


Modeling Non-Linear Psychological Processes: Reviewing and Evaluating (Non-) parametric Approaches and Their Applicability to Intensive Longitudinal Data

Jan I. Failenschmid, Leonie V.D.E Vogelsmeier, Joris Mulder, and Joran Jongerling
Tilburg University

Author Note

Jan I. Failenschmid  <https://orcid.org/0009-0007-5106-7263>

Correspondence concerning this article should be addressed to Jan I. Failenschmid,
Tilburg School of Social and Behavioral Sciences: Department of Methodology and Statistics,
Tilburg University, Warandelaan 2, 5037AB Tilburg, Netherlands. E-mail:
J.I.Failenschmid@tilburguniversity.edu

Abstract

Here could your abstract be!

Keywords: Here could your keywords be!

Modeling Non-Linear Psychological Processes: Reviewing and Evaluating (Non-) parametric Approaches and Their Applicability to Intensive Longitudinal Data

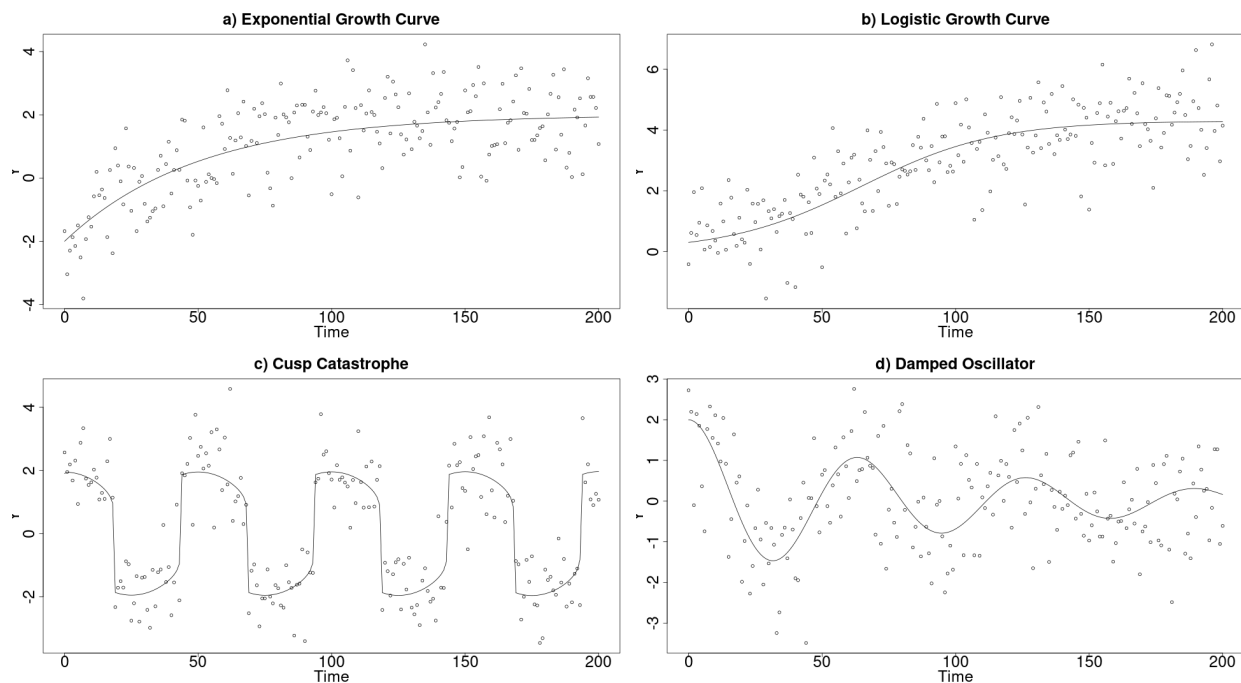
Psychological constructs are increasingly viewed as components of complex dynamic systems (Nesselroade & Ram, 2004; Wang et al., 2012). This perspective emphasizes that these constructs fluctuate over time and within individuals. To study these variations and the underlying processes, researchers are increasingly collecting intensive longitudinal data (ILD) using ecological momentary assessment (EMA), experience sampling, or similar methods (Fritz et al., 2023). In these studies several individuals are assessed at a high frequency (up to multiple times per day) using brief questionnaires or passive measurement devices. This rich data allows researchers to examine variations in latent psychological variables within an ecologically valid context and to explain them through (between-person differences) in within-person processes.

Many psychological phenomena and processes have been discovered during the recent years. Clear examples of this are the learning and growth curves observed in intellectual and cognitive development (Kunnen, 2012; McArdle et al., 2002). In these cases, an individual's latent ability increases over time from a (person specific) starting point towards an (person specific) asymptote, which reflects their maximum ability. Additional examples of asymptotic growth over the shorter time spans that are typically studied with ILD include motor skill development (Newell et al., 2001) and second language acquisition (De Bot et al., 2007). Figure 1 illustrates these general growth curves through an exponential growth function (a) and a logistic growth function (b), both of which are common model choices for these kinds of processes.

Another commonly observed non-linear phenomenon involves the construct under study switching between multiple distinct states, which often correspond to different means. This occurs, for example, during the sudden perception of cognitive flow, where individuals abruptly switch from a 'normal' state to a flow state and back (Ceja & Navarro, 2012). Another example is alcohol use relapse, where patients suddenly switch from an abstinent state to a relapsed state (Witkiewitz & Marlatt, 2007). This sudden switching behavior is exemplified in Figure 1 (c) through a cusp catastrophe model. This model, drawn from catastrophe theory, naturally leads to

Figure 1

Examples of non-linear processes demonstrated to occur in psychological time series.



Note. This figure shows four demonstrated psychological non-linear processes. Panels (a) and (b) show exponential and logistic growth curves, respectively, which have been shown to describe intellectual and cognitive development (Kunnen, 2012; McArdle et al., 2002), motor skill learning (Newell et al., 2001), and second language acquisition (De Bot et al., 2007). Panel (c) shows a cusp model from catastrophe theory that exhibits apparent jumps between two stable states. As such, it has been used to describe the perception of mental flow (Ceja & Navarro, 2012) or alcohol use relapse (Witkiewitz & Marlatt, 2007). Lastly, panel (d) shows a damped oscillator, which describes the return of a perturbed system to baseline, which has been proposed as a model for affect regulation (Chow et al., 2005).

mean level switches when varying one of its parameters (Chow et al., 2015; van der Maas et al., 2003).

As a final example, one may consider (self-) regulatory systems, which maintain a desired state by counteracting external perturbations. These systems regulate adaptively, meaning that the regulation strength depends on the distance between the current and desired states. The common autoregressive model describes such a system in which the regulation strength depends linearly

on this distance. However, this relationship may also be non-linear, such that the regulatory force increases disproportionately with larger mismatches. Such (self-) regulatory models have been proposed to describe for example emotion regulation (Chow et al., 2005). Figure 1 (d) illustrates a (self-) regulatory system exemplified as a damped oscillator.

Although initial evidence for non-linearity in psychological research exists, theories about the nature and form of non-linear psychological processes remain scarce (Tan et al., 2011). This gap is largely due to the lack of advanced statistical methods that are flexible enough to study the complex behavior of these processes adequately. As a result, researchers are often ill-equipped to infer the functional characteristics of non-linear processes from ILD, which hinders the development and evaluation of guiding theories for subsequent studies. Due to this lack of adequate available statistical methods, non-linear trends are most often addressed in psychology through polynomial regression or regression splines.

Polynomial regression (Jebb et al., 2015) uses higher-order terms (e.g., squared or cubed time) as predictors in a standard multiple linear regression model. While effective for relatively simple non-linear relationships, particularly those that can be represented as polynomials, this method has significant limitations and likely leads to invalid results when applied to more complex latent processes, such as mean switching or (self-) regulatory systems (e.g., Figure 1 c & d). In these cases, polynomial approximations require many higher-order terms to capture the process's high variability, which raises the problem of over- or underfitting the data, causes model instability, and leads to nonsensical inferences (e.g., interpolating scores outside the scale range; Boyd and Xu (2009) and Harrell (2001)).

An alternative approach is spline regression, which constructs a complex non-linear trend by joining multiple simple piecewise functions at specific points, called knots (e.g., combining multiple cubic functions into a growth curve with plateaus; Tsay and Chen (2019)). However, spline regression requires a careful, manual selection of the optimal piecewise functions and knot locations. This can be problematic in practice because, as mentioned, precise guiding theories about the functional form of most psychological processes are lacking (Tan et al., 2011). This

absence of clear guidance can easily lead to misspecified models and invalid results.

These limitations in current practices underscore the need for alternative statistical methods to study non-linear processes. Various such advanced statistical methods, such as kernel regression, Gaussian processes, smoothing splines, and latent change score models, are available outside of psychology. However, these methods have rarely been applied in psychology because they have not been reviewed for an applied audience, nor have their assumptions and inference possibilities been evaluated in the context of ILD. As a result, psychological researchers struggle to select the most suitable method for a specific context. This challenge is further complicated by the fact that the ideal statistical method may depend on the characteristics of the underlying non-linear process, which are generally unknown. Especially, since the smooth processes for which many of these methods were originally developed are unlikely to occur in psychological research.

To address this important gap, this article reviews four advanced non-linear analysis methods and evaluates their applicability to typical ILD. Specifically, we compare how well each method can recover different latent processes under common ILD conditions in a simulation study. We also demonstrate the conclusions that can be drawn from each method by applying them to an existing dataset. The methods reviewed in this article range from data-driven non-parametric techniques to a flexible parametric modeling framework, which were selected to accommodate varying degrees of prior knowledge, as precise theories about the nature of non-linear psychological processes are often missing (Tan et al., 2011). Further, to introduce these methods accessibly and apply them under conditions where software implementations are available, this article focuses primarily on the univariate single-subject design.

