficiently large number of people to make individual behavioral predictions plausible. In the history of psychology, the failure of personality variables to predict individual behaviors led Walter Mischel to reject the notion that general personality traits exist in favor of context-dependent if-then contingencies (Mischel, 1968). More broadly, forecasting of any kind is often rife with high-profile failures and misplaced confidence. A poll from the *Literary Digest* famously forecasted the United States presidential election in 1936 to be a decisive victory for Alf Landon, whom they predicted to receive 57% of the popular vote; to the contrary Franklin Roosevelt won in a landslide with 60% of the popular vote (Squire, 1988). In macroeconomics, forecasts are similarly suspect (e.g., missing the 2007 global recession, Culbertson & Sinclair, 2014). Even in cases where forecast accuracy has immensely improved over the last 50 years, the public continues to regard many forecasts with intense doubt (see, e.g., severe weather and tornado forecasts, Ripberger et al., 2014; Brooks, 2004; Simmons & Sutter, 2009; Lazo, Morss, & Demuth, 2009). These difficulties of forecasting future data only from past observations led Harvey (1989, p. xi) to compare the task to “driving a car blindfolded while following directions given by a person looking out the back window.” With these caveats in mind, we propose to apply two well-developed methods of forecasting from the time series literature to psychological behavior.

The work was supported by DA032582 (NIDA)

Correspondence should be made to Michael D. Hunter, Department of Human Development and Family Studies, Pennsylvania State University, 119 Health and Human Development Building, University Park, PA16802, USA. Email: [mhunter.ou@gmail.com](mailto:mhunter.ou@gmail.com)

Intensive longitudinal data (ILD) afford the possibility of creating behavioral forecasts, but methods for forecasting ILD have generally been absent from the behavioral literature. Forecasting ILD presents several challenges and opportunities. A critical challenge of any forecast is the calibration of the forecast error: the forecast must include an estimate of its accuracy. No forecast will ever predict behavior perfectly, but a calibrated forecast will have a known, quantifiable degree of accuracy and be correct in assessing its own amount of error. The forecasting methods we apply are paired with estimates of their forecast error. We emphasize the meaning of this forecast error and recognize the limits of behavioral forecasting with our current level of knowledge of human behavior.

The opportunities for forecasting ILD abound, but to make skilled predictions, it is important to recognize human behavior is a complex system and thus certain intrinsic limits may exist for these predictions due to the inherent system dynamics. For instance, shifts in affect appear to be chaotic and are difficult to predict (Fredrickson & Losada, 2005). The same applies to behaviors associated with other psychological processes: recent literature on substance use has acknowledged that use and relapse are complex processes following a nonlinear trajectory (Hufford, Witkiewitz, Shields, Kodya, & Caruso, 2003; Witkiewitz & Marlatt, 2004). Likewise, other impulsive or pathological behaviors such as suicide attempts, non-suicidal self-injury, engagement in violence and crime are most likely chaotic, dynamic processes which are difficult to predict with commonly used methods in psychological research. Time series data on behavioral, psychological, and physiological states now allows us to apply modern forecasting methods to realistically predict change. Being able to forecast psychological and behavioral processes has important implications in studying treatment process and measuring progress in psychotherapy (Hayes, Laurenceau, Feldman, Strauss, & Cardaciotto, 2007), neuronal responses (Friston, Harrison, & Penny, 2003; Stephan et al., 2008; Havlicek, Friston, Jan, Brazdil, & Calhoun, 2011), interpersonal dynamics (Boker & Laurenceau, 2006), and the evolution of organizational behavior (Allen, Strathern, & Baldwin, 2008) just to name a few areas of application.

When considering forecasting methods for ILD, it is prudent to first examine methods that already exist. Perhaps the simplest method of forecasting is to carry the last observation forward. This method is simple enough to be conceived without statistical foundations, but can be derived from Martingale theory (Hall & Heyde, 1980, p. 1) and random walks (West & Harrison, 1997, pp. 26–27, 70). The limitations of this naive forecast are numerous (see Mandelbrot, 1971, for additional limitations in the economic context). First, the underlying model of behavior is, in essence, whatever the person just did, they will continue to do forever. Second, only the last observation, rather than the full history of observations, is accounted for in any way. Thus, no slope or patterned trajectory of behavior is involved. Third, there is no *intrinsic* notion of forecast error with the carry-forward forecast, but it can be augmented with one: usually a random walk (e.g., Hyndman & Athanasopoulos, 2018, Ch. 3). When buttressed with the notion of a random walk to induce a notion of forecast error, the naive forecast suggests linearly increasing forecast error variance over time (Hyndman & Athanasopoulos, 2018, Ch. 3; Mandelbrot, 1971; Gregson, 1983, p. 45). Fourth, there is no necessary statistical basis for the carry-forward forecast. This lack of intrinsic statistical basis is at the heart of the other three problems: an oversimplified behavioral model, a lack of statistical consistency leading the forecast to improve in quality as the number of observations increases, and the absence of any reasonable properties for the error distribution. The problems of the naive carry-forward forecast make it useful for comparison with more theoretically motivated methods, as we will see later in this paper.

A slightly more sophisticated forecast results from a latent growth curve model (e.g., McArdle & Epstein, 1987). Latent growth curves fit polynomial or nonlinear template curves to the observed data. Individuals randomly differ in their coefficients for these template curves. For instance, a linear latent growth curve model estimates a global mean intercept and a global mean slope along with the variances and covariances of the intercepts and slopes across people. Each person has

their own intercept and slope, but these are considered random samples from the bivariate normal mean and slope distribution.

There are several important weaknesses of the latent growth curve approach to forecasting ILD. First, latent growth curve models fit template curves directly to the observed data, rather than estimate intrinsic dynamic characteristics of the data with the observed trajectories following from these intrinsic dynamics. That is, latent growth curves fit a *global* curve pattern to the data rather than indicating the *local* rules governing the change process. Such a forecast may under-emphasize more recent observations, and over-weight more temporally distant observations.<sup>1</sup> Second, latent curve forecasts necessarily extrapolate in the time variable along the shape of the template curve. Because many latent growth curve models have polynomial form, the polynomial forecast will either increase to positive infinity or decrease to negative infinity as time grows large. Thus, many latent growth curve models produce implausible forecasts for typically bounded behavioral variables. Third, latent curve modeling approaches either omit measurement noise or confound measurement noise with process noise. This distinction will be elaborated later in this paper. Fourth, latent curve models are best suited to studying between-person differences around average trajectories rather than person-centered processes. Put another way, latent curves tend to be nomothetic models, whereas the methods applied in this paper—and models most suitable to ILD—are principally idiographic (see Molenaar, 2004, for a classic paper extolling these distinctions). Again, the limitations of forecasts from latent growth curve models make them apt for comparison with other methods, as we will illustrate in this paper.

Instead of the previously discussed forecasting methods, we emphasize methods of forecasting pioneered in the time series literature (e.g., Box & Jenkins, 1976). These time series forecasting methods have an easy interpretation within the context of Kalman filtering and standard models of dynamical systems (Harvey, 1989). Time series forecasting methods may be more appropriate for ILD due to the similar number of observations. Generically, there is no minimum number of time points for a time series analysis<sup>2</sup> and there is no maximum number of time points for a longitudinal analysis. However, as the number of time points increases, the research questions and goals tend to shift into alignment with time series models (see Baltes, Reese, & Nesselrode, 1977; Molenaar, 1997, for further discussion of these shifting questions and goals). In our experience, time series forecasts are routinely applied to data with 10 to 100,000,000 time points, whereas carry-forward and latent growth curve forecasts are more typical for less than 10 time points per person.

Time series forecasting methods also solve many of the most pronounced shortcomings of latent growth curves. Although some time series forecasts have pre-supposed shapes (e.g., the drift model to be discussed later), many do not. The template global curves of latent growth curves are a necessary feature of the method, whereas they are an infrequent and optional part of more conventional time series methods. So, extending the forecast of most time series models does not necessarily pre-suppose a continued linear—or other polynomial—trajectory. Time series methods generally do not fit template *global* curve patterns to the observed data. Instead, time series methods use *local* and recursive rules to govern the observed patterns of change in non-polynomial forms idiographically.

When using time series methods like the Kalman filter to analyze intensive longitudinal data, there are two common and standard methods of forecasting. The first method is called Kalman forecasting, usually applied to linear dynamical systems. The second method is called ensemble forecasting, usually applied to nonlinear dynamical systems. These methods are not new to the

<sup>1</sup>For example, many psychological time series exhibit “burn-in” or habituation periods where early observations may have different behavior than later observations. Equally weighting early and late observations would produce a worse forecast than a weighting scheme that emphasized the more recent observations (Gregson, 1983).

<sup>2</sup>Under ergodic conditions (i.e., Hannan, 1970, p. 201), time series are appropriate even for single-occasion data from multiple subjects.

fields of statistics or time series, but to our knowledge have not been previously applied in a behavioral setting.

The goal of the present study is to develop, implement, apply, and evaluate Kalman and ensemble forecasting for intensive longitudinal data in a behavioral setting. First, we provide some background on modeling dynamical systems in general. Next, we give further details on the Kalman and ensemble forecasting methods used for these dynamical systems. Third, we briefly describe the programming interfaces for newly implemented forecasting functions in free and open-source programs for modeling dynamical systems. Fourth, we illustrate the use of these dynamical systems models and their forecasts using data relating to substance use; furthermore, we evaluate the Kalman and ensemble forecasts and compare them to simpler alternatives like the carry-forward and growth curve methods previously described.

## 1. General Dynamic Modeling in Discrete and Continuous Time

We begin our exposition of general dynamic modeling with the discrete time linear dynamical system, then progress to its continuous time analogue, finally generalizing these to their nonlinear versions. The linear, discrete time dynamical system (1) is simple compared to the other versions, (2) may be more familiar to many researchers in the behavioral sciences, and (3) is a critical part of fitting all of these models. A first step in fitting nonlinear or continuous time models with the methods we describe is to approximate them with linear, discrete time models. Therefore, we consider the linear discrete time model paramount for our purposes. After considering dynamical systems generally, we augment these with observed data, measurement, and forecasting.

A linear dynamical system in discrete time generically takes the form (cf. Kalman, 1960, Equation 15)

$$\boldsymbol{\eta}_t = \mathbf{B}_d \boldsymbol{\eta}_{t-1} + \boldsymbol{\Gamma}_d \mathbf{x}_t + \boldsymbol{\zeta}_t \quad (1)$$

where  $\boldsymbol{\eta}_t$  is the vector-valued, time-evolving, latent state of the dynamical system at time  $t$ ,  $\mathbf{B}_d$  is a matrix that describes how the state changes from the previous state  $\boldsymbol{\eta}_{t-1}$ ,  $\mathbf{x}_t$  is a vector of observed disturbances to the state of the dynamical system with instantaneous effects and disturbance regression weights given by  $\boldsymbol{\Gamma}_d$ , and  $\boldsymbol{\zeta}_t$  is a vector of unmeasured disturbances to the state with covariance  $\boldsymbol{\Psi}_d$ . Equation 1 is a standard model of any linear change process in discrete time. When paired with assumptions about the distributions of  $\boldsymbol{\eta}_t$  and  $\boldsymbol{\zeta}_t$ , Eq. 1 becomes a statistical model with many useful properties which we exploit later. Most often, we assume that (1)  $\boldsymbol{\eta}_t$  and  $\boldsymbol{\zeta}_t$  are Gaussian, (2)  $\boldsymbol{\zeta}_t$  has mean  $\mathbf{0}$  for all times  $t$ , (3)  $\boldsymbol{\zeta}_t$  are independent and identically distributed over time but allowing contemporaneous covariances, and (4)  $\boldsymbol{\eta}_t$  and  $\boldsymbol{\zeta}_t$  are uncorrelated contemporaneously and at all time lags:  $Cov(\boldsymbol{\eta}_t, \boldsymbol{\zeta}_\tau) = \mathbf{0}$  for all times  $t$  and  $\tau$ .

A parallel definition of a linear dynamical system in continuous time is possible and the primary methods applied also hold in this case (cf. Kalman, 1960, Equation 12).

$$\frac{d\boldsymbol{\eta}(t)}{dt} = \mathbf{B}_c \boldsymbol{\eta}(t) + \boldsymbol{\Gamma}_c \mathbf{x}(t) + \boldsymbol{\zeta}(t) \quad (2)$$

Equation 2 is a continuous time version of Eq. 1.<sup>3</sup> Although Eq. 1 contains only first-order lags, Hamilton (1994, pp. 3043–3044) and Hunter (2018, p. 317, Equation 31) showed higher-order lags for discrete time models using only Eq. 1. Similarly, Eq. 2 shows only a first-order

<sup>3</sup>Note that to ease readability, we state this equation in its derivative form rather than its differential form. Although this lacks some rigor, we feel it increases clarity.

stochastic differential equation, but introductory textbooks on differential equations show higher-order differential equations as systems of first-order derivatives (e.g., Edwards & Penney, 2004; V. I. Arnold, 1973; Hirsch & Smale, 1974; Hirsch, Smale, & Devaney, 2003). So, first-order lags and derivatives are sufficient to cover any finite lag or derivative order.

In the continuous time linear dynamical system, the discrete time latent state  $\eta_t$  is replaced with the continuous time latent state  $\eta(t)$ . Similarly, the discrete time matrices  $\mathbf{B}_d$ ,  $\mathbf{\Gamma}_d$ , and  $\mathbf{\Psi}_d$  are replaced by their continuous time analogues,  $\mathbf{B}_c$ ,  $\mathbf{\Gamma}_c$ , and  $\mathbf{\Psi}_c$ . Importantly, the algebraic forms of Eqs. 1 and 2 are parallel, but the meaning of the matrices has non-trivially changed. The continuous time matrix  $\mathbf{B}_c$  can always be nonlinearly transformed into a discrete time matrix  $\mathbf{B}_d$  given some fixed time step, but the reverse transformation from discrete time to continuous time is often not possible (Hamerle, Nagl, & Singer, 1991; Brockwell, 1995; He & Wang, 1989; Huzii, 2007; Chan & Tong, 1987). This lack of reversibility means that there exist discrete time models with no continuous time analogue.

