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 understood 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 one or more individuals are assessed at a high frequency (multiple times per day) using brief questionnaires or passive measurement devices. These rich data allow researchers to examine complex temporal variations in the underlying psychological variables within an ecologically valid context and to explain them through (between-person differences) in within-person processes.

Due to these ILD studies, many non-linear psychological phenomena and processes have been discovered during 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, following an intricate non-linear trajectory from a (person specific) starting point towards an (person specific) asymptote, which reflects the inidividual's 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 shows common model choices for these kinds of processes in the form of an exponential growth function (a) and a logistic growth function (b).

Another common non-linear phenomenon is that the construct of interest switches between distinct states, which often correspond to different mean levels. 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 &

Figure 1

Examples of non-linear processes demonstrated to occur in psychological time series.
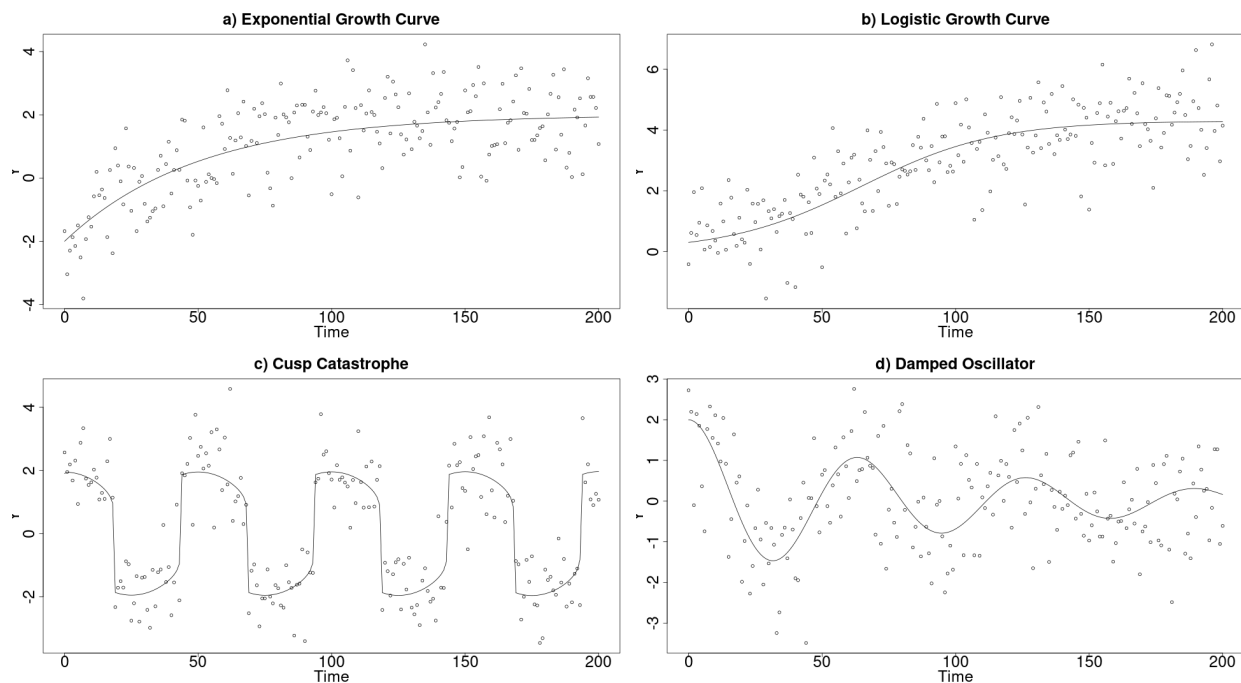
Note. This figure shows four demonstrated psychological non-linear processes. Panels (a) and (b) show exponential and logistic growth curves, respectively. Panel (c) shows a cusp catastrophe model. Lastly, panel (d) shows a damped oscillator.

Marlatt, 2007). This sudden switching behavior has been modelled using a cusp catastrophe model. This dynamic model, drawn from catastrophe theory, naturally leads to mean level switches when varying one of its parameters (Chow et al., 2015; van der Maas et al., 2003) and has been exemplified in Figure 1 (c).

As a final example, one may consider (self-) regulatory systems, which maintain a desired state by counteracting external perturbations. In these systems the regulatory force often depends on the distance between the current and the 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 changes disproportionately with larger mismatches. Such a (self-) regulatory model has been used to model, for example, emotion regulation (Chow et al., 2005) using a damped oscillator model.

This model is exemplified in Figure 1 (d).

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). Frequently, psychological theories are too general (Oberauer & Lewandowsky, 2019) to predict specific non-linear dynamics and ILD studies could have the potential to refine these theories through a nuanced understanding of how the involved psychological variables interact over time. Such refined theories could, for instance, take the form of formal dynamic models, such as differential equation (Boker, 2012) or state-space models (Durbin & Koopman, 2012), describing how a given process changes over time. However, in order to develop these types of theories it is first necessary to identify phenomena in ILD, which are replicable and empirically observable features of the underlying processes, such as the described state switching or regulatory oscillations. Formal theories about the underlying process should then be able to explain these phenomena and different candidate theories can be compared on their success to do so (Borsboom et al., 2021). While the study of non-linear phenomena in ILD is receiving increasingly more attention in psychology and different statistical techniques are developed to explore these phenomena (Cui et al., 2023; Humberg et al., 2024), researchers are currently still limited in their ability to infer non-linear phenomena from ILD due to a lack of advanced statistical methods that are flexible enough to adequately capture and explore these processes, which hinders the development and evaluation of guiding theories. Due to this lack of adequate available statistical methods, non-linear trends are most often addressed in psychology through polynomial regression or spline regression.

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 the currently available methods underscore the need for more sophisticated statistical methods to study and explore non-linear processes. Various such advanced statistical methods, such as kernel regression, Gaussian processes, and smoothing splines 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 assumed 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 three advanced non-linear analysis methods and evaluates their applicability to typical ILD scenarios (Section 1). Specifically, we compare how well each method can recover different latent processes under common ILD conditions in a simulation study (Section 2). We also demonstrate the conclusions that can be drawn from each method by applying them to an existing dataset (Section 3). The specific methods reviewed in this article are semi- and non-parametric regression techniques that are able

infer non-linear functions from data while accommodating varying degrees of prior knowledge. Further, to introduce these methods accessibly and apply them under conditions where software implementations are available, this article focuses on the univariate single-subject design.

1   Non-linear analysis methods

1.1   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 function to approximate the entire process, LPR estimates smaller, local polynomials at any 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, 1995a; Ruppert & Wand, 1994).