Method

Data structure

Generally, any psychological construct under study follows a (possibly non-linear) function over time, as represented by the lines in Figure 1. However, since these psychological

processes are typically unobservable or latent, they are measured through observable indicators, such as questionnaire items or passive measurements. The observations on these indicators (Figure 1, dots) differ from the true values of the latent process due to measurement error, which may come from an imperfect measurement instrument. For this introduction, we assume that all time-point-specific measurement errors are independent and normally distributed. The model for the observations of a single indicator can then be written as:

$$Y_t = f(t) + \varepsilon_t; \quad \varepsilon_t \sim N(0, \sigma_\varepsilon^2) \quad (1)$$

where $f(t)$ represents the potentially non-linear latent process, and ε_t represents the time-point-specific measurement error.

In this context, researchers typically aim to infer the underlying process and draw conclusions about its functional form. The following sections will introduce four methods to achieve these goals, two non-parametric techniques, one semi-parametric approach, and one parametric modeling framework, using the (self-) regulatory process depicted in Figure 1 (d) as a running example.

Local polynomial regression

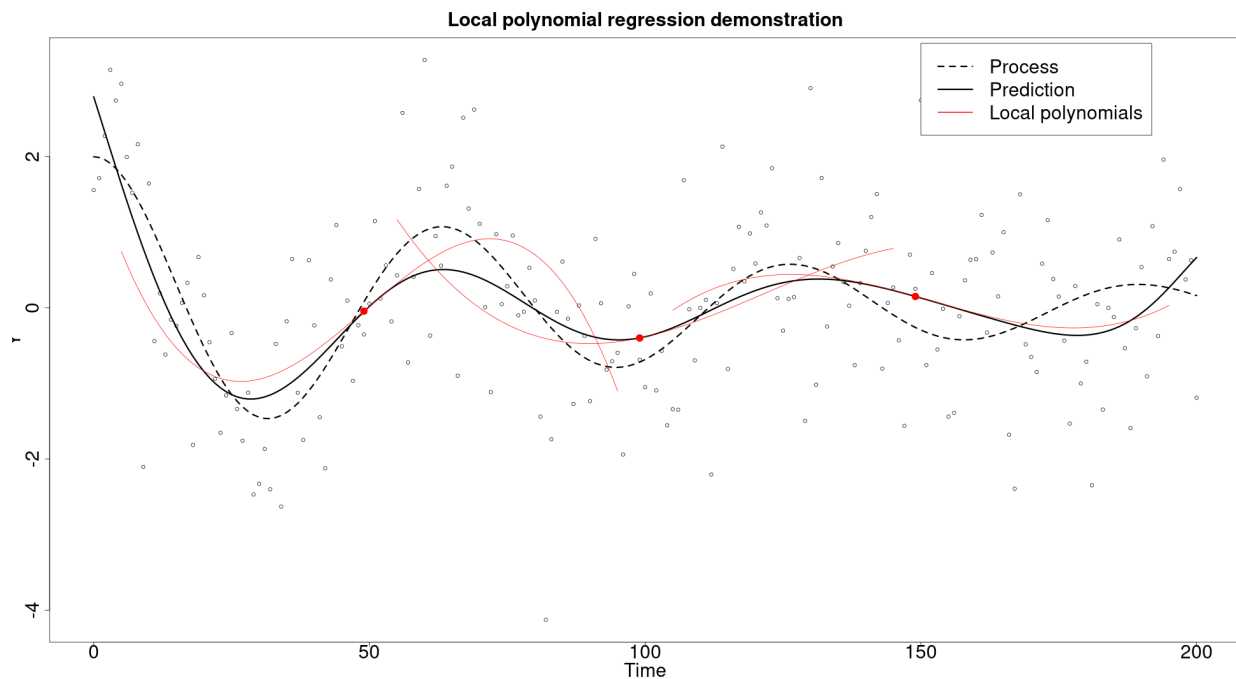
The first technique is called local polynomial regression (LPR). Similarly to regular polynomial regression, LPR approximates the process using polynomial basis functions (e.g., squared or cubed time). However, instead of using one large polynomial to approximate the entire process, LPR estimates smaller, local polynomials at each point in time. These local polynomials are then combined into a single non-linear function over the entire set of observations (Fan & Gijbels, 2018; Fan & Gijbels, 1995; Ruppert & Wand, 1994).

To determine the value of the LPR at a specific time point, the data is centered around that point (by shifting the data along the time axis so that the chosen time point is at zero), and a low-order polynomial is fitted around it. Additionally, to account for the idea that data points closer in time are more related, a weighting function is applied during the polynomial estimation.

This function assigns weights to each data point based on its distance from the point of interest. The value of the LPR at the chosen time point is then given, by the intercept of the locally weighted polynomial. To find the LPR value at a different time point, the same procedure is repeated, centering the data around the new point of interest. Since it is theoretically possible to repeat this process at infinitely many time points, LPR is a non-parametric technique. Figure 2 shows the estimated LPR for the example process depicted in Figure 1 (d). In this figure, three truncated examples of local cubic regressions are shown in red, which contribute to the overarching LPR for this process .

Figure 2

Demonstration of how local polynomial regression (solid black) estimates the underlying process (dotted black). Here, three local cubic functions (red) are shown as examples at the time points 50, 100, and 150, which provide the values of the local polynomial regression at these time points.



When fitting an LPR, several decisions must be made regarding the degree of the local polynomials and the optimal weighting of the data. Typically, the degree of these local

polynomials is kept low and odd. This choice reflects a bias-variance tradeoff, where higher-order polynomials reduce bias but increase variance, when transitioning from an odd to an even power (Ruppert & Wand, 1994). The data weighting in an LPR is achieved through a kernel function, which is usually centered and symmetric to assign weights based on the distance from the origin. Common kernel choices include the Gaussian and symmetric Beta distributions. However, most commonly used kernel functions yield similar inferences and a specific kernel can be selected to minimize a chosen criterion function. The kernel is further defined by a bandwidth parameter, which determines its width and effectively controls the influence of more distant data points. The bandwidth parameter, in practice, represents the wiggleness of the estimated process. Several methods are available to find the optimal bandwidth by optimizing a data-dependent criterion function, such as cross-validation or the mean integrated squared error (Debruyne et al., 2008; Köhler et al., 2014).

Due to its non-parametric nature, LPR makes minimal assumptions about the data. However, it does require that the underlying process is smooth or differentiable, which is a necessary condition for polynomial approximation. Another key assumption is that the process has constant wiggleness, represented by a single bandwidth parameter. However, this assumption may be relaxed by using a time-varying bandwidth.

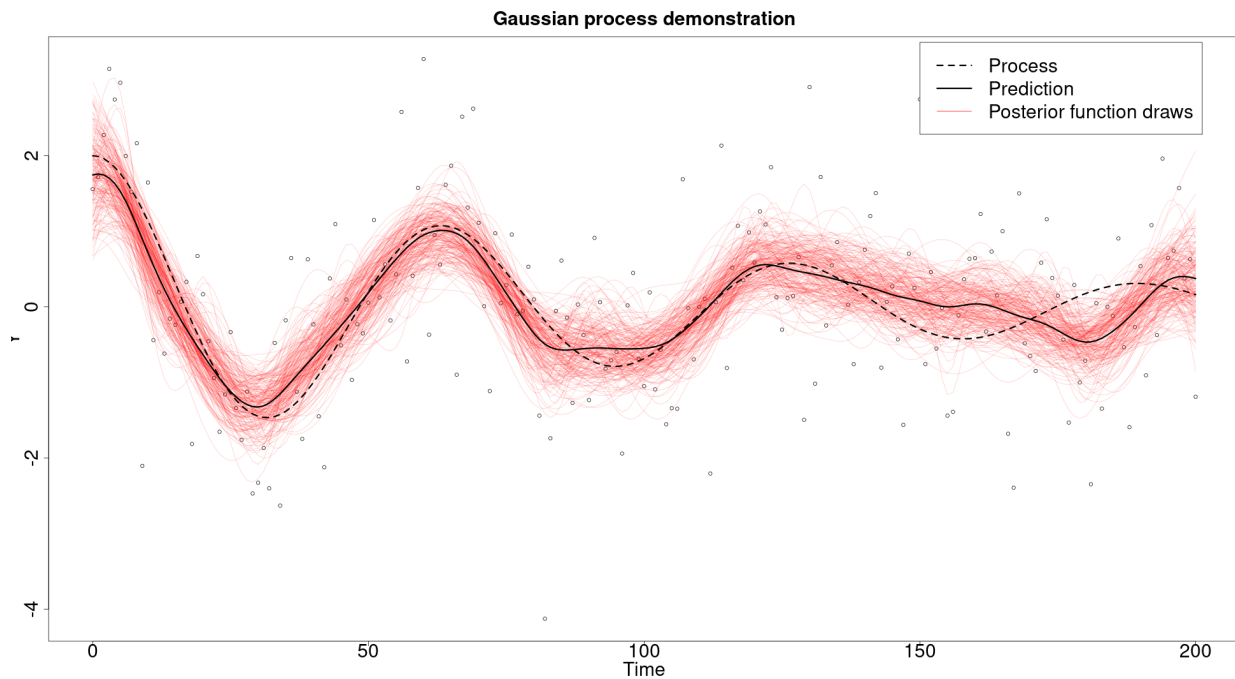
Gaussian process regression

The second non-parametric technique is Gaussian process (GP) regression, a Bayesian approach that directly defines a probability distribution over an entire family of non-linear functions flexible enough to capture many complex processes effectively (Betancourt, 2020; Rasmussen & Williams, 2006; Roberts et al., 2013). Unlike regular probability distributions (e.g., normal distributions) that specify the plausibility of single values, Gaussian processes determine the plausibility of entire (non-linear) functions. In a Bayesian framework, one can use a GP to define a prior distribution for the latent process. This prior is then combined with an appropriate likelihood for the observed data to obtain a posterior distribution for the latent process given the

observed data. This posterior distribution represents an updated belief about which functions describe the latent process well, allowing one to draw inferences about the process itself. Figure 3 illustrates such a posterior distribution for the running example process. The red lines represent a sample of non-linear functions drawn from the posterior distribution, with the pointwise average of these functions providing a mean estimate for the underlying process.