The matrix  $\mathbf{B}_d$  maps the latent state forward in time by one, uniform, discrete step. By contrast, the matrix  $\mathbf{B}_c$  maps the latent state at some time on to the rate of change in the latent state at the same time. Given this rate of change, some finite interval of time, and certain regularity conditions, the stochastic differential equation in Eq. 2 can be solved for the predicted latent state. Generically, the solution of Eq. 2 takes the form (Harvey, 1989, p. 492, Equation 9.3.1)

$$\eta_{t_i} = \int_{t_{i-1}}^{t_i} \mathbf{B}_c \eta(t) + \mathbf{\Gamma}_c \mathbf{x}(t) + \zeta(t) dt \quad (3)$$

$$\eta_{t_i} = \underbrace{e^{\mathbf{B}_c(t_i - t_{i-1})}}_{\mathbf{B}_d} \eta(t_{i-1}) + \underbrace{\int_{t_{i-1}}^{t_i} e^{\mathbf{B}_c t} dt \mathbf{\Gamma}_c \mathbf{x}(t)}_{\mathbf{\Gamma}_d} \quad (4)$$

where we are solving Eq. 2 for  $\eta(t_i)$  at time  $t_i$  given some initial state  $\eta(t_{i-1})$  at time  $t_{i-1}$ . To obtain Eq. 4 from Eq. 3, we further assume that the exogenous covariates  $\mathbf{x}(t)$  are constant during this interval, and that  $\zeta(t)$  has Itô integral zero (see L. Arnold, 1974, p. xii–xiii). As shown by Oud and Jansen (2000), the assumption that the covariates are constant between observations can be relaxed. We use the discrete time notation for the latent state  $\eta_{t_i}$  to make the parallel between Eqs. 4 and 1 more evident. In essence, the continuous time dynamical system is solved to transform it into the discrete time system. Again, Eq. 2 is a standard model of any linear change process in continuous time. Later we will use Eq. 4 to turn a continuous time model into its analogous discrete time model (Eq. 1) to fit a continuous time model to data. When paired with the proper statistical assumptions, Eq. 2 becomes a useful statistical model. Most often, we assume that  $\eta(t)$  is Gaussian and  $\zeta(t)$  follows a Wiener process, making Eq. 2 a stochastic differential equation (L. Arnold, 1974) and requiring the integrals in Eqs. 3 and 4 to be stochastic Itô integrals.

The analogue of Eq. 1 for a nonlinear system replaces the matrices  $\mathbf{B}_d$  and  $\mathbf{\Gamma}_d$  by a general nonlinear function  $f_d(\eta_{t-1}, \mathbf{x}_t)$ , yielding

$$\eta_t = f_d(\eta_{t-1}, \mathbf{x}_t) + \zeta_t \quad (5)$$

where we assume that  $f_d(\eta_{t-1}, \mathbf{x}_t)$  is a continuous and once differentiable function of  $\eta_{t-1}$  (Bar-Shalom, Ti, & Kirubarajan, 2001, p. 382, Equation 10.3.1-1). The continuous time version of Eq. 5 follows similarly.

$$\frac{d\eta(t)}{dt} = f_c(\eta(t), \mathbf{x}(t)) + \zeta(t) \quad (6)$$

where we must now add the assumption that  $f_c(\eta(t), \mathbf{x}(t))$  is also a continuous function of time (cf. Kalman, 1963, p. 155). Just as with the linear case of Eqs. 1 and 2, the functions  $f_d()$  and  $f_c()$  from Eqs. 5 and 6 are parallel in structure but not identical in meaning. The function  $f_d()$  directly maps the latent state forward in time, whereas  $f_c()$  maps the latent state on to its rate of change. Again, the differential equation in 6 can be solved to create a difference equation as in 5, but the reverse transformation is not always possible.<sup>4</sup> Also similar to the linear case of Eq. 2, for the purposes of model estimation the nonlinear system in Eq. 6 must be solved as in Eq. 3; however, the solution must be found numerically for the nonlinear case because no analytic solution like Eq. 4 exists for generic nonlinear systems (Hirsch et al., 2003).

So far, Equations 1 through 6 are purely mathematical representations of almost any change process. When paired with certain distributional assumptions, regularity conditions (including the stochastic Itô integral), and measurement models, these equations become general models of virtually any human change process. In the linear case these models are estimable in the `OpenMx` (Neale et al., 2016; Hunter, 2018) R package, and in the linear and nonlinear case they are estimable in the `dynr` R package (Ou, Hunter, & Chow, 2019) as well as several others. Both programs use Kalman filter one-step-ahead forecasts to compute the log likelihood of the data given free parameter values using a multivariate Gaussian density function in a procedure called prediction error decomposition (Schweppe, 1965; de Jong, 1988). There also exist ways to use ensemble forecasts to estimate the parameters of dynamical systems (e.g., J. L. Anderson, 1996; 2001; J. L. Anderson & Anderson, 1999). Although the estimation of the parameters of dynamical systems models is intimately related to their forecasts, the remainder of this paper focuses purely on the forecasting procedures.

## 2. Forecasting Dynamical Systems

There are two common and standard methods of forecasting when using time series methods like the Kalman filter to analyze intensive longitudinal data. Kalman forecasting is usually applied to linear dynamical systems, and ensemble forecasting is usually applied to nonlinear dynamical systems. Neither of these methods are new to the fields of statistics or time series (e.g., Box & Jenkins, 1976; Harvey, 1989; West & Harrison, 1997; Durbin & Koopman, 2001; Hyndman & Athanasopoulos, 2018), but to our knowledge they have not been previously applied in a behavioral setting.

### 2.1. Kalman Forecasting

Equation 1 gives the state equation in discrete time. In continuous time, Equation 2 or 6 must be solved for  $\eta$ . In the case of Equation 2, the analytic solution is known as the hybrid continuous discrete Kalman–Bucy filter (Kalman & Bucy, 1961). It is a hybrid in that measurement occurs in discrete time but the process is in continuous time. In the case of Equation 6, the hybrid continuous discrete *extended* Kalman–Bucy filter (Kulikov & Kulikova, 2014) makes a local linear approximation of the nonlinear function  $f_c()$ .

The (linear) measurement model for both discrete and continuous time is

$$\mathbf{y}_t = \mathbf{\Lambda}\eta_t + \mathbf{K}\mathbf{x}_t + \epsilon_t \quad (7)$$

where  $\mathbf{y}_t$  is the vector of observed variables,  $\mathbf{\Lambda}$  is a matrix of regression weights that maps the latent variables on to the observed variables akin to factor loadings,  $\mathbf{K}$  is a matrix of regression

<sup>4</sup>This result necessarily follows from the linear case (e.g., Brockwell, 1995) because the linear dynamics are a special case of the nonlinear dynamics.



weights for the observed exogenous covariates  $\mathbf{x}_t$ , and  $\boldsymbol{\epsilon}_t$  is a vector of unmeasured disturbances to the measurement process with covariance  $\boldsymbol{\Theta}$ . Note that the observations  $\mathbf{y}_t$  are always made at discrete (if sometimes unequally spaced) times. Thus, Eq. 7 applies to both continuous time and discrete time dynamical models.

An important aspect of Eqs. 1–6 and 7 is the two distinct kinds of disturbances in the dynamical systems we are discussing. There are dynamic disturbances  $\boldsymbol{\zeta}_t$  and  $\boldsymbol{\zeta}(t)$  on the one hand, and there are measurement disturbances  $\boldsymbol{\epsilon}_t$  on the other hand. Both dynamic disturbances and measurement disturbances are estimated simultaneously, but can be distinguished by their effects. Dynamic noise affects the latent process itself, whereas measurement noise only influences the observations. As such, dynamic noise carries forward in time, but measurement noise only impacts a single occasion and does not carry forward across time. An example may make this distinction more clear. A participant filling out a daily mood survey may have something happen to them that was not measured by the researchers and that negatively impacts their mood. Such an event would contribute to dynamic noise because it (1) impacts the participant's true mood rather than just the measurement of mood, and (2) would be expected to further influence later mood states. Alternatively, the wording on a particular mood item may be somewhat ambiguous, leading the participant to respond with some degree of inconsistency on that item. Such an event would contribute to measurement noise because it (1) strictly impacts the measurement of mood rather than the true, underlying mood state, (2) would be expected not to influence later mood states, and (3) only impacts the individual item with the ambiguous wording rather than spreading over the entire mood scale. Furthermore, dynamic and measurement noise impacts the ability to make forecasts differently.

A forecast can be made for both the latent variables ( $\boldsymbol{\eta}_t$  or  $\boldsymbol{\eta}(t)$ ) and for the observed variables ( $\mathbf{y}_t$ ), but the latter is dependent on the former. The forecasts discussed in this paper generally proceed recursively. A forecast for the latent state progresses from some initial estimate of the latent state. Subsequent forecasts are made sequentially based on previous forecasts. Suppose we have some initial latent state estimate which we denote  $\boldsymbol{\eta}_{t-1|t-1}$  to mean the estimate of the true  $\boldsymbol{\eta}_{t-1}$  at time  $t-1$  given all the information up to and including measurements at time  $t-1$ . Note that  $\boldsymbol{\eta}_{t-1|t-1}$  is not properly a forecast because it uses information up to and including time  $t-1$ . Rather,  $\boldsymbol{\eta}_{t-1|t-1}$  is akin to a regression-based factor score from factor analysis. Indeed, Priestley and Subba Rao (1975) showed under quite general circumstances that Kalman filter estimates of the latent states,  $\boldsymbol{\eta}_{t-1}$ , are equal to regression-based factor scores. Suppose furthermore that we have some estimate of the error covariance of  $\boldsymbol{\eta}_{t-1|t-1}$ . We denote our estimate of the error covariance at time  $t-1$  given all the information up to and including the measurements at time  $t-1$  by  $\mathbf{P}_{t-1|t-1}$ . Given these estimates of the initial latent state ( $\boldsymbol{\eta}_{t-1|t-1}$ ) and its error covariance ( $\mathbf{P}_{t-1|t-1}$ ), a Kalman forecast is constructed by mapping these forward in time according to the dynamic model (Kalman, 1960, Equation 23).

$$\boldsymbol{\eta}_{t|t-1} = \mathbf{B}_d \boldsymbol{\eta}_{t-1|t-1} + \boldsymbol{\Gamma}_d \mathbf{x}_t \quad (8)$$

Equation 8 is the expected value of Eq. 1 given the estimate  $\boldsymbol{\eta}_{t-1|t-1}$  of the true  $\boldsymbol{\eta}_{t-1}$ . That is,  $\boldsymbol{\eta}_{t|t-1}$  is the one-step-ahead forecast for the latent state at time  $t$ . The Kalman forecast error then becomes (Kalman, 1960, Equation 24)

$$\mathbf{P}_{t|t-1} = \mathbf{B}_d \mathbf{P}_{t-1|t-1} \mathbf{B}_d^\top + \boldsymbol{\Psi}_d \quad (9)$$

Equation 9 is the covariance of  $\boldsymbol{\eta}_{t|t-1} - \boldsymbol{\eta}_t$ . That is,  $\mathbf{P}_{t|t-1}$  is the one-step-ahead forecast error covariance matrix at time  $t$ .

The Kalman forecast is recursive: each forecast is built on the current latent state estimate. The Kalman forecast makes a prediction for the latent state one time step ahead of the current time; however, a chain of forecasts can readily be strung together to create an  $n$ -step ahead forecast. In the case of a linear, time-invariant, stable dynamical system, the long-range forecast for the latent state ( $\boldsymbol{\eta}_{t+n-1|t-1}$ ) always approaches the zero vector in the appropriate dimensional space (B. D. O. Anderson & Moore, 1979, Section 4.4) and the forecast error ( $\mathbf{P}_{t+n-1|t-1}$ ) approaches a steady state error covariance matrix (Harvey, 1989, p. 121, Equation 3.3.21).

Because the Kalman forecast is recursive, it requires initial estimates of the latent state and error covariance. If the forecast is made after several observations, then the most recent estimates of the latent state and error covariance are used for forecasting (Harvey, 1989, p. 120). If, however, no observations have yet been incorporated, then either asymptotic or diffuse latent state and error covariance matrices initialize the filter (Harvey, 1989, p. 121, Equation 3.3.22).

For a nonlinear dynamical system, the same one-step ahead forecast is possible. The straightforward expected value of Eq. 5 replaces Eq. 8 (Bar-Shalom et al., 2001, p. 383, Equation 10.3.2-4), and a local linear approximation of  $f_d(\boldsymbol{\eta}_{t-1}, \mathbf{x}_t)$  replaces  $\mathbf{B}_d$  in Eq. 9 (Bar-Shalom et al., 2001, p. 384, Equation 10.3.2-6). However, concatenating multiple one-step ahead forecasts to create an  $n$ -step ahead forecast becomes less tenable (Bar-Shalom et al., 2001, pp. 385–387). The forecast value continues to share many of the same properties as the linear case, but the forecast error becomes less accurate. The forecast value can easily be mapped forward in time exactly according to the nonlinear dynamics. Thus, the desirable properties of the forecast value persist in the nonlinear case. However, the forecast error requires a linear approximation and each repeated time step compounds the approximation errors. Thus, the forecast error quality degrades over repeated applications.

In continuous time linear and nonlinear systems, the same basic forecasting procedure follows the pattern of Eqs. 8 and 9 with one additional intervening step: the differential equations for the latent state and the dynamic error must be solved which reduces them to the discrete time case previously discussed. Before the standard forecast is possible for continuous time models, the continuous time latent state must be solved as in Eqs. 3–4. Additionally, the continuous time error creates a differential equation for the forecast error that can be solved similarly. In the linear case, these solutions are analytically known (see Kalman & Bucy, 1961) and computable (see Van Loan, 1978). In the nonlinear case, the solutions must be computed numerically (see Ou et al., 2019). Repeated forecasts for the continuous time case need not be strung together. Rather, the system of differential equations can be solved for the single specific forecast time. Instead of concatenating several forecasts together in discrete time steps, the forecast is constructed directly for the targeted time.