To determine the value that the LPR predicts at a specific time point, the data is first 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 fact that the polynomial approximation is more accurate for data points closer in time, a weighting function is applied during the polynomial estimation, which assigns weights to each data point based on its distance from the point of interest. The value that the LPR predicts for the chosen time point is then given by the intercept of the locally weighted polynomial at this point in time. This procedure can be written as:

$$y_t = f(t) + \varepsilon_t$$

$$\mathbf{X} = \begin{bmatrix} 1 & (t_1 - t^*)^1 & \dots & (t_1 - t^*)^p \\ \vdots & \vdots & \ddots & \vdots \\ 1 & (t_n - t^*)^1 & \dots & (t_n - t^*)^p \end{bmatrix}$$

$$\mathbf{W} = \begin{bmatrix} w_{1,1} & & \\ & \ddots & \\ & & w_{n,n} \end{bmatrix}$$

(1)

$$\hat{f}(t^*) = Intercept((\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} s \mathbf{X}^T \mathbf{W} \mathbf{y})$$

where a univariate process f is inferred at the time point t^*. Then, $\mathbf{X}$ is the model matrix of the local polynomial[1] centered around the chosen time point and $\mathbf{W}$ is a diagonal matrix comprised of the weights associated with each datum. To find the value that the LPR predicts for 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 an example process introduced 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.

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 only when transitioning from an odd to an even power (Ruppert & Wand, 1994). The data weighting in an LPR is achieved through a kernel function of the form $w_{i,i} = K(\frac{t_i - t^*}{h})$, which is usually centered and symmetric to assign weights based on the distance from the origin. Common kernel choices include the Gaussian and Epanechnikov, with the latter being optimal in many applications and aspects (Fan et al., 1997). Each kernel is further defined by a bandwidth parameter h, which determines its width and

---

[1] Note that the terms in this matrix correspond to a Taylor expansion with p terms around t^*.

Figure 2

Demonstration of a local polynomial regression
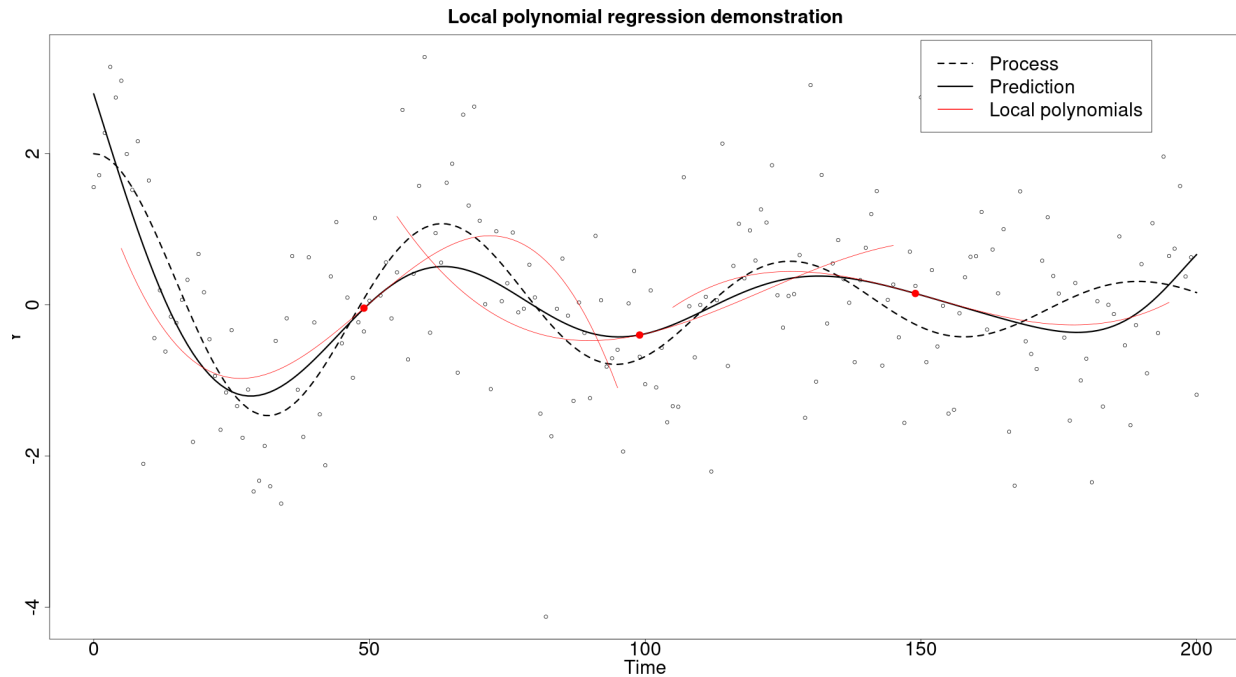
Note. This figure shows how LPR (solid black) estimates the underlying process (dotted black). Here, three local cubic functions (red) are shown as examples at the time points 50, 100, and 150, which provide the values of the LPR at these time points. The complete LPR was plotted by evaluating these local cubic regressions at 200 points along the time axis.

effectively controls the influence of more distant data points. The bandwidth parameter represents the wiggliness of the estimated process in practice. Several methods are available to find the optimal bandwidth by optimizing a data-dependent criterion function, such as the cross-validation error 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 p times differentiable, which is a necessary condition for local polynomial approximation (since it relies on Taylor's theorem). It is further noteworthy that the polynomial approximation is biased unless the process follows a polynomial with a maximum degree of p. However, this bias is usually negligible and there are methods available to correct for it (Calonico et al., 2019). Another key assumption is that the process has

constant wiggliness, represented by a single bandwidth parameter. However, this assumption may be relaxed by using a time-varying bandwidth (Fan & Gijbels, 1995b) or polynomial degree (Fan & Gijbels, 1995a), but this extension is beyond the scope of this introduction.

1.2 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, which is 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) continuous functions. In a Bayesian framework, one can use a GP to define a prior distribution for the latent process, as $P(f) \sim GP$. 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.

$$P(f \,|\, \mathbf{y}) \propto P(\mathbf{y} \,|\, f)P(f) \tag{2}$$

This posterior distribution represents an updated belief about which functions describe the latent process well (Kruschke, 2011), 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.