Figure 3

Demonstration of how Gaussian process regression estimates the underlying process (dotted black). Here, a sample of functions drawn from the posterior Gaussian process probability distribution is shown (red). The predicted value for the underlying process is then obtained by averaging the drawn functions.



The GP prior is parameterized by a mean function and a covariance function, which are continuous extensions of the mean vector and covariance matrix of a multivariate normal distribution. These functions can be selected based on domain knowledge or through data-driven model selection (Abdessalem et al., 2017; Richardson et al., 2017). In practice, the mean function

is often set to zero when no specific prior knowledge is available, which does not constrain the posterior mean to zero but instead indicates a lack of prior information about its deviations from zero. The covariance function is typically based on a kernel function, which assigns covariances between time points only depending on their distance (e.g., quadratic exponential, Matern class). Finally, appropriate hyperpriors for the parameters of the mean and covariance functions are used to generate the corresponding posterior distribution. These hyperpriors reflect prior beliefs about the hyperparameters and can be used to constrain them to sensible values.

The functional behavior of a GP is entirely determined by the posterior mean and covariance function. Thus, to accurately capture a process, it is crucial that the functional family generated by a chosen mean and covariance function is similar to the actual process. Common choices for the covariance function result in smooth and covariance-stationary GPs with constant wiggleness. Unlike LPR, GP regression provides more interpretable information about the process through the posterior distributions of the hyperparameters. One typical hyperparameter, the characteristic length scale of the covariance function, is analogous to the bandwidth parameter in LPR, as it also describes the wiggleness of the estimated process. However, GP regression can include additional hyperparameters, allowing for more specific theories to be tested through model comparison.

Generalized additive models

Generalized additive models (GAM) are a semi-parametric modeling framework that builds on smooth terms, which are non-linear functions that are inferred from the data through smoothing splines (Hastie & Tibshirani, 1999; Wood, 2006, 2020). Smoothing splines are an extension of the regular spline regression introduced above, which circumvent the knot placement issue by placing a knot at each observation (Tsay & Chen, 2019). This is equivalent to a basis function regression with as many basis functions as there are observations. However, using this many knots usually leads to overfitting. To prevent this, smoothing splines use a penalty term, similar to the penalty used in a lasso or ridge regression, which controls the smooth term's

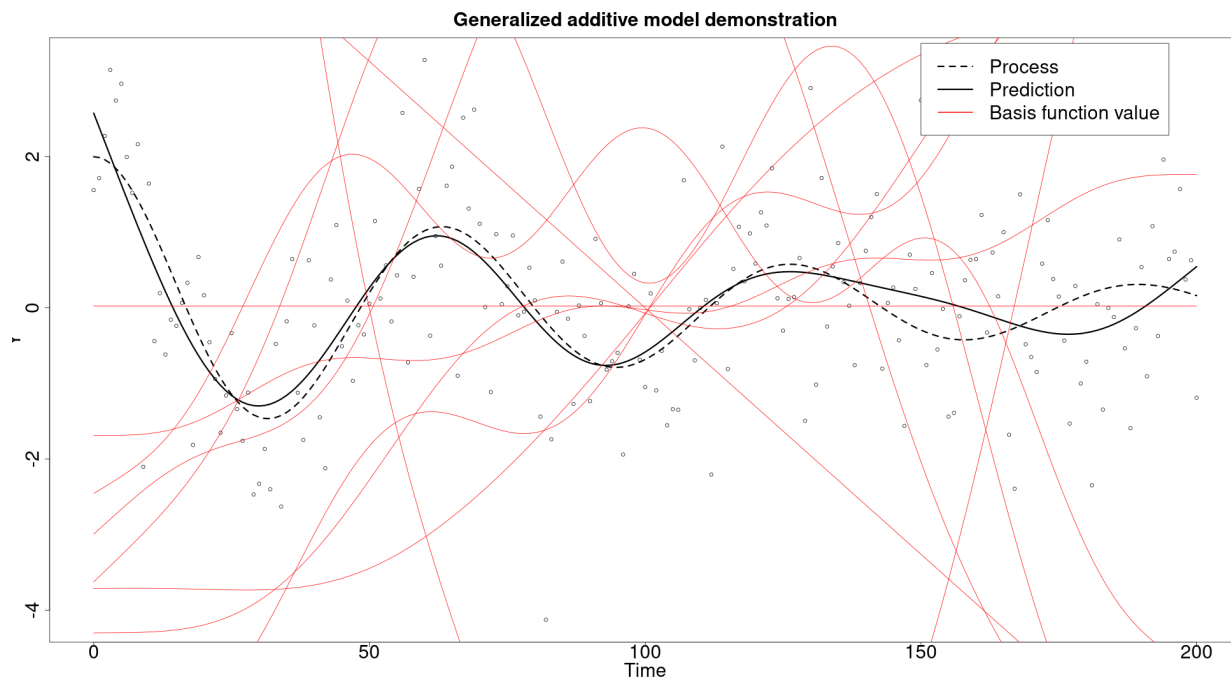
wigglyness (Gu, 2013; Wahba, 1980). This penalty makes it possible to balance the flexibility and fitting of the smooth term, ensuring that the model captures the process accurately without excessive complexity. In practice, the optimal weight of the penalty is determined by minimizing a criterion function, such as the generalized cross-validation criterion (Golub & von Matt, 1997; Wood, 2006). While researchers must still choose a set of basis functions, cubic (Tsay & Chen, 2019) and thin plate spline bases (Wood, 2003) are optimal for many applications.

These smooth terms can then be combined in an additive regression model, where each smooth term essentially functions as a predictor within a regular regression analysis. In this model, smooth terms may be multiplied by covariates and summed into a single overall non-linear function to explain the data. This approach makes it possible to formulate models such as a time-varying autoregressive model, where the intercept and autoregressive parameters are smooth terms of time (Bringmann et al., 2015; Bringmann et al., 2017). By integrating non-parametric smooth terms into a broader parametric model, GAMs become semi-parametric models that are well-suited for testing specific hypotheses while keeping the flexibility needed to accurately capture the latent process. Figure 4 illustrates a simple GAM with a single smooth term for time, fitted to the example process. Some of the scaled thin-plate smoothing spline basis functions that make up the smooth term are shown in red.

Compared to the previous methods, GAMs offer a more accessible modeling framework, enabling the specific modeling and testing of partial theories. For instance, a GAM can model a linear trend with an added smooth term around it to capture non-linear deviations. This makes it possible to gain insight into both the linear trend and the necessity of the smooth term, which may be examined through model comparison. Additionally, GAMs also provide an estimate of the wigglyness of the process through the weight that is assigned to the smoothing penalty. In contrast, to LPR and GP this penalty weight does not assume constant wigglyness. Lastly, the basis function coefficients within each smooth term can be interpreted, but the specific interpretations depend on the spline basis used.

Figure 4

Demonstration of how generalized additive models (solid black) estimate the underlying process (dotted black). Here, the predicted values for the process at any point in time correspond to the weighted average of the basis functions (red).



Parametric models

Finally, several methods to parametrically model non-linear processes will be introduced. The most direct approach for this involves defining a non-linear regression model for the process itself. As with all parametric models, this approach requires a thorough prior theory about the functional form of the process, which is often difficult or even impossible to find in practice due to the complexity of the underlying dynamic systems. However, in the case of the running example, such a formulation exists. Due to the oscillatory nature of this (self-) regulatory process, a cosine function is a natural candidate to model it. Additionally, to account for the oscillations diminishing over time, the model needs to describe a decreasing amplitude. An example of such a model is:

$$f(t) = Ae^{-ct} \cos(\omega t - \delta)$$

$$\omega = \sqrt{k - \frac{c^2}{4}}$$
(2)

where the parameters control the initial amplitude of the wave function (A), the frequency (ω), the rate at which the amplitude decreases (c), and the phase (δ). Even for this relatively simple system, this parametric form is already quite complex and for more intricate processes, finding an appropriate model of this form can be nearly impossible.

Instead it is often more practical to define a model for how a process changes over time using a differential equation model (Boker, 2012). These models describe the relationship between the current value of the process and its instantaneous rate of change. By combining a model of the process's change with information about its initial state, it is then possible to infer the entire trajectory of the process. Thus, differential equations form a general class of models capable of representing various dynamic processes. For the running example, we used a damped oscillator model. This is a classic differential equation model originally designed to describe the behavior of an oscillating spring with resistance. This model can be expressed as:

$$\frac{\partial f}{\partial t} = v$$

$$\frac{\partial v}{\partial t} = -cv + -kf$$
(3)

This model relates the rate of change of the position of the spring f to its velocity v . In turn, the rate of change of the velocity is determined by the velocity itself through the damping coefficient c and by the position of the spring through the spring constant k . Here, the spring constant is related to the frequency of the oscillations since a stiffer spring oscillates faster and the damping coefficient is related to how quickly the spring returns to its basis position after being perturbed. Chow et al. (2005) show how this model can be applied in a psychological context to model emotional self-regulation. Such that a perturbed emotional state is regulated back to the desired baseline faster the larger the perturbation is.

