

**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](mailto: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**

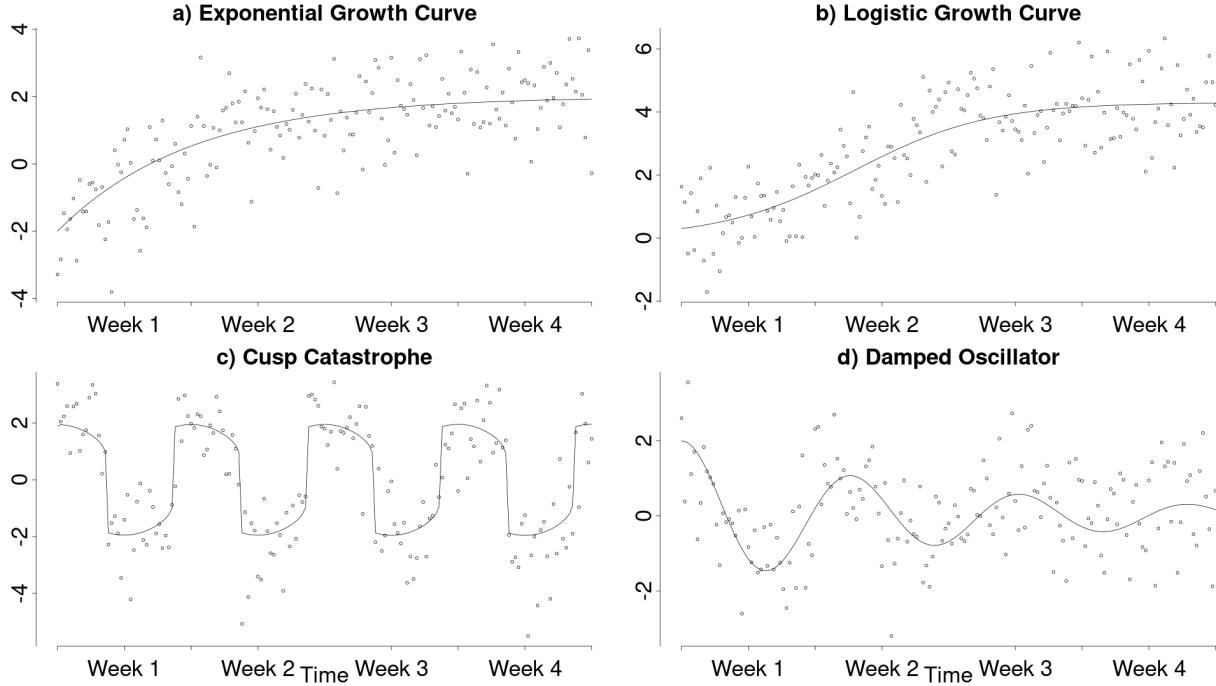
Psychological constructs are increasingly understood as components of complex dynamic systems (Nesselroade & Ram, 2004; Wang et al., 2012). This perspective emphasizes that these constructs fluctuate over time and within individuals. To study these variations and the underlying processes, researchers 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 in 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 a (person-specific) asymptote, which reflects the individual's maximum ability. Additional examples of asymptotic growth over 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 switching between seemingly distinct states that differ, for instance, in their 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

**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.

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 degree of regulation 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 degree of regulation changes disproportionately with larger distances. 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 to result in specific hypotheses (Oberauer & Lewandowsky, 2019), such as, for example, about the form of specific non-linear dynamics. ILD studies could potentially help 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 dynamic models, such as differential equation (Boker, 2012) or state-space models (Durbin & Koopman, 2012), that describe how a given process changes over time. However, in order to develop these types of theories it is first necessary to identify empirical phenomena in ILD, which can be thought of "stable and general features of the world that scientists seek to explain" (Borsboom et al., 2021). Examples of such phenomena could be the described state switching or regulatory oscillations, if they generalize beyond individual data sets and contexts. 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. One reason for this is 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 the absence of adequate statistical methods, non-linear trends in psychology are often addressed through polynomial or piecewise 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 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 complex trajectory,

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), Harrell (2001), and Jianan et al. (2023)).

An alternative approach is piecewise 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 quickly 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 like 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, it is difficult for psychological researchers 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 ideal smooth processes, as depicted in Figure 1, for which many of these methods were originally developed, are unlikely to occur in psychological research. This is because psychological constructs are (especially in EMA research) measured in an environment in which external influences constantly affect and perturb the construct. These constant perturbations can quickly result in the non-smooth or rough processes that are typically seen in psychological time series. It is for instance quite easy to imagine that someone's happiness may not decrease smooth and gradually, after stubbing their toe.

To address this important gap, this article reviews three advanced non-linear analysis

methods and evaluates their applicability to typical ILD scenarios (Section 1). The methods reviewed in this article are different semi- and non-parametric regression techniques, available in the open-source software R (R Core Team, 2024), which are able to infer non-linear functions from data while accommodating varying degrees of prior knowledge. In a simulation study, we compare how well each method recovers different non-linear processes under common ILD conditions (Section 2). Lastly, we demonstrate how the best-performing method can be applied to analyze an existing dataset (Section 3). 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

In the following paragraphs the three selected semi- and non-parametric regression techniques will be introduced.

### 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 to the data around it. Since 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 time point.

Formally this procedure can be expressed using the following set of equations:

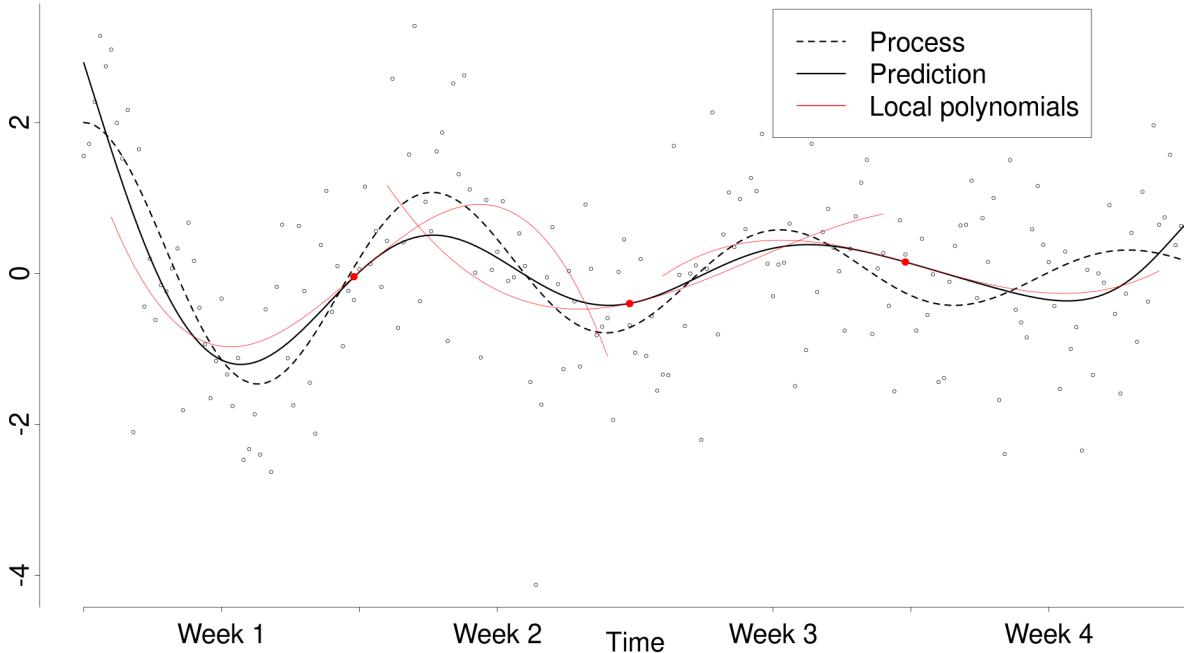
$$\begin{aligned}
 y_t &= f(t) + \varepsilon_t \\
 \mathbf{X} &= \begin{bmatrix} 1 & (t_1 - t^*)^1 & \dots & (t_1 - t^*)^p \\ \vdots & \vdots & \ddots & \vdots \\ 1 & (t_n - t^*)^1 & \dots & (t_n - t^*)^p \end{bmatrix} \\
 \mathbf{W} &= \begin{bmatrix} w_{1,1} \\ \ddots \\ w_{n,n} \end{bmatrix} \\
 \hat{f}(t^*) &= Intercept((\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W} \mathbf{y})
 \end{aligned} \tag{1}$$

where a univariate process  $f$  is inferred at the chosen time point  $t^*$ . Then  $\mathbf{X}$  is the model matrix of a multiple linear regression, such that the columns correspond to polynomial transformations up to degree  $p$  of the data centered around  $t^*$ <sup>1</sup>. Specifically,  $\mathbf{X}$  contains polynomial transformations of centered time (i.e., the first column is filled with ones, the second column contains the centered time points, the third column the centered time points squared, and so on). Further,  $\mathbf{W}$  is a diagonal matrix containing the weights associated with each datum. The last equation is a normal equation solving for the coefficients of a weighted multiple linear regression. The intercept of this regression gives the estimated value of the LPR at  $t^*$ .