The GP prior is parameterized by a mean function $m(t)$ and a covariance function $cov(t, t)$, 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

Figure 3

Demonstration of a Gaussian process regression

Gaussian process demonstration
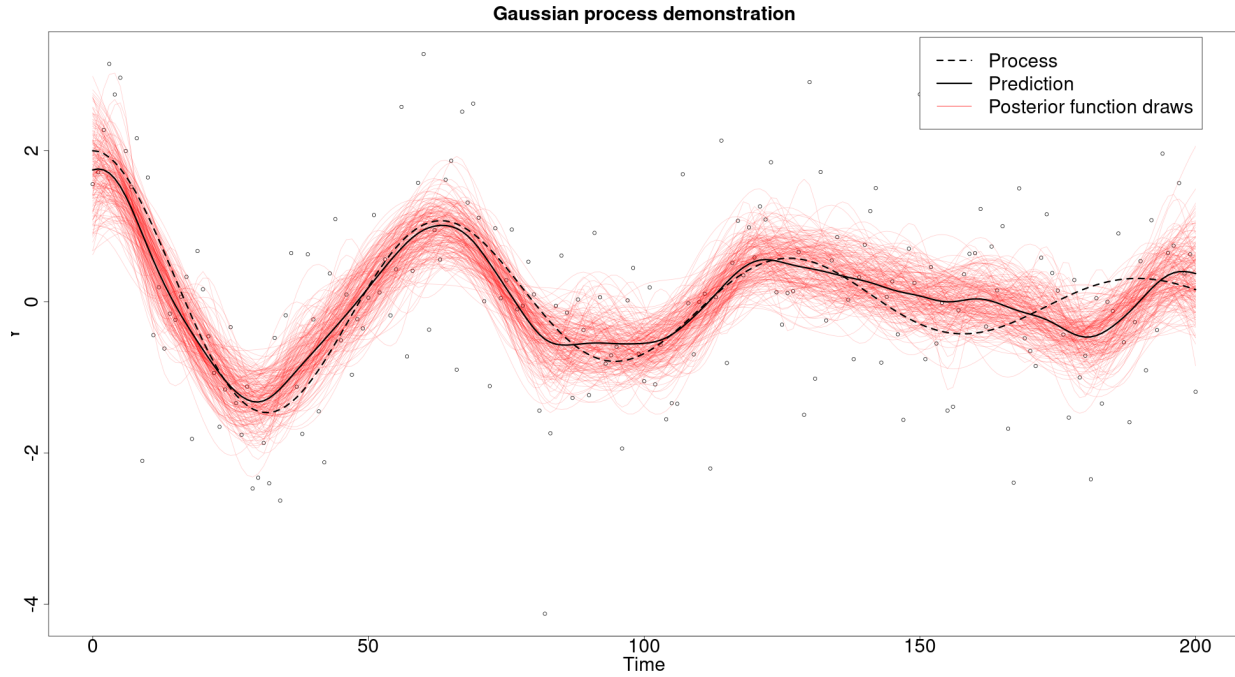
Note. This figure shows how a Gaussian process regression estimates the underlying process (dotted black). Here, a sample of functions drawn from the posterior Gaussian process probability distribution with a squared exponential kernel is shown (red). The predicted value for the underlying process is then obtained by averaging the drawn functions.

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 kernel), such that

$$cov(t_i, t_j) = k(|t_i - t_j|) \tag{3}$$

Sometimes, is possible to derive an analytic posterior distribution for a GP, as for example in the case of a prior GP with a zero mean function and a Gaussian likelihood with a zero mean and a standard deviation of σ. Then the posterior process values at a selected set of time points $\mathbf{t}^*$ are described by a multivariate normal distribution with a mean vector and covariance matrix of:

$$f(\mathbf{t}^* \mid \mathbf{y}) \sim MvN(\mu(\mathbf{t}^*), \Sigma(\mathbf{t}^*))$$

$$\mu(\mathbf{t}^*) = K(\mathbf{t}^*, \mathbf{t})[K(\mathbf{t}, \mathbf{t}) + \sigma^2 \mathbf{I}]^{-1} \mathbf{y} \tag{4}$$

$$\Sigma(\mathbf{t}^*) = K(\mathbf{t}^*, \mathbf{t})[K(\mathbf{t}, \mathbf{t}) + \sigma^2 \mathbf{I}]^{-1} K(\mathbf{t}, \mathbf{t}^*)$$

where $\mathbf{K}$ is a matrix collecting the covariances given by $cov(t, t)$ at the observation time points $\mathbf{t}$ and the evaluation time points $\mathbf{t}^*$ (which may also be the same).

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 corresponding posterior hyperparameter distributions also provide more interpretable information than the LPR about the process. Most common covariance kernels build on a characteristic lengthscale and a marginal standard deviation hyperparameter. The characteristic lengthscale determines the wigglyness of the estimated process, similarly to the bandwidth parameter in an LPR. The marginal standard deviation describes the spread of the functions which are described by the GP at any point in time. However, GP regression can also include additional hyperparameters, allowing for more specific theories to be tested through model comparison.

The functional behavior of a GP is entirely determined by the posterior mean and covariance function. Therefore, the assumptions made in a GP regression are directly tied to the chosen mean and covariance functions and the functional family that they are generating. For a GP to accurately capture a process, it is crucial that this functional family is similar to the actual process. Most common choices for the covariance function result in smooth and covariance-stationary GPs with constant wiggliness.

1.3   Generalized additive models

Generalized additive models (GAMs) are a class of semi-parametric models 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 extend regular spline

regression to avoid the knot placement problem by providing enough flexibility to the spline to overfit the data. One approach to this is to use cubic splines, which are piecewise cubic polynomials with knots placed at each data point. Alternatively, thin-plate splines can be used, which entirely avoid knots and instead utilize increasingly wiggly basis functions that are defined over the entire range of the data (Wood, 2003). These basis functions are combined in a regression model similar to polynomial regression. Using as many basis functions as there are data points is analogous to placing a knot at each data point. To prevent the resulting overfitting, smoothing splines add an additional penalty term, similar to those used in a lasso or ridge regression, to control the smooth term's flexibility during the estimation (Gu, 2013; Wahba, 1980). This penalty balances the flexibility and fit of the smooth term, ensuring the model captures the underlying process accurately without introducing unnecessary complexity. The optimal weight of the penalty is typically determined by minimizing a criterion function, such as the generalized cross-validation criterion (Golub & von Matt, 1997; Wood, 2006).

A smoothing spline for a single smooth term β may then be written as

$$\hat{\beta}(t) = \operatorname*{argmin}_{\alpha} P(\mathbf{y} \mid \beta(t)) + \lambda \int (\beta(t)'')^2 dt$$

$$\beta(t) = \sum_{k=1}^{K} \alpha_k R_k(t)$$

(5)

where the first part of the equation describes the likelihood of the data given the smooth term and the second part of the equation corresponds to the penalty term. This illustrates nicely, how the smoothing spline balances data fit, in the form of the likelihood, and the complexity or wigglyness of the estimate, measured by the integrated squared second derivative of the smooth term. Here, λ denotes the weight assigned to the penalty term. Lastly, the smoothing spline is comprised of the spline basis functions $R_k(t)$ and their respective regression coefficients.