Alternatively, when working with equally spaced time points, one can model dynamic

local changes in discrete time using difference equations (Durbin & Koopman, 2012). These equations express the current value of the latent construct as a function of its previous value. A common example of this approach is the classic autoregressive model. While the smooth nature of the running example process cannot be perfectly captured by difference equations, it can be approximated by these equations:

$$\begin{aligned} f_t &= f_{t-1} + v_{t-1} \\ v_t &= (1 - c)v_{t-1} + -kf_{t-1} \end{aligned} \tag{4}$$

Lastly, when modeling a process through local dynamics, two types of errors should be taken into account. The first is dynamic error, which describes any external perturbations to the latent construct that are carried forward over time. For instance, if a participant experiences an unusually pleasant conversation that elevates their true positive affect, this change represents an error effect if it is not accounted for by the model. However, since the true positive affect level has increased, this will influence future measurements due to emotional inertia. Dynamic errors can be incorporated into the model in different ways, but the most common approach is to add a normally distributed error term to the deterministic dynamic change models described above.

The second type of error that should be considered is measurement error, which represents the difference between observed measurements and the latent construct values, which may be introduced by an imperfect measurement instrument. These errors are typically modeled using a factor or item response model that links the observations to the latent construct. Combining a dynamic change model with a dynamic error component and a factor model yields a state space model (discrete time; Durbin & Koopman, 2012) or a stochastic differential equation model (continuous time). These models can then be used to infer the process and estimate the parameters of the dynamic equations using the Kalman filter and its extensions (Chow et al., 2007).

Simulation

Problem

A simulation study was conducted to assess the effectiveness of the introduced parametric and non-parametric methods in recovering different non-linear processes, which may be encountered in EMA research (Figure 1). To apply the methods under the conditions described in the introduction, and within the constraints of available software implementations, the simulation focused on a univariate single-subject design. Hence, the simulated data represented repeated measurements of a single variable for one individual.

Design and Hypotheses

For the non-parametric analysis methods (i.e., LPR, GP), we expect that each method in its default configuration will most accurately infer processes that are (a) continuous (i.e., without sudden jumps), (b) have constant wigglyness (i.e., constant second derivative), and (c) are smooth (i.e., differentiable). We expect this, because both methods by default produce continuous, smooth estimates with a single constant bandwidth or lengthscale. For the GAMs, we expect that only criteria (a) and (c) will influence the performance, as GAMs do not assume constant wigglyness. The parametric modeling approach is expected to provide the most accurate inferences, serving as a benchmark for comparison with the other methods. We also expect that the performance of parametric models will decrease as the model complexity increases (e.g., with jumps or reduced smoothness), though likely less so than with non-parametric and semi-parametric models. Additionally, we expect that larger sample sizes will lead to more accurate inferences, with both (d) the overall length of the sampling period and (e) the sampling frequency being varied.

To conduct the simulation with processes that might be encountered in real EMA studies, we selected the exemplar processes illustrated in Figure 1 as a basis. These include two growth curves, modeled as an exponential and a logistic growth curve, a mean-level switching process, modeled as a cusp catastrophe, and a self-regulatory process, represented by a damped oscillator. These processes make it possible to test the impact of (a) sudden jumps and (b) changing

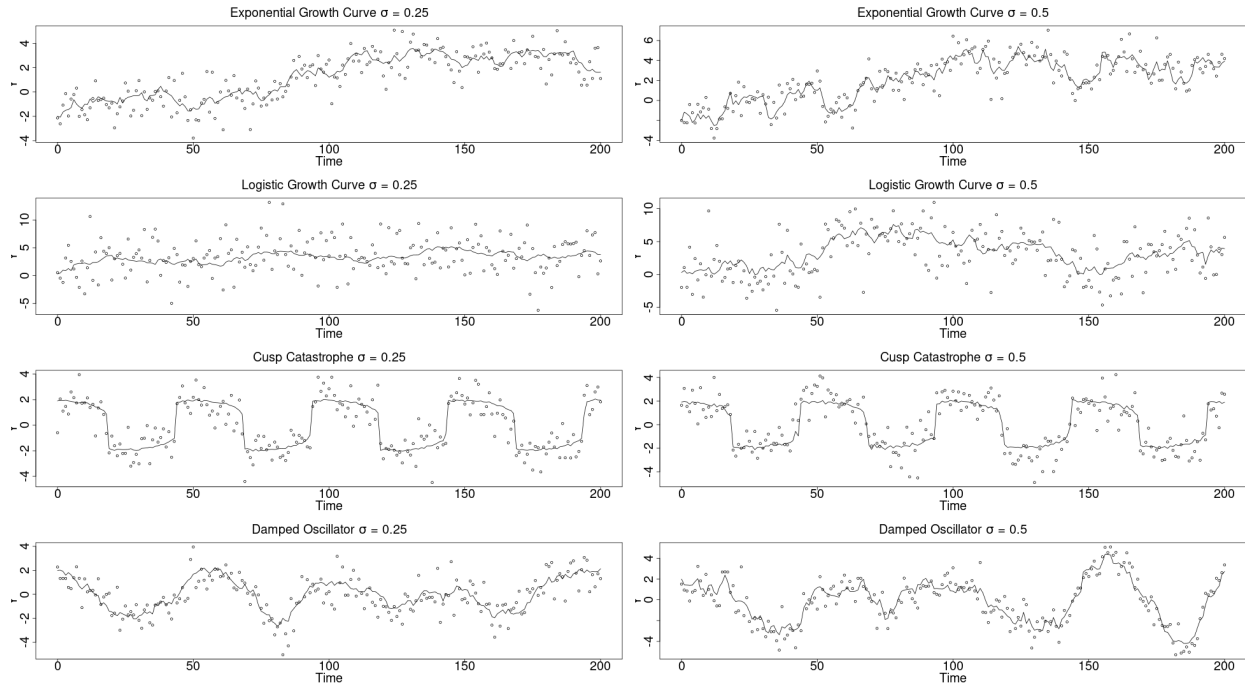
wigglyness on the four methods. First, we hypothesize that the cusp catastrophe model, which is the only process featuring jumps, will be least accurately inferred by all methods. Second, all four processes exhibit changes in wigglyness (i.e., changes in the second derivative) over time. However, while the wigglyness of the exponential and logistic growth functions and the damped oscillator decreases monotonically, the cusp catastrophe's wigglyness changes cyclically. Therefore, we hypothesize that longer sampling periods for the exponential and logistic growth curves and the damped oscillator will reduce the inference accuracy of the LPR and the GP, as the single bandwidth or lengthscale parameter becomes increasingly inadequate to capture the changing wigglyness over time. We do not expect this effect to occur for the cusp catastrophe process, or when using GAMs or parametric models.

To manipulate the (c) smoothness of the processes, a dynamic noise component was added to the data-generating models. This perturbed the processes at each point in time by a normally distributed error, resulting in non-smooth (i.e., non-differentiable or rough) trajectories. The degree of roughness was controlled by the variance of these dynamic errors and we considered variances of 0.5, 1, and 2 reasonable relative to the process range. Figure 5 illustrates one possible realization of the exemplar processes with dynamic noise. Importantly, we intentionally omitted a condition without dynamic noise from this simulation, as dynamic noise is reasonably expected to be present in all psychological intensive longitudinal data (ILD).

Additionally, the sample size was varied during the simulation by manipulating both (d) the sampling period and (e) the sampling frequency, as these distinct methodological choices are expected to impact the performance of the analysis methods differently. Specifically, for the LPR and the GP, which rely only on data in local neighborhoods during the estimation, we expected that extending the sampling period beyond this neighborhood will not increase the inference accuracy. In fact, if the process exhibits changing wigglyness over the extended period, as previously discussed, increasing the sampling period might even negatively affect the inference accuracy. In contrast, we expected GAMs and parametric models, which incorporate the entire dataset in their estimations, to perform better with a longer sampling period. Since there is no

Figure 5

One possible realization of the non-linear exemplar processes with dynamic errors



inherent scaling to the time axis in this simulation, we chose to simulate the first half of each process in one condition and the full process in another, referred to as sampling periods of one or two weeks. This scaling is arbitrary and could be changed to any time frame. Lastly, we expected that increasing the sampling frequency will generally improve the inference accuracy across all methods, as it provides more information about the latent processes. We tested sampling frequencies of three, six, and nine measurements per day, to align these choices with typical experience sampling study designs (Wrzus & Neubauer, 2023).

Procedure

To simulate the data, each exemplar process was represented as a generative stochastic differential equation model. This means that, at each time point, the rate of change of the process was determined by a (non-linear) function of its current value combined with an additive dynamic error component modeled by a Wiener process. Data from these generative models were then

simulated using the Euler-Maruyama method. This method draws samples from stochastic differential equation models with an arbitrary degree of accuracy, while making it possible to manipulate the sampling frequency and dynamic error variance. Measurement errors were added later to the latent process data at each time point from a standard normal distribution, generating the final sets of observations. For a detailed technical explanation of the data generation process, see Appendix A.

