

**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

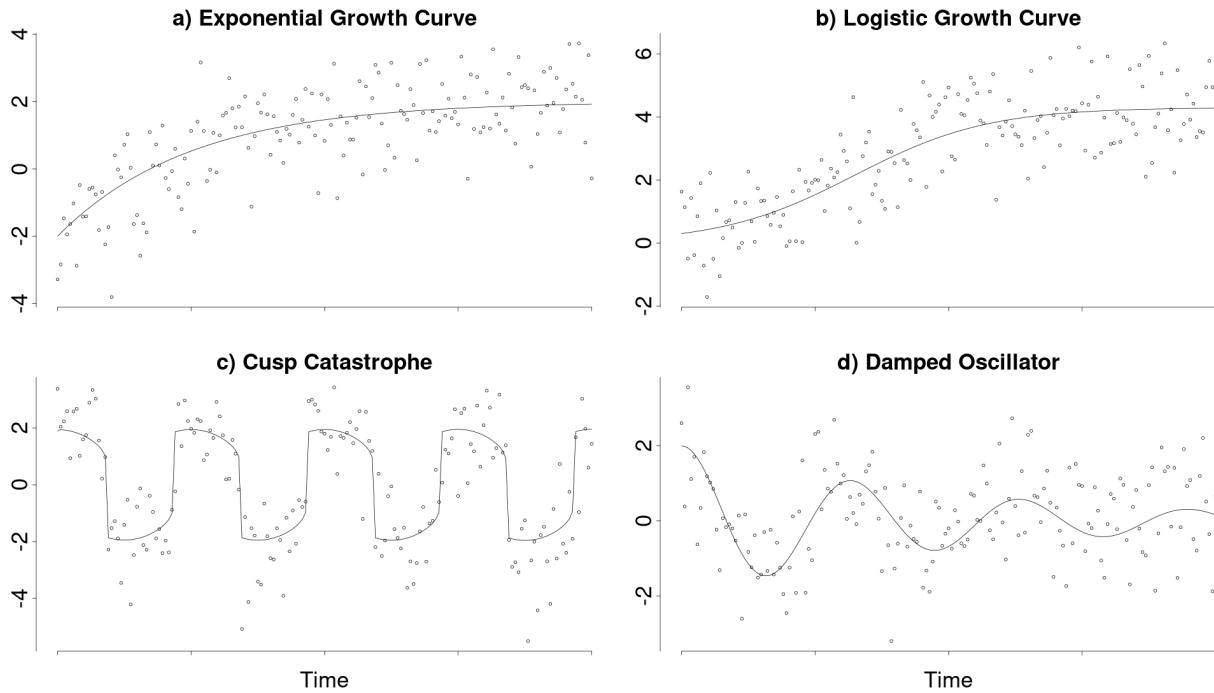
Psychological constructs are increasingly understood as components of complex dynamic systems (Nesselroade & Ram, 2004; Wang et al., 2012). This perspective emphasizes that these constructs fluctuate over time and within individuals. To study these variations and the underlying processes, researchers 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.

Through 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 of this is alcohol use relapse, where patients suddenly switch from an abstinent state to a relapsed state (Witkiewitz & Marlatt, 2007). This sudden

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.

switching behavior has been modelled using a cusp catastrophe model (Chow et al., 2015; van der Maas et al., 2003). This dynamic model 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 specific form of non-linear dynamics. ILD studies can potentially help to refine such theories by providing a nuanced understanding of how psychological variables interact over time. These refined theories could, for instance, take the form of parametric 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 more specific theories (about the form of non-linear change), it is first necessary to empirically uncover, observe, and study phenomena such as the mentioned state switching or regulatory oscillations in ILD, and determine 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 based on their success in doing 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, in addition to a linear effect of 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 together at specific points, called knots (e.g., connecting multiple cubic functions end-to-end into an overarching growth curve; Tsay and Chen (2019)). However, piecewise 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. Moreover, the ideal statistical method may depend on the characteristics of the underlying non-linear process (which, as yet, are generally unknown). Especially, since the types of non-linear processes for which the different methods were developed match the types of processes occurring in psychological research (e.g., the processes depicted in Figure 1) to varying degrees, particularly with regard to the assumed smoothness of the non-linear process. This means that the method most suitable for the relatively smooth process in Figure 1a is not necessarily also the most suitable method for rougher changes.

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, three semi- and non-parametric regression techniques will be introduced; local polynomial regression, Gaussian processes, and generalized additive models.

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 every 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:

$$y_t = f(t) + \varepsilon_t \quad (1)$$

$$\mathbf{X} = \begin{bmatrix} 1 & (t_1 - t^*)^1 & \dots & (t_1 - t^*)^p \\ \vdots & \vdots & \ddots & \vdots \\ 1 & (t_n - t^*)^1 & \dots & (t_n - t^*)^p \end{bmatrix} \quad (2)$$

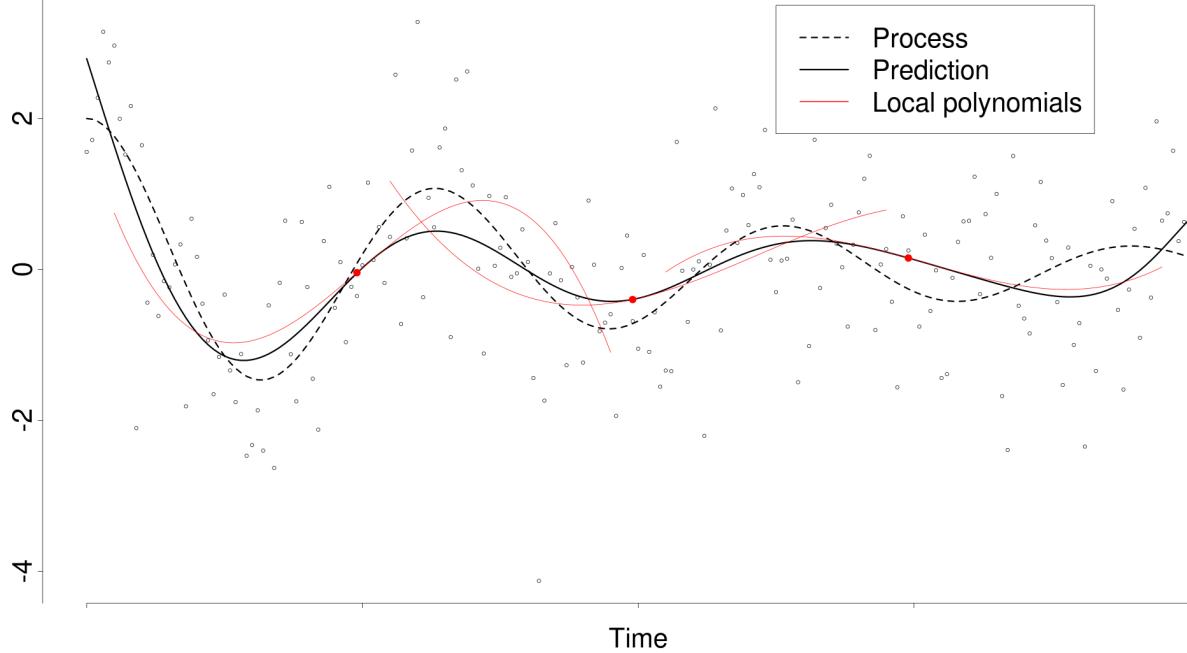
$$\mathbf{W} = \begin{bmatrix} w_{1,1} & & \\ & \ddots & \\ & & w_{n,n} \end{bmatrix} \quad (3)$$