Regardless of the linearity of the dynamics or whether they occur in discrete or continuous time, the measurement takes place at discrete (if unevenly spaced) times. The one-step-ahead forecast latent state and covariance are used with Eq. 7 to create the forecasts for the observations.

$$\mathbf{y}_{t|t-1} = \mathbf{\Lambda} \boldsymbol{\eta}_{t|t-1} + \mathbf{K} \mathbf{x}_t \quad (10)$$

$$\mathbf{S}_{t|t-1} = \mathbf{\Lambda} \mathbf{P}_{t|t-1} \mathbf{\Lambda}^\top + \boldsymbol{\Theta} \quad (11)$$

$\mathbf{y}_{t|t-1}$  and  $\mathbf{S}_{t|t-1}$  are the forecast mean and forecast error covariance, respectively, for the raw data. We note that  $\mathbf{y}_{t|t-1}$  and  $\mathbf{S}_{t|t-1}$  are estimates rather than true values and use information up to time  $t - 1$  to make predictions about time  $t$ . When raw data observations of later time points become available, the latent forecasts are updated based on them by orthogonal projections (see Kalman, 1960).



$$\boldsymbol{\eta}_{t|t} = \boldsymbol{\eta}_{t|t-1} + \mathbf{P}_{t|t-1} \boldsymbol{\Lambda}^T \mathbf{S}_{t|t-1}^{-1} (\mathbf{y}_t - \mathbf{y}_{t|t-1}) \quad (12)$$

$$\mathbf{P}_{t|t} = \mathbf{P}_{t|t-1} - \mathbf{P}_{t|t-1} \boldsymbol{\Lambda}^T \mathbf{S}_{t|t-1}^{-1} \boldsymbol{\Lambda} \mathbf{P}_{t|t-1} \quad (13)$$

The updated latent state estimates,  $\boldsymbol{\eta}_{t|t}$ , and their error covariance,  $\mathbf{P}_{t|t}$ , are then used to create the forecasts for the next time point with  $\mathbf{P}_{t|t}$  being the covariance of  $\boldsymbol{\eta}_{t|t} - \boldsymbol{\eta}_t$ . Equations 12 and 13 optimally combine information from the forecasts and the observations to update the forecasts when new data are available (Brookner, 1998, Ch. 1 & 2). When no data are available, the update equations simply leave the original forecasts unchanged. Importantly, basic properties of the multivariate Gaussian distribution imply that only first- and second-order moments are needed to specify the forecast distributions (Eqs. 8–9 and 10–11) for linear models (Rao, 2001, Ch. 8; Kalman, 1960, p. 45, Theorem 5(A)). For nonlinear models, the forecast distributions are not necessarily Gaussian, but are approximations up to the second order moments (Kalman, 1960, p. 45, Theorem 5(C)).

## 2.2. Ensemble Forecasting

In essence the ensemble forecasting method simulates many trajectories from a dynamic model over the desired time period and then averages these simulated trajectories to create the mean forecast. The spread among the simulated trajectories indicates the precision of the forecast. Importantly, the ensemble forecast makes the individual trajectories available, not just their summary statistics. This is critical for nonlinear dynamical systems because they may exhibit behavior that drastically deviates from linear expectations. Depending on the model, small deviations in initial conditions may lead to arbitrarily large differences in long-term outcomes. In the literature on chaos, this is called sensitive dependence on initial conditions (e.g., Hirsch et al., 2003; Cvitanovic, Artuso, Mainieri, Tanner, & Vattay, 2017). Moreover, the forecast distributions for linear dynamics are always necessarily Gaussian (Rao, 2001, Ch. 8; Kalman, 1960, p. 45, Theorem 5(A)); however, the forecast distribution for nonlinear dynamics may have any distribution. Thus, obtaining the entire ensemble and using it as a sample approximation of the arbitrary distribution affords much more informed forecasts and forecast error estimates.

The origin of ensemble forecasting is generally traced to weather prediction in meteorology (Epstein, 1969; Leith, 1974) where the dynamics are high dimensional and highly nonlinear. Thus, the integral in the nonlinear version of Eq. 3 generally has no analytic solution and is computationally expensive. Consequently, simulation-based Monte Carlo methods of forecasting became expedient. There are numerous related but distinct versions of ensemble forecasting, and the reader is directed to modern texts (e.g., Warner, 2014; Coiffier, 2011) for details on these variations.

For the purposes of forecasting ILD, we propose to apply a basic perturbation method for creating trajectories (e.g., Katzfuss, Stroud, & Wikle, 2016, Equation 10). We choose this method because it is among the most basic ensemble methods and avoids many of the complexities encountered when forecasting atmospheric dynamics (e.g., J. L. Anderson & Anderson, 1999). In the perturbation method, the dynamic noise disturbances are randomly generated and added to the forecast values of the latent states. In discrete time, these disturbances have an easily specified distribution. Each of the  $k$  disturbances is Gaussian-distributed according to Eqs. 1 and 5.

$$\boldsymbol{\zeta}_{k,t} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Psi}_d) \quad (14)$$

In the continuous time case, the form is similar but the dynamic noise distribution must be integrated for each forecast time. For the linear continuous time case, the distribution takes the

form below which is the discretized dynamic noise from Eq. 2 (Harvey, 1989, p. 484, Equation 9.1.20b).

$$\boldsymbol{\zeta}_{k,t} \sim \mathcal{N}\left(\mathbf{0}, \int_{t_{i-1}}^{t_i} e^{\mathbf{B}_c(t_i-t_{i-1})} \boldsymbol{\Psi}_c e^{\mathbf{B}_c^\top(t_i-t_{i-1})} dt\right) \quad (15)$$

For the nonlinear continuous time case, the integral does not have a closed form but can be solved numerically as the discretization of the dynamic noise from Eq. 6. In both Eqs. 14 and 15 we assume a multivariate Gaussian distribution for the perturbations to yield results that asymptotically approach the Kalman forecasting results in the large ensemble limit (Katzfuss et al., 2016, p. 352).

In the perturbation ensemble, the  $k$ th member of an ensemble with  $K$  total members is created by adding a perturbation to the predicted latent state.<sup>5</sup>

$$\boldsymbol{\eta}_{1,t|t-1} = \mathbf{f}_d(\boldsymbol{\eta}_{t-1}, \mathbf{x}_t) + \boldsymbol{\zeta}_{1,t} \quad (16)$$

$$\boldsymbol{\eta}_{2,t|t-1} = \mathbf{f}_d(\boldsymbol{\eta}_{t-1}, \mathbf{x}_t) + \boldsymbol{\zeta}_{2,t} \quad (17)$$

$$\vdots = \vdots$$

$$\boldsymbol{\eta}_{K,t|t-1} = \mathbf{f}_d(\boldsymbol{\eta}_{t-1}, \mathbf{x}_t) + \boldsymbol{\zeta}_{K,t} \quad (18)$$

Note that  $\mathbf{x}_t$  are instantaneous observed disturbances which we assume are observed even for the targeted forecasting time  $t$ . If these disturbances are missing at time  $t$ , then missing data techniques should be paired with the chosen forecasting method (e.g., multiple imputation; see Li et al., 2019). At the next time step, each of the  $K$  members of the ensemble will also be perturbed when predicting forward in time. Thus, for single-subject forecasts,  $K$  ensemble members are forecast forward at each time; for multi-subject forecasts,  $K$  ensemble members are forecast forward at each time and for each person. Each person has their own  $K$ -member ensemble. We only show the discrete time case in Eqs. 16–18 because the continuous time case is always solved for each (possibly unevenly spaced) desired time point which effectively discretizes the continuous time model. The discretized dynamic errors for continuous time models evolve according to Eq. 15 for different amounts of time between forecasts.

Katzfuss et al. (2016) reviewed basic properties of the ensemble and Kalman forecasts which we describe next. In the linear case, the ensemble at each time will maintain a Gaussian distribution with the ensemble mean asymptotically approaching the Kalman forecast mean as the ensemble size  $K$  becomes large. That is, the mean of the left-hand side of Eqs. 16 through 18 approaches the left-hand side of Eq. 8 for the linear case and a large ensemble. Similarly for the linear case, the ensemble variance asymptotically approaches the Kalman forecast variance as  $K$  increases. That is, the covariance matrix of  $\boldsymbol{\eta}_{1,t|t-1}$  through  $\boldsymbol{\eta}_{K,t|t-1}$  approaches  $\mathbf{P}_{t|t-1}$  of Eq. 9 for the linear case and a large ensemble.

In the nonlinear case, the ensemble will generally be non-Gaussian, and the forecasts need not agree with the Kalman forecasts. The ensemble mean and variance can still be computed, but these may be less meaningful or important depending on the shape of the non-Gaussian distribution. Critically, the entire forecast distribution is available in the ensemble forecast. The availability of the entire ensemble distribution allows forecasts with known possibilities for forecast errors even though the forecast error distribution may be non-Gaussian. Nonparametric methods like sample

<sup>5</sup>Although the dynamical system we consider may be nonlinear, all of them are stable. Stable systems in discrete time and continuous time show no sensitive dependence on initial conditions (Arrowsmith & Place, 1990, Ch. 3). Instead, these systems have continuous dependence on initial conditions and perturbations (Hirsch et al., 2003, Ch. 17; V. I. Arnold, 1988, Ch. 3–4).

quantiles can create metrics for forecast errors regardless of the distribution. Thus, if the ensemble distribution is composed of two sharp peaks separated by a wide gap, the researcher may expect values near the centers of the peaks but nowhere in between them and may ignore the mean and variance of the ensemble in favor of ensemble histograms and quantiles.

Just as with the Kalman filter discussed previously, we have shown in Eqs. 16–18 the linear discrete time forecast procedure. The nonlinear and continuous time procedures follow as generalizations of the linear discrete time method. As with the Kalman filter, the continuous time model is solved for the target forecast time, thus creating a discrete time model for each desired forecast time. The nonlinear case replaces the linear forecasting function (Eq. 1 or 2) with the nonlinear forecasting function (Eq. 5 or 6). Thus, the ensemble forecast is similarly applicable across all types of dynamic system discussed in the present paper.

### 2.3. Graphical Comparison of Kalman and Ensemble Forecasts

To better understand the relation between the Kalman and ensemble forecasts, Fig. 1 shows two-dimensional examples of linear and nonlinear dynamical forecasts in discrete time. Suppose the last observation occurred at time  $t - 1$  and happened to be forecast forward with a spherical forecast error distribution. The left circle of each panel in Fig. 1 shows the forecast point estimate  $x_t$ , but most importantly shows the error distribution around the point estimate as a black circle. We take the black circle to be an isocontour of equal probability density. The shape on the right of each panel shows how the isocontour is deformed by forecasting it forward. For the linear systems the shapes become elliptical: a necessary feature of linear Gaussian systems under affine transformations (B. D. O. Anderson & Moore, 1979, Section 5.2). Panels A and B of Fig. 1 show two examples of the isocontours induced by linear dynamics. As discussed previously, the isocontours for Fig. 1 panels A and B are asymptotically the same for the Kalman and ensemble forecasts. In the Kalman forecast case, the forecast point estimate is given by Eq. 8 and the shape of the isocontour is given by the Gaussian distribution implied by the forecast error covariance matrix in Eq. 9. In the ensemble forecast, the forecast point estimate is the sample mean of the ensemble members in Eqs. 16 through 18 and the shape of the isocontour is given by the sample covariance matrix of the same ensemble members. More generally, sample quantiles better represent the nonparametric distribution of the ensemble forecast distribution.

The nonlinear examples in panels C and D of Fig. 1 show the differences between the Kalman and the ensemble methods. The black lines still show the isocontour from the ensemble members, but the gray line now shows the local linear approximation from the extended Kalman filter. The ensemble isocontours can be calculated from the two-dimensional quantiles of the ensemble members, whereas the isocontours for the Kalman filter are linearized approximations of these from the (now approximate) Kalman forecast error in Eq. 9 from locally linearizing the dynamics for use in the matrix expression (see, e.g., Bar-Shalom et al., 2001, Ch. 10).

The degree of approximation error between the ensemble and Kalman methods depends on the amount and kind of the nonlinearity. In Fig. 1 panel C, our subjective judgment is that error induced by the Kalman forecast's linear, Gaussian approximation is minimal. However, for Fig. 1 panel D, important features of the ensemble distribution are not represented in the Kalman forecast distribution. Forecasts of very different properties result from the ensemble and the Kalman forecasts in panel D. Thus, some nonlinear dynamics are well-approximated by linear dynamics (e.g., Fig. 1 panel C), whereas others are not (e.g., Fig. 1 panel D).

## 3. Software Implementation of Kalman and Ensemble Forecasting

To increase the utility of the forecasting methods discussed in this paper, we have implemented them in freely available open-source software. In particular, the authors have added `predict`

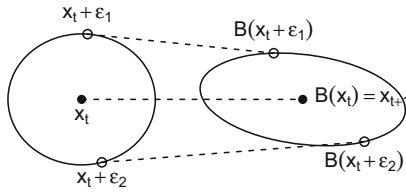
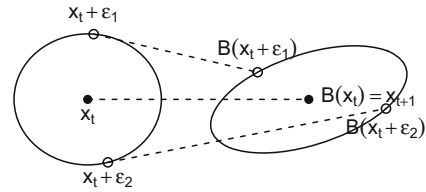
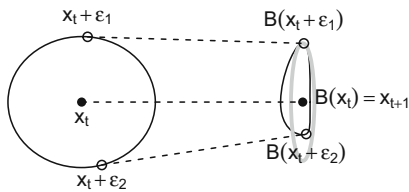
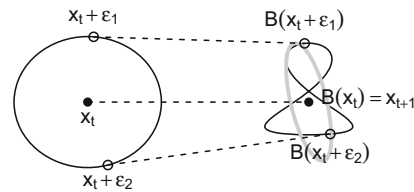
**A: Example Effect of Linear Dynamics****B: Example Effect of Linear Dynamics****C: Example Effect of Nonlinear Dynamics****D: Example Effect of Nonlinear Dynamics**

FIGURE 1.