To determine the required number of data sets per condition, a power simulation was conducted based on an initial pilot sample of 30 generated data sets per condition. Based on the pilot sample, the outcome measures (e.g., MSE, GCV, and confidence interval coverage scores) and their corresponding standard deviations were calculated by condition. These standard deviations were then used to predict the Monte Carlo standard errors of the means of each outcome measure across increasing sample sizes. These Monte Carlo standard errors reflect the expected variation in the outcome statistics due to random processes within the simulation. We selected the number of data sets per condition for the full simulation, so that the maximum expected Monte Carlo error across all outcome measures and conditions was 0.05. This criterion was met with N data sets per condition.

Model estimation

After simulating the data, all introduced methods were applied to each data set using the statistical software R (R Core Team, 2024). First, the LPR was estimated using the `nprobust` package (Calonico et al., 2019), which allows to correction for the bias inherent in LPRs. Second, GPs were estimated in STAN (Gabry et al., 2024) with a zero mean and a squared exponential covariance function, following common practice. Third, GAMs with a single smooth term for time were fitted using the `MGCV` package (Wood, 2011). Finally, parametric differential equation models corresponding to the true data-generating models were created and estimated using the `Dynr` package (Ou et al., 2019). While the same non-parametric models were used across all conditions, the parametric models were tailored to each specific data-generating process. After

fitting each model to the data, they were used to obtain point and interval estimates (i.e., 95% confidence and credible intervals) for the latent process at each time point. A detailed description of each model fitting procedure is provided in Appendix B.

To ensure reliable model fit and reasonable inferences, the fitting procedures for each method were validated on pilot samples within each condition. After this, an initial run of the simulation was performed, which revealed that the GAMs over- or linearly underfit some data sets and that the parametric models overfit some data sets. To prevent this, the fitting procedures of both methods were adjusted and the simulation rerun. Further, if any models failed to converge during the simulation, the corresponding outcome measures were excluded from the following analyses.

Outcome measures

To evaluate and compare the performance of the different analysis methods, we focused on three outcome measures. The first two assessed each method's accuracy in predicting the process values at or between the observed time points. These predictive accuracy measures indicate how well each method captures the underlying non-linear process. The third outcome measure evaluated the accuracy of the uncertainty estimates provided by each method. Specifically, whether the confidence or credible intervals produced by each method correctly included the true state value the expected proportion of times.

Capturing the non-linear process

To assess how effectively each method captured the non-linear process at the observed time points, we calculated the mean squared error (MSE) between the estimated and generated process values. Additionally, to evaluate how well each method would predict unobserved process values within the process range, we computed the generalized cross-validation ((GCV; Golub et al., 1979)) criterion for each method and data set. The GCV is a more computationally efficient and rotation-invariant version of the ordinary leave-one-out cross-validation criterion, with a similar interpretation. The latter is calculated by removing one data point, refitting the model

while keeping certain parameters fixed, predicting the left-out observation, and calculating the squared prediction error. By repeating this procedure for each data point and averaging the squared errors, leave-one-out cross-validation provides an estimate of how accurately the model predicts unobserved values within the design range.

Subsequently, two ANOVAs were conducted to analyze differences in MSE and GCV values across simulation conditions and analysis methods. To simultaneously identify both factors that likely do and do not affect the MSE and GCV values, an exhaustive model search was performed. In this search, AIC and BIC model weights were used to determine the model, which provides the optimal balance between data fit and model complexity. Notably, AIC weights can be interpreted as conditional model probabilities when the true model is among the tested models (Wagenmakers & Farrell, 2004). The specific hypotheses stated earlier were tested using post hoc tests and marginal comparisons with appropriate corrections for multiple comparisons. Model assumptions were assessed on the final selected model. Given the large sample sizes in this simulation, the ANOVAs were robust to moderate violations of normality, and any potential violations of homoscedasticity were addressed using heteroscedasticity-consistent standard errors. However, if both assumptions were severely violated, results were instead presented descriptively, using means and standard errors for the respective conditions.

Uncertainty quantification

To evaluate the uncertainty estimates provided by each method, we recorded whether the true generated process was located within the confidence or credible intervals at each time point. Subsequently, the average confidence interval coverage proportion for each method and data set was obtained, by averaging over all time points. Given that all confidence or credible intervals were set at a 95% confidence level, the expected coverage proportion should ideally be close to 95%. Due to Monte Carlo error in the simulation, average coverage proportions between 93% and 97% were also deemed acceptable. Average coverage proportions above 97% suggested overestimated standard errors, while those below 93% indicated either a poor approximation of

the underlying process or an underestimation of the standard errors.

Results (Template)

Capturing the non-linear process

For both outcome measures, the AIC and BIC model weights indicated the most complex model, which included all interactions, as the best-fitting model with a model weight of 1. Although the residuals for both models showed considerable deviations from normality, characterized by being platykurtic, the residual distributions were unimodal and approximately symmetric. Therefore, given the large sample sizes in this simulation, the ANOVAs are expected to be robust against these deviations. However, a Breusch-Pagan test indicated heteroscedasticity in the residuals for both outcome measures, which was subsequently corrected. After applying these corrections and using Bonferroni adjustments for conducting two separate ANOVAs, the type-III ANOVA for MSE revealed significant main and interaction effects, except for one four-way interaction and the five-way interaction. The type-III ANOVA for the GCV values indicated that all effects were significant. Although most higher-order interactions had partial- η^2 values close to zero for both outcome measures. Table 1 presents the partial- η^2 values for all effects that showed at least a small effect size, highlighting that the largest effects were associated with the main effects and interactions that include the analysis method and the latent process. Therefore, the following section will focus on analyzing these effects. For a comprehensive overview of all effects in the model, see Appendix C.

Figure 6 (a) shows the mean effect of different analysis methods for each process, averaged across sampling periods, frequency, and dynamic error variances. As expected, the parametric models consistently show the lowest MSE and GCV across all processes. Among the semi- and non-parametric techniques, the GAM performed the best in terms of MSE and GCV, followed by the LPR and the GP infer the latent processes with the least accuracy. Additionally, there are some differences in how accurately both growth curves and the damped oscillator were inferred. However, contrary to our expectations, the cusp-catastrophe process was inferred with a

Table 1

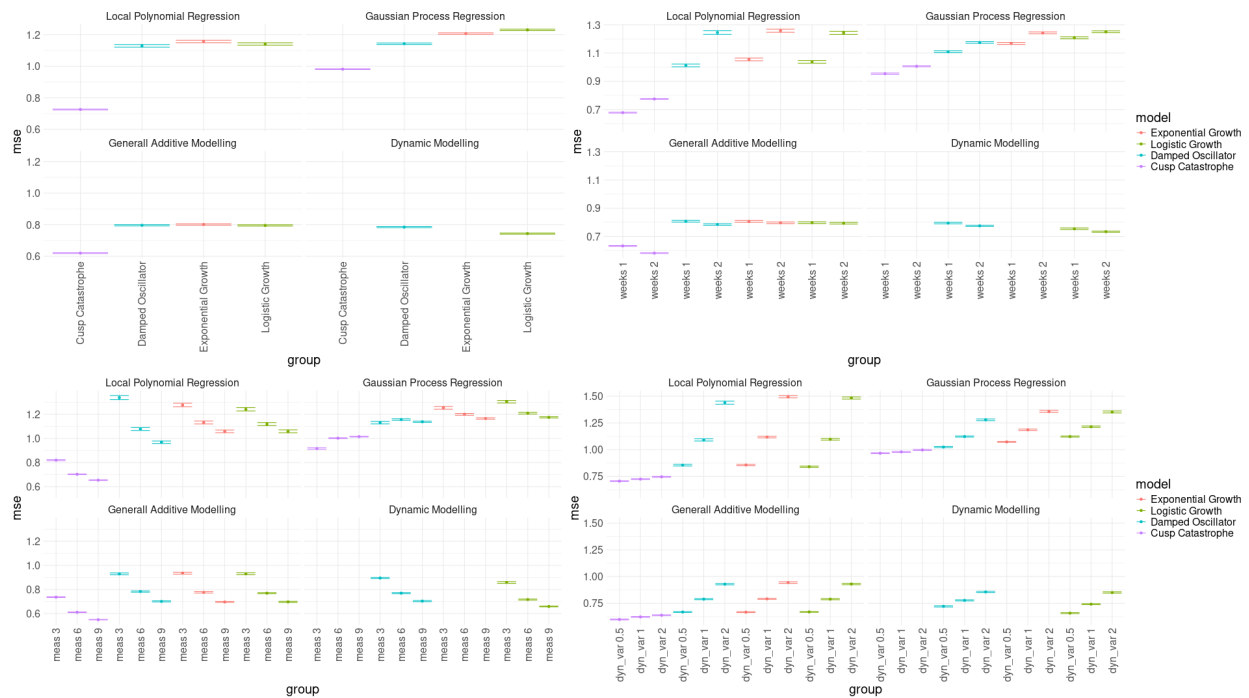
All MSE and GCV ANOVA effects that have at least a small effect size in terms of the partial- η^2 (> 0.01)

Factor	MSE partial- η^2
Method	0.65
Process	0.43
Sampling period (SP)	0.07
Sampling frequency (SF)	0.23
Dynamic error variance (DE)	0.47
Method:Process	0.11
Method:SP	0.10
Method:SF	0.08
Method:DE	0.18
Process:SF	0.02
Process:DE	0.18
SF:DE	0.02
Method:Process:SF	0.03
Method:Process:DE	0.03

lower MSE and GCV by all methods.

Figure 6

Average MSE values illustrating the simulation results. Panel (a) shows the effect of the analysis methods for each latent process. The other three panels show the effects of measurement period (b), measurement frequency (c), and dynamic error variance (d) for each analysis method and latent process.