The same procedure can be repeated to find the value that the LPR predicts for the process at a different time point by centering the data around the new point of interest and reestimating the local polynomial. 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 damped oscillator example process introduced in Figure 1 (d). This figure shows three examples of the 200 local cubic regressions that were evaluated to draw the overarching LPR for this process. To highlight the local nature of the polynomials, each cubic function was shortened to a

---

<sup>1</sup> Note that the polynomial terms in this model matrix are derived from a Taylor series approximation around  $t^*$  with  $p$  derivative terms.

**Figure 2***Demonstration of a local polynomial regression*

*Note.* This figure shows how LPR (solid black) estimates the underlying process (dotted black). Here, three examples of the 200 local cubic polynomials that provide the predicted values for the LPR are shown in red. The cubic polynomials were shortened to a region where the kernel assigned weights larger than .02, to highlight their localized nature.

region where the kernel assigned weights larger than .02.

When fitting an LPR, several decisions must be made regarding the optimal weighting of the data and the degree of the local polynomial. The data weighting in an LPR is achieved through a kernel function of the form

$$w_{i,i} = K\left(\frac{t_i - t^*}{h}\right) \quad (2)$$

which, in this case, is a mathematical equation that determines the influence of different data points during the local polynomial estimation. Because these kernels are usually symmetric around the origin, they assign weights to data points depending on how far away they are from the point of interest  $t^*$ . Common choices for kernels include the Gaussian and Epanechnikov kernels. Both assign higher weights to data points in closer proximity. However, while the Gaussian kernel

assigns small weights to all distant data points, the Epanechnikov kernel assigns zero weights beyond a certain distance. Additionally, the Epanechnikov kernel has also been shown to be optimal in many aspects, particularly when estimating points where the entire width of the kernel lies within the data range (Fan et al., 1997). Each kernel is further defined by a bandwidth parameter  $h$ , which determines the kernel's width and effectively controls the influence of more distant data points. The bandwidth parameter represents the wigginess 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). Most standard software packages offer automated default procedures for this task. However, the optimization criterion may also be selected based on specific research objectives (e.g., optimizing a cross-validation error may be most attractive, if the primary research interest is out of sample prediction).

In practice researchers also need to choose the degree of the local polynomials, which additionally reflects an assumption about how smooth the underlying process is. Specifically, for a first-degree LPR, the process should not exhibit any corners, discontinuities, or vertical sections. This ensures that the processes rate of change (i.e., derivative), approximated by the first-order polynomial term, is well behaved. Higher-order local polynomials require this smoothness for increasingly complex rates of change. For instance, a second-degree LPR requires that the rate of change itself is smooth, in turn ensuring that its rate of change is well behaved. This property should hold for all  $p$  rates of change of a process when using an LPR with  $p$  degrees (under Taylor's theorem). Typically, the degree of the local polynomials is chosen to be 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 second assumption made by the LPR is that the process has constant wigginess with respect to the chosen polynomial degree. This assumption is made, because a single bandwidth parameter is used to capture the process at any point in time. For example, a linear LPR, estimates the linear polynomials at any point in time using weights derived from a kernel with the same

bandwidth. This bandwidth should then be optimal across the entire process. However, this assumption is less strict for higher order polynomials, due to their increased flexibility and may be relaxed more by using a time-varying bandwidth (Fan & Gijbels, 1995b) or polynomial degree (Fan & Gijbels, 1995a), but this extension is beyond the scope of this paper.

Lastly, it is noteworthy, that unless the underlying process follows a polynomial trajectory of at most degree  $p$  the approximation with local polynomials is biased. For instance, whereas a process that follows a quadratic trajectory can be accurately inferred at any point in time by a local quadratic regression (and all higher-order LPRs), a process that follows an exponential trajectory cannot be inferred with perfect accuracy by any LPR with a finite degree. Instead, there will be a small bias in the estimate, that decreases for larger local polynomial degrees. However, this bias is usually negligible and there are methods available to correct for it (Calonico et al., 2019).

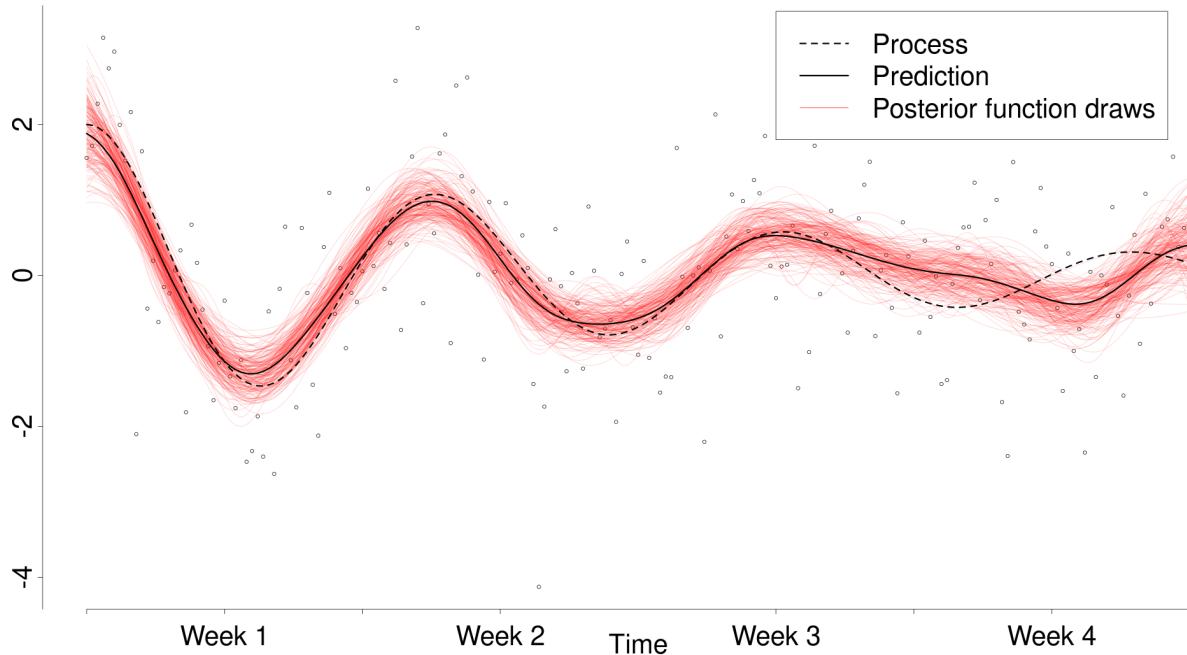
## 1.2 Gaussian process regression

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 distribution) that specify the likelihood of single values, Gaussian processes determine how likely entire (non-linear) functions are. A GP is defined indirectly, such that, if the functions it describes are evaluated at any finite set of time points, the resulting sample of function values would follow a multivariate normal distribution. 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.

$$P(f | \mathbf{y}) \propto P(\mathbf{y} | f)P(f) \quad (3)$$

This posterior distribution represents an updated belief about which functions describe the latent

**Figure 3**  
*Demonstration of a Gaussian process regression*



*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.

process well (Kruschke, 2011), making it possible to draw inferences about the process. 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, such that the pointwise average of these functions provides 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. This 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 the function values at different time points depending on their

distance.

$$\text{cov}(t_i, t_j) = k(|t_i - t_j|) \quad (4)$$

The types of processes that can be captured by a GP regression primarily depend on the chosen form of the covariance kernel and, to a lesser extent, the mean function. This is because the covariance kernel largely dictates the behavior of the functions described by the GP. The default covariance kernel in most standard software is the squared exponential, which produces a Gaussian process that is covariance stationary and very smooth. Therefore, using the squared exponential covariance kernel imposes a stricter smoothness assumption about the underlying process than the smoothness assumption introduced for LPR. However, many other covariance kernels are available, each resulting in GPs with different behaviors and assumptions about the underlying process. A notable example is the Matérn class of kernels, which relaxes the strict smoothness assumption made by the squared exponential kernel. This makes it possible to model rougher processes. One particularly interesting Matérn kernel gives rise to the covariance of the continuous-time autoregressive process.