$$\hat{f}(t^*) = Intercept((\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W} \mathbf{y}) \quad (4)$$

where a univariate process f is inferred at the chosen time point t^* . Matrix \mathbf{X} is the model matrix of a multiple linear regression model, such that the columns correspond to the predictors used to predict f at time point t^* . 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 contains the centered time points squared, and so on up to the chosen degree p^1 of the polynomial). Further, \mathbf{W} is a diagonal matrix containing the weights associated with each datum. Lastly, Equation 4 is a normal equation solving for the coefficients of a weighted multiple linear regression model. The intercept of this regression gives the estimated value of the LPR at t^* . This procedure is repeated for all time points of interest and the estimated LPR values are subsequently connected into an overall estimate of the nonlinear process.

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.

¹ 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.

When fitting an LPR, two decisions must be made that pertain to the optimal weighting of the data around the time point at which an LPR value is estimated and the degree of the local polynomial. First, a specific kernel function must be chosen. This kernel function is a mathematical equation of the form

$$w_{i,i} = K\left(\frac{t_i - t^*}{h}\right), \quad (5)$$

which determines the influence of different data points (i.e., the data weighting) during the local polynomial estimation. Common choices for this kernel function are the Gaussian and Epanechnikov kernels. Both assign higher weights to data points in closer proximity to the point of interest t^* , however, while the Gaussian kernel assigns small weights to all distant data points, the Epanechnikov kernel assigns zero weights beyond a certain distance. Because the Epanechnikov kernel has also been shown to be optimal in many situations (Fan et al., 1997), it

constitutes a good default choice if there are no strong preferences for other kernels. The exact amount of weight given to different data points by a kernel is further defined by a bandwidth parameter h which determines the kernel's width, and as such, the wiggleness of the estimated process, such that a smaller bandwidth is associated with wigglier estimates. 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). This optimization criterion may be selected based on specific research objectives. For instance, optimizing the cross-validation error may be most attractive if the primary research interest is out of sample prediction. However, most standard software packages offer automated default procedures to find the optimal bandwidth, which is especially attractive for researchers new to LPR. Note that the fact that the LPR uses a single value for the bandwidth parameter implies that the method assumes that the underlying process has constant wiggleness (with respect to the degree of the local polynomials). This does not make the LPR too rigid, especially not for higher order polynomials, as these have a substantial amount of flexibility with which to approximate the non-linearity of a process. Moreover, this constant wiggleness assumption can be relaxed by using a time-varying bandwidth (Fan & Gijbels, 1995b) or a time-varying polynomial degree (Fan & Gijbels, 1995a), but these extensions are beyond the scope of this paper.

Second, researchers must also choose the degree of the local polynomials, which 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 process's rate of change (i.e., first 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, which ensures 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 (e.g., local linear or local cubic polynomials). This choice reflects a bias-variance tradeoff, where higher-order

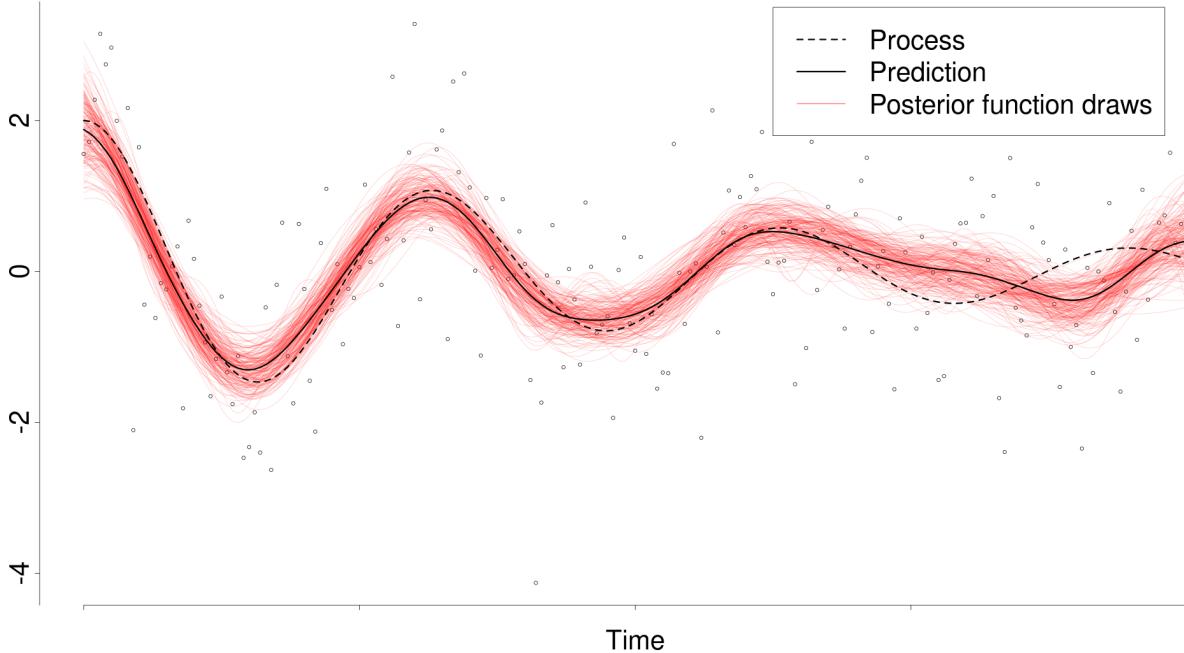
polynomials reduce bias but increase variance only when transitioning from an odd to an even power (Ruppert & Wand, 1994). It is noteworthy, that unless the underlying process follows a polynomial trajectory with at most the same degree as the LPR the approximation with local polynomials is biased. For 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, GPs 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 will 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:

$$P(f | \mathbf{y}) \propto P(\mathbf{y} | f)P(f) \quad (6)$$

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

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.

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.

When fitting a GP, researchers need to decide what mean function $m(t)$ and covariance function $cov(t, t)$ to use. These functions are continuous extensions of the mean vector and covariance matrix of a multivariate normal distribution. 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. However, if there is prior information available about the trend of the non-linear process (such as in a developmental context, where a learning curve could be expected to have a general upwards trend) it would be possible to include this in the GP by adding, for example, a linear mean function.