Example two-dimensional forecasting distributions for linear and nonlinear dynamical systems. The gray lines for the nonlinear systems show a linear approximation to the nonlinear dynamics.

methods to two R (R Development Core Team, 2020) packages that fit the dynamical systems models discussed in this paper. The `predict` methods were added to the `OpenMx` (Neale et al., 2016, version 2.18) package and the `dynr` (Ou et al., 2019, version 0.1.16) package. More details of these methods and working examples are in online supplementary materials.

#### 4. Application to Drug Use

Of the 20 million individuals aged 18 and older who suffer from substance use disorders (SUDs) in the United States (Substance Abuse and Mental Health Services Administration, 2018; Lipari & Van Horn, 2017), approximately 7.6% will seek (or be forced to seek) treatment (Lipari & Van Horn, 2017), and up to 60% of those relapse within 1 year (McLellan, Lewis, O'Brien, & Kleber, 2000; Bowen et al., 2014). In this section, we apply time series forecasting methods to substance use. The goals of the application are fourfold. First, we aim to construct a theoretically plausible model of substance use. Such a model is not intended to be the “true” or “best” model for the data, but needs to account for the binary nature of the most easily collected daily substance use data: namely, whether or not the person used drugs or alcohol on a particular day. Second, we aim to fit this model to time series data from multiple individuals’ actual substance use behavior. Third, we aim to create Kalman and ensemble forecasts from the estimated model, comparing them to each other and to several simpler alternatives. Fourth, we aim to evaluate the forecasts using hold-out data reserved for this purpose. The evaluation scheme for forecasts is generally a variation on cross-validation via a holdout sample (see Harvey, 1989; Box & Jenkins, 1976; West & Harrison, 1997; Warner, 2014; Coiffier, 2011, for details). We assess the accuracy, calibration, and precision of the advocated forecasts and several alternative forecasts. However, note that such

a forecasting evaluation is necessarily model and data specific, and does not necessarily reflect the quality of the forecasting methods for all purposes.

#### 4.1. Participants and Design

Data were obtained from 354 participants (36% male; 79% Caucasian; mean age = 34.51 [SD = 10.78]) undergoing a randomized clinical trial set within an inpatient substance use treatment facility in a southeast US metropolitan area (see Bornovalova, Gratz, Daughters, Hunt, & Lejuez, 2012, for a full description). The majority (82.6%) of our participants were involved in the criminal justice system and were court-mandated to treatment. Retrospective daily drug and alcohol use data were collected at approximately 1, 3, and 6 months after community reintegration via the drug and alcohol Timeline Follow-Back (TLFB; Robinson, Sobell, Sobell, & Leo, 2014; Sobell et al., 2001). Of the original sample, daily drug and alcohol use data were available for 261 participants (74% of the original sample; 66% male; 66% Caucasian non-Hispanic; mean age = 34.83 [SD = 10.70]) with between 26 and 275 days of data per person ( $M = 165.5$ ,  $SD = 55.3$ ). These binary daily drug and alcohol use data form the basis of the subsequent model fitting and forecasting.

#### 4.2. Model Development

Based on both a theoretical and empirical understanding of substance use, we expect at least two “modes” of behavior in this sample: use and non-use. Underlying the binary use data, we expect there may exist a continuous and time-varying latent inclination to use substances that is subject to unmeasured disturbances (e.g., Maisto, Hallgren, Roos, & Witkiewitz, 2018). A dynamic system with two stable modes of operation is said to exhibit bistability. One of the simplest models for bistability in continuous time is given by the following equation:

$$\frac{d\eta}{dt} = -a\eta(1-\eta)(\eta-b) + \zeta(t) \quad (19)$$

Equation 19 is a nonlinear continuous time dynamical systems model, a special case of Eq. 6. This model has three free parameters ( $a$ ,  $b$ , and  $\sigma_\zeta^2$ ), the interpretation of which follows.

If we omit for the moment consideration of the stochastic term  $\zeta(t)$ , Eq. 19 is an example of a gradient dynamical system (see Hirsch et al., 2003, Ch. 9). The term “gradient dynamical system” refers to the right hand side of Eq. 19 being the gradient of some “potential” function that acts like a potential energy function of a physical system. Among the class of nonlinear differential equations, gradient dynamical systems are particularly well-understood and well-behaved (Smale, 1961). Their fixed points and other dynamical characteristics are often easy to determine. For example, Eq. 19 has three fixed points: 0 which is stable/attracting, 1 which is also stable/attracting, and  $b$  which is unstable/repelling. Moreover, the behavior of the solutions of gradient dynamical systems can be understood by using the metaphor of a ball rolling in a potential well. The potential well determines the shape of a “hill,” and the ball simply rolls on the surface of this hill. The potential field of the gradient dynamical system of Eq. 19 is a fourth-order polynomial. The shape of the potential function for Eq. 19 has two wells, so we refer to it as the double-well potential model in the present work.

Several example potential fields are shown in Fig. 2. As illustrated in the figure, the  $a$  parameter determines the total stability of the two stable fixed points: the taller the barrier between the two local minima, the more difficult it is to move from one fixed point attractor to the other. The  $b$  parameter determines the relative stability of the two minima: thus, a transition from 1 (use) to 0 (non-use) may be much more difficult than the reverse. As the  $b$  parameter approaches 1, the well at 0 remains stable but the well at 1 becomes less stable. As the  $b$  parameter approaches 0, the well at 1 remains stable but the well at 0 becomes less stable. In the limiting case of  $b = 1$ ,

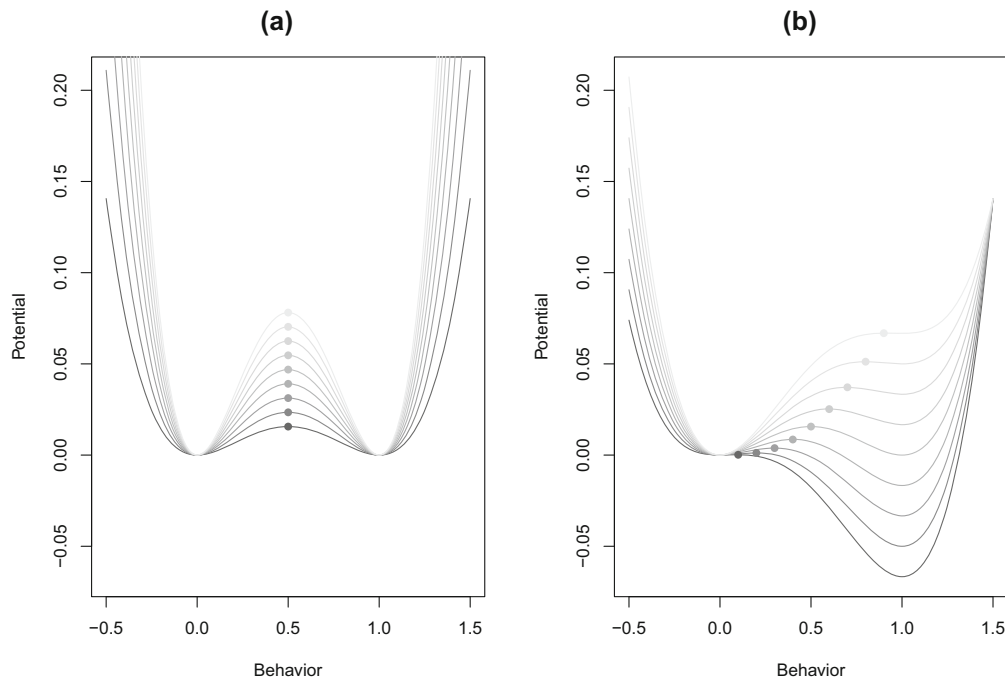


FIGURE 2.

Example potential field for substance use. There are two stable fixed points: one at each of the local minima at 0 and 1. There is one unstable fixed point indicated by the solid dot. (a) Potential field varying the  $a$  parameter for total stability from  $a = 1$  in dark gray to  $a = 5$  in light gray by steps of 0.5. (b) Potential field varying the  $b$  parameter for relative stability of the two stable fixed points from  $b = 0.1$  in dark gray to  $b = .9$  in light gray.

there is no well at one. Similarly, in the limiting case of  $b = 0$ , there is no well at zero. Thus, the  $b$  parameter is bounded between 0 and 1 and determines the relative stability of the two wells.

Now incorporating the stochastic term  $\zeta(t)$ , the mental representation of this model is of a ball rolling in the potential well shown in Fig. 2 while simultaneously being constantly, randomly perturbed. Once in the neighborhood of a well, each well is stable and thus self-sustaining. Some disturbance is then required to move it from its current stable behavior. The variance  $\sigma_\zeta^2$  gives the variance of the stochastic shocks to the ball in the potential well. Depending on the depth of the two wells and the size of the shocks, the system will vacillate between the two wells. Whenever a disturbance of sufficient strength impacts the ball (exceeding the separation energy of the wells), the ball moves from one well to the other.

We pair the dynamical model in Eq. 19 with the simple measurement model

$$y_t = \eta_t + \epsilon_t \quad (20)$$

which is a special case of Eq. 7. For the purposes of this data application, we constrain the measurement noise variance  $\sigma_\epsilon^2 = 0.5$ . This decision has both practical and theoretical considerations behind it. On the practical side, in our experience it is often numerically difficult to simultaneously estimate the measurement noise and the dynamic noise for models with only one indicator,<sup>6</sup> but no such model converged adequately for our purposes. On the theoretical side, we are expressing

<sup>6</sup>Unlike in structural equation modeling, it is possible to estimate both the measurement noise and the dynamic noise for single indicator models (Hunter, 2018). Such models are identified and routinely estimated in some settings (e.g., radar tracking, Brookner, 1998)



a fixed degree of uncertainty in the measurements by setting their error variance to  $\sigma_\epsilon^2 = 0.5$ . Psychometrically, we are saying the standard error of measurement is  $\sqrt{.5} \approx 0.707$ , or about 85% confidence in any given use/non-use response.<sup>7</sup>

Finally, we freely estimate the initial mean and variance assuming a Gaussian distribution.

$$\eta_1 \sim \mathcal{N}(\mu_\eta, \sigma_\eta^2) \quad (21)$$

Given the length of the time series modeled, the initial conditions do not greatly constrain the model in any way. Models such as these become independent of their initial conditions exponentially fast (Harvey, 1989).

### 4.3. Model Results

We estimate the parameters of Eqs. 19 and 21 using the `dynr` program (Ou et al., 2019), noting that all the parameters of Eq. 20 are fixed, and that a single model is estimated simultaneously for all participants with the same parameter values. `dynr` uses maximum likelihood prediction error decomposition (Schweppe, 1965; de Jong, 1988) for its continuous discrete extended Kalman filtering (Kulikov & Kulikova, 2014). The method assumes a Gaussian distribution for the measurement residuals (i.e.,  $\epsilon_t$  of Eq. 20): an assumption we are certainly violating given our binary measurements. We discuss consequences of this violation under the ‘‘Model Assessment’’ online supplement.

Bergmeir and colleagues (Bergmeir & Benítez, 2012; Bergmeir, Costantini, & Benítez, 2014; Bergmeir, Hyndman, & Koo, 2018) suggested evaluating time series models by using an initial portion of the data for training, and reserving the later times as a hold-out sample. Therefore, we fit the model to the first 80% of each person’s data, retaining the remaining 20% as a hold-out sample for later forecasting evaluation.

Table 1 shows the resulting parameter estimates which are the same across all people. All parameter estimates are statistically significantly different from zero, likely due to the large sample size in both people and time. As noted previously, the  $a$  and  $b$  parameters define the shape of the potential well for the dynamical system. A visual depiction of the gradient function and the corresponding potential well is shown in Fig. 3a and b, respectively. The slope (i.e., gradient) of the curve in Fig. 3b is given by the curve in Fig. 3a. The estimated parameters indicate that the non-use state is more stable than the use state because its attractor basin is deeper. The parameters also indicate that it is not difficult to pass from use to non-use in this sample because neither attractor basin is particularly deep. The dynamic noise  $\sigma_\zeta = \sqrt{0.0023} \approx 0.048$  suggests that the typical magnitude of a few days of shocks could easily send a person from one attractor basin to the other. Finally, the initial conditions imply that most people start near the non-use basin ( $\mu_\eta = 0.0759$ ), but vary substantially  $\sigma_\eta \approx 0.216$ .

For some of the forecast evaluation, we trained the model only on the first 30% of each person’s time series. The resulting parameter estimates were quite close to those reported in Table 1 with the exception that the dynamic noise variance,  $\sigma_\zeta^2$ , was estimated at nearly zero. We discuss further model assessment regarding assumed homogeneity of subjects, unmodeled binary observations, and un-estimated measurement error in an online supplement.

<sup>7</sup>If the standard error of measurement is .707, then an 85% confidence interval would be  $\pm .707 \cdot 1.44 \approx 1.00$  assuming asymptotic normality.

TABLE 1.  
Parameter estimates from candidate nonlinear dynamical systems model estimated on first 80% of each person's data.

	Estimate	Std. Error	<i>t</i> -value	2.5% CI	97.5% CI	Pr(>  <i>t</i>  )
<i>a</i>	0.1398	0.0169	8.2682	0.1067	0.1730	0.0000
<i>b</i>	0.4943	0.0238	20.7706	0.4476	0.5409	0.0000
$\sigma_{\zeta}^2$	0.0020	0.0003	7.3059	0.0014	0.0025	0.0000
$\mu_{\eta}$	0.0724	0.0201	3.6041	0.0330	0.1118	0.0002
$\sigma_{\eta}^2$	0.0404	0.0073	5.5377	0.0261	0.0546	0.0000

Note. Std. Error=Standard Error. CI= Confidence Interval Bound.