The effect of the sampling period is illustrated in Figure 6 (b), showing that extending the sampling period from one to two weeks decreased the mean MSE and GCV values for the parametric models and GAM, but increased these values for the LPR and GP. This partially aligns with our expectations, as we predicted this effect would be absent for the LPR and GP when inferring the cusp catastrophe model, which is not supported by the data. Increasing the sampling frequency (Figure 6 c) generally led to lower mean MSE and GCV values, for all methods and processes. Lastly, larger dynamic error variances (Figure 6 d) resulted in higher mean MSE and GCV values across all processes and analysis methods, with this effect being least

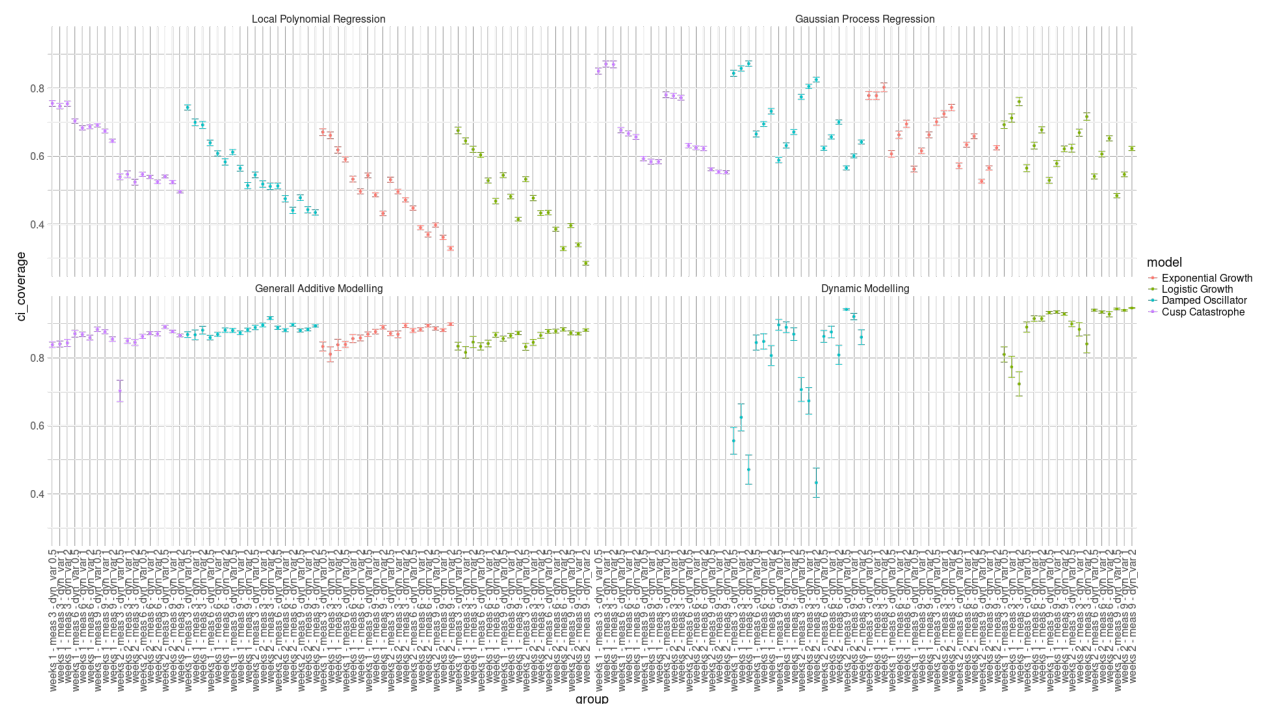
pronounced in the cusp-catastrophe model.

Uncertainty quantification

Figure 7 shows the average confidence interval coverage proportion for each condition in the simulation. It can be seen that the only analysis method which produces an average confidence interval coverage within the prespecified accuracy interval of 93% to 97% is the parametric modelling. All three non and semi-parametric analysis methods produced average confidence interval coverages considerably below 93%. However, among these methods the average confidence interval coverages achieved by the GAM are more accurate and more stable across the different conditions than the coverages of the LPR and GP.

Figure 7

Average confidence interval coverage across all processes, analysis methods, and simulation conditions



An Empirical Example

In the following, the four analysis methods previously introduced were applied to depression data from the Leuven clinical study. This study used experience sampling measures to study the dynamics of anhedonia in individuals with major depressive disorder (Heininga et al., 2019). This study was selected for its heterogeneous sample, which includes participants with major depressive disorder, borderline personality disorder, and healthy controls. This diversity increases the likelihood of the data exhibiting a range of (possibly non-linear) dynamics and processes. Specifically, Houben et al. (2015) found in their meta-analysis that individuals with lower psychological well-being tend to experience greater emotional variability, less emotional stability, and higher emotional inertia. Although, this finding did not replicate in an analysis of positive affect within the Leuven clinical study (Heininga et al., 2019). Further, emotional inertia, the extent to which an emotional state carries over across time points, has been shown to vary within individuals over time, which makes it likely that the processes underlying this data are non-stationary. Lastly, this data also makes it possible to explore the previously introduced theory that emotion regularization may work as a (self-) regulatory system through parametric modelling.

To maintain consistency with how the methods were introduced and to avoid using measurement models with multiple indicators, we analyzed momentary depression, which was measured using a single item. This item was chosen over affect measures because it displays sufficient variability, has a relatively low proportion of participants with strong floor or ceiling effects, and is measured on a broad response scale (0 to 100), making it ideal for illustrating the introduced methods.

Sample and data description

The participants in the clinical sample of the Leuven clinical study were screened by clinicians during the intake in three Belgian psychiatric wards (Heininga et al., 2019). Patients who met the DSM criteria for mood disorders or borderline personality disorder during the intake were eligible for enrollment, while those presenting with acute psychosis, mania, addiction, or

(neuro-)cognitive symptoms were excluded. Following the screening, 90 patients enrolled in the study. Additionally, 44 control participants were matched to the clinical sample by gender and age, resulting in a total sample size of 134. Within the clinical sample, three patients withdrew during the baseline assessment, two were excluded due to faulty devices, and seven were removed for responding to less than half of the EMA measures. In the control sample, one participant was excluded for responding to less than half of the EMA measures, and three were removed for meeting the criteria for a current psychiatric disorder. Consequently, the final published data set included 78 participants in the clinical sample and 40 participants in the control sample.

During the study, all participants completed a baseline assessment, followed by seven days of semi-random EMA assessments, with 10 equidistant assessments per day. However, the starting date of the EMA measures varied between people. During each assessment, participants responded to 27 questions covering emotions, social expectancies, emotion regulation, context, and psychiatric symptoms. This analysis focused on the item assessing momentary depressive mood (i.e., "How depressed do you feel at the moment?") rated on a scale from 0 to 100. One additional participant was removed from the analysis for consistently reporting a score of zero across all assessments, resulting in a final sample of 117 participants. For a more thorough sample and data description, see Heininga et al. (2019).

The published data set was obtained from the EMOTE database (Kalokerinos et al., n.d.). The initial study procedure was approved by the KU Leuven Social and Societal Ethics Committee and the KU Leuven Medical Ethics Committee. This secondary data analysis was approved by the Ethics Review Board of the Tilburg School of Social and Behavioral Sciences.

Analysis Plan

Exploratory idiographic analysis

First, the LPR, GP, and GAM were applied to explore the idiographic latent processes underlying the data. Each method was applied separately to the time-series of each participant, using the same specifications as in the simulation study (Appendix B). However, for the LPR,

only local cubic polynomials were considered to keep the interpretation of the bandwidth consistent across participants. Since all participants were assessed over seven days, but not during the same period, the time-series for each participant was centered so that the first measurement time point served as the zero point. The LPR bandwidth, GP lengthscale, and GAM smoothing parameter were then analyzed to assess the wigglyness of the idiographic processes. Additionally, the GCV values produced by each method were evaluated to determine which method provided the most accurate interpolations. Lastly, the mean squared error was calculated for each method and data set to estimate the expected measurement error.

Multilevel analysis

After the idiographic analysis, mixed-effects GAMs were fitted to the data to assess the extent to which common and idiographic smooth are necessary for accurately modeling the data. The baseline model was a random intercept model, where each participant's data was modelled by a person-specific flat line. The second model fitted a random smooth, in which each participant's data was modelled by an individual smooth function with a common smoothing parameter through factor-smooth interactions. To determine if there was evidence of a common non-linear trend, this model was compared to an extended model, which combined a common non-linear smooth with individual random non-linear deviations. These deviations were again modelled using factor-smooth interactions with a common smoothing parameter. Finally, to test potential heterogeneity in the smoothing parameters across participants, a fourth model was fitted, which allowed for individual smooth terms with varying smoothing parameters.

Parametric analysis

Lastly, three parametric differential equation models were applied to explore the data. The first model is a random-walk model, where the rate of change in the latent depression state does not depend on its current value, such that future values are only influenced by random perturbations in form of the dynamic errors. The second model is a continuous-time autoregressive model, where the depression score reverts to an individual's mean at a rate linearly

dependent on the distance from this mean. Both the individual mean and autoregressive effect were treated as random effects by incorporating them as time-invariant state variables, similar to how random effects are included in dynamic structural equation models. The third model is the damped oscillator model discussed earlier. To adapt this model to the data structure, a random mean was added to account for between-person differences in the baseline level around which the depression score oscillates. Random effects were also included for the spring constant and damping coefficient to accommodate individual differences in oscillation patterns. All three models included components for measurement and dynamic error variances. Lastly, the model fit was compared between the models using the AIC and BIC.