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,

$$\text{cov}(t_i, t_j) = k(|t_i - t_j|) \quad (7)$$

The types of processes that can be captured by a GP regression primarily depend on the chosen form of the covariance kernel. 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 (such that the described covariances between the function values only depend on their distance) and very smooth. Therefore, using the squared exponential covariance kernel imposes a stricter smoothness assumption on 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, such as the popular continuous-time autoregressive process.

As was the case with the LPR, the kernels (and thus the covariance function) used in a GP have parameters, called hyperparameters, which determine the shape of the estimated non-linear process. 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.

The hyperparameters of the GP are not estimated by optimizing a specific criterion, but through Bayesian analyses. This means that researchers need to provide input on the amount of wigginess and spread in the process through prior distributions, which might be hard given the mentioned lack of information on the shape of processes in Social Sciences. Fortunately, researchers can always provide vague prior distributions which basically state that all values of

the parameters are equally likely a-priori. An advantage of this Bayesian estimation is that GP regression provides a natural approach to uncertainty quantification for the wiggliness (and spread) of the process. Like LPR however, GP regression does also assume that the amount of wigginess of the underlying process is constant.

Note that, whereas LPR is a mainly data-driven procedure, GP regression is more model based as the form of the GP prior (i.e., the chosen mean and covariance function) generates a family of functions to which the process is assumed to belong. This makes it possible to include additional hyperparameters in a GP construction, to test specific (incomplete) theories and hypotheses through, for example, model selection.

1.3 Generalized additive models

Generalized additive models (GAM) are a class of semi-parametric models that can be seen as an extension of regression models which do not use variables as predictors of an outcome, but so-called smooth terms.

$$Y_t = \beta_0 + \sum_{i=1}^I \beta_i(x_i) \quad (8)$$

Each smooth term β_i reflects a non-linear function across a predictor (e.g., time), and a GAM combines these different smooth terms into an overall estimate of a non-linear process (Hastie & Tibshirani, 1999; Wood, 2006, 2020). The smooth terms are then inferred from the data through smoothing splines. These smoothing splines are very similar to the piecewise spline regression introduced earlier in that they also take the form of a basis function regression. This basis function regression does not use the raw predictor values directly. Instead, it combines multiple predefined functions of the predictor $R_{ki}(x_i)$ (e.g., the piecewise functions in a piecewise spline regression) and their corresponding weights α_{ki} (regression coefficients) in a regression model.

$$\beta_i(x_i) = \sum_{k=1}^K \alpha_k R_k(x_i) \quad (9)$$

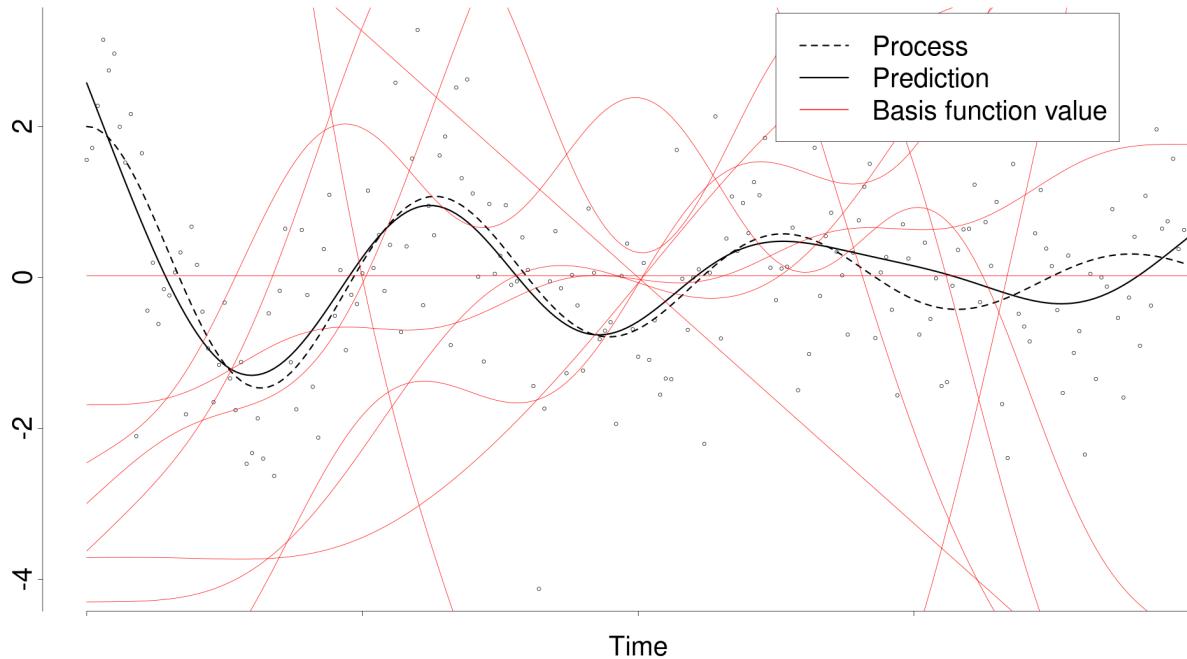
Generally, using more basis functions (and thus coefficients) results in a smoothing spline that is more flexible. Meaning, that it can fit the data closer during the estimation. Smoothing splines rely on using enough basis functions to overfit on the data, which ensures that they are flexible enough to capture the underlying process accurately.

There are many ways available to construct such a smoothing spline, the most common ones being cubic and thin-plate splines, which are different ways generate the basis functions that constitute the smoothing spline. Cubic splines, on the one hand, are piecewise third-degree polynomials that are joined together smoothly at knots. When a knot is placed at every data point, cubic splines can optimally interpolate and thus overfit the data. Thin-plate splines, on the other hand, avoid knots entirely and instead use automatically generated and increasingly wiggly basis functions (e.g., the first basis function is constant, the second is linear, the third has one extremum, the fourth has two, and so on) that are defined over the entire range of the data. When using as many basis functions as data points this guarantees that thin-plate spline perfectly interpolates and overfits the data.

To counteract this overfitting, smoothing splines have 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. This penalty balances the complexity and fit of the smooth term, ensuring the model captures the underlying process accurately without overfitting (Gu, 2013; Wahba, 1980).

While GAMs can combine multiple smooth terms, in this paper we will focus on GAMs consisting of a single smooth term. 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 used in this example are shown in red.

When using GAMs, researchers theoretically need to choose the spline basis used for smooth term construction first. However, the presented cubic and thin-plate splines have been found to be optimal in many aspects (and most common spline bases do not result in considerably different estimates either). This leaves the choice on the number of basis functions to use in

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).

constructing the smooth term, where using more basis functions allows researchers to capture more complex processes. In addition, the number of basis functions should at least be large enough to enable sufficient model fit and at most be the same as the number of data points. Models with large amounts of basis functions can be computationally intensive. Fortunately, the selection of lower rank basis functions is automated in most standard software (Wood, 2011), which makes GAMs accessible to use.