GAMs extend on the smoothing spline approach by making it possible to combine multiple smooth terms in an overarching 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 x and summed into a single overall non-linear function f,

which estimates the process.

$$\hat{f}(t) = \underset{\alpha}{\text{argmin}} \, P(\mathbf{y} \,|\, f(t)) + \sum_{j=1}^{J} \lambda_j \int (\beta_j(t)'')^2 dt$$

$$f(t) = \sum_{j=1}^{J} \beta_j x_j \tag{6}$$

$$\beta_j(t) = \sum_{k=1}^{K} \alpha_{j,k} R_{j,k}(t)$$

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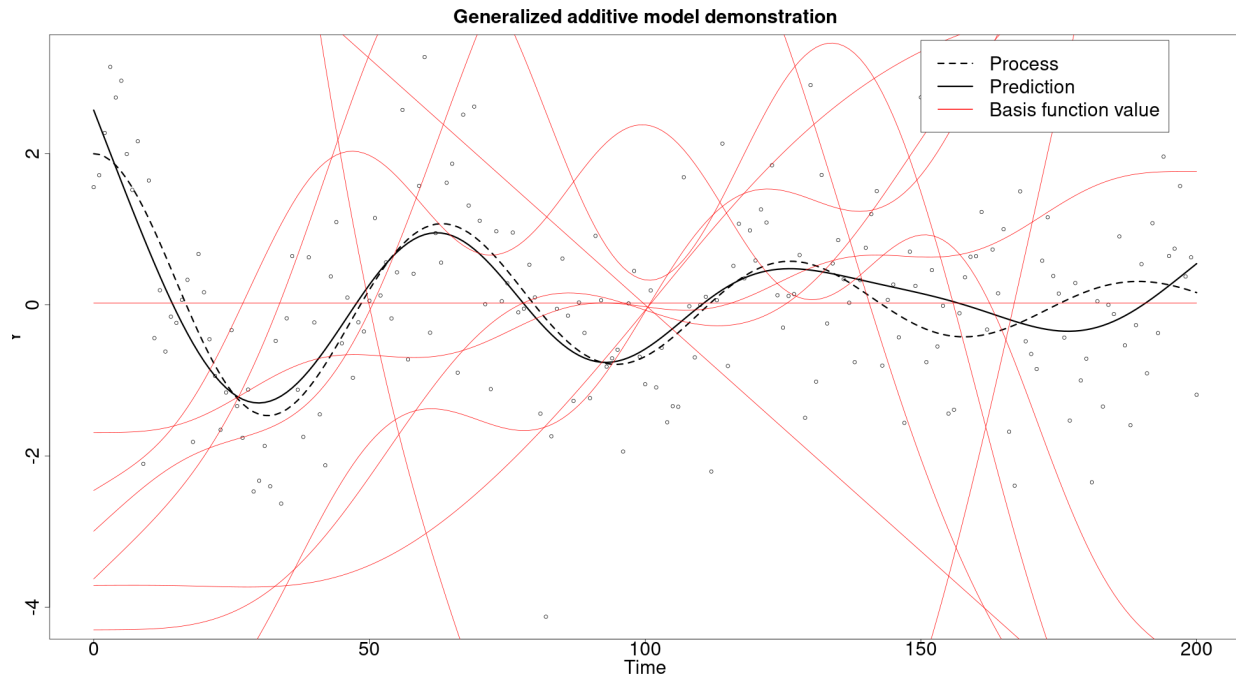
Note. This figure shows how generalized additive models (solid black) estimate the underlying process (dotted black). Here, the predicted values for the process at any point in time correspond to the weighted average of the basis functions (red).

2   Simulation

2.1   Problem

A simulation study was conducted to assess the effectiveness of the introduced methods in recovering different non-linear processes, which may be encountered in EMA research (Figure 1). In addition to that, the introduced methods were also compared to the true, parametric data-generating models, which served as a benchmark for how accurately the processes can be recovered, if the dynamic form of the process is known. 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.

2.2   Design and Hypotheses

For the LPR and the GP, we expect that both method in their 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. Additionally, we expect that the overall sample size will influence the accuracy of the 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 wiggliness 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.

To manipulate the (c) smoothness of the processes a dynamic error component was added

to them. These dynamic errors reflect 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. To add these errors, each process was perturbed by a normally distributed error at each point in time, 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 errors. 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).

Figure 5

One possible realization of the non-linear exemplar processes with dynamic errors
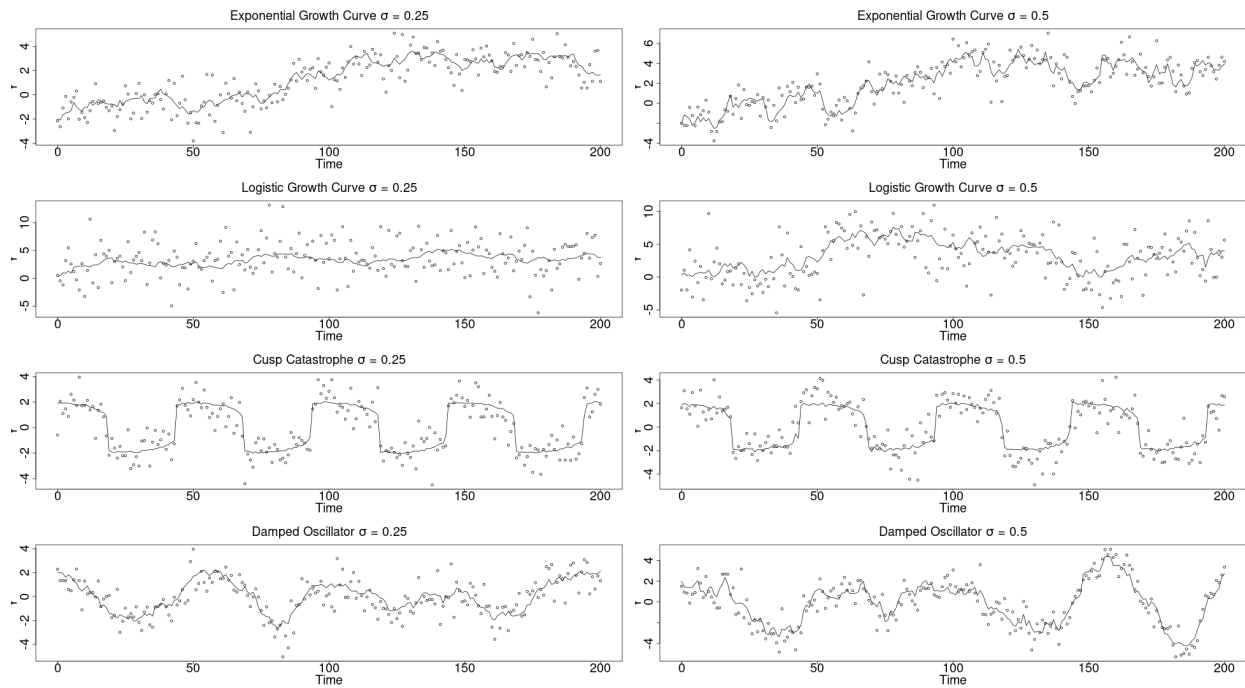
Additionally, the sample size was varied during the simulation by manipulating both (d) the sampling period and (e) the sampling frequency, as these distinct methodological choices are expected to impact the performance of the analysis methods differently. Specifically, for the LPR and the GP, which 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 to this, we expected the GAMs, which incorporate the entire dataset in their estimation, to perform better with a longer sampling period. Since within the simulation there is no inherent scaling to the time variable, we simulated either only the first half of each process or the entire process to represent different sampling periods. For the ease of reading and to correspond to typical EMA conditions, this will be referred to as sampling over either one or two weeks. However, this scaling is arbitrary and could be changed to any other time frame. Lastly, we expected that increasing the sampling frequency will generally improve the inference accuracy across all methods, since there is more information about the latent process available. Relative to the introduced weekly scale, we tested sampling frequencies of three, six, and nine measurements per day, to cover typical EMA sample sizes (Wrzus & Neubauer, 2023).