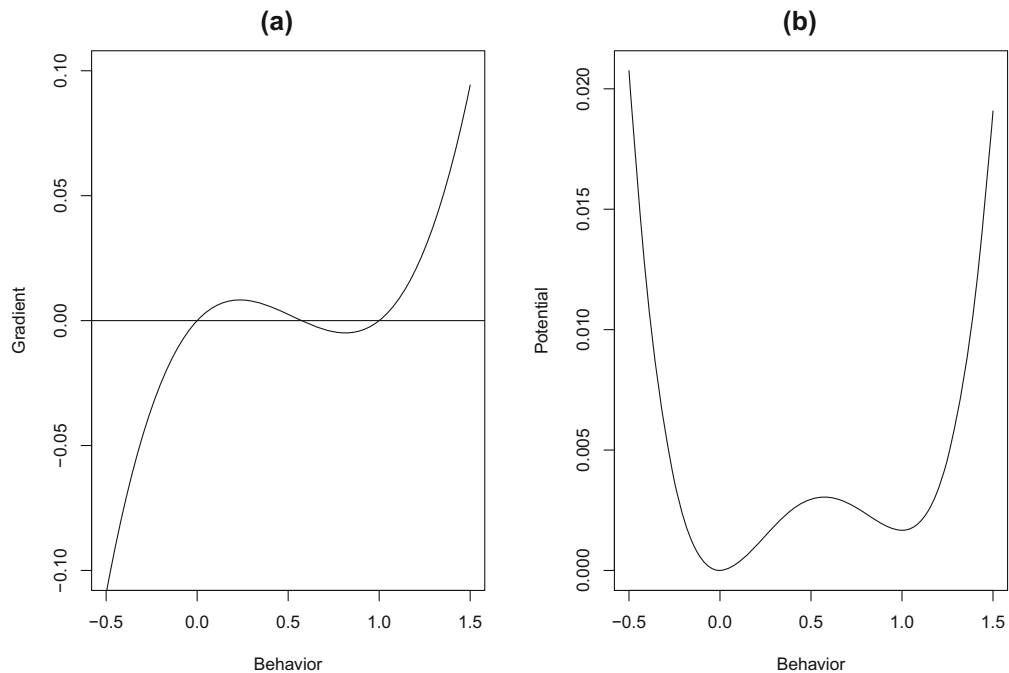


FIGURE 3.

Estimated gradient function and its corresponding potential function. (a) Gradient function. The intersections of the curve with the zero line give the local minima and maxima of the potential function. (b) Potential function. A person's binary substance use behavior is modeled as a ball rolling along the curve shown.

#### 4.4. Forecasts

To best describe the similarities and differences between forecasting methods, we construct both Kalman forecasts and ensemble forecasts from the fitted model. R code to construct Kalman and ensemble forecasts for the double-well potential model is shown in online supplementary materials as well as on the Open Science Framework page <https://osf.io/5q8z9/> (Ou et al., 2019; Hunter, 2018). Figure 4 shows several examples of Kalman and ensemble forecasts, the latter of which uses 1,000 ensemble members for each person at each time point. The solid lines in Fig. 4 give the forecasts, with the dashed lines and shaded regions giving the 95% confidence intervals. Additionally, 20 of the 1,000 ensemble members are also shown for the ensemble forecasts in Fig. 4.

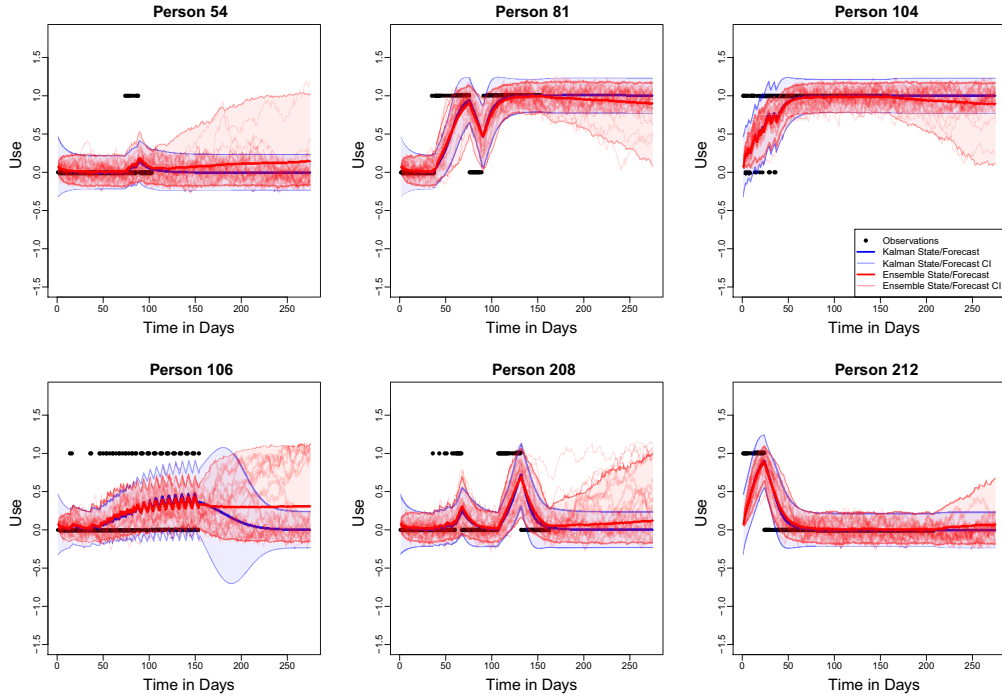


FIGURE 4.

Example observations and forecasts using the gradient model in Eq. 19. CI = 95% Confidence Interval.

Several features of Fig. 4 are noteworthy. First, when observed data are present, the two methods necessarily agree. However, when making forecasts (i.e., predictions when no observations are present) for nonlinear systems, the two methods can diverge and in these examples they do diverge. Second, the forecast confidence intervals generally also behave differently. The ensemble confidence intervals may be asymmetric, and can be either wider or narrower than the Kalman-based confidence intervals which according to its assumptions must always be symmetric and Gaussian. The ensemble confidence intervals tend to expand and sometimes include both use at 1 and non-use at 0, whereas the Kalman-based confidence intervals are often narrower and exclude one or the other. One of these methods may be overly confident or overly liberal, and we investigate this further in the forecasting evaluation. Third, related validation work is needed to address differences in predicted long-term stability. The Kalman forecast in each case predicts long-term stability at either daily use or no use whatsoever. The ensemble forecast mean, by contrast, tends toward the middle between use and non-use. This centralizing tendency is not a general feature of ensemble forecasts, but rather of these forecasts paired with the fitted model. Some of the ensemble members cluster around 0, whereas others cluster around 1: the tightness of the clustering is related to the  $a$  parameter for total stability. Aggregating over the entire ensemble produces a mean that is not particularly representative. The entire ensemble distribution is useful in this regard. Although we use the ensemble mean as our forecast value, other summary statistics from the ensemble distribution are possible and the subject of future work.

Figure 5 shows a further demonstration of the non-representativeness of the mean. In Fig. 5, we show the histograms of the entire ensemble for person 106 at various times. The distribution initially appears Gaussian but quickly becomes skewed, somewhat diffuse, and then bimodal. This may reflect a general breakdown in predictability over time and is not reflected in the Kalman forecasts which continue to exclude 1 for this person throughout time.

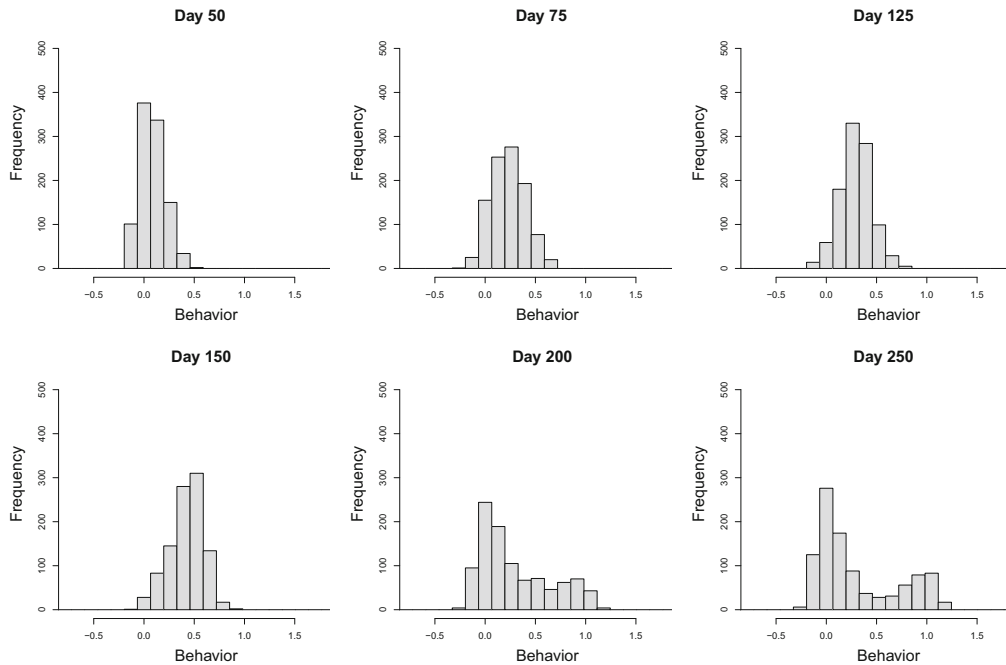


FIGURE 5.  
Histograms of the ensemble forecast members at various times for person 106.

#### 4.5. Forecast Evaluation

To check the performance of the applied forecasting procedures we evaluate the forecasts in three ways. First, we evaluate the forecast *accuracy*; that is, we measure how close the predicted values are to the observed values. Second, we evaluate the forecast error *calibration*; that is, we measure how close the forecast error estimates are to the actual errors. An ideal forecast will have high accuracy (i.e., small prediction errors) and be well-calibrated. Third, we evaluate the forecast *precision* by examining the absolute width of the forecast intervals. Narrower intervals indicate greater forecast precision.

To observe forecast performance under a variety of settings, we create 1-day, 7-day, 30-day, and 91-day forecasts. These forecast lengths correspond to useful timescales for making predictions about drug and alcohol use: 1-day, 1-week, 1-month, and 1-season, respectively. The 1-, 7-, and 30-day forecasts used models that were trained on the first 80% of each person's time series. For example, if a person had 165 days of observation, then the models were trained on the first 132 days of that person's data. The predictions for this person are then for the 133rd, 139th, and 162nd days. Each person may have a different number of days of observation and consequently their forecasts occur on different days, but the time lags for the forecasts are the same across all people.

For the 91-day forecasts, models were trained on the first 30% of each person's time series instead of the first 80%. Sample size and missing data were the primary drivers of this decision. We wanted to balance a sufficiently large sample size for effective model training while maintaining enough non-missing hold-out data for effective testing. For the 80% training data, 100% of people had 1-day ahead observations (261 out of 261), 98% had 7-day ahead observations (255 out of 261), and 72% had 30-day ahead observations (189 out of 261). For the 30% training data, 76% had 91-day ahead observations (199 out of 261).

In addition to the Kalman and ensemble forecasts from the double-well potential model, we selected nine simple forecasting models for accuracy comparison. A subset of these were also used for forecast calibration and precision. The mean and mode across all training observations were the first two simple models. The third and fourth models were linear and logistic latent growth curves using the `lme4` package (Bates, Mächler, Bolker, & Walker, 2015)

$$y_{ij} = b_{0j} + b_{1j}Time_{ij} + e_{ij} \quad (22)$$

$$b_{0j} = \gamma_{00} + u_{0j} \quad (23)$$

$$b_{1j} = \gamma_{10} + u_{1j} \quad (24)$$

where  $y_{ij}$  is the binary drug/alcohol use variable for person  $j$  and time  $i$ . To aid model convergence, the  $Time_{ij}$  variable was rescaled such that 0 was the first possible observation and 1 was the last possible observation. The logistic version of Eq. 22 adds the logistic transformation to the same basic model, adjusts the level-1 residual distribution, and uses Laplace approximation for the numerical integration. The fifth and sixths simple forecasting models were the within-person means and modes.

The final three simple forecast models were more conventionally from the time series literature (e.g., Hyndman & Athanasopoulos, 2018) and were fit idiographically as separate models for each person. The seventh forecasting model was the naive carry-forward method. Statistically, we instantiated this model as a random walk without drift. Equation 25 gives the more general random walk with drift model.

$$y_i = c + y_{i-1} + e_i \quad (25)$$

The naive random walk model without drift results from setting the drift parameter  $c$  to zero. The eighth forecast model allowed the drift parameter to be nonzero. The final simple forecast model was an automatically selected ARIMA model. All of these last three forecasts along with the within-person mean forecast were created with functions from the `forecast` package (Hyndman & Khandakar, 2008).

**4.5.1. Forecast Accuracy** The primary metric for forecast accuracy was the root mean squared error (RMSE) between the forecast point estimate and the true observation from the hold-out data. Figure 6 shows the forecast accuracy for the simple forecasting methods along with the Kalman and ensemble forecasts based on the double-well potential model.

For the 1-day and 7-day forecasts, the simplest conventional time series forecasts are clearly superior: the naive, drift, and automatically selected ARIMA model have the smallest RMSE. For the 30-day forecasts, the differences between forecast methods are generally diminished but with slightly better performance from the same three time series methods and also from the Kalman and ensemble methods. For the 91-day forecasts, the linear latent growth model and the drift model show extremely poor performance. Any small trend in the first 30% of the data is assumed to continue for the remaining 91 days. Implausible forecasts result from these linear trends when they are assumed to continue for too long.

In addition to the 30% trained 91-day forecasts, we produce a modified forecast for the double-well potential model. We use the parameters from the 80% training data in the double-well potential model, but forecast from the 30% training location. These modified forecasts are depicted in Fig. 6 as 'x's for the 91-day forecasts. The modified forecasts capitalize on greater model parameter precision when making long-range forecasts, but still face the same long-timeline challenges as the other methods. For the 91-day forecasts, the ensemble method trained on 80% of the data is slightly superior to all other forecasting methods.