Note that, while the number of basis functions determines, in part, the complexity of a GAMs smooth term, GAMs do not have a parameter like the bandwidth parameter of the LPR or the lengthscale of GPs that reflects the wiggliness of a process. Instead, GAMs provide an indication of wiggliness through their effective degrees of freedom (EDF), a measure of the model's effective complexity (where an EDF of one corresponds to a linear model). Unfortunately, since the EDF are not a parameter, GAMs do not provide a measure of the

uncertainty in the wiggleness of the process like the GP does. However, unlike the LRP and GP, GAMs do not assume constant wiggleness of the underlying process.

While GAMs work more or less “out-of-the-box”, they are still a more model based approach, just like GPs. So, similarly to a GP, one could construct a GAM that combines a linear trend with a non-linear smooth term to capture deviations from the linear trend, provided this model aligns with one’s domain knowledge.

1.4 Differences between methods

A key difference between the three non-linear methods introduced in this paper is the information they provide about the wiggleness 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 respective 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. Other important differences between the methods are summarized in Table 1.

2 Simulation

2.1 Problem

A simulation study was conducted to assess the effectiveness of the introduced methods in recovering different non-linear processes that 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 (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 in line with how they were introduced and within the

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

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

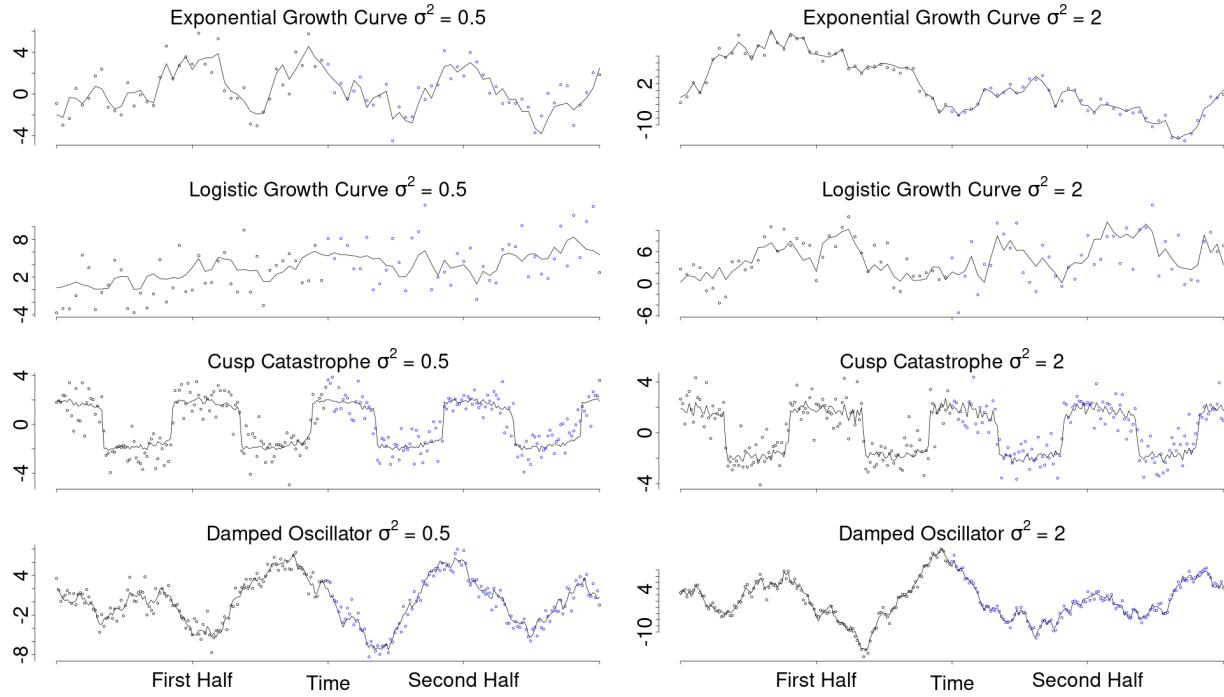
constraints of available software implementations, the simulation focused on a univariate single-subject design. Hence, the simulated data represented repeated measurements of a single variable for one individual.

2.2 Design

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).

Since real live process are likely affected by external influences (i.e., context effects), we account for these influences in the simulation by adding dynamic errors to each process. These dynamic errors reflect external perturbations (or other forms of unaccounted influence) 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. We considered variances of 0.5, 1, and 2 reasonable relative to the process range. Figure 5 provides an overview of the different simulation conditions. To demonstrate the varying roughness of the processes, the figure presents two possible realizations of each process, one with a dynamic error variance of 0.5 (left) and another with a variance of 2 (right). Importantly, we intentionally omitted a condition without dynamic errors from this simulation, as dynamic errors are reasonably expected to be present in all psychological ILD.

Additionally, we varied both the sampling period (i.e., the duration over which the process

Figure 5*Simulation conditions*

Note. This figure illustrates the different conditions that were manipulated in the simulation. It shows two possible realizations of each exemplar process with dynamic error variances of 0.5 (left) and 2 (right). Further, it shows how the sampling period was manipulated by sampling either only over the first half (black dots) or over the entire period of each process (black and blue dots). Lastly, the sampling frequency was manipulated. The top four panels display samples with one observation per time step, whereas the bottom four panels display samples with three observations per time step.

is measured) and the sampling frequency (i.e., the number of measurements taken during that period) for each process. Since the data in the simulation is generated, it does not have an inherent time scale. This means that one time step in the simulated data could represent an hour, a day, or a year, and that the time scaling of the simulated processes is only meaningful relative to the scale at which each process exhibits its characteristic behavior. For instance, if the chosen time scale is too long, both growth curves (Figure 1 a and b; with the chosen parameters) would display a brief period of growth followed by a long, nearly flat phase approaching the asymptote. To ensure consistency, all processes were simulated to display their characteristic behavior over the same period.

To simulate measuring each process for a shorter or longer time, thus capturing different

behaviors of each process, data was either generated over only the first half (Figure 5, black dots) or over the entire period (Figure 5, black and blue dots). For example, sampling the two growth curves only over the first half of the period would mean that the process is not measured close to the asymptote.

Lastly, we manipulated the sampling frequency by varying the number of equidistant measurements taken per time step. Sampling frequencies of one, two, and three observations per time step were tested to reflect typical EMA sample sizes (Wrzus & Neubauer, 2023). This resulted in overall sample sizes ranging between 42 observations (one observation per time step for half the period) to 252 observations (three observations per time step for the entire period). The top four panels of Figure 5 show examples with one observation per time step, while the bottom four panels illustrate three observations per time step, simulating a scenario where measurements are taken three times more frequently.