2.3   Data generation

To simulate data for each process, they must first be represented as parametric generative models that replicate the structure of psychological time series data. In such a time series, any psychological construct follows a (potentially non-linear) function over time, as depicted 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. The observed values 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 simulation, 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 expressed as follows:

$$Y_t = f(t) + \varepsilon_t; \quad \varepsilon_t \sim N(0, \sigma_\varepsilon^2) \tag{7}$$

In this model, $f(t)$ represents a potentially non-linear latent process (like the ones presented in Figure 1, although other processes are also possible), and ε_t represents the time-point-specific measurement error. The most direct way to define a parametric model for $f(t)$ is as a (non-linear) function of time. Unfortunately, many processes have functional forms that are too complex for this representation, and this approach does not allow for the modeling of dynamic errors.

Instead, each process was represented as a generative stochastic differential equation model. These dynamic models describe the relationship between the process's current value and its instantaneous rate of change. By combining these models with information about the initial state of the construct, it becomes possible to infer the entire trajectory of the process. Differential equations are a widely used class of dynamic models because they can capture complex processes in much simpler parametric forms. Nevertheless, they do require considerable theoretical knowledge about the process for applied modelling. For instance, a differential equation model representing the introduced logistic growth process can be expressed as follows:

$$\frac{dy}{dt} = ry(1 - \frac{y}{k}) \tag{8}$$

This model relates the rate of change of y to its current value and to how far away the current value is from the asymptote k through a growth rate constant r.

There are different ways to add dynamic errors to differential equation models. However, the most common modelling choice is to add an additive Wiener process to the deterministic model.

$$dy = ry(1 - \frac{y}{k})dt + \sigma dW_t \tag{9}$$

The Wiener process is a continuous non-differentiable stochastic process, which describes normally distributed dynamic errors over any given discrete time interval. These errors have a mean of zero and a variance dependent on the length of the time interval and σ, making them optimal for this simulation. Importantly, these dynamic errors continuously influence the rate of change of the process and are propagated forward in time through the deterministic dynamics of the model.

All four processes were modeled as stochastic differential equations by substituting their respective deterministic dynamics into Equation 8, as detailed in Appendix B. Latent process data were simulated using the Euler-Maruyama method, which approximates stochastic differential equations with an arbitrarily high accuracy by linearizing them over small discrete time intervals. The resulting high-resolution data were then subsampled to achieve the desired sampling frequency. Finally, measurement errors were added to the latent process data at each time point from a standard normal distribution, generating the final sets of observations

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 (Siepe et al., 2023). 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.

2.4   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, the parametric differential equation models corresponding to the true data-generating models were 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.

2.5   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.

2.5.1   *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, differences in MSE and GCV values across simulation conditions and analysis methods were analyzed using ANOVAs. To simultaneously identify factors that likely do and do not affect the MSE and GCV values, an exhaustive ANOVA model search was performed. In this search, AIC and BIC model weights were used to select a final model based on a 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 (Blanca et al., 2017), and any potential violations of homoscedasticity were addressed using heteroscedasticity-consistent standard errors. However, if both assumptions were severely violated, the results were instead presented descriptively, using means and standard errors for the respective conditions.

2.5.2   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.

2.6   Results (Template)

2.6.1   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.

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

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 lower MSE and GCV by all methods.

Figure 6

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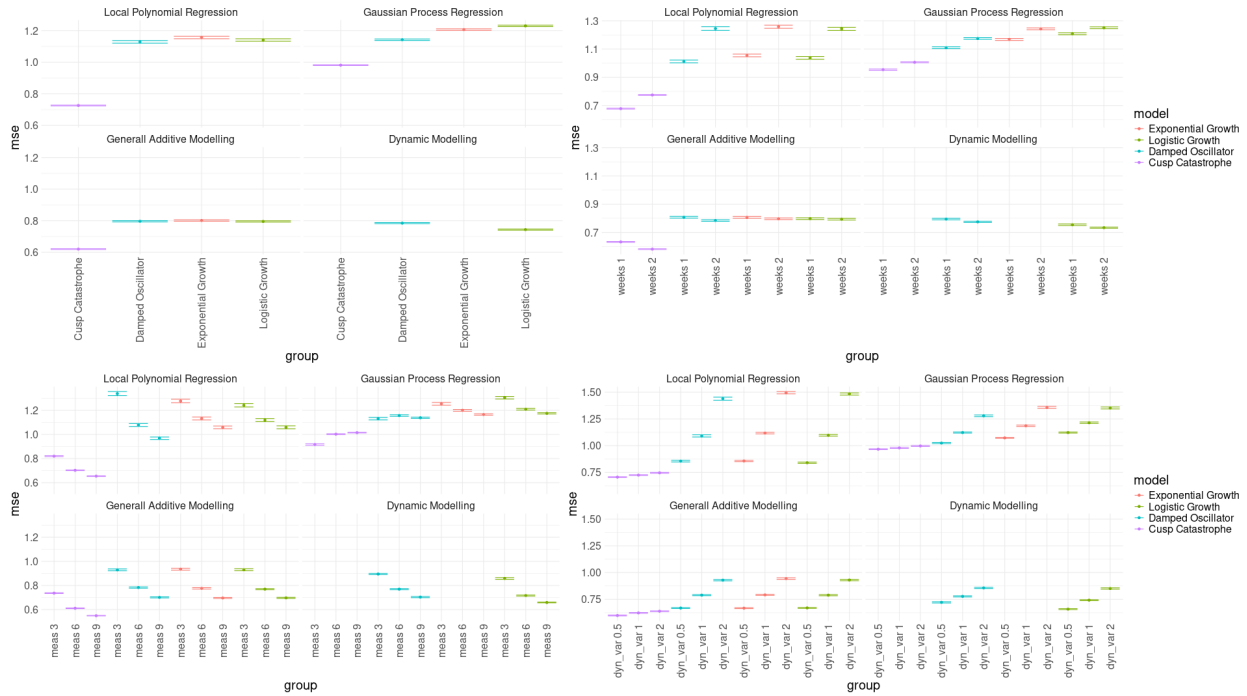
Note. Panel (a) shows the effect of the analysis methods for each latent process. The other three panels show the effects of measurement period (b), measurement frequency (c), and dynamic error variance (d) for each analysis method and latent process.