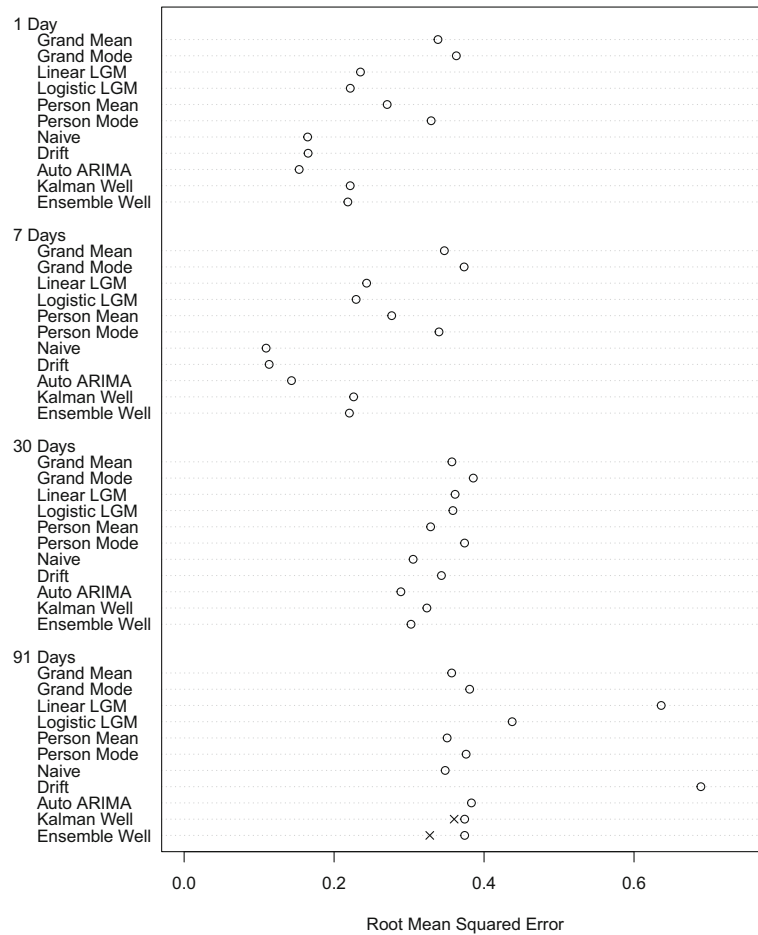


FIGURE 6.

Accuracy of forecasts as measured by root mean squared error between observed and predicted values. 1-day, 7-day, and 30-day forecasts were trained on the first 80% of each person's data and all people. The 91-day forecasts were trained on the first 30% of each person's data and all people. The 'x's for the Kalman and ensemble forecasts at 91 days indicate results from the parameters of the 80% training model, but forecast from the 30% starting point. LGM=Latent Growth Model; ARIMA=Autoregressive Integrated Moving Average; Well=Double-Well Potential Model.

Overall, no forecasting method we tested had particularly high accuracy. The RMSE is consistently in the 0.2 to 0.4 range with few exceptions. Of course, the poor forecast accuracy is not necessarily a reflection on the forecasting method *per se*, but rather on the model used to make the forecasts and that model's fidelity to the data. The simplest models (naive carry-forward, drift, and automatic ARIMA) are clearly performing best for the short-range 1-day and 7-day forecasts. These forecast models are quite similar. The naive method simply carries the previous observation forward unchanged. The drift method adds an estimated trend parameter  $c$  (Eq. 25), but 95% of estimated drift parameters were between  $-0.007$  and  $+0.007$ . The automatic ARIMA method conducts a model search across a range of autoregressive and moving average parameters along with adding differencing ("integration") between observations. The four most commonly selected ARIMA models were ARIMA(0, 0, 0) (168/261 = 64%), ARIMA(0, 1, 0) (33/261 = 13%), ARIMA(0, 1, 1) (12/261 = 5%), and ARIMA(0, 0, 1) (7/261 = 3%). Together, the four most commonly selected ARIMA models include 84% of the total number of people.



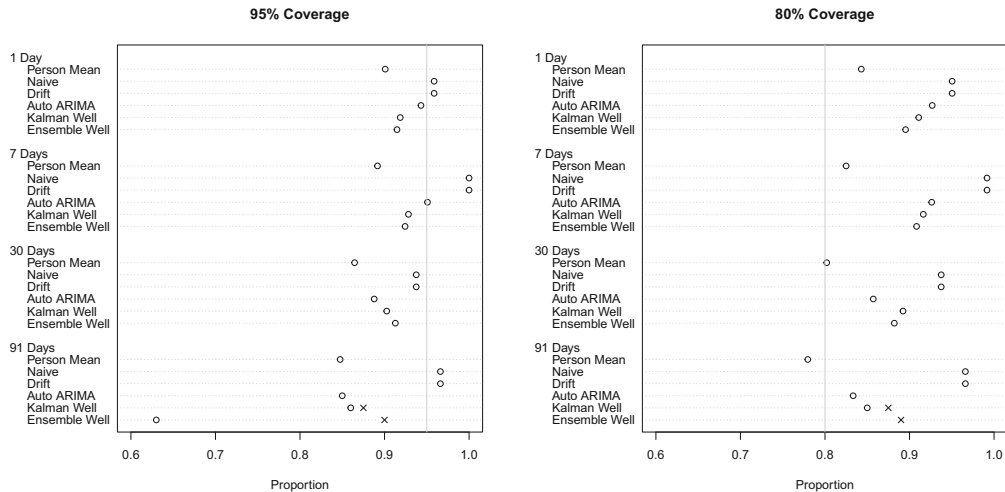


FIGURE 7.

Calibration of forecast errors as measured by coverage: the proportion of times the observed point is included within the forecast error distribution of a given confidence level. 1-day, 7-day, and 30-day forecasts were trained on the first 80% of each person's data and all people. The 91-day forecasts were trained on the first 30% of each person's data and all people. The 'x's for the Kalman and ensemble forecasts at 91 days indicate results from the parameters of the 80% training model, but forecast from the 30% starting point. The coverage for ensemble 91-day forecast of the 30% trained model is 20% and is not shown. ARIMA=Autoregressive Integrated Moving Average; Well=Double-Well Potential Model.

The success of the simple methods of forecasting may be due to two factors. First, the naive and drift methods rely heavily on persistence: whatever a person is doing at time  $t$  is predicted to continue at time  $t + h$  for any lag  $h$ . The double-well potential model that is the basis for our Kalman and ensemble forecasts has no built-in persistence. In fact, at every time the model supposes a Gaussian-distributed dynamic noise shock with standard deviation  $\sqrt{\hat{\sigma}_\eta^2} = 0.201$ . A 2.5 standard deviation shock would easily send a person out of one potential well and into the other: from use to non-use or vice versa. The second factor in favor of the simple methods is their large number of free parameters. All of the simple forecasting methods are fit idiographically: separate models for each person. Even the naive method has two free parameters per person (the last observation and the within-person standard deviation), yielding 522 total parameters. By contrast, the double-well potential model has just 5 free parameters because it is fit nomothetically. The first factor suggests that adding more persistence to the double-well potential model may improve its forecast accuracy. The second factor suggests that fitting more idiographically may be necessary for improved forecast performance.

**4.5.2. Forecast Calibration** Beyond forecast accuracy, some of the evaluated forecast methods also come paired with estimates of their own forecast errors. We now seek to examine the accuracy of these forecast errors, a property called calibration. Figure 7 shows the coverage of 95% and 80% forecast confidence intervals for a subset of the methods used to evaluate forecast accuracy.

We chose to evaluate the forecast calibration of the within-person mean, the naive carry-forward random walk,<sup>8</sup> the random walk with drift, the automatically selected ARIMA model, and the Kalman and ensemble forecasts from the double-well potential model. The simpler forecasting methods were chosen because (1) they performed reasonably well on forecast accuracy, (2) they

<sup>8</sup>The naive carry-forward forecast method has no necessary forecast error, but we augment it with one by using a random walk, as is common (e.g., Hyndman & Athanasopoulos, 2018, Ch. 3).

span a set of simple alternative forecasting methods, and (3) they have readily available forecast error functions.

We define coverage as the proportion of times that the observed value falls within the forecast error range. Ideally, a 95% forecast confidence interval will include the observed value 95% of the time, and exclude it 5% of the time. A forecast error distribution that includes the observed value at the nominal error rate is well-calibrated. In the context of forecasting, coverage that is much higher than the nominal rate implies that the forecast error distribution is wider than it should be, leading to forecasts that are much less precise than they could be. By contrast, coverage that is much lower than the nominal rate implies that the forecast error distribution is narrower than it should be, leading to forecasts that are overly precise. Both kinds of miscalibration are problems. Depending on the context, under-confidence or over-confidence might have more severe consequences.

In Fig. 7 it appears that most of the 95% forecasts are over-confident. The true coverage rates are often substantially lower than the nominal rate, suggesting that the forecast error distributions are too narrow. The notable exception to this over-confidence is the naive and drift methods, which are under-confident. The forecast error distributions for the naive and drift methods are too large. For the naive and drift methods, the forecast error distributions grow linearly without bound, resulting in under-confident miscalibrated forecasts.

The 80% forecasts in Fig. 7 are generally under-confident. The true coverage rates are often substantially higher than the nominal rate, suggesting that the forecast error distributions are too wide. Even among the under-confident 80% forecasts, the naive and drift methods are the most egregiously miscalibrated. Although the naive and drift forecasts showed some of the best forecast precision in Fig. 6, they also showed some of the worst calibration in Fig. 7. The automatically selected ARIMA model might be the best balance of forecast accuracy and coverage. However, the automatically selected ARIMA model is entirely idiographic, and consequently has a very large total number of parameters. The Kalman and ensemble forecasts are strong contenders for a good combination of forecast accuracy and coverage when accounting for parsimony.

An important note about the 91-day ensemble forecasts is needed. Just as with the forecast accuracy, two versions of forecasts were created for the 91-day task. The 'o's in Fig. 7 show the forecast calibration for models trained on the first 30% of the data. For the Kalman and the ensemble forecasts, an additional forecast was made that used the parameters from the 80% trained model, but forecast from the 30% time point. These forecasts are shown as 'x's in Fig. 7. The ensemble forecast coverage is particularly negatively affected by the 30% training data. The estimates of the dynamic noise variance were much smaller for the 30% trained model than for the 80% trained model. Consequently, the ensemble method—which relies on the dynamic noise variance to create perturbations—had very narrow forecast error distributions and led to 63% coverage for the nominally 95% forecasts and 20% coverage for the nominally 80% forecasts. The latter is not shown in Fig. 7. This finding highlights (1) the reliance of the ensemble method on accurate dynamic noise variance estimation, and (2) the possibility of non-stationary dynamic noise processes in the observed data. Essentially, in the first 30% of the data, there are many fewer disturbances that would cause variation in drug and alcohol use behavior.

*4.5.3. Forecast Precision* Complementing the forecast accuracy and the forecast calibration is an examination of the absolute forecast interval width: a measure of forecast precision. We define the width of the forecast interval as the upper bound minus the lower bound. Figure 8 shows box plots of forecast widths for six forecasting methods that were analyzed for forecast calibration: within-person mean, naive carry-forward random walk, random walk with drift, automatically selected ARIMA, Kalman forecasts from the double-well potential model, and ensemble forecasts from the double-well potential model. The box plot shows the distribution of the forecast widths across people because each person has their own forecast width under each method and for each lag.

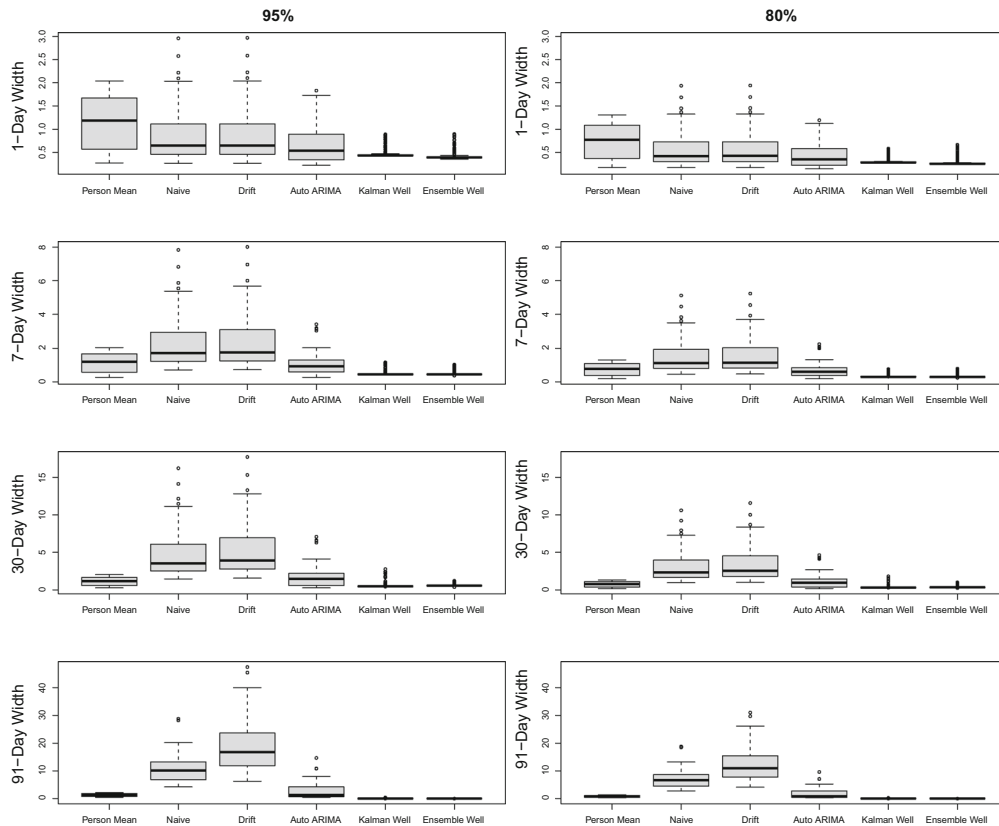


FIGURE 8.

Box plots of widths of forecast error intervals. Interval width is the upper bound minus the lower bound. The left column is for 95% confidence forecasts; the right column is for 80% confidence forecasts. The rows show 1-day, 7-day, 30-day, and 91-day ahead forecasts. 1-day, 7-day, and 30-day forecasts were trained on the first 80% of each person's data and all people. The 91-day forecasts were trained on the first 30% of each person's data and all people. The vertical axis is the same across each row, but differs for different rows. ARIMA=Autoregressive Integrated Moving Average; Well=Double-Well Potential Model.