Lastly, the mean and covariance functions also include parameters, called hyperparameters, which can be estimated by assigning appropriate priors to them. Most common covariance kernels (such as the introduced squared exponential or the Matérn class kernels) build on a characteristic lengthscale and a marginal standard deviation parameter. The characteristic lengthscale effectively determines the wigginess of the estimated process, by quantifying how quickly the covariance decreases with increasing distances between time points. 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, for example, model comparison. Most standard software provides default GP priors. However, it is also possible to choose or construct a specific GP prior to fit a specific context or research objective.

Whereas LPR is a mainly data driven procedure GP regression is more model based.

Specifically, the form of the GP prior generates a family of functions to which the process is assumed to belong. This makes it possible to include theoretical domain knowledge and specific hypotheses of interest in the kinds of function that are modelled by a GP. If one for example expects that a non-linear process has a general upwards trend, it would be possible to model a linear mean function with non-linear deviations that are captured by a GP. Another advantage of the GP regression is that the Bayesian estimation provides a natural approach to uncertainty quantification, especially for the wigglyness of the process. While there is no direct way of quantifying the uncertainty associated with the bandwidth of the LPR, the uncertainty of the lengthscale of the GP is captured in its posterior distribution.

### 1.3 Generalized additive models

Generalized additive models (GAM) are a class of semi-parametric models that build on so called 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 address the knot placement problem. Instead of placing knots at meaningful locations, smoothing splines use a very large number of knots to guarantee overfitting the data. One approach to this is to use cubic splines, which are piecewise cubic polynomials. By placing a knot at each data point these splines perfectly interpolate the data. 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 then combined in a regression model similar to polynomial regression (i.e., each basis function functions as predictors in the regression equation). Using as many basis functions as 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 how closely the smooth term fits the data during the estimation (Gu, 2013; Wahba, 1980). This penalty balances the complexity and fit of the smooth term, ensuring the model captures the underlying process accurately without overfitting.

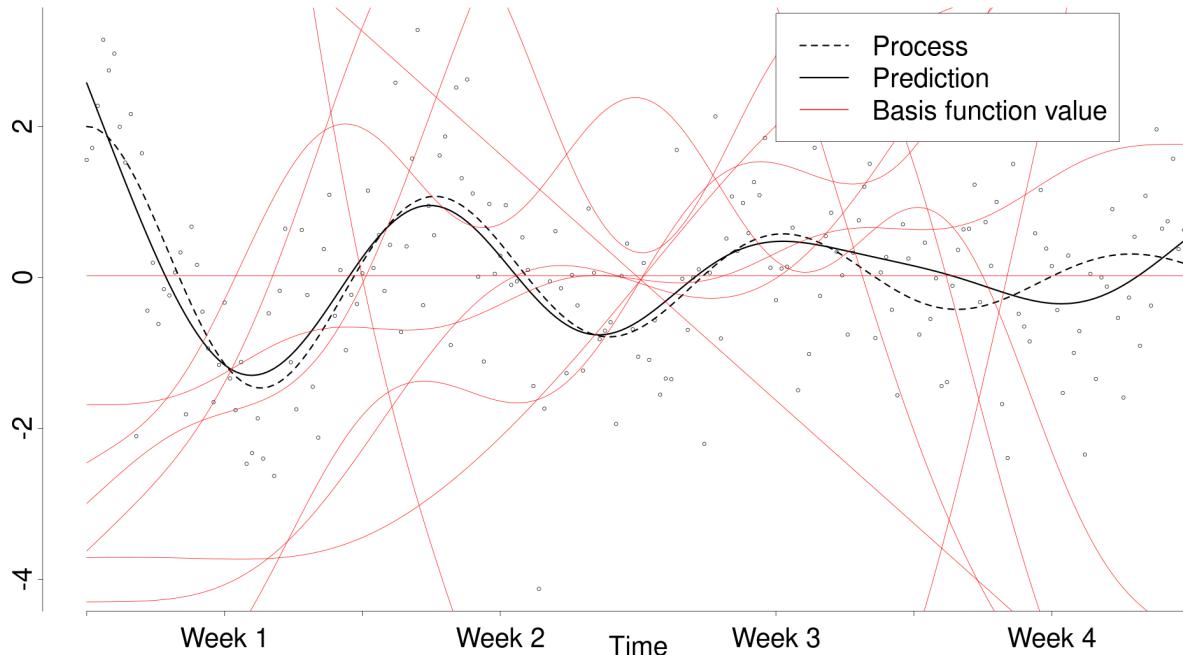
The optimal weight of the penalty is typically determined by minimizing a criterion function, such as the generalized cross-validation criterion or through likelihood maximization, which is performed automatically in standard software (Golub & von Matt, 1997; Wood, 2006). While researchers still need to select the spline basis and the number of basis functions, thin-plate splines are optimal in many aspects and most common spline bases do not result in considerably different estimates. The number of basis functions should be chosen at least large enough to enable sufficient model fit. This selection is also automated in most standard software (Wood, 2011)

A smoothing spline for a single smooth term  $\beta$  may then be written as

$$\begin{aligned}\hat{\beta}(t) &= \underset{\alpha}{\operatorname{argmin}} P(\mathbf{y} | \beta(t)) + \lambda \int (\beta(t)')^2 dt \\ \beta(t) &= \sum_{k=1}^K \alpha_k R_k(t)\end{aligned}\tag{5}$$

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, how the smoothing spline balances data fit, in the form of the likelihood, and the complexity or wiggleness of the estimate, measured by the overall squared curvature of the smooth term. Here,  $\lambda$  denotes the weight assigned to the penalty term that is optimized over. Lastly, the smoothing spline is comprised of spline basis functions  $R_k(t)$ , such as the introduced cubic or thin-plate splines and their respective regression coefficients  $\alpha_k$ .

GAMs extend on the smoothing spline approach further by making it possible to combine multiple smooth terms (of potentially different input variables) in an overarching additive regression model, where each smooth term essentially functions as a (input-varying) coefficient within a regular regression analysis. 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

**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).

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 construction with a single smooth term for time, fitted to the example process. The ten thin-plate spline basis functions that were used to construct this smooth term are shown in red.

Similar to GPs, GAMs constitute a more model based approach to inferring non-linearity. Indeed, GAMs and GP regression are closely related modelling frameworks, both of which can combine partial theories with data driven non-linear smooth terms. In this way, GAMs can also, for instance, model a linear function with non-linear deviations, which are captured by a smooth term. GAMs also provide an estimate of the wiggliness of the process through the effective degrees of freedom (EDF) of the model. These are a measure of the models effective complexity, where an EDF of one corresponds to a linear model. Unfortunately, GAMs do not provide a measure of the uncertainty of the EDF. While all three introduced methods provide some

information about the wiggliness of the underlying process, each method measures different aspects of the wiggleness and the interpretation of the wiggleness estimates of each method depends on the chosen configurations. For instance, the interpretation of the bandwidth of an LPR changes depending on the selected polynomial degree and kernel. Similarly, the interpretation of the lengthscale of a GP regression changes depending on the chosen mean function and covariance kernel. This makes it almost impossible to compare the wiggleness estimates between the presented methods and even between different configurations of the same method. However in contrast to the LPR and the GP, the GAMs do not assume that the process has constant wiggleness (with respect to how each method quantifies wigglyness).

Table 1 summarizes the similarities and differences between the three introduced methods.