Results

Exploratory idiographic analysis

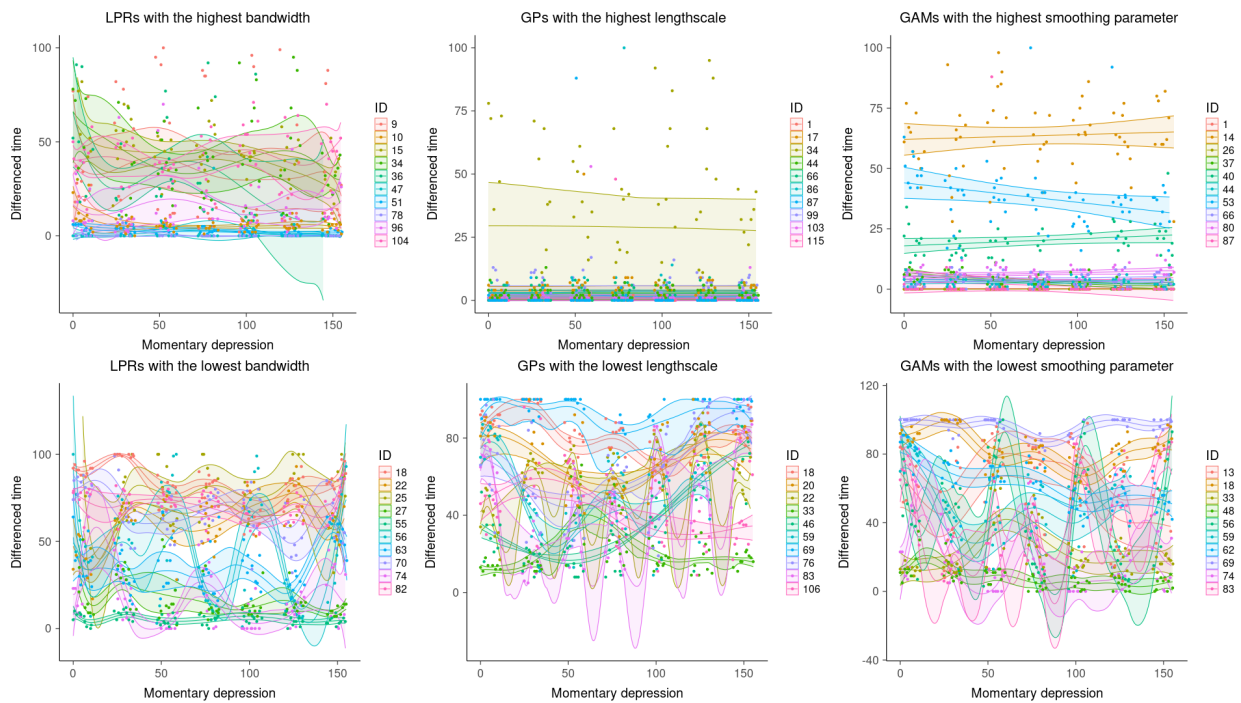
First, the LPR, GP, and GAM were used to estimate the individual latent depression processes. For the local cubic regression, the median optimal bandwidth was 21.28 hours (*IQR*: 5.52). For the GP, the median optimal length scale was 22.57 standard deviations (*IQR*: 15.09). Lastly, for the GAMs, the median optimal smoothing parameter was $8.18 * 10^9$ (*IQR*: $1.76 * 10^{10}$). Unfortunately, these three measures of wiggleness are not only on different scales, but there is also only a moderate correlation between the bandwidth and lengthscale parameters ($r = 0.33$). Further, the smoothing parameter of the GAMs shows little to no correlation with the other two measures (bandwidth: $r = -0.03$; length scale: $r = -0.08$). This discrepancy arises because, while all three parameters reflect the wiggleness of the estimate, they capture different aspects of it. For example, in data with a linear trend, the bandwidth of the local cubic regression and the smoothing parameter of the GAMs would theoretically be infinite, while the length scale parameter of the GP would have a finite value. Additionally, the interpretation of each wiggleness parameter depends in the model configurations chosen and would change for different configurations of these methods.

Because of this, there is not much value in interpreting the absolute values of these

parameters. Instead, we explored the range of functional behaviours inferred by the most extreme values of each parameters. Figure 8 shows the ten least and most wiggly processes inferred by each method. This figure reveals considerably heterogeneity in the functional behaviour inferred by each method. Most interestingly, the least wiggly processes inferred by both the GPs and the GAMs are linear trends, indicating the absence of any dynamic errors for these individuals. In contrast to this, the processes with the highest inferred wigglyness, display either large dynamic errors around their respective person means or in addition to this a different non-linear dynamic.

Figure 8

The ten least and most wiggly idiographic latent depression processes as inferred by the LPRs, GPs, and GAMs



Lastly, a cross-validation was conducted using the generalized cross-validation criterion to investigate which method predicted the latent processes most accurately. The median GCV for the GAMs was 125.29 (*IQR*: 201.04), for the GP it was 248.27 (*IQR*: 578.00), and for the LPR it was 131.57 (*IQR*: 213.73). In addition to this, the mean squared error was calculated between the

predictions generated by each method and the data. Here the median MSE of the GAM was 114.04 (*IQR*: 171.59), for the GP it was 197.54 (*IQR*: 388.40), and for the LPR it was 101.74 (*IQR*: 153.77). Together with the GCV this indicates that the GAM inferred the latent processes most accurately, whereas the LPR slightly overfit the data and the GPs tended to underfit the data.

Multilevel analysis

First, the baseline model with only the random intercepts had an adjusted R^2 of 0.75 ($GCV = 231$). Extending the model to include random smooths for each individual with a common smoothing penalty increased the adjusted R^2 to 0.81 ($GCV = 194.89$). However, adding a common smooth to this model did not change the model fit. Lastly, using person-specific smooth terms for the random smooth increased the adjusted R^2 further to 0.82 ($GCV = 184.29$). Table 2 compares the four models using the AIC and BIC. The AIC indicates the most complex model, providing evidence for the necessity of individual smooth parameters. However, the BIC, which given the sample size uses a heavier penalty, indicates the random intercept model as the best fitting model, highlighting that the majority of the variance in the data is due to mean level differences. Overall, there is conflicting evidence regarding which model fits the data best.

Table 2

Model information criteria for four mixed effects GAMs

Model	Degrees of freedom	AIC	BIC
Random intercept	117.3599	58997.31	59803.73
Random smooth w. common penalty	708.5276	57712.79	62581.35
Common smooth w. deviations	708.0516	57712.76	62578.04
Random smooth w. individual penalties	625.9867	57331.22	61632.60

Parametric analysis

Not working yet

Discussion

Conclusion

References

- Abdessalem, A. B., Dervilis, N., Wagg, D. J., & Worden, K. (2017). Automatic Kernel Selection for Gaussian Processes Regression with Approximate Bayesian Computation and Sequential Monte Carlo. *Frontiers in Built Environment*, 3, 52.
<https://doi.org/10.3389/fbuil.2017.00052>
- Betancourt, M. (2020). Robust Gaussian Process Modeling. Retrieved June 9, 2023, from https://betanalpha.github.io/assets/case_studies/gaussian_processes.html
- Boker, S. M. (2012). Dynamical systems and differential equation models of change. In H. Cooper, P. M. Camic, D. L. Long, A. T. Panter, D. Rindskopf, & K. J. Sher (Eds.), *APA handbook of research methods in psychology, Vol 3: Data analysis and research publication*. (pp. 323–333). American Psychological Association.
<https://doi.org/10.1037/13621-016>
- Boyd, J. P., & Xu, F. (2009). Divergence (Runge Phenomenon) for least-squares polynomial approximation on an equispaced grid and Mock–Chebyshev subset interpolation. *Applied Mathematics and Computation*, 210(1), 158–168.
<https://doi.org/10.1016/j.amc.2008.12.087>
- Bringmann, L. F., Ferrer, E., Hamaker, E., Borsboom, D., & Tuerlinckx, F. (2015). Modeling Nonstationary Emotion Dynamics in Dyads Using a Semiparametric Time-Varying Vector Autoregressive Model. *Multivariate Behavioral Research*, 50(6), 730–731.
<https://doi.org/10.1080/00273171.2015.1120182>
- Bringmann, L. F., Hamaker, E. L., Vigo, D. E., Aubert, A., Borsboom, D., & Tuerlinckx, F. (2017). Changing dynamics: Time-varying autoregressive models using generalized additive modeling. *Psychological Methods*, 22(3), 409–425.
<https://doi.org/10.1037/met0000085>
- Calonico, S., Cattaneo, M. D., & Farrell, M. H. (2019). nprobust: Nonparametric kernel-based estimation and robust bias-corrected inference. *Journal of Statistical Software*, 91(8), 1–33. <https://doi.org/10.18637/jss.v091.i08>