We show 95% interval widths in the left column of Fig. 8 and 80% widths in the right column. The rows show different forecasting lags: 1 day, 7 days, 30 days, and 91 days ahead. As before, the 1-, 7-, and 30-day models were trained on the first 80% of each person's time series, whereas the 91-day models were trained on the first 30%. In every case, the Kalman and the ensemble widths are on average the narrowest. Moreover, the forecasts widths differ across people much less for the Kalman and the ensemble methods than for any of the simpler methods. The narrow forecast widths for the Kalman and the ensemble methods are notable because the accuracy and the coverage of these methods were comparable to the simpler methods. In particular, note that although the 1-day and 7-day forecasts were considerably more accurate for the naive and drift methods than for the Kalman and ensemble methods, the forecast widths were much wider. So, the naive and drift methods provided accurate but imprecise forecasts.

A forecast interval width wider than 1.0 is not useful for the binary drug and alcohol use data because it encompasses both use and non-use. Figure 9 shows the proportion of all forecasts with widths less than 1.0. We call the proportion of forecast intervals with widths less than 1.0 the forecast "utility." Across all conditions, the Kalman and the ensemble forecasts have the highest utility. In all cases, the utility of the Kalman and the ensemble forecasts is near one. By contrast,

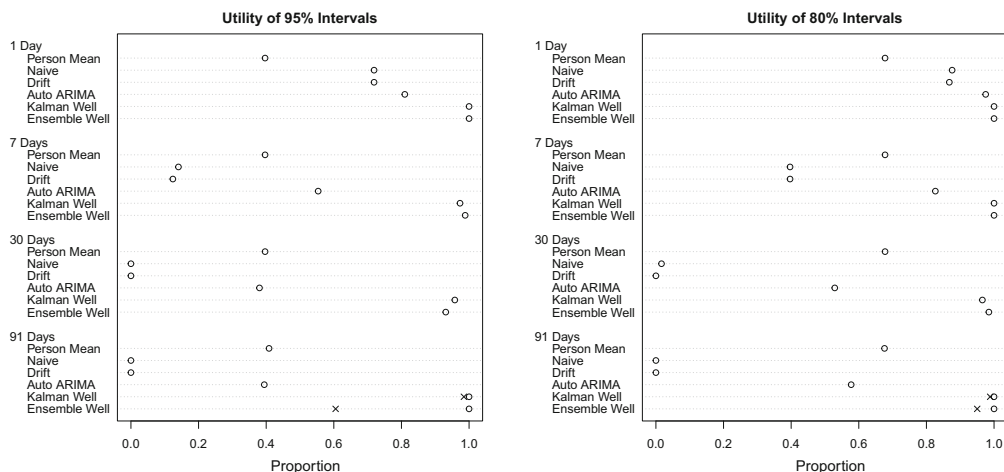


FIGURE 9.

Utility of forecasts as measured by the proportion of forecasts with interval width less than 1.0. 1-day, 7-day, and 30-day forecasts were trained on the first 80% of each person's data and all people. The 91-day forecasts were trained on the first 30% of each person's data and all people. The 'x's for the Kalman and ensemble forecasts at 91 days indicate results from the parameters of the 80% training model, but forecast from the 30% starting point. ARIMA=Autoregressive Integrated Moving Average; Well=Double-Well Potential Model.

in the 7-day, 30-day, and 91-day forecasts, the utility of the naive and the drift forecasts is the lowest of the methods compared. Similarly, in the 1-day forecasts, the naive and drift methods only have higher utility than the within-person mean. The lack of utility of the naive and drift forecasts tempers positive conclusions about their accuracy. The drift and naive forecasts have high accuracy, but show poor coverage performance (miscalibration, see Fig. 7), and Fig. 9 shows they have little utility beyond 1-day forecasts because their forecast intervals rapidly expand to include both use and non-use behaviors.

## 5. Discussion

In this paper, we have discussed two methods of forecasting intensive longitudinal data (ILD). Both methods begin with the estimation of parameters for a time series model of the data. We argue that the time series models considered are sufficiently general to encompass almost any desired model for ILD. After model estimation, the two methods differ in how they make forecasts from those models. The first forecasting method is based on the analytic properties of the time series model and the Kalman filter. The second method is based on the stochastic properties of the time series model and a Monte Carlo simulated ensemble. On analytic grounds, the Kalman prediction was expected to perform optimally for linear Gaussian models, whereas the ensemble prediction was expected to perform better for nonlinear non-Gaussian models. We graphically demonstrated differences between the Kalman and ensemble forecast methods in a series of linear and nonlinear models. In the application of a nonlinear model to substance use data, we found differences in forecast properties between the Kalman and ensemble methods, and compared their performance to simpler alternatives. Both the Kalman and ensemble methods outperformed simpler alternatives with regard to forecast interval width by having much narrower intervals. The ensemble forecast method had better accuracy than the Kalman forecast method in the nonlinear model of substance use for long-range forecasts, but depended heavily on the estimated dynamic noise variance used for perturbation.

The primary contributions of the present paper are fourfold. First, we reviewed two common methods of forecasting and advocated their application for ILD: an analytic Kalman forecasting method and a stochastic ensemble forecasting method. These forecasting methods have firm foundations in time series analysis and the representation of change processes as dynamical systems. Second, we implement these forecasting methods in freely available open-source software. Thus, the forecasting methods we propose are readily available to anyone using those implementations and all of their computational details are available for inspection via their GitHub pages (OpenMx: <https://github.com/OpenMx/OpenMx> and dynr: <https://github.com/mhunter1/dynr>). Third, we developed a nonlinear double-well potential model for drug and alcohol use. Fourth, we extensively evaluated the advocated forecasting methods and compared them to several simpler alternatives, finding the advocated methods were comparable to simpler methods in forecast accuracy and calibration but far superior in forecast precision.

Along with the aforementioned contributions, there remains much to understand and evaluate about ILD forecasting methods. Although we evaluated the Kalman and ensemble forecasting methods using a time-based hold-out sample (e.g., Bergmeir et al., 2014), we encourage further study on best practices for forecast evaluation of ILD, particularly pulling from the literature on machine learning (e.g., Hastie, Tibshirani, & Friedman, 2009; Haykin, 2008) and on time series analysis (e.g., Box & Jenkins, 1976; Harvey, 1989).

As a more methodologically focused paper, we do not fully analyze the data on substance use in the application. The model fitted in the application is cursory and—although adequate for the purposes of illustration—fails to capture some important features of the data. The binary nature of the data implies some misspecification in assuming the observations have a Gaussian distribution conditional on the latent state. Methods for non-Gaussian observations of time series (e.g., Durbin, 1997; Durbin & Koopman, 2000) exist (see Helske, 2017), but are generally not available for multisubject time series like those frequently found in the behavioral sciences. Relating to the multiple subjects, we made an assumption of homogeneity across people for the purposes of the modeling: all people in the sample were assumed to follow the same dynamics. The homogeneity assumption—which is necessary but not sufficient for ergodicity (Hannan, 1970)—may not hold, but could be relaxed by allowing random effects in the global and relative stability parameters across people by using methods described by Ou, Hunter, Lu, Stifter, and Chow (under review).

The quality of the data itself may validly be questioned. The timeline follow-back method has been previously shown to be reliable and valid for collecting daily drug and alcohol use information, corresponding well with daily reports of use gathered from experience sampling methods (Simons, Wills, Emery, & Marks, 2015). Indeed, the biological verification in these data indicated there was 89% overall agreement with self-report, and 95% agreement at 1-month follow-up. However, a recent study disentangling between- and within-person effects and comparing the two methods has found the within-person agreement to be somewhat lower (Lucas, Wallsworth, Anusic, & Donnellan, 2020). Moreover, participants—especially those for whom use carries a legal penalty—may be reluctant to report accurately on their own substance use behavior.

Even acknowledging the limitations of the present study, the results are consistent with patterns seen in similar samples (Bowen et al., 2014). Moreover, further development of ILD methods hold a great deal of promise. As ILD become more prevalent, researchers will naturally want to make predictions about future behavior from them. At best, forecasts from ILD may provide levels of predictive precision not previously thought possible, and may revolutionize many facets of our lives. At worst, a high-quality forecast for ILD should fail gracefully by accurately reflecting its uncertainty without giving incorrect precision or its corresponding false sense of security. By using well-established forecasting methods like Kalman prediction and stochastic ensemble prediction, the best case for forecasting ILD may soon be within reach.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## References

- Allen, P. M., Strathern, M., & Baldwin, J. (2008). Complexity: The integrating framework for models of urban and regional systems. In S. Albeverio, D. Andrey, P. Giordano, & A. Vancheri (Eds.), *The dynamics of complex urban systems* (pp. 21–41). Physica-Verlag HD. [https://doi.org/10.1007/978-3-7908-1937-3\\_2](https://doi.org/10.1007/978-3-7908-1937-3_2)
- Anderson, B. D. O., & Moore, J. B. (1979). *Optimal filtering*. Prentice-Hall.
- Anderson, J. L. (1996). A method for producing and evaluating probabilistic forecasts from ensemble model integrations. *Journal of Climate*, 9(7), 1518–1530. [10.1175/1520-0442\(1996\)009<1518:amfpae>2.0.co;2](https://doi.org/10.1175/1520-0442(1996)009<1518:amfpae>2.0.co;2)
- Anderson, J. L. (2001). An ensemble adjustment Kalman filter for data assimilation. *Monthly Weather Review*, 129(12), 2884–2903. [10.1175/1520-0493\(2001\)129<2884:aeakff>2.0.co;2](https://doi.org/10.1175/1520-0493(2001)129<2884:aeakff>2.0.co;2)
- Anderson, J. L., & Anderson, S. L. (1999). A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. *Monthly Weather Review*, 127(12), 2741–2758. [10.1175/1520-0493\(1999\)127<2741:amciot>2.0.co;2](https://doi.org/10.1175/1520-0493(1999)127<2741:amciot>2.0.co;2)
- Arnold, L. (1974). *Stochastic differential equations: Theory and applications*. John Wiley & Sons.
- Arnold, V. I. (1973). *Ordinary differential equations*. Cambridge, MA: MIT Press. Translated from the Russian by R. A. Silverman.
- Arnold, V. I. (1988). *Geometrical methods in the theory of ordinary differential equations* (2nd ed.). Springer.
- Arrowsmith, D. K., & Place, C. M. (1990). *An introduction to dynamical systems*. Cambridge University Press.
- Baltes, P. B., Reese, H. W., & Nesselroade, J. R. (1977). *Life-span developmental psychology: Introduction to research methods*. Belmont, CA: Wadsworth.
- Bar-Shalom, Y., Ti, X. R., & Kirubarajan, T. (2001). *Estimation with applications to tracking and navigation* (Vol. 45). New York: Wiley. <https://doi.org/10.1037/a0014170> (No. 1).
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1). <https://doi.org/10.18637/jss.v067.i01>
- Bergmeir, C., & Benítez, J. M. (2012). On the use of cross-validation for time series predictor evaluation. *Information Sciences*, 191, 192–213. <https://doi.org/10.1016/j.ins.2011.12.028>
- Bergmeir, C., Costantini, M., & Benítez, J. M. (2014). On the usefulness of cross-validation for directional forecast evaluation. *Computational Statistics & Data Analysis*, 76, 132–143. <https://doi.org/10.1016/j.csda.2014.02.001>
- Bergmeir, C., Hyndman, R. J., & Koo, B. (2018). A note on the validity of cross-validation for evaluating autoregressive time series prediction. *Computational Statistics & Data Analysis*, 120, 70–83. <https://doi.org/10.1016/j.csda.2017.11.003>
- Boker, S. M., & Laurenceau, J. P. (2006). Dynamical Systems Modeling: An Application to the Regulation of Intimacy and Disclosure in Marriage. In T. A. Walls & J. L. Schafer (Eds.), *Models for Intensive Longitudinal Data* (pp. 195–218). Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780195173444.003.0009>
- Bornoalova, M. A., Gratz, K. L., Daughters, S. B., Hunt, E. D., & Lejuez, C. (2012). Initial RCT of a distress tolerance treatment for individuals with substance use disorders. *Drug and Alcohol Dependence*, 122(1–2), 70–76. <https://doi.org/10.1016/j.drugalcdep.2011.09.012>
- Bowen, S., Witkiewitz, K., Clifasefi, S. L., Grow, J., Chawla, N., Hsu, S. H., & Larimer, M. E. (2014). Relative Efficacy of Mindfulness-Based Relapse Prevention, Standard Relapse Prevention, and Treatment as Usual for Substance Use Disorders: A Randomized Clinical Trial. *JAMA Psychiatry*, 71(5), 547–556. <https://doi.org/10.1001/jamapsychiatry.2013.4546>
- Box, G. E. P., & Jenkins, G. M. (1976). *Time series analysis: Forecasting and control* (Revised). San Francisco: Holden-Day.
- Brockwell, P. J. (1995). A note on the embedding of discrete-time ARMA processes. *Journal of Time Series Analysis*, 16(5), 451–460. <https://doi.org/10.1111/j.1467-9892.1995.tb00246.x>
- Brookner, E. (1998). *Tracking and Kalman filtering made easy*. New York: Wiley.
- Brooks, H. E. (2004). Tornado warning performance in the past and future: a perspective from signal detection theory. *Bulletin of the American Meteorological Society*, 85(6), 837–844. <https://doi.org/10.1175/bams-85-6-837>
- Chan, K. S., & Tong, H. (1987). A note on embedding a discrete parameter ARMA model in a continuous parameter ARMA model. *Journal of Time Series Analysis*, 8(3), 277–281. <https://doi.org/10.1111/j.1467-9892.1987.tb00439.x>
- Coiffier, J. (2011). *Fundamentals of numerical weather prediction*. Cambridge New York: Cambridge University Press.
- Culbertson, D. S., & Sinclair, T. M. (2014). The failure of forecasts in the great recession. *Challenge*, 57(6), 34–45. <https://doi.org/10.2753/0577-5132570603>
- Cvitanović, P., Artuso, R., Mainieri, R., Tanner, G., & Vattay, G. (2017). *Chaos: Classical and quantum*. Copenhagen: Niels Bohr Institute. <http://ChaosBook.org>
- de Jong, P. (1988). The likelihood for a state space model. *Biometrika*, 75(1), 165–169. <https://doi.org/10.2307/2336450>
- Durbin, J. (1997). Monte carlo maximum likelihood estimation for non-Gaussian state space models. *Biometrika*, 84(3), 669–684. <https://doi.org/10.1093/biomet/84.3.669>