## 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 this simulation, the three methods were not only compared against each other, but also to a polynomial regression model (which is the current most used method to model non-linear trends in psychology) and to parametric models that accurately specify the non-linear process. These (data generating) parametric models were added to serve as a benchmark for non-linear process recovery. To apply the introduced 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

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

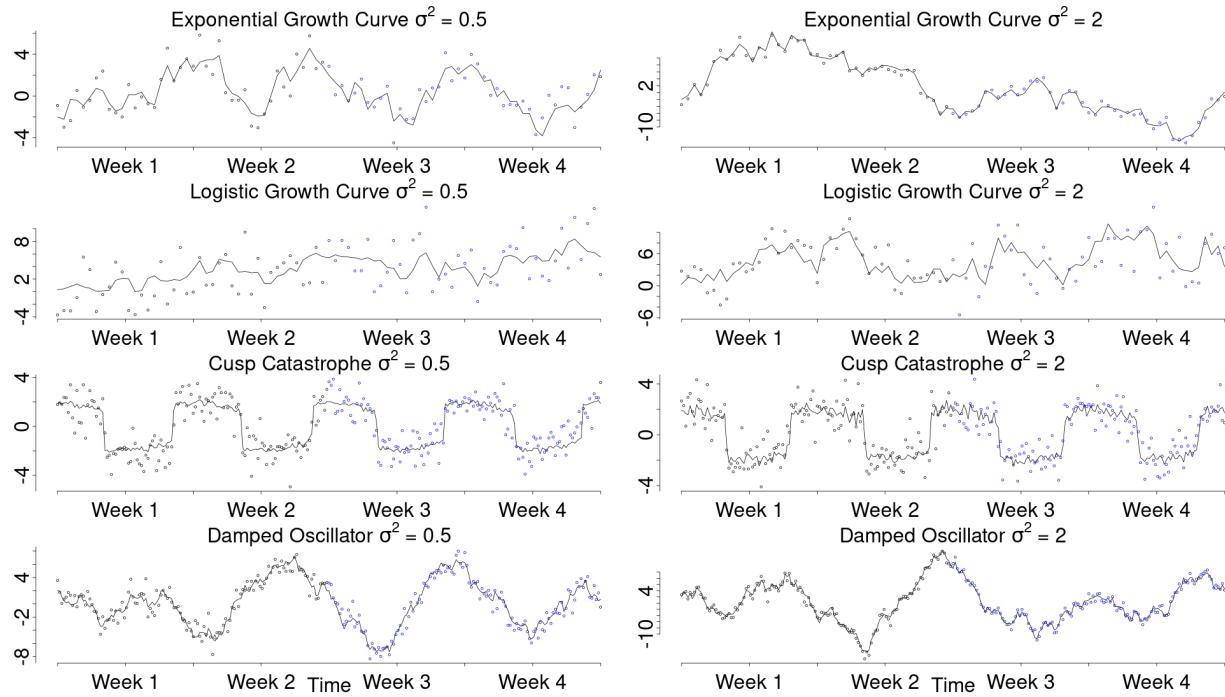
**Table 1***A comparison of LPR, GP regression and GAMs*

	LPR	GP	GAM
Advantages	<ul style="list-style-type: none"> <li>• Intuitive theory</li> <li>• Completely data driven</li> </ul>	<ul style="list-style-type: none"> <li>• Most interpretable parameters</li> <li>• Natural uncertainty quantification</li> <li>• Flexible modelling framework can incorporate prior theory</li> </ul>	<ul style="list-style-type: none"> <li>• Intuitive theory</li> <li>• Some interpretable parameters</li> <li>• Flexible modelling framework can incorporate prior theory</li> </ul>
Disadvantages	<ul style="list-style-type: none"> <li>• Least interpretable parameters</li> <li>• Biased for most processes</li> <li>• No uncertainty estimate for the wigginess</li> </ul>	<ul style="list-style-type: none"> <li>• Unintuitive theory</li> <li>• Difficult to specify in practice</li> </ul>	<ul style="list-style-type: none"> <li>• No uncertainty estimate for the wigginess</li> </ul>
Required Choices	<ul style="list-style-type: none"> <li>• Polynomial degree</li> <li>• Kernel</li> <li>• Optimization criterion</li> </ul>	<ul style="list-style-type: none"> <li>• Covariance kernel</li> <li>• Mean function</li> <li>• Hyperpriors</li> </ul>	<ul style="list-style-type: none"> <li>• Spline basis</li> <li>• Optimization Criterion</li> </ul>
Key assumptions	<ul style="list-style-type: none"> <li>• P-times differentiable process</li> <li>• Constant wigginess</li> </ul>	<ul style="list-style-type: none"> <li>• Assumptions depend on chosen specifications</li> </ul>	<ul style="list-style-type: none"> <li>• Smooth process</li> <li>• Homoscedasticity</li> </ul>
Estimation	<ul style="list-style-type: none"> <li>• OLS</li> </ul>	<ul style="list-style-type: none"> <li>• Bayesian</li> </ul>	<ul style="list-style-type: none"> <li>• OLS, MLE, Bayesian</li> </ul>
Key sources of information	<ul style="list-style-type: none"> <li>• Fan and Gijbels (2018)</li> </ul>	<ul style="list-style-type: none"> <li>• Rasmussen and Williams (2006)</li> </ul>	<ul style="list-style-type: none"> <li>• Wood (2006)</li> </ul>

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.

Because these smooth processes are unlikely to appear in psychology, due to external influences, we manipulated the smoothness of the processes, by adding a dynamic error component. These dynamic errors reflect external perturbations to the latent construct that are carried forward in 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 for example 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 the different simulation conditions. Regarding the smoothness of the processes, it shows two possible realizations of each process, with a dynamic error variance of 0.5 (left) and 2 (right). 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).

For each of the processes, we further varied the sampling period (i.e., how long each process is measured), the sampling frequency (i.e., how often each process is measured). Since the time series data was simulated, it does not have an inherent time scaling. This means that, one time step in the simulated data could correspond to one hour, day, or year and that the time scaling of the simulated processes is only meaningful with respect to the scale at which each process exhibits its characteristic behaviour. For example, if the chosen time scale is too long, both growth curves (with the chosen parameters) would display a very brief growth period followed by a lengthy almost flat period close to the asymptote. To maintain consistency, all processes were simulated to display their characteristic behaviour over the same time range. To simulate sampling from each process over different time periods, capturing different behaviours

**Figure 5***Simulation conditions*

*Note.* This figure illustrates the different conditions that were manipulated in the simulation. It shows two possible realization of each exemplar processes with dynamic errors variances of 0.5 (left) and 2 (right). Further, the sampling period was manipulated by sampling either only from the first half of each process (black dots) or from the entire process (blue dots). Lastly, the sampling frequency was manipulated. Here the top four panels display samples of three observations per day, whereas the bottom four panels display samples of 9 observations per day.

of each process, data was either simulated from only the first half of the process (Figure 5 black dots), or from the entire process (Figure 5 black and blue dots). For the ease of reading and to correspond to typical EMA conditions, this will be referred to as sampling over either two or four weeks. However, this scaling is arbitrary and could be changed to any other time frame. Lastly, the sampling frequency was manipulated by sampling from each process at different rates. 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). The top four panels of Figure 5 illustrate three observations per day, whereas the bottom four panels display nine observations per day. This resulted in total sample sizes between 42 (3 measurements per day over two weeks) and 252 (9 measurements over four weeks).

Regarding the LPR and the GP, we expect that both methods, using the configurations in which they are most often applied and implemented in standard software, will struggle to infer processes that are not continuous (i.e., with sudden jumps), have varying wiggleness, and are not 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 the continuity and smoothness of the process will influence the performance, since GAMs do not assume constant wiggleness. The parametric modeling approach is expected to provide the most accurate inferences, serving as a benchmark for comparison with the other methods.

Therefore, we first hypothesized that the cusp catastrophe model, which is the only process featuring apparent jumps, will be least accurately inferred by all methods. Second, all four processes exhibit changes in wiggleness (as defined by each of the non-parametric methods respectively) over time. However, while the wiggleness of the exponential and logistic growth functions and the damped oscillator decreases monotonically, the cusp catastrophe's wiggleness changes cyclically (i.e., low wiggleness during the plateau phases and very high wiggleness during the jumps). Therefore, we hypothesized 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 wiggleness over time. We do not expect this effect to occur for the cusp catastrophe process, or when using GAMs. Third, we hypothesized that larger dynamic error variances will lead to less accurate inference by all methods.

Regarding the sample size, we expected that varying the sampling period and frequency will 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 wiggleness 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. Lastly, we expected that increasing the sampling frequency will generally improve the inference accuracy across all methods, as there is more information about the latent process available.

## 2.3 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.

### 2.3.1 *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 omitted 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 the same interpretation. Both, the GCV and the leave-one-out cross-validation criterion, describe how accurately the model predicts omitted data points within the design range, which provides information about how well the model interpolates the process.

The results of the mean MSE and GCV values in each condition will be presented in figures. Additionally, to efficiently summarize the high dimensional results and formalize the analysis, two separate ANOVAs were fit to the MSE and GCV values. Both ANOVAs included all possible main and interaction effects of the data generating processes, the analysis methods, and simulation conditions (i.e., sampling period, frequency, and dynamic error variance). In the

ANOVAs only the three introduced non-parametric methods and the polynomial regression were compared, because the parametric models only serve as a benchmark and estimating the true data generating models is not viable in practice. Since even very small effects can lead to statistical significance at the large sample sizes considered in this simulation, we focused on the effects that showed at least a small effect size according to the partial  $\eta^2$  ( $> 0.01$ ). The assumptions of the ANOVAs were tested. However, given the large sample sizes in this simulation, the ANOVAs were assumed to be robust to moderate violations of normality (Blanca et al., 2017), and any potential violations of homoscedasticity were addressed using heteroscedasticity-consistent standard errors.