2.3 Hypotheses

Since the parametric models matched the true data-generating models, we expected them to infer the underlying processes most accurately, serving as a benchmark for comparison with the other methods. We also anticipated that the (global) polynomial regression would infer all processes less accurately than the three introduced approaches, due to the limitations outlined in the introduction. However, we did not have specific hypotheses regarding the relative performance of the three non- and semi-parametric methods.

Additionally, we anticipated that all methods, except the parametric models, would struggle to accurately infer the cusp-catastrophe process, as each method produces continuous estimates that may be inadequate for capturing the apparent jumps in the process. Similarly, because all methods (except the parametric models) produce smooth estimates using the default configurations in which they are most often applied and implemented in standard software, we expected the performance of all methods to decline with larger dynamic error variances.

Our expectations regarding the sampling period were more nuanced. The LPR and GP

assume that the wiggliness (as defined by each non-parametric method respectively) of the process remains constant over time. However, all four processes exhibit changes in wiggliness. For example, the wiggliness of the exponential and logistic growth functions and the damped oscillator decreases monotonically, while the cusp-catastrophe's wiggliness fluctuates cyclically (i.e., low wiggliness during plateau phases and high wiggliness during jumps). Therefore, we hypothesized that longer sampling periods for the exponential and logistic growth curves and the damped oscillator would reduce the inference accuracy of the LPR and GP, as the single bandwidth or lengthscale parameter becomes increasingly inadequate to capture the changing wiggliness over time. Furthermore, the LPR was not expected to benefit from the additional data provided by measuring the process for a longer time, since it relies mostly on data in its local neighborhood during the estimation. Therefore, extending the sampling period beyond these local neighborhoods was not expected to increase the inference accuracy. In contrast, the GAMs and parametric models, which do not assume constant wiggliness and utilize the entire dataset during the estimation, were expected to infer all processes more accurately with larger sampling periods.

Lastly, we predicted that increasing the sampling frequency would generally improve the inference accuracy across all methods by providing more information about the underlying processes.

2.4 Outcome measures

To evaluate and compare the performance of the different analysis methods, we focused on three outcome measures: the mean squared error (MSE), the generalized cross-validation criterion (GCV), and the confidence interval coverage. 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 at the expected frequency.

To assess how effectively each method captured the non-linear process at the observed time points, we calculated the 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 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 are presented in figures. To efficiently summarize the high-dimensional results and formalize the analysis, two separate ANOVAs were conducted for the MSE and GCV values. We focused on effects that demonstrated at least a small effect size, defined as a partial η^2 larger than 0.01. Both ANOVAs included all main and interaction effects of the data-generating processes, analysis methods, and simulation conditions (i.e., sampling period, frequency, and dynamic error variance). Note that the ANOVAs compared only the three non-parametric methods and the polynomial regression and not the parametric models. The latter served only as a benchmark because estimating the true data-generating models is not viable in practice.

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 ($\max(seCIC) \approx 0.03$), 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% may indicate (but may not only be due to) underestimated uncertainty.

2.5 Data generation

Each process in Figure 5 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:

$$dy = ry(1 - \frac{y}{k})dt + \sigma dW_t \quad (10)$$

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 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. Simulating the other processes was achieved by simply using other equations for the deterministic dynamics in Equation 10. The precise equations used for the other methods are 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 $se_{CIC} \approx 0.03$.

2.6 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 can correct 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 Dynr 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.

2.7 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.

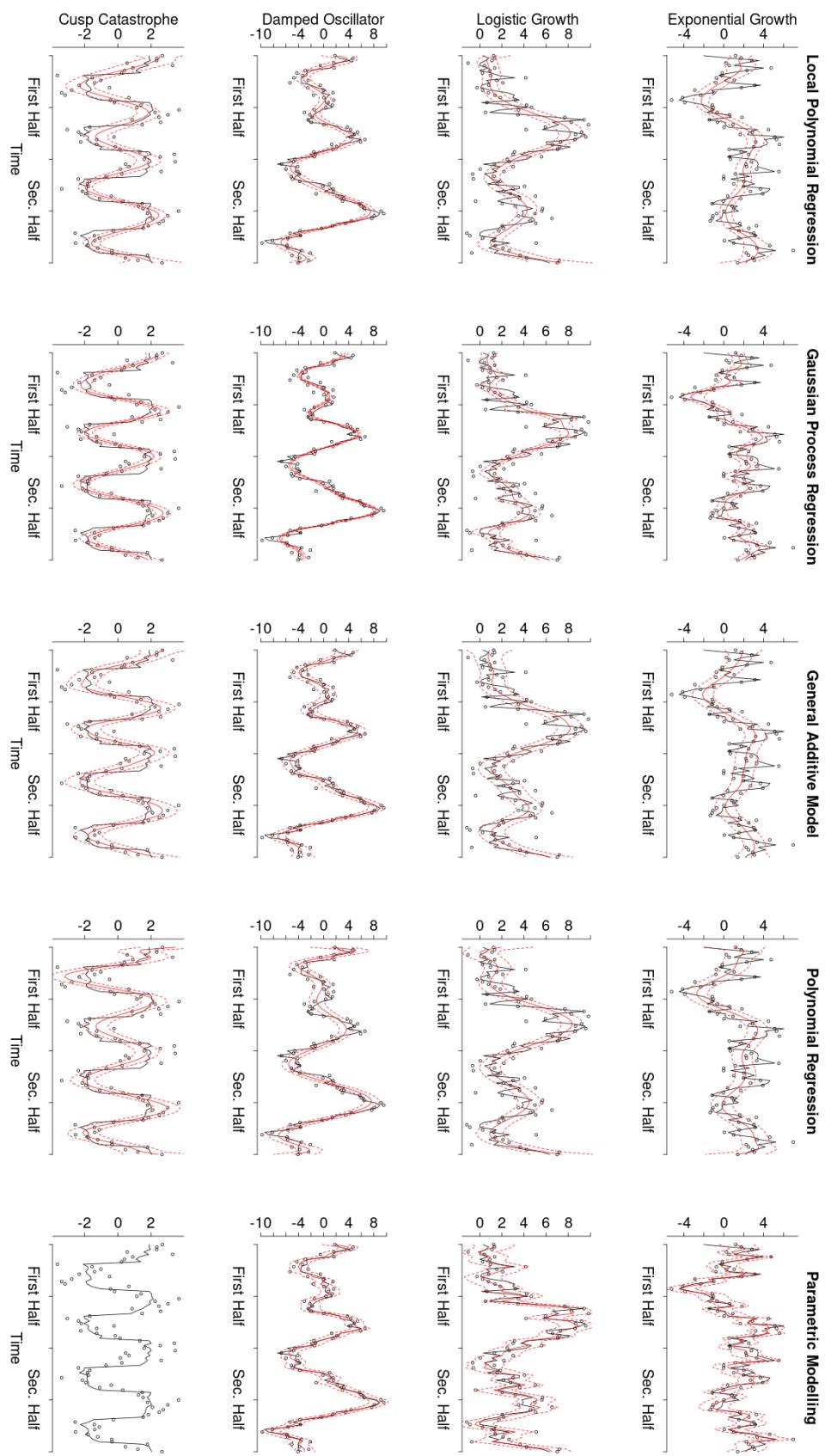
2.7.1 *Capturing the non-linear process*