- Durbin, J., & Koopman, S. J. (2000). Time series analysis of non-Gaussian observations based on state space models from both classical and bayesian perspectives. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 62(1), 3–56. <https://doi.org/10.1111/1467-9868.00218>
- Durbin, J., & Koopman, S. J. (2001). *Time series analysis by state space methods*. Oxford University Press.
- Edwards, C. H., & Penney, D. E. (2004). *Differential equations and boundary value problems: Computing and modeling* (3rd ed.). Upper Saddle River, NJ: Pearson Education.
- Epstein, E. S. (1969). Stochastic dynamic prediction. *Tellus*, 21(6), 739–759. <https://doi.org/10.1111/j.2153-3490.1969.tb00483.x>
- Fredrickson, B. L., & Losada, M. F. (2005). Positive affect and the complex dynamics of human flourishing. *American Psychologist*, 60(7), 678–686. <https://doi.org/10.1037/0003-066x.60.7.678>
- Friston, K. J., Harrison, L., & Penny, W. (2003). Dynamic causal modelling. *NeuroImage*, 19(4), 1273–1302. [https://doi.org/10.1016/s1053-8119\(03\)00202-7](https://doi.org/10.1016/s1053-8119(03)00202-7)
- Gregson, R. A. M. (1983). *Time series in psychology*. Hillsdale, NJ: Lawrence Erlbaum.
- Hall, P., & Heyde, C. C. (1980). *Martingale limit theory and its application*. New York: Academic Press.
- Hamerle, A., Nagl, W., & Singer, H. (1991). Problems with the estimation of stochastic differential equations using structural equations models. *Journal of Mathematical Sociology*, 16(3), 201–220.
- Hamilton, J. D. (1994). State-space models. In R. F. Engle & D. L. McFadden (Eds.), *Handbook of econometrics* (Vol. 4, p. 3039–3080). Elsevier. <http://www-stat.wharton.upenn.edu/~stine/stat910/>
- Hannan, E. J. (1970). *Multiple time series*. New York: Wiley.
- Harvey, A. C. (1989). *Forecasting, structural time series models and the Kalman filter*. Cambridge: Cambridge University Press.
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning*. Springer.
- Havlicek, M., Friston, K. J., Jan, J., Brazdil, M., & Calhoun, V. D. (2011). Dynamic modeling of neuronal responses in fMRI using cubature Kalman filtering. *NeuroImage*, 56(4), 2109–2128. <https://doi.org/10.1016/j.neuroimage.2011.03.005>
- Hayes, A. M., Laurenceau, J. P., Feldman, G., Strauss, J. L., & Cardaciotto, L. (2007). Change is not always linear: The study of nonlinear and discontinuous patterns of change in psychotherapy. *Clinical Psychology Review*, 27(6), 715–723. <https://doi.org/10.1016/j.cpr.2007.01.008>
- Haykin, S. O. (2008). *Neural networks and learning machines* (3rd ed.). Upper Saddle River, NJ: Pearson Education.
- He, S. W., & Wang, J. G. (1989). On embedding a discrete-parameter ARMA model in a continuous-parameter ARMA model. *Journal of Time Series Analysis*, 10(4), 315–323. <https://doi.org/10.1111/j.1467-9892.1989.tb00031.x>
- Helske, J. (2017). KFAS: Exponential family state space models in r. *Journal of Statistical Software*, 78(10). <https://doi.org/10.18637/jss.v078.i10>
- Hirsch, M. W., & Smale, S. (1974). *Differential equations, dynamical systems and bifurcations of vector fields*. Academic Press.
- Hirsch, M. W., Smale, S., & Devaney, R. (2003). *Differential equations, dynamical systems, and an introduction to chaos* (2nd ed.). Burlington, MA: Academic Press.
- Hufford, M. R., Witkiewitz, K., Shields, A. L., Kodya, S., & Caruso, J. C. (2003). Relapse as a nonlinear dynamic system: Application to patients with alcohol use disorders. *Journal of Abnormal Psychology*, 112(2), 219–227. <https://doi.org/10.1037/0021-843X.112.2.219>
- Hunter, M. D. (2018). State space modeling in an open source, modular, structural equation modeling environment. *Structural Equation Modeling: A Multidisciplinary Journal*, 25(2), 307–324. <https://doi.org/10.1080/10705511.2017.1369354>
- Huzii, M. (2007). Embedding a Gaussian discrete-time autoregressive moving average process in a gaussian continuous-time autoregressive moving average process. *Journal of Time Series Analysis*, 28(4), 498–520. <https://doi.org/10.1111/j.1467-9892.2006.00520.x>
- Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: Principles and practice*. OTexts.
- Hyndman, R. J., & Khandakar, Y. (2008). Automatic time series forecasting: The forecast package for R. *Journal of Statistical Software*, 27(3). <https://doi.org/10.18637/jss.v027.i03>
- Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Basic Engineering*, 82, 35–45. <https://doi.org/10.1115/1.3662552>
- Kalman, R. E. (1963). Mathematical description of linear dynamical systems. *Journal of the Society for Industrial and Applied Mathematics. Series A (Control)*, 1, 152–192.
- Kalman, R. E., & Bucy, R. S. (1961). New results in linear filtering and prediction theory. *Transactions of the ASME, Series D, Journal of Basic Engineering*, 83, 95–108. <https://doi.org/10.1115/1.3658902>
- Katzfuss, M., Stroud, J. R., & Wikle, C. K. (2016). Understanding the ensemble Kalman filter. *The American Statistician*, 70(4), 350–357. <https://doi.org/10.1080/00031305.2016.1141709>
- Kulikov, G. Y., & Kulikova, M. V. (2014). Accurate numerical implementation of the continuous-discrete extended Kalman filter. *IEEE Transactions on Automatic Control*, 59(1), 273–279. <https://doi.org/10.1109/tac.2013.2272136>
- Lazo, J. K., Morss, R. E., & Demuth, J. L. (2009). 300 billion served. *Bulletin of the American Meteorological Society*, 90(6), 785–798. <https://doi.org/10.1175/2008bams2604.1>
- Leith, C. E. (1974). Theoretical skill of Monte Carlo forecasts. *Monthly Weather Review*, 102(6), 409–418. [10.1175/1520-0493\(1974\)102<0409:tsomcf>2.0.co;2](https://doi.org/10.1175/1520-0493(1974)102<0409:tsomcf>2.0.co;2)
- Li, Y., Ji, L., Oravec, Z., Brick, T. R., Hunter, M. D., & Chow, S. M. (2019). dyn.mi: An R program for multiple imputation in dynamic modeling. *International Journal of Computer, Electrical, Automation, Control and Information Engineering*, 13(5), 302–311. <http://waset.org/Publicationsp=149>

- Lipari, R. N., & Van Horn, S. L. (2017). Trends in substance use disorders among adults aged 18 or older. *The CBHSQ Report*, 1–10.
- Lucas, R. E., Wallsworth, C., Anusic, I., & Donnellan, M. B. (2020). A direct comparison of the day reconstruction method (DRM) and the experience sampling method (ESM). *Journal of Personality and Social Psychology*. <https://doi.org/10.1037/pspp0000289>
- Maisto, S. A., Hallgren, K. A., Roos, C. R., & Witkiewitz, K. (2018). Course of remission from and relapse to heavy drinking following outpatient treatment of alcohol use disorder. *Drug and Alcohol Dependence*, 187, 319–326. <https://doi.org/10.1016/j.drugalcdep.2018.03.011>
- Mandelbrot, B. B. (1971). When can price be arbitrated efficiently? A limit to the validity of the random walk and martingale models. *The Review of Economics and Statistics*, 53(3), 225. <https://doi.org/10.2307/1937966>
- McArdle, J. J., & Epstein, D. (1987). Latent growth curves within developmental structural equation models. *Child Development*, 58, 110–133.
- McLellan, A. T., Lewis, D. C., O'Brien, C. P., & Kleber, H. D. (2000). Drug dependence, a chronic medical illness: implications for treatment, insurance, and outcomes evaluation. *Journal of the American Medical Association*, 284(13), 1689–1695.
- Mischel, W. (1968). *Personality and assessment*. Lawrence Erlbaum.
- Molenaar, P. C. M. (1997). Time series analysis and its relationship with longitudinal analysis. *International Journal of Sports Medicine*, 18(S 3), S232–S237. <https://doi.org/10.1055/s-2007-972720>
- Molenaar, P. C. M. (2004). A manifesto on psychology as idiographic science: Bringing the person back into scientific psychology, this time forever. *Measurement*, 2, 201–218. [https://doi.org/10.1207/s15366359mea0204\\_1](https://doi.org/10.1207/s15366359mea0204_1)
- Neale, M. C., Hunter, M. D., Pritikin, J. N., Zahery, M., Brick, T. R., Kirkpatrick, R. M., & Boker, S. M. (2016). OpenMx 2.0: Extended structural equation and statistical modeling. *Psychometrika*, 80(2), 535–549. <https://doi.org/10.1007/s11336-014-9435-8>
- Ou, L., Hunter, M. D., & Chow, S. (2019). What's for dynr: A package for linear and nonlinear dynamic modeling in R. *The R Journal*, 11(1). <https://doi.org/10.32614/RJ-2019-012>
- Ou, L., Hunter, M. D., Lu, Z., Stifter, C. A., & Chow, S. (under review). Estimation of nonlinear mixed-effects continuous-time models using the continuous-discrete extended Kalman filter.
- Oud, J. H. L., & Jansen, R. A. R. G. (2000). Continuous time state space modeling of panel data by means of SEM. *Psychometrika*, 65(2), 199–215.
- Priestley, M., & Subba Rao, T. (1975). The estimation of factor scores and Kalman filtering for discrete parameter stationary processes. *International Journal of Control*, 21(6), 971–975. <https://doi.org/10.1080/00207177508922050>
- R Development Core Team. (2020). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from <http://www.R-project.org> (ISBN 3-900051-07-0)
- Rao, C. R. (2001). *Linear statistical inference* (2nd ed.). John Wiley & Sons.
- Ripberger, J. T., Silva, C. L., Jenkins-Smith, H. C., Carlson, D. E., James, M., & Herron, K. G. (2014). False alarms and missed events: The impact and origins of perceived inaccuracy in tornado warning systems. *Risk Analysis*, 35(1), 44–56. <https://doi.org/10.1111/risa.12262>
- Robinson, S. M., Sobell, L. C., Sobell, M. B., & Leo, G. I. (2014). Reliability of the timeline followback for cocaine, cannabis, and cigarette use. *Psychology of Addictive Behaviors*, 28(1), 154–162. <https://doi.org/10.1037/a0030992>
- Schweppe, F. C. (1965). Evaluation of likelihood functions for Gaussian signals. *IEEE Transactions on Information Theory*, IT-11, 61–70.
- Simmons, K. M., & Sutter, D. (2009). False alarms, tornado warnings, and tornado casualties. *Weather, Climate, and Society*, 1(1), 38–53. <https://doi.org/10.1175/2009wcas1005.1>
- Simons, J. S., Wills, T. A., Emery, N. N., & Marks, R. M. (2015). Quantifying alcohol consumption: Self-report, transdermal assessment, and prediction of dependence symptoms. *Addictive Behaviors*, 50, 205–212. <https://doi.org/10.1016/j.addbeh.2015.06.042>
- Smale, S. (1961). On gradient dynamical systems. *The Annals of Mathematics*, 74(1), 199. <https://doi.org/10.2307/1970311>
- Sobell, L. C., Agrawal, S., Annis, H., Ayala-Velazquez, H., Echeverria, L., Leo, G. I., & Ziolkowski, M. (2001). Cross-cultural evaluation of two drinking assessment instruments: Alcohol timeline followback and inventory of drinking situations. *Substance Use & Misuse*, 36(3), 313–331. <https://doi.org/10.1081/ja-100102628>
- Squire, P. (1988). Why the 1936 Literary Digest poll failed. *Public Opinion Quarterly*, 52(1), 125. <https://doi.org/10.1086/269085>
- Stephan, K. E., Kasper, L., Harrison, L. M., Daunizeau, J., den Ouden, H. E. M., Breakspear, M., & Friston, K. J. (2008). Nonlinear dynamic causal models for fMRI. *NeuroImage*, 42(2), 649–662. <https://doi.org/10.1016/j.neuroimage.2008.04.262>
- Substance Abuse and Mental Health Services Administration. (2018). Key substance use and mental health indicators in the United States: Results from the 2017 National Survey on Drug Use and Health (HHS Publication No. SMA 18-5068, NSDUH Series H-53). Rockville, MD: Center for Behavioral Health Statistics and Quality, Substance Abuse and Mental Health Services Administration. [2020-04-16]<https://www.samhsa.gov/data/sites/default/files/cbhsq-reports/NSDUHFFR2017/NSDUHFFR2017.pdf>

- Van Loan, C. (1978). Computing integrals involving the matrix exponential. *IEEE Transactions on Automatic Control*, 23(3), 395–404. <https://doi.org/10.1109/tac.1978.1101743>
- Warner, T. T. (2014). *Numerical weather and climate prediction*. Cambridge University Press.
- West, M., & Harrison, P. (1997). *Bayesian forecasting and dynamic models*. New York: Springer.
- Witkiewitz, K., & Marlatt, G. A. (2004). Relapse prevention for alcohol and drug problems: That was Zen, this is Tao. *American Psychologist*, 59(4), 224–235. <https://doi.org/10.1037/0003-066X.59.4.224>

*Manuscript Received: 30 JUN 2020*  
*Final Version Received: 22 OCT 2021*  
*Published Online Date: 22 JAN 2022*