### **2.3.2 *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 ( $\max(se_{CIC}) \approx 0.03$ ) individual average confidence interval coverages are expected to deviate from the ideal 95%. Because of this, average coverage proportions between 89% and 100% were also deemed acceptable. Average coverage proportions below 89% indicated either a poor approximation of the underlying process or underestimated uncertainty.

## **2.4 Data generation**

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. Combined with information about the initial state of the process this makes it possible to describe the entire process indirectly. For instance, the stochastic differential equation model used to represent the introduced logistic growth process can be expressed as

follows:

$$dy = ry(1 - \frac{y}{k})dt + \sigma dW_t \quad (6)$$

The first half of this model defines the deterministic dynamics of the process. It 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  $r$ . The second part of the model accounts for the dynamic errors in the form of a Wiener process. 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 depending on the length of the time interval and  $\sigma$ , 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 the first half of Equation 6, as detailed in the online supplementary material. The resulting processes were then simulated using the Euler-Maruyama method, which approximates stochastic differential equation systems 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 independently from a standard normal distribution, generating the final sets of observations,

Based on an initial pilot sample of 30 data sets per condition, we determined the number of replications needed to achieve a Monte Carlo standard error of less than 0.03 for the confidence interval coverage (Siepe et al., 2023). These Monte Carlo standard errors reflect the expected variation in the outcome statistics due to random processes within the simulation. This analysis showed that 100 replications per condition would result in maximal Monte Carlo standard errors of approximately  $se_{MSE} \approx 0.05$ ,  $se_{GCV} \approx 0.38$ , and  $se_{CIC} \approx 0.03$ .

## 2.5 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). The polynomial regressions were estimated using base R, with correlated (i.e., standard) polynomial terms. Finally, the parametric stochastic differential equation models corresponding to the true data-generating models were estimated using the Dyrn package (Ou et al., 2019). 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 the online supplementary material. Further, if any models failed to converge during the simulation, the corresponding outcome measures were excluded from the following analyses.

## 2.6 Results

In the simulation a small proportion of GP regressions and parametric models did not converge. This is most likely due to the small sample sizes considered and the automated model fitting in the simulation. Most notably, the parametric models were not able to infer the cusp catastrophe from the small simulated data sets due to the complexity of the model. The performance measures of the methods that did not converge were removed from the following analysis. Additionally, the parametric models appear to have overfit for a small number of data sets. The complete simulation results and data are available in the online supplementary material.

### 2.6.1 *Capturing the non-linear process*

First, it is noteworthy that all considered methods visually performed well in mean predicting the simulated processes. Figure 6 illustrates an example from each process being

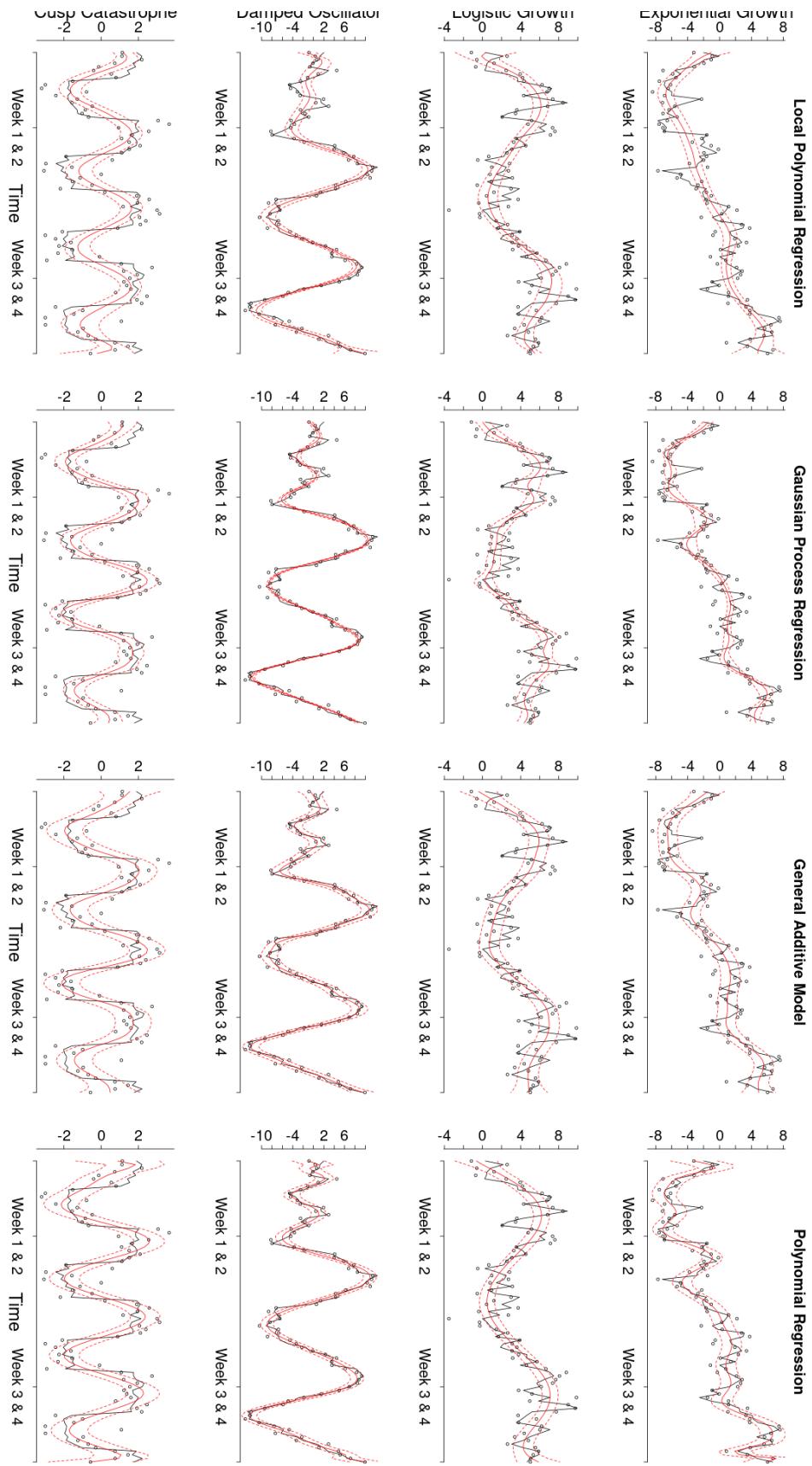
inferred by all methods. The predicted means produced by each method closely followed the simulated processes, although the LPR appears to underfit for some data sets. Additionally, near the boundaries (i.e., the ends) of the simulated time series, the GP regression sometimes tended towards zero, which is a known characteristic of the squared exponential kernel (this can, for example, be seen by comparing the GP inference in Figure 3 to the estimates in Figure 4 or 2). *Further, the polynomial regression appears to overfit near the boundary, resulting in excessive uncertainty, which is a known behavior of polynomial regressions.* However, there is considerable variation and overlap in the accuracy of the different methods across the various data sets, which highlights the need for a more formal analysis of the performance of each method.

To summarize the high dimensional results efficiently, two separate ANOVAs were fitted to the MSE and GCV values, including all possible main and interaction effects. Although the residuals for both models showed considerable deviations from normality, which were mainly characterized by being leptokurtic, the residuals were unimodal and approximately symmetric. Given the large sample sizes in this simulation, we thus assumed that the ANOVAs were robust to these deviations. Further, a Breusch-Pagan test indicated heteroscedasticity in the residuals for both outcome measures, which was corrected for using heteroscedasticity-consistent standard errors. Lastly, Bonferroni corrections were applied to adjust the p-values for conducting two separate ANOVAs.

The type-III ANOVAs for both outcome measures indicated that all main and first order interaction effects were significant. Additionally, some higher order interaction terms were significant, which differed between the two outcome measures. However, due to the large sample size, even very small effects can lead to statistical significance. Therefore, Table 2 presents all effects for which the partial- $\eta^2$ , which gives the proportion of variance explained by an effect after partialling out all other effects, indicates at least a small effect size for either the MSE or the GCV. The following sections will focus on describing some of these effects and a comprehensive overview of all effects in the models can be found in the online supplementary material.