It is noteworthy that the point estimates produced by all considered methods were able to visually follow the simulated processes adequately. Figure 6 illustrates an example of each process being inferred by all methods. While the predicted means produced by each method all follow the overall trajectory of the processes there are several observations that can be made based on this figure. First, the local and the global polynomial regression appear to underfit (i.e., oversmooth) for some of the processes. Second, the GP regression produced confidence intervals that are notably narrower than the uncertainty estimates produced by the other methods. Third, while the global polynomial regression generally underfit the processes, it appears to overfit near the boundary (i.e., the beginning and end of the timeseries) resulting in excessive uncertainty. This is a known problematic behavior of the global polynomial regression. 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. Table 2 presents all effects for which the partial- η^2 , indicates at least a small effect size for either the MSE or the GCV. The most influential effects for both the MSE and GCV belonged to the analysis method, the data-generating process, and the dynamic error variance. The main effects of the sampling period and sampling frequency were considerably smaller and comparable in magnitude to some of the first-degree interaction effects. 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.

The mean MSE results are depicted in Figure 7. Panel (a) shows the average MSE with which each method inferred each of the processes, when averaging over all other conditions.

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- η^2 MSE	partial- η^2 GCV
Method	0.39	0.25
Process	0.54	0.40
SP	0.17	0.08
SF	0.15	0.23
DEV	0.56	0.48
Method:Process	0.22	0.11
Method:SP	0.16	0.08
Process:SP	0.06	0.04
Method:SF	0.05	0.01
Process:SF	0.01	0.04
Method:DEV	0.18	0.10
Process:DEV	0.27	0.23
SP:DEV	0.02	0.02
SF:DEV	0.01	0.05
Method:Process:SP	0.07	0.05
Method:Process:SF	0.02	0.01
Method:SP:SF	0.01	0.00
Method:Process:DEV	0.08	0.05
Method:SP:DEV	0.02	0.03
Process:SP:DEV	0.01	0.01
Process:SF:DEV	0.00	0.02
Method:Process:SP:DEV	0.01	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.

There appear to be clear differences in the average MSE with which each method infers all processes on average, illustrating the main effect of the analysis method. 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 MSE values. Similarly, the main effect of the process can be seen in that some processes were inferred with a lower average MSE by all methods. Most notably, the cusp catastrophe was inferred with lower MSE values than the other processes across all methods. Lastly, panel (a) also depicts the interaction effect between the methods and the processes, since the mean MSE differences between the

different analysis methods are not equal across the different processes. For example, the polynomial regression displayed a noticeably larger mean MSE for all processes except the cusp catastrophe in comparison to the more advanced statistical methods.

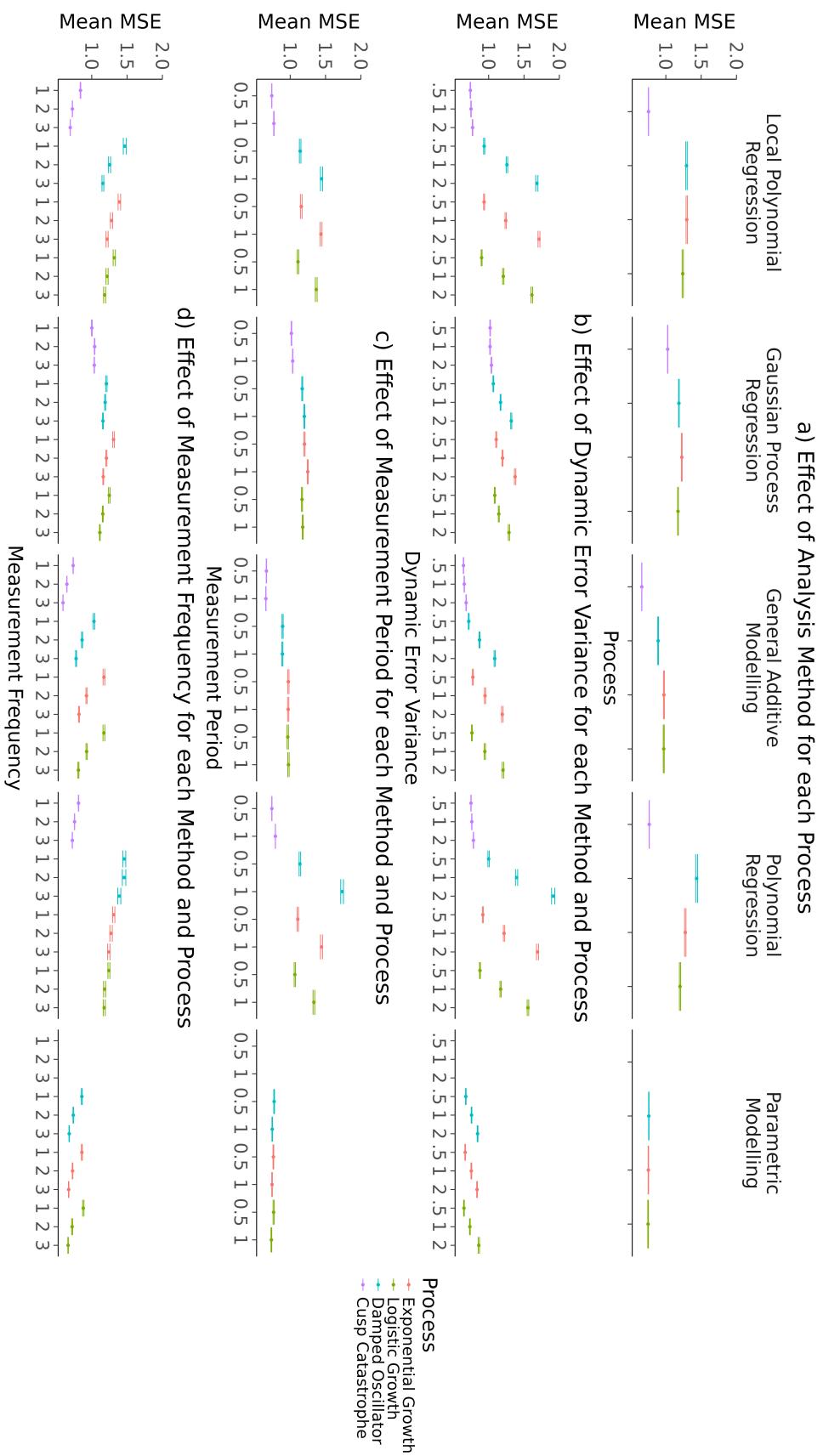
Figure 7 (b) illustrates the effect of increasing dynamic error variances for all methods and processes (averaged over all sampling periods and frequencies), which is the largest main effect among the simulation conditions. A larger dynamic error generally lead to larger average MSE values. However, this effect was considerably less pronounced for the parametric models, GPs, GAMs, and the cusp-catastrophe process, highlighting the interaction between the dynamic error variance and the method or process respectively. Panel (c) shows the effect of the sampling period for each method and process (averaged over all dynamic error variances and sampling frequencies). Here it can be seen that sampling over the entire period, rather than just the first half, resulted in larger mean MSE values for the local and global polynomial regression for all processes except the cusp-catastrophe. This effect was (almost) absent for the GP regression, absent for the GAMs, and reversed for the parametric models. Lastly, Figure 7 (d) shows that the mean MSE generally decreased with larger sampling frequencies for each method and process (averaged over all dynamic error variances and sampling periods).