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.

2.6.2   *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
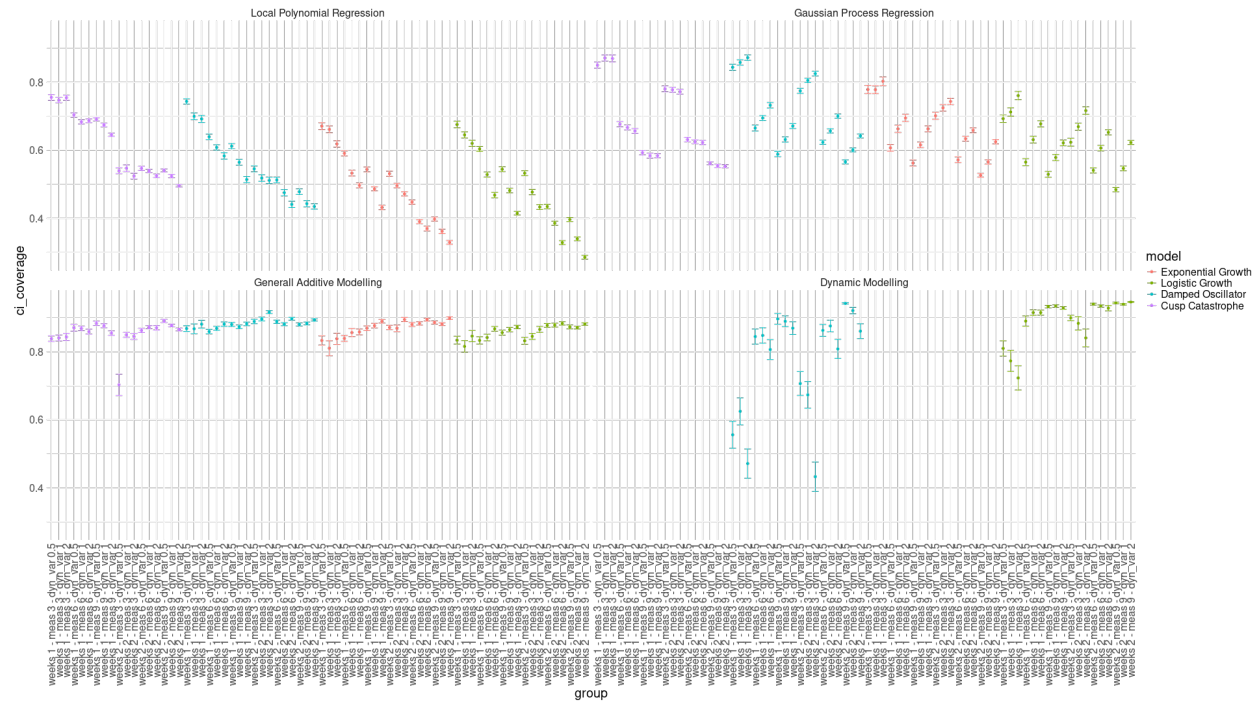
2.7   Conclusion

This simulation showed that the GAMs inferred the latent processes more accurately than the LPR and GP across all conditions, as indicated by the MSE, GCV, and confidence interval coverage. Furthermore, the GAMs were nearly as accurate as the true data-generating parametric models in the simulation. This result was unexpected, as we anticipated that the smooth estimates produced by GAMs might not be well-suited for inferring the rough (i.e., non-differentiable) processes in the simulation. However, the LPR and the default GP configurations that were used also produce smooth estimates, leading to similar misspecifications as the GAMs. This suggests that the accuracy of the GAMs might have been least affected by the roughness of the latent process. However, another factor potentially influencing the accuracy of the LPR and GP is their assumption of constant wiggliness, as defined by their respective bandwidth and lengthscale parameters, which was violated in the tested processes.

It is important to note that the observed results may be attributable to the specific configurations used for each method rather than the methods themselves. These configurations were chosen to reflect how each method is commonly applied in its most basic form and not to optimally infer the types of processes simulated. Consequently, different configurations and extensions might improve the performance of the LPR and GP. Importantly, it has been shown that the standard GAM formulation is a (potentially improper) GP (Wahba, 1978), and the specific GAM used in this simulation can be expressed as a GP with a linear mean and a non-stationary covariance function (Rasmussen & Williams, 2006). This implies that there are certain GP formulations which perform at least as well as the GAMs did, although these formulations deviate from how GPs are most commonly applied. Regarding the LPR, several optimality results have been obtained for differentiable processes, indicating that LPR should be at least as accurate as the GAMs and GPs for these processes (Fan et al., 1997). It is however unclear whether these results generalize to the non-differentiable processes that we suspect can be found most often in psychology.

Contrary to our expectations, the simulation indicated that the cusp catastrophe process

was inferred most accurately by the LPR, GP, and GAM. We had anticipated that the smooth,

continuous estimates produced by these methods would struggle to adapt to the apparent jumps

exhibited by this process. However, this effect seems to have been mitigated by the cusp

catastrophes strong resilience to external perturbations. This property is highlighted in Figure 5,

where dynamic errors with the same variance have been applied to all four processes. Despite the

perturbations having an equal variance, the cusp-catastrophe model appears to be the least affected

and even closely resembles the unperturbed process (Figure 1). Further evidence of this can be

seen in the simulation, where the effect of increasing the dynamic error variance was weakest for

the cusp process. Due to this, the simulation was rerun with considerably smaller dynamic error

variances, and under these conditions, the cusp model was indeed inferred with the least accuracy.