- Ceja, L., & Navarro, J. (2012). 'Suddenly I get into the zone': Examining discontinuities and nonlinear changes in flow experiences at work. *Human Relations*, 65(9), 1101–1127. <https://doi.org/10.1177/0018726712447116>
- Chow, S.-M., Ferrer, E., & Nesselroade, J. R. (2007). An Unscented Kalman Filter Approach to the Estimation of Nonlinear Dynamical Systems Models. *Multivariate Behavioral Research*, 42(2), 283–321. <https://doi.org/10.1080/00273170701360423>
- Chow, S.-M., Ram, N., Boker, S. M., Fujita, F., & Clore, G. (2005). Emotion as a Thermostat: Representing Emotion Regulation Using a Damped Oscillator Model. *Emotion*, 5(2), 208–225. <https://doi.org/10.1037/1528-3542.5.2.208>
- Chow, S.-M., Witkiewitz, K., Grasman, R., & Maisto, S. A. (2015). The cusp catastrophe model as cross-sectional and longitudinal mixture structural equation models. *Psychological Methods*, 20(1), 142–164. <https://doi.org/10.1037/a0038962>
- De Bot, K., Lowie, W., & Verspoor, M. (2007). A Dynamic Systems Theory approach to second language acquisition. *Bilingualism: Language and Cognition*, 10(01), 7. <https://doi.org/10.1017/S1366728906002732>
- Debruyne, M., Hubert, M., & Suykens, J. A. K. (2008). Model Selection in Kernel Based Regression using the Influence Function. *Journal of Machine Learning Research*, 9(78), 2377–2400. <http://jmlr.org/papers/v9/debruyne08a.html>
- Durbin, J., & Koopman, S. J. (2012). *Time series analysis by state space methods* (Second edition). Oxford University Press
- Hier auch später erschienene, unveränderte Nachdrucke Literaturverzeichnis: Seite [326]-339.
- Fan, J., & Gijbels, I. (2018). *Local Polynomial Modelling and Its Applications* (1st ed.). Routledge. <https://doi.org/10.1201/9780203748725>
- Fan, J., & Gijbels, I. (1995). Adaptive Order Polynomial Fitting: Bandwidth Robustification and Bias Reduction. *Journal of Computational and Graphical Statistics*, 4(3), 213–227. <https://doi.org/10.1080/10618600.1995.10474678>

- Fritz, J., Piccirillo, M., Cohen, Z. D., Frumkin, M., Kirtley, O. J., Moeller, J., Neubauer, A. B., Norris, L., Schuurman, N. K., Snippe, E., & Bringmann, L. F. (2023). So you want to do ESM? Ten Essential Topics for Implementing the Experience Sampling Method (ESM). <https://doi.org/10.31219/osf.io/fverx>
- Gabry, J., Češnovar, R., Johnson, A., & Bröder, S. (2024). *Cmdstanr: R interface to 'cmdstan'* [R package version 0.8.0, <https://discourse.mc-stan.org>]. <https://mc-stan.org/cmdstanr/>
- Golub, G. H., Heath, M., & Wahba, G. (1979). Generalized Cross-Validation as a Method for Choosing a Good Ridge Parameter. *Technometrics*, 21(2), 215–223. <https://doi.org/10.1080/00401706.1979.10489751>
- Golub, G. H., & von Matt, U. (1997). Generalized Cross-Validation for Large-Scale Problems [Publisher: [American Statistical Association, Taylor & Francis, Ltd., Institute of Mathematical Statistics, Interface Foundation of America]]. *Journal of Computational and Graphical Statistics*, 6(1), 1–34. <https://doi.org/10.2307/1390722>
- Gu, C. (2013). *Smoothing spline ANOVA models* (2nd ed) [OCLC: ocn828483429]. Springer
Introduction – Model construction – 3. Regression with Gaussian-type responses – More splines – Regression with responses from exponential families – Regression with correlated responses – Probability density estimation – Hazard rate estimation – Asymptotic convergence – Penalized pseudo likelihood – R package gas – Conceptual critiques.
- Harrell, F. E. (2001). General Aspects of Fitting Regression Models [Series Title: Springer Series in Statistics]. In *Regression Modeling Strategies* (pp. 11–40). Springer New York. https://doi.org/10.1007/978-1-4757-3462-1_2
- Hastie, T., & Tibshirani, R. (1999). *Generalized additive models*. Chapman & Hall/CRC
Originally published: London ; New York : Chapman and Hall, 1990.
- Heininga, V. E., Dejonckheere, E., Houben, M., Obbels, J., Sienaert, P., Leroy, B., Van Roy, J., & Kuppens, P. (2019). The dynamical signature of anhedonia in major depressive disorder:

- Positive emotion dynamics, reactivity, and recovery. *BMC Psychiatry*, 19(1), 59.
<https://doi.org/10.1186/s12888-018-1983-5>
- Houben, M., Van Den Noortgate, W., & Kuppens, P. (2015). The relation between short-term emotion dynamics and psychological well-being: A meta-analysis. *Psychological Bulletin*, 141(4), 901–930. <https://doi.org/10.1037/a0038822>
- Jebb, A. T., Tay, L., Wang, W., & Huang, Q. (2015). Time series analysis for psychological research: Examining and forecasting change. *Frontiers in Psychology*, 6.
<https://doi.org/10.3389/fpsyg.2015.00727>
- Kalokerinos, E. K., Russo-Batterham, D., Koval, P., Moeck, E. K., Grewal, K. K., Greenaway, K. H., Shrestha, K. M., Garrett, P., Michalewicz, A., Garber, J., & Kuppens, P. (n.d.). *The EMOTE Database: An open, searchable database of experience sampling data mapping everyday life* (Manuscript in preparation).
- Köhler, M., Schindler, A., & Sperlich, S. (2014). A Review and Comparison of Bandwidth Selection Methods for Kernel Regression [Publisher: [Wiley, International Statistical Institute (ISI)]]. *International Statistical Review / Revue Internationale de Statistique*, 82(2), 243–274. Retrieved November 28, 2023, from
<https://www.jstor.org/stable/43299758>
- Kunnen, S. E. (2012). *A Dynamic Systems Approach of Adolescent Development* [OCLC: 823389367]. Taylor; Francis.
- McArdle, J. J., Ferrer-Caja, E., Hamagami, F., & Woodcock, R. W. (2002). Comparative longitudinal structural analyses of the growth and decline of multiple intellectual abilities over the life span. *Developmental Psychology*, 38(1), 115–142.
<https://doi.org/10.1037/0012-1649.38.1.115>
- Nesselroade, J., & Ram, N. (2004). Studying Intraindividual Variability: What We Have Learned That Will Help Us Understand Lives in Context. *Research in Human Development*, 1(1), 9–29. https://doi.org/10.1207/s15427617rhd0101&2_3

- Newell, K. M., Liu, Y.-T., & Mayer-Kress, G. (2001). Time scales in motor learning and development. *Psychological Review*, 108(1), 57–82.
<https://doi.org/10.1037/0033-295X.108.1.57>
- Ou, L., Hunter, M. D., & Chow, S.-M. (2019). What's for dynr: A package for linear and nonlinear dynamic modeling in R. *The R Journal*, 11, 1–20.
- R Core Team. (2024). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria. <https://www.R-project.org/>
- Rasmussen, C. E., & Williams, C. K. I. (2006). *Gaussian processes for machine learning* [OCLC: ocm61285753]. MIT Press.
- Richardson, R. R., Osborne, M. A., & Howey, D. A. (2017). Gaussian process regression for forecasting battery state of health. *Journal of Power Sources*, 357, 209–219.
<https://doi.org/10.1016/j.jpowsour.2017.05.004>
- Roberts, S., Osborne, M., Ebdon, M., Reece, S., Gibson, N., & Aigrain, S. (2013). Gaussian processes for time-series modelling [Publisher: Royal Society]. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 371(1984), 20110550. <https://doi.org/10.1098/rsta.2011.0550>
- Ruppert, D., & Wand, M. P. (1994). Multivariate Locally Weighted Least Squares Regression. *The Annals of Statistics*, 22(3). <https://doi.org/10.1214/aos/1176325632>
- Tan, X., Shiyko, M., Li, R., Li, Y., & Dierker, L. (2011). A Time-Varying Effect Model for Intensive Longitudinal Data. *Psychological methods*, 17, 61–77.
<https://doi.org/10.1037/a0025814>
- Tsay, R. S., & Chen, R. (2019). *Nonlinear time series analysis*. John Wiley & Sons
Includes index.
- van der Maas, H. L. J., Kolstein, R., & van der Pligt, J. (2003). Sudden Transitions in Attitudes. *Sociological Methods & Research*, 32(2), 125–152.
<https://doi.org/10.1177/0049124103253773>

- Wagenmakers, E.-J., & Farrell, S. (2004). AIC model selection using Akaike weights. *Psychonomic Bulletin & Review*, 11(1), 192–196. <https://doi.org/10.3758/BF03206482>
- Wahba, G. (1980). Spline bases, regularization, and generalized cross validation for solving approximation problems with large quantities of noisy data. In *Approximation Theory III* (pp. 905–912). Academic Press.
- Wang, L. (, Hamaker, E., & Bergeman, C. S. (2012). Investigating inter-individual differences in short-term intra-individual variability. *Psychological Methods*, 17(4), 567–581. <https://doi.org/10.1037/a0029317>
- Witkiewitz, K., & Marlatt, G. A. (2007). Modeling the complexity of post-treatment drinking: It's a rocky road to relapse. *Clinical Psychology Review*, 27(6), 724–738. <https://doi.org/10.1016/j.cpr.2007.01.002>
- Wood, S. N. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society (B)*, 73(1), 3–36.
- Wood, S. N. (2003). Thin Plate Regression Splines. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 65(1), 95–114. <https://doi.org/10.1111/1467-9868.00374>
- Wood, S. N. (2006). *Generalized additive models: An introduction with R* [OCLC: ocm64084887]. Chapman & Hall/CRC.
- Wood, S. N. (2020). Inference and computation with generalized additive models and their extensions. *TEST*, 29(2), 307–339. <https://doi.org/10.1007/s11749-020-00711-5>
- Wrzus, C., & Neubauer, A. B. (2023). Ecological Momentary Assessment: A Meta-Analysis on Designs, Samples, and Compliance Across Research Fields. *Assessment*, 30(3), 825–846. <https://doi.org/10.1177/10731911211067538>