Figure 7 (a) illustrates the mean MSE values with which each method inferred each

**Figure 6**  
*Example processes inferred by each of the introduced methods*



*Note.* This figure shows how each of the introduced methods inferred an example of each of the processes from the simulation.

**Table 2**  
*Effect sizes from the MSE and GCV ANOVAs*

Effect	partial- $\eta^2$ MSE	partial- $\eta^2$ GCV
Method	0.40	0.26
Process	0.55	0.41
SP	0.18	0.09
SF	0.15	0.24
DEV	0.56	0.49
Method:Process	0.23	0.13
Method:SP	0.18	0.11
Process:SP	0.07	0.05
Method:SF	0.05	0.01
Process:SF	0.01	0.05
Method:DEV	0.16	0.09
Process:DEV	0.27	0.24
SP:DEV	0.02	0.03
SF:DEV	0.01	0.05
Method:Process:SP	0.08	0.07
Method:Process:SF	0.02	0.01
Method:Process:DEV	0.07	0.05
Method:SP:DEV	0.02	0.03
Process:SP:DEV	0.01	0.01
Process:SF:DEV	0.01	0.02
Method:Process:SP:DEV	0.02	0.02

*Note.* This table shows all effects from the MSE and GCV ANOVA that had at least a small effect partial- $\eta^2 \geq 0.01$  on either outcome.  
 SP: Sampling period; SF: Sampling frequency; DEV: Dynamic error variance.

process, averaged across sampling periods, frequencies, and dynamic error variances. This figure highlights the main effect of the analysis method, showing clear differences in the mean MSE with which each method inferred all processes. Specifically, the parametric modelling showed the lowest average MSE, followed closely by the GAMs, whereas the GP regression, LPR, and the polynomial regression had larger average MSEs. Additionally, Figure 7 (a) illustrates the main effect of the processes, as each process was inferred by all the methods with different mean MSE values. Most notably, the cusp catastrophe was inferred with lower MSE values than the other processes by all methods. Further, one can see that there is an interaction between the analysis

methods and the processes, since the differences in how accurately each process was inferred differ between the methods. For example, the difference between the MSEs for the cusp catastrophe and the MSEs for the other processes is larger for the LPR than for the other considered methods. Notably, the polynomial regression displayed a larger mean MSE for all processes except the cusp catastrophe in comparison to the more advanced statistical methods.

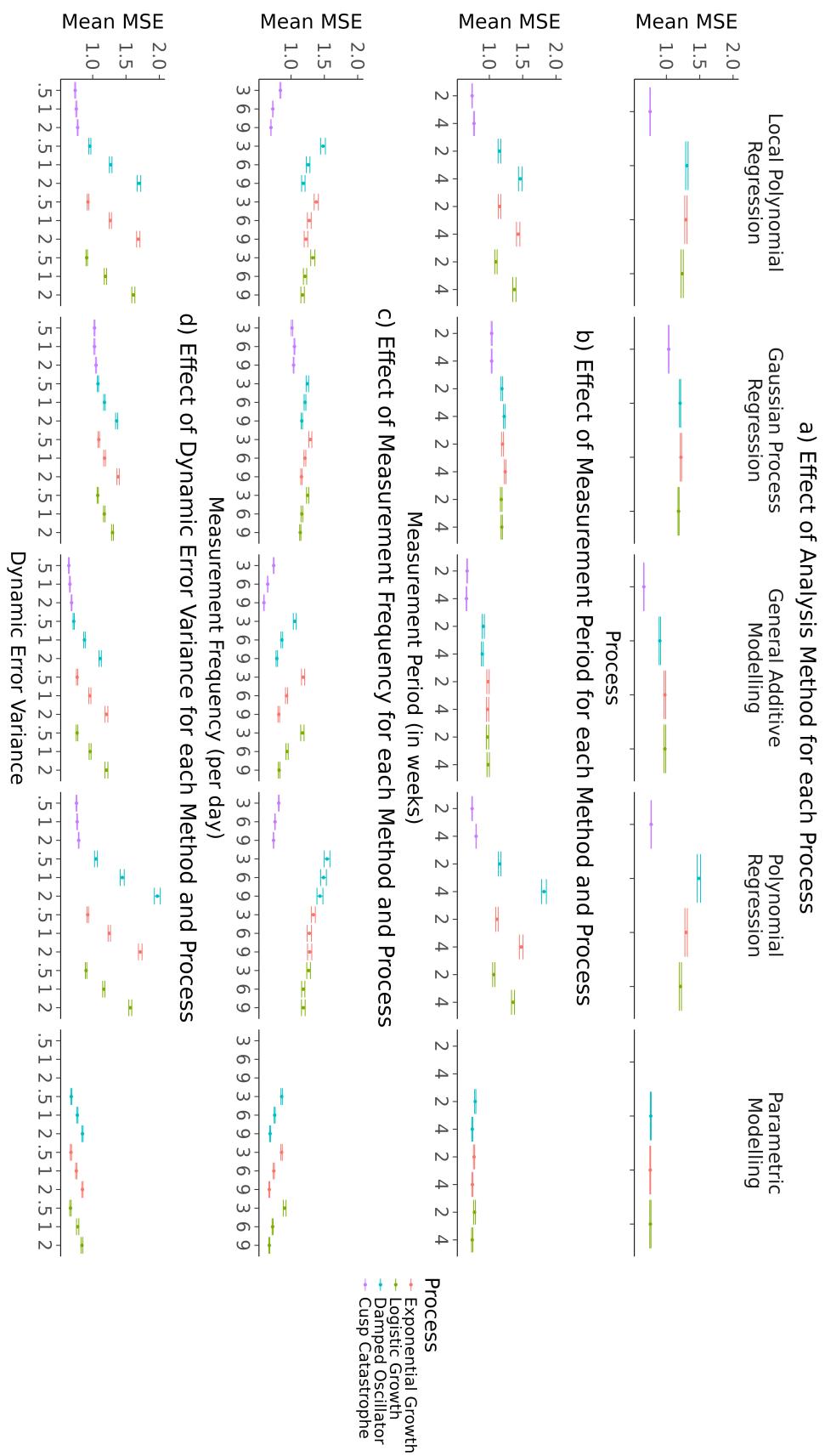
Figure 7 (b) shows the average MSE over different measurement periods for each analysis method and process, averaged over measurement frequencies and dynamic error variances. The results indicate that sampling over the entire process, rather than just the first half, led to higher average MSE values for both local and global polynomial regression across all processes. This effect was much less pronounced for the GP regression, absent for the GAMs, and reversed for the parametric models. Further, Figure 7 (c) illustrates that the mean MSE generally decreased with larger sampling frequencies for each method and process, while averaging over the sampling periods and dynamic error variances. Lastly, Figure 7 (d) demonstrates that larger dynamic error variances increased the mean MSE values across all methods and processes, when averaged over sampling periods and frequencies. This effect was, however, least pronounced for the cusp catastrophe.

Figure 8 displays the corresponding effects for the mean GCV values. Similar to the MSE results, the GAMs show a mean GCV value closest to the benchmark parametric models. However, in terms of the mean GCV, the GP regressions perform equally well as the GAMs. The local and global polynomial regressions show considerably larger mean GCV values for all processes except the cusp catastrophe. The effects of measurement period, frequency, and dynamic error variance on the mean GCV appear to follow largely the same patterns that were observed for the mean MSE.