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 as 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 the dynamic error variance, measurement period, and frequency on the mean GCV appear to follow largely the same patterns that were observed for the mean MSE.

2.7.2 *Uncertainty quantification*

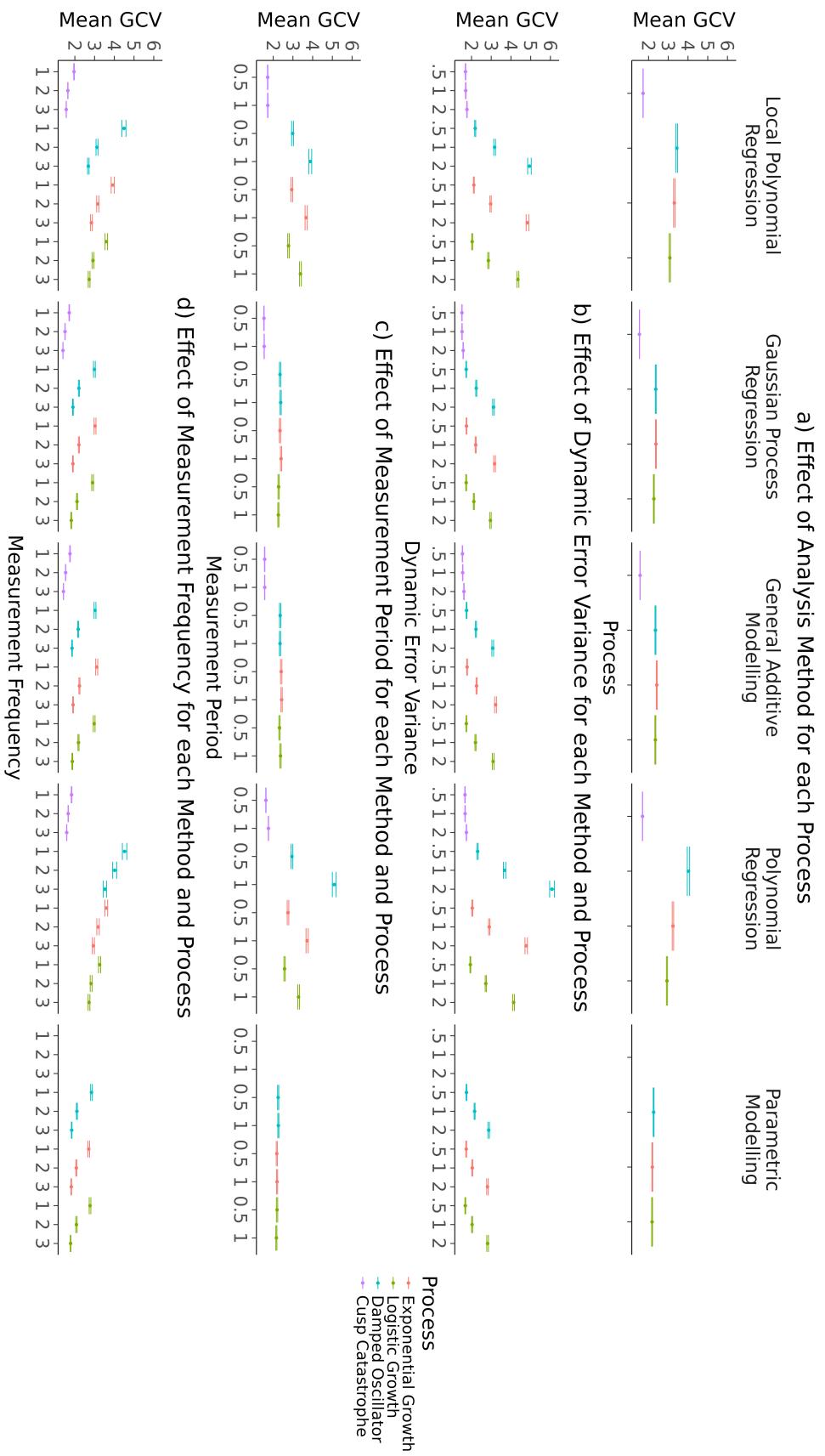
Figure 9 shows the average confidence interval coverage proportion for the conditions described above. The grey area represents an average confidence interval coverage between 89%

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.