The results indicate that measuring at a higher frequency increased the inference accuracy

of all considered methods. Therefore, it is generally advantageous from a statistical point of view

to measure as frequently as possible. However, in practice, this must be balanced against

considerations such as participant burden and fatigue, which can adversely affect data quality if

measurements are taken too often. Similarly, when selecting the sampling period, it is essential to

use domain knowledge about the scale of the underlying dynamics to ensure that the

measurements capture sufficient variation in the latent process. Beyond this, the simulation

showed that extending the sampling period improved the inference accuracy of the GAMs and

parametric models but may decrease the accuracy of the LPR and GP. This reduction in accuracy

could however possibly be mitigated by using extensions for a variable bandwidth, polynomial

degree, or lengthscale respectively. Lastly, the simulation revealed that larger dynamic error

variances decreased the accuracy of all methods. Therefore, reducing the magnitude of dynamic

errors is advisable in practice. This could, for example, be achieved by measuring context

variables and other sources of perturbations and incorporating them into the model.

3 An Empirical Example

In the following, the three 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.

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.

3.1 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. For a more thorough sample and data description,

see Heininga et al. (2019). The final data set used for this analysis contained 77 participants in the clinical sample and 40 participants in the control sample, who were matched to the clinical sample by gender and age, resulting in a total sample size of 117 [2].

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.

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 (TSB RP FT16).

3.2   Analysis Plan

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

---

[2] The original published data set contained one additional participant who was removed for this analysis since they had a depression score of zero across all assessments.

the most accurate interpolations. Lastly, the mean squared error was calculated for each method and data set to estimate the expected measurement error.

3.3   Results (Template)

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*: $1.76 * 10^{10}$). These three measures of wiggliness 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 wiggliness 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 wiggliness 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 behaviors 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 behavior 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.

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

Figure 8

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

GPs, and GAMs
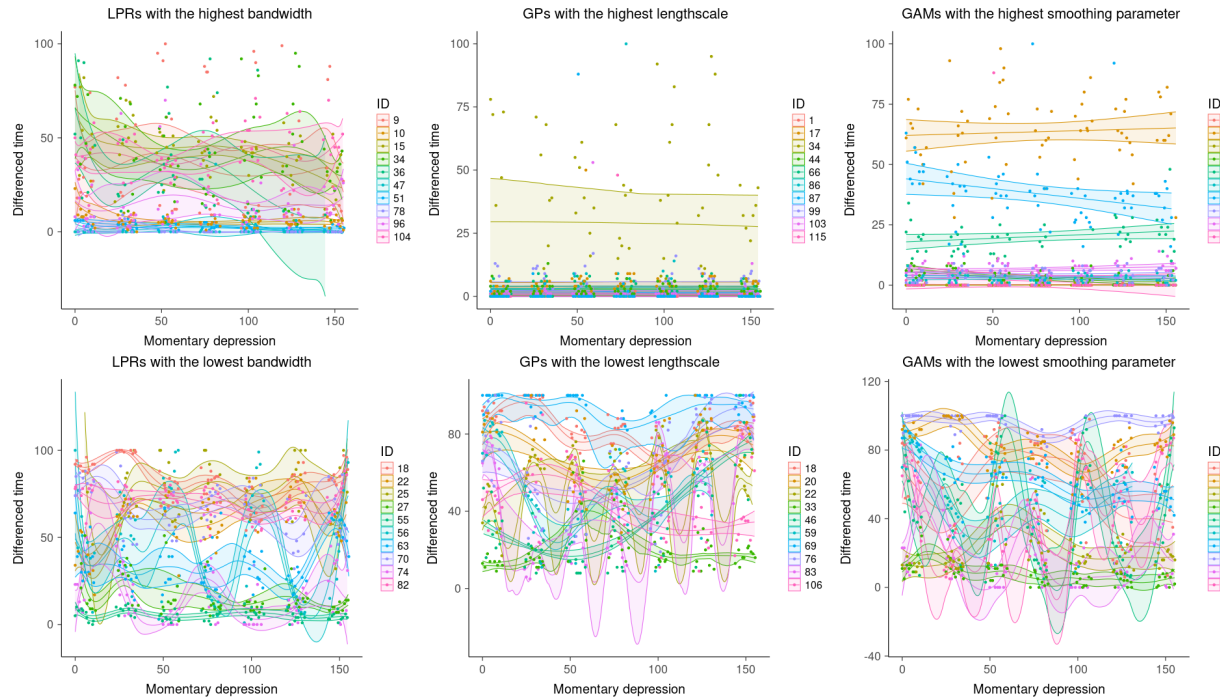
GAMs was 125.29 (*IQR*: 201.04), for the GP it was 248.27 (*IQR*: 578.00), and for the LPR it was 131.57 (*IQR*: 213.73). In addition to this, the mean squared error was calculated between the predictions generated by each method and the data. Here the median MSE of the GAM was 114.04 (*IQR*: 171.59), 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.

4   Conclusion

WIP

5   Discussion

This article has several important limitations. Since, only a limited selection of processes was used it is possible that the presented results may not generalize to other processes. However, the used processes already constitute violations to the smoothness assumption made by the LPR, GP, and GAM to which these methods demonstrated some robustness. In addition to this, as explained earlier, using different configurations for the LPR, GP, and GAM is likely to change the presented results. Further, we focused on introducing these methods by inferring time dependent non-linear processes from univariate, single subject data with independent normally distributed measurement errors. This is a statistically idealized setting, which does not address many of the goals and challenges researchers are facing when working with ILD. Frequently, ILD contains measurements for many individuals on several psychological constructs. Classically, this enables researchers to study how these constructs vary and interact over time and how these dynamics differ between people. In addition to that, each construct is frequently measured using multiple indicators with ordinal measurement errors, which are modelled using different psychometric models (e.g., factor models, item response models).

There are fortunately many ways in which the presented methods can be adapted to these more complex data structures. The GP and GAM can be easily adapted to work with multilevel data without substantially extending the statistical theory underlying both methods. For the GAM this extension is already implemented in some software. Another approach may be to study between person differences in the latent processes using functional data analysis. In this analysis the individual latent processes are first estimated using one of the presented data driven techniques. Subsequently, the inferred processes are treated as function valued data, which can be analyzed to find for example group differences in a functional ANOVA (Kaufman & Sain, 2010) or to find the functions which account for the maximum between person variation in a functional principal component analysis (Aue et al., 2015).