### **2.6.2 *Uncertainty quantification***

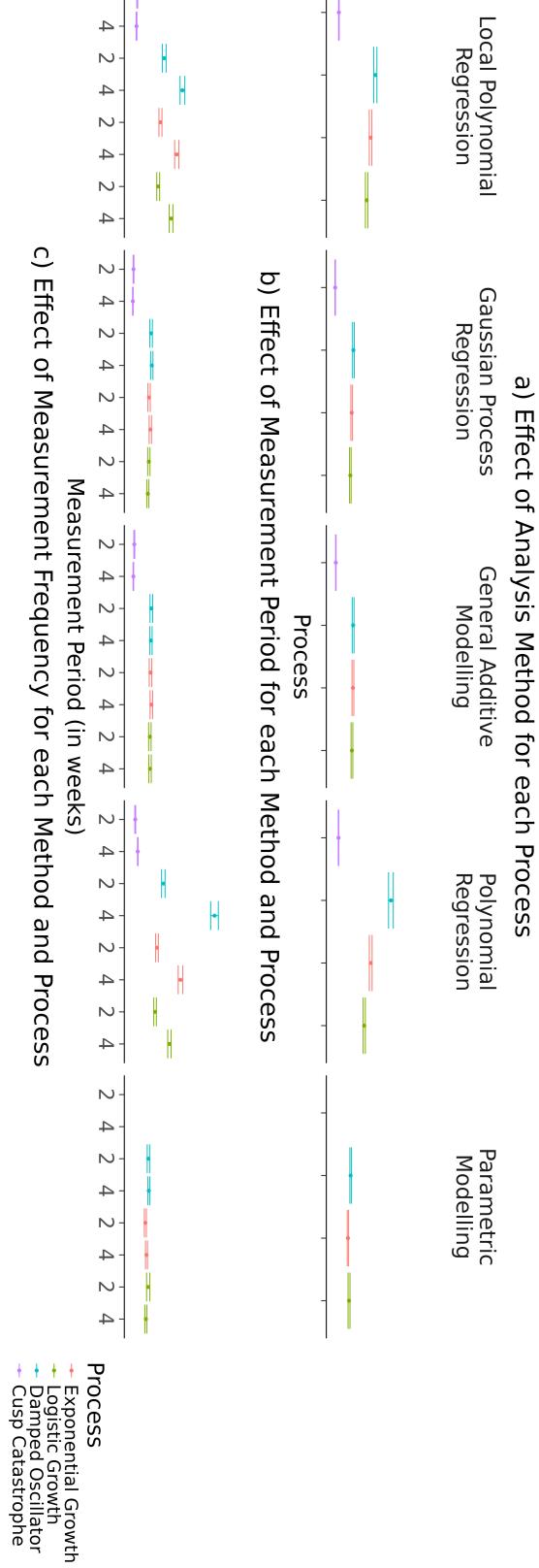
Figure 9 shows the average confidence interval coverage proportion for the conditions described above. The grey area represents an average confidence interval coverage between 89%

**Figure 7**  
*Mean MSE effects across all processes, analysis methods, and simulation conditions*



**Note.** Panel (a) shows the effect of the analysis method for each 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.

**Figure 8**  
*Mean GCV effects across all processes, analysis methods, and simulation conditions*



*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.

and 100%, which indicates no considerable deviation from the ideal 95% given the Monte Carlo error of the simulation. Only the parametric models produced some mean confidence interval coverages that fell within this area. Among the other methods, the GAMs, produced the largest average confidence interval coverage, followed by the global and then the local polynomial regression. The GP regression appears to result in the smallest confidence interval coverage.

## 2.7 Conclusion

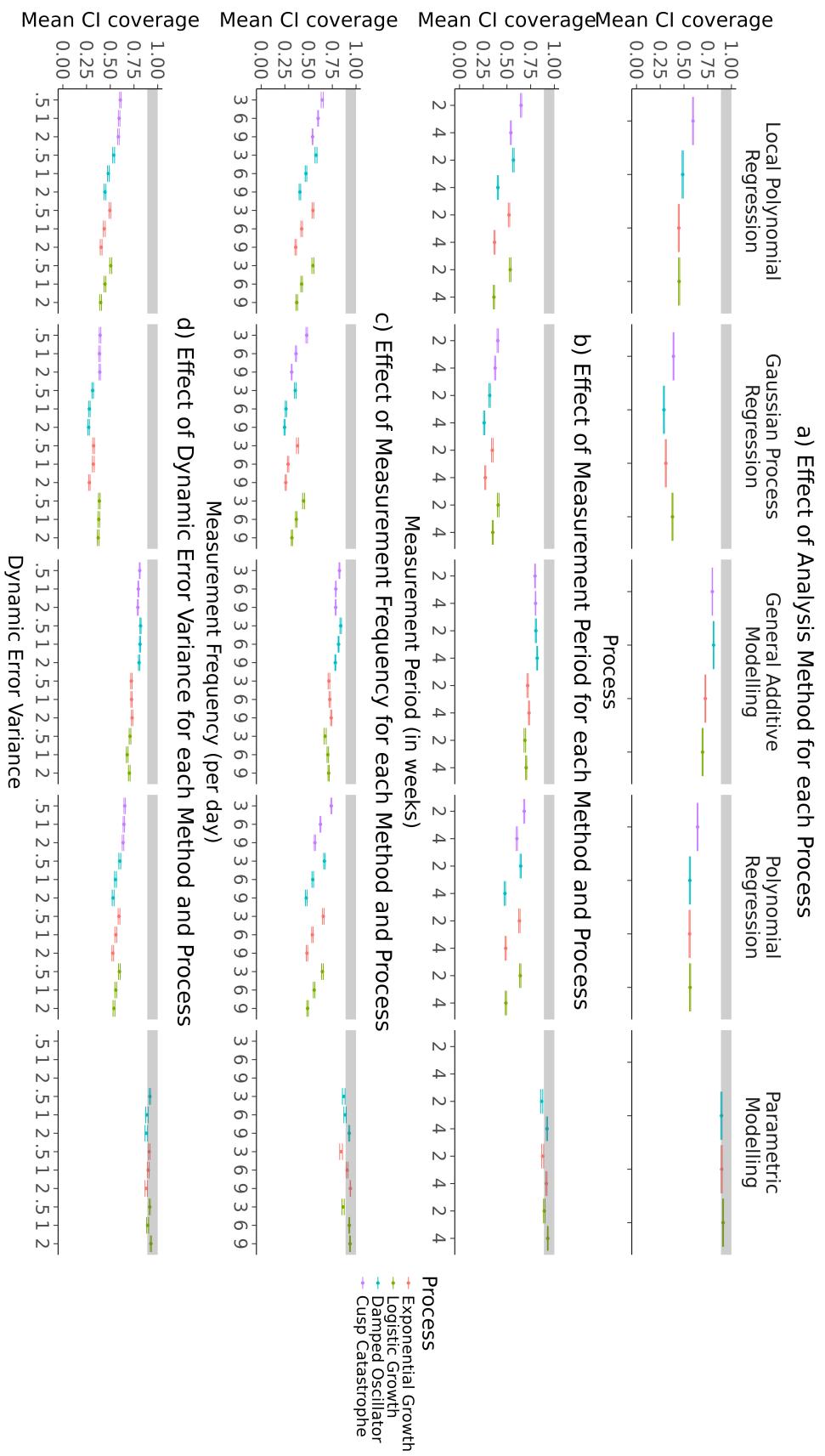
This simulation showed that among the considered methods the GAMs inferred all processes with the most accuracy as indicated by the mean MSE, GCV and confidence interval coverage. The GAMs performed closest to the true data generating models, followed by the GP (with regards to the MSE and GCV, as the GP had the lowest confidence interval coverage), and then the local and global polynomial regression. This result is unexpected, as we anticipated that the smooth inferences produced by GAMs may be ill-suited for inferring the rough (i.e., non-differentiable) processes in the simulation. However, as Figure 6 shows, all methods except the parametric models produced smooth inferences. This suggests that the additional flexibility that the GAMs provide over the LPR, by not assuming constant wigginess, and over the GP, by effectively relying on improper priors, enabled the GAMs to infer the processes more accurately in this simulation. From this follows, that GAMs are an attractive starting point for modelling unknown processes in practice. Especially, when there is little prior theory about the functional form of the process.

However, it is important to note that the observed results may be most attributable to the specific configurations that were used rather than the general modelling approaches underlying each method. The specific configurations of each method were chosen to reflect how each method is most commonly applied in practice and not to optimally infer the simulated processes. Consequently, different configurations and extensions are likely to improve the performance of the LPR and GP<sup>2</sup>. For example using a GP with a Matern class kernel, which makes less strict

---

<sup>2</sup> Especially, since the GAMs used in this simulation can be expressed as a GP with a linear mean function (with improper priors) and a non-stationary covariance function (Rasmussen & Williams, 2006; Wahba, 1978). This

**Figure 9**  
*Average confidence interval coverage across all processes, analysis methods, and simulation conditions*



*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.

smoothness assumption about the data, could be expected to improve the accuracy for rough processes, like the ones considered in the simulation. Therefore, if there is already some prior knowledge about the form of the process available, GPs are still a very interesting modelling approach in practice, as they yield more interpretable inferences. However, this likely still requires fine tuning the configurations of the GP to one's specific conditions.