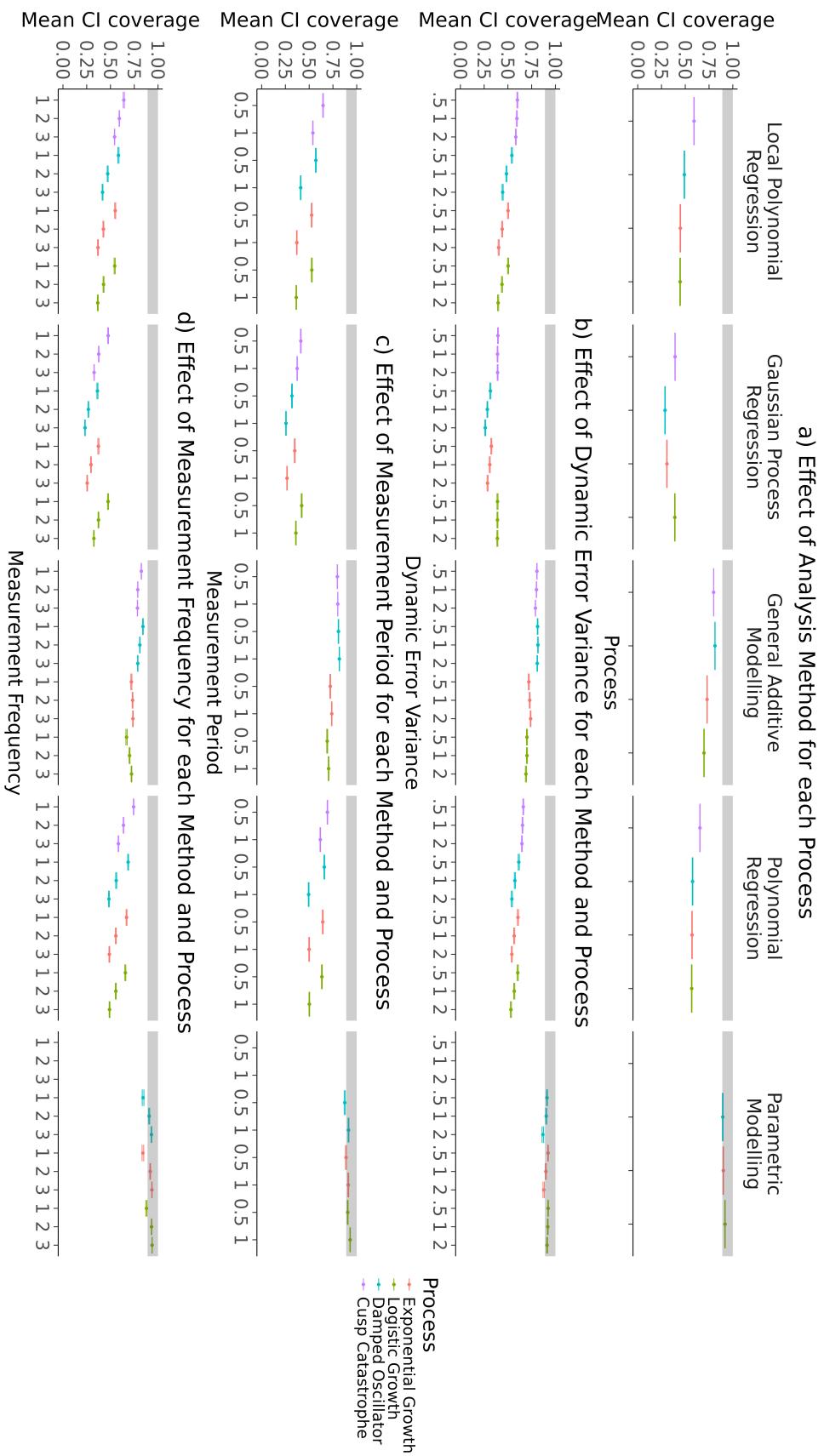
and 100%, which indicates no considerable deviation from the ideal 95% given the Monte Carlo error of the simulation. Only the parametric models produced some mean confidence interval coverages within this area. Among the other methods, the GAMs, produced the largest average confidence interval coverage,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

2.8 Conclusion

The simulation showed that the GAMs inferred all processes with the most accuracy among the considered methods, in their default configuration, 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. From this follows, that GAMs are an attractive starting point for modelling unknown non-linear processes in practice. Especially, when there is little prior theory about the functional form of the process.

Additionally, the simulation revealed that larger dynamic error variances reduced the accuracy of all methods. Therefore, addressing sources of dynamic error in practice is recommended. This can be achieved, for example, by measuring and accounting for contextual variables and other sources of external disturbances. The results also indicate that increasing the sampling frequency improved the performance of all methods, making it generally advantageous to measure processes more frequently. However, in practice, this must be weighed against considerations like participant burden and fatigue, which can negatively impact data quality. Lastly, the simulation showed that extending the sampling period enhanced the performance of the parametric models but may decrease the accuracy of LPR and polynomial regression. This reduction in accuracy for the LPR could potentially be mitigated by using extensions that allow for variable bandwidths or polynomial degrees.

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.

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 117². 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 depression³.

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

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.

³ 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.

underlying the data. On the one hand, Figure 10 (a) shows the ten estimated processes with the lowest EDF, which appear to be essentially linear. In fact, 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.

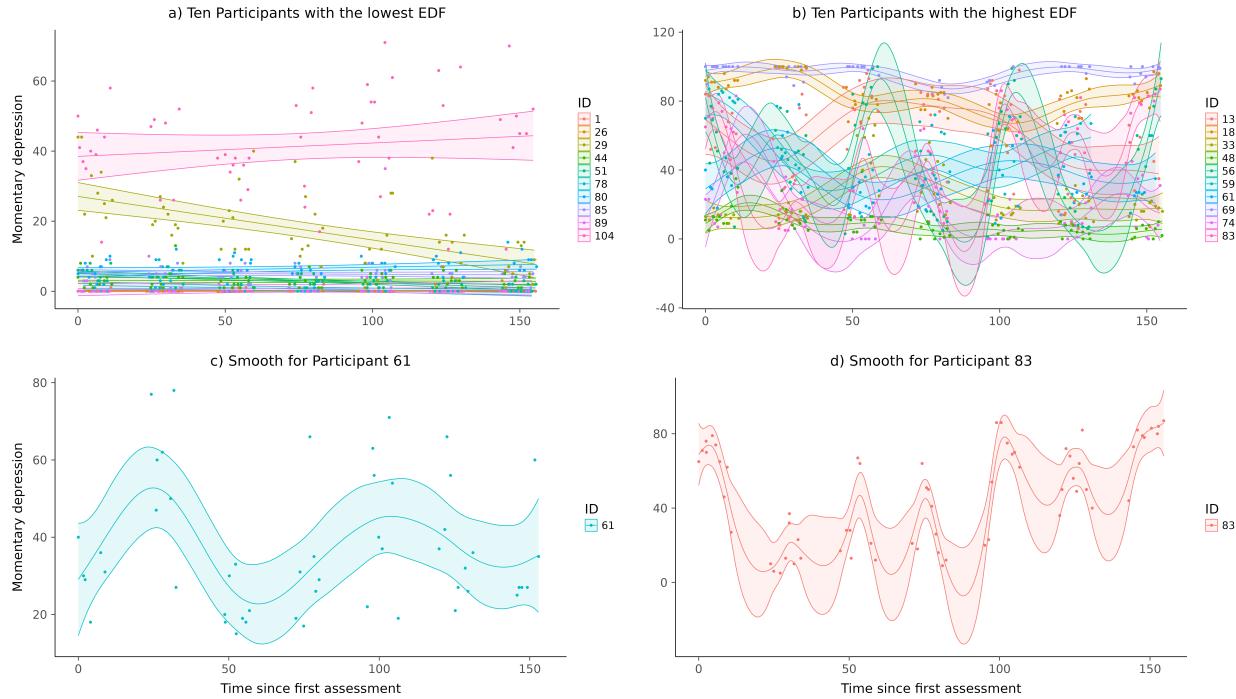
Figure 10 (b), on the other hand, shows the ten participants with the wiggliest estimated processes, indicated by the largest EDF. For these participants, the estimated processes clearly deviates from a linear trend. Additionally, visual inspection of the estimated non-linear processes, did not reveal any common behaviors across the processes of different participants. Nevertheless, the idiographic processes still displayed interesting behaviors. 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, Figure 10 (d) illustrates the inferred process of another participant that may be switching between a low and a high depression state, with considerable oscillations within each state.

3.2 Conclusion

This example demonstrates how GAMs can be used to explore potentially non-linear processes and identify possible empirical phenomena. In this depression data, the most prominent finding is that the GAMs inferred a linear trend for the majority of participants. For some