Similarly, the GP and GAM are already equipped to estimate latent variables in general and they are not limited to single indicators. Therefore, it is possible to extend them to

incorporate typical psychometric measurement models, to accurately capture more complex measurement error distributions. Unfortunately, the software implementations of this may be computationally very intensive and are currently only available in a very limited form (Clark & Wells, 2023). The parametric models introduced already include a factor model for the observed variables, which can incorporate multiple indicators. In addition to this, these indicators can also be non-normally distributed.

Lastly, there are many ways in which the presented methods can be used to study multivariate data. This is because, even though all methods were used to infer time dependent non-linear processes in this paper, they can in theory be used to estimate many smooth and continuous functions (and even non-smooth and non-continuous functions to the degree presented in this paper). This makes it possible to for example infer non-linear cross- and autoregressive relationships from data in discrete time (Eleftheriadis et al., 2017; Rasmussen & Williams, 2006; Wood, 2006) and non-linear differential equation models in continuous time (Yildiz et al., 2018). For the presented processes in particular this should even be more appropriate, since they were generated using differential equation models. These models also present the exciting possibility to combine partial parametric models with non-linear data driven functions. Another possibility is to infer an unobserved input variable to a parametric differential equation model (Álvarez et al., 2009; Nayek et al., 2019). Lastly, it is possible to infer non-linear seasonality and cyclic components on indicator variables (Clark & Wells, 2023).

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

Álvarez, M., Luengo, D., & Lawrence, N. D. (2009). Latent Force Models [ISSN: 1938-7228]. *Proceedings of the Twelth International Conference on Artificial Intelligence and Statistics*, 9–16. Retrieved September 11, 2023, from https://proceedings.mlr.press/v5/alvarez09a.html

Aue, A., Norinho, D. D., & Hörmann, S. (2015). On the Prediction of Stationary Functional Time Series. *Journal of the American Statistical Association*, *110*(509), 378–392. https://doi.org/10.1080/01621459.2014.909317

Betancourt, M. (2020). Robust Gaussian Process Modeling. Retrieved June 9, 2023, from https://betanalpha.github.io/assets/case_studies/gaussian_processes.html

Blanca, M. J., Alarcón, R., & Arnau, J. (2017). Non-normal data: Is ANOVA still a valid option? *Psicothema*, (29.4), 552–557. https://doi.org/10.7334/psicothema2016.383

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

Borsboom, D., Van Der Maas, H. L. J., Dalege, J., Kievit, R. A., & Haig, B. D. (2021). Theory Construction Methodology: A Practical Framework for Building Theories in Psychology. *Perspectives on Psychological Science*, *16*(4), 756–766. https://doi.org/10.1177/1745691620969647

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., 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

Clark, N. J., & Wells, K. (2023). Dynamic generalised additive models (<span

style="font-variant:small-caps;">DGAMs) for forecasting discrete ecological

time series. *Methods in Ecology and Evolution*, *14*(3), 771–784.

https://doi.org/10.1111/2041-210X.13974

Cui, J., Hasselman, F., & Lichtwarck-Aschoff, A. (2023). Unlocking nonlinear dynamics and multistability from intensive longitudinal data: A novel method. https://doi.org/10.31234/osf.io/wjzg2

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.

Eleftheriadis, S., Nicholson, T. F. W., Deisenroth, M. P., & Hensman, J. (2017). Identification of Gaussian Process State Space Models [Publisher: arXiv Version Number: 2]. https://doi.org/10.48550/ARXIV.1705.10888

Fan, J., & Gijbels, I. (2018). *Local Polynomial Modelling and Its Applications* (1st ed.). Routledge. https://doi.org/10.1201/9780203748725

Fan, J., Gasser, T., Gijbels, I., Brockmann, M., & Engel, J. (1997). Local Polynomial Regression: Optimal Kernels and Asymptotic Minimax Efficiency. *Annals of the Institute of Statistical Mathematics*, *49*(1), 79–99. https://doi.org/10.1023/A:1003162622169

Fan, J., & Gijbels, I. (1995a). 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

Fan, J., & Gijbels, I. (1995b). Data-Driven Bandwidth Selection in Local Polynomial Fitting: Variable Bandwidth and Spatial Adaptation. *Journal of the Royal Statistical Society Series*

B: Statistical Methodology, *57*(2), 371–394.

https://doi.org/10.1111/j.2517-6161.1995.tb02034.x

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., & Bronder, 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

Humberg, S., Grund, S., & Nestler, S. (2024). Estimating nonlinear effects of random slopes: A comparison of multilevel structural equation modeling with a two-step, a single-indicator, and a plausible values approach. *Behavior Research Methods*, *56*(7), 7912–7938. https://doi.org/10.3758/s13428-024-02462-9

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).

Kaufman, C. G., & Sain, S. R. (2010). Bayesian functional {ANOVA} modeling using Gaussian process prior distributions. *Bayesian Analysis*, *5*(1). https://doi.org/10.1214/10-BA505

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

Kruschke, J. K. (2011). *Doing bayesian data analysis: A tutorial with R and BUGS* [OCLC: ocn653121532]. Academic Press.

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

Nayek, R., Chakraborty, S., & Narasimhan, S. (2019). A Gaussian process latent force model for joint input-state estimation in linear structural systems. *Mechanical Systems and Signal Processing*, *128*, 497–530. https://doi.org/10.1016/j.ymssp.2019.03.048

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

Oberauer, K., & Lewandowsky, S. (2019). Addressing the theory crisis in psychology. *Psychonomic Bulletin & Review*, *26*(5), 1596–1618. https://doi.org/10.3758/s13423-019-01645-2

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., Ebden, 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

Siepe, B. S., Bartoš, F., Morris, T. P., Boulesteix, A.-L., Heck, D. W., & Pawel, S. (2023).

Simulation Studies for Methodological Research in Psychology: A Standardized Template

for Planning, Preregistration, and Reporting. https://doi.org/10.31234/osf.io/ufgy6

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.

Wahba, G. (1978). Improper Priors, Spline Smoothing and the Problem of Guarding Against

Model Errors in Regression. *Journal of the Royal Statistical Society Series B: Statistical*

Methodology, *40*(3), 364–372. https://doi.org/10.1111/j.2517-6161.1978.tb01050.x

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

Yildiz, C., Heinonen, M., Intosalmi, J., Mannerström, H., & Lähdesmäki, H. (2018). Learning Stochastic Differential Equations With Gaussian Processes Without Gradient Matching [arXiv:1807.05748 [cs, stat]]. Retrieved June 5, 2024, from http://arxiv.org/abs/1807.05748

Comment: The accepted version of the paper to be presented in 2018 IEEE International Workshop on Machine Learning for Signal Processing