Regarding the LPR, several optimality results have been found indicating that LPR should be at least as accurate as GAMs and GPs for processes that satisfy the smoothness assumption of the LPR (Fan et al., 1997). It is however unclear whether these results generalize to the rough processes that we suspect can be found most often in psychology. Lastly, the tested polynomial regression inferred the underlying processes least accurately. This appeared to be due to numerical model instability that resulted in underfitting, as not enough higher-order polynomial terms could be included in the models, to capture the process accurately. Another issue with the polynomial regression was that the estimates diverged towards the ends of the timeseries. The numerical model instability can be partially mitigated by using orthogonal polynomials, which are however more difficult to interpret. Generally, there appears to be no reason to use a global polynomial regression instead of a GAM, in any situation in which prior theory does not strongly suggest that the process follows a low order polynomial trajectory.

Contrary to our expectation, the simulation indicated that the cusp catastrophe process was inferred most accurately by all methods. 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 overshadowed 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 being of 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

---

implies that there exists a GP configuration which performs at least as well as the GAMs.

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 further 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 parametric models but may decrease the accuracy of the LPR and the polynomial regression. This reduction in accuracy could however possibly be mitigated by using extensions for a variable bandwidth, or polynomial degree in an LPR. 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, we applied GAMs to depression data from the Leuven clinical study, which were obtained from the EMOTE database (Kalokerinos et al., n.d.). We selected the data for their heterogeneous sample, which contains momentary depression scores of participants who met the DSM criteria for mood disorders or borderline personality disorder during an intake, as well as depression scores for healthy controls. For a more thorough sample and data description, including screening protocols, see Heininga et al. (2019). The diversity in the study population makes it likely to find non-linear processes for at least some participants.

The dataset used for this application 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<sup>3</sup>. The participants completed seven days of semi-random EMA assessments, with 10 equidistant assessments per day. During each assessment, participants responded to the item ‘How depressed do you feel at the moment?’ on a scale ranging from 0 to 100 to assess their momentary depression<sup>4</sup>.

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).

### 3.1 Analysis and Results

To maintain consistency with how the GAMs were introduced in this article, we applied them to explore the idiographic processes underlying the data of each individual. Each GAM was fit using the same specifications as in the simulation, with a single smooth term for time. Inspecting the estimated processes revealed a clear picture of the heterogeneity in the processes underlying the data. On one hand, Figure 10 (a) shows the ten estimated processes with the lowest EDF, which appear to be essentially linear. Indeed, for 72 out of 117 participants, the process inferred by the idiographic GAMs is effectively a linear trend ( $EDF < 1.001$ ). For some of these participants, the linear estimates appear to be due to strong floor effects where participants repeatedly indicate depression scores close to zero. However, this explanation does not hold for all participants, as some participants also received linear estimates without displaying floor effects. Interestingly, the linear estimates also indicate an effective absence of dynamic errors, which would be expected to result in deviations from the linear trend, for these participants.

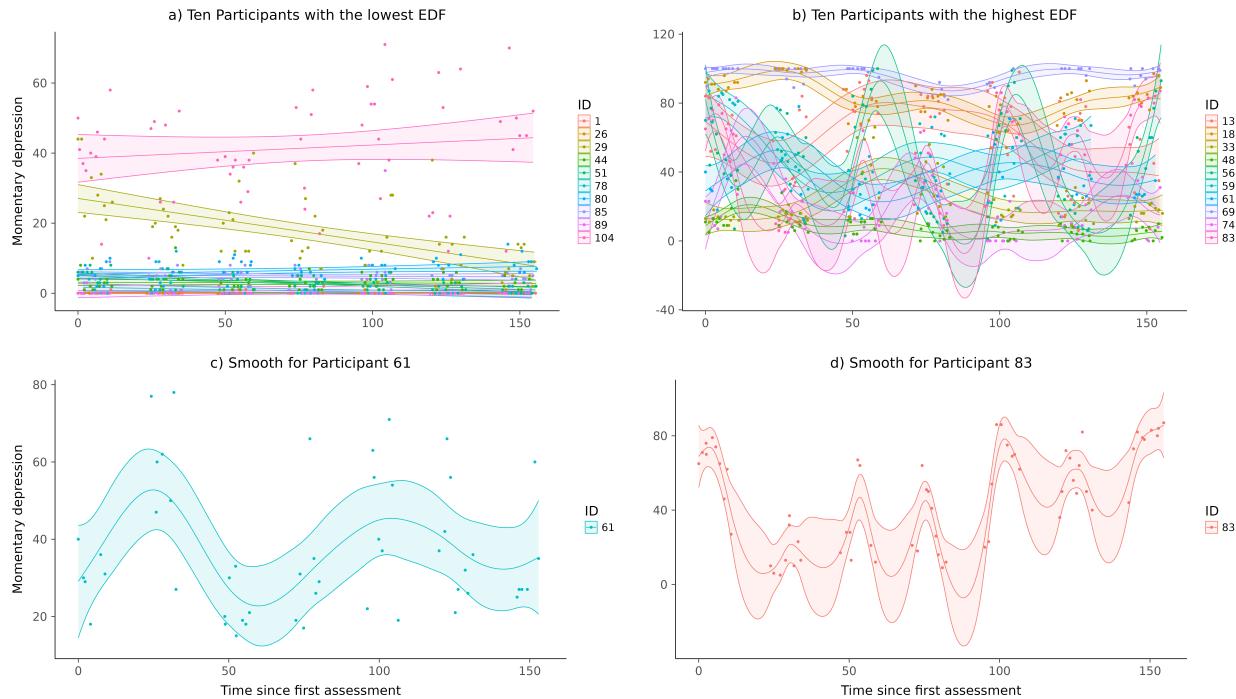
Figure 10 (b) shows the ten participants with the wiggliest estimated processes, indicated by the largest EDF. For these participants, the estimated processes clearly deviates from a linear

---

<sup>3</sup> 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.

<sup>4</sup> Note that this was only one of 27 items that were assessed at every measurement occasion. The other items (questions about emotions, social expectancies, emotion regulation, context, and psychiatric symptoms) are, however, not relevant to this application and, therefore, not further described.

**Figure 10**  
*Estimated depression processes*



trend, which may either be due to the presence of non-linearity or the presence of dynamic errors. It is not possible to distinguish which of these possibilities is true using this analysis. Additionally, visual inspection of the estimated non-linear processes, did not reveal any common behaviours across the processes of different participants. Nevertheless, the idiographic processes still displayed interesting behaviours. With respect to the processes introduced in this article, two participants show particularly interesting trajectories. First, Figure 10 (c) shows an estimated process that closely resembles the decreasing oscillations exhibited by a damped oscillator. Second, the estimated process of another participant illustrated in Figure 10 (d) may switch between a state low and a state of high depression, with considerable oscillations within each state.

### 3.2 Conclusion

#### 4 Discussion

This results presented in this article have 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 going to change their performances. 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 in practice. 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 need to be modelled using different psychometric models (e.g., factor models, item response models).

There are fortunately many ways in which the presented methods could be adapted to these more complex data structures. The GP and GAM can be naturally adapted to incorporate 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 by extension the GAMs can model latent variables underlying for

example a factor model, which naturally extends these methods to a setting with multiple indicator variables. This raises the possibility of extending these models to typical psychometric measurement models, to accurately capture more complex measurement error distributions. However, it is unclear how well these extensions work in practice, beyond the Gaussian process factor models, which are already implemented in software (Clark & Wells, 2023). The parametric models introduced already include a factor model for the observed variables, which can incorporate multiple indicators, which may 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 autoregressive relations. These models also present the exciting possibility to combine partial parametric models with non-linear data driven functions, estimated by the presented methods. Lastly, the presented methods could be used to infer unobserved input variables to parametric dynamic models (Álvarez et al., 2009; Nayek et al., 2019), which could for example take the form of seasonal and cyclic influences to indicator variables (Clark & Wells, 2023). While these extensions and possibilities exist in theory and are in isolation implemented in software, most are currently not available and implemented in a form which would allow applied researchers to flexibly adapt these methods to the characteristics of specific ILD. Therefore it is important that future research, combines and extends the presented methods and implements these extensions in software that makes them accessible for applied modelling.

## 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 Twelfth 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](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., & Broder, S. (2024). *Cmdstanr: R interface to 'cmdstan'* [R package version 0.8.1, <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](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., Sienraert, 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>
- 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>
- Jianan, Z., Yiyi, W., Hongyi, D., & Qingyang, L. (2023). A case study of the Lunger phenomenon based on multiple algorithms [arXiv:2311.11253 [cs, math]]. Retrieved September 12, 2024, from <http://arxiv.org/abs/2311.11253>  
Comment: 13 Figures 9 Pages. After first submission, there was a revision of the authorship order, which was the result of joint discussions
- 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](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., Shiyo, 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>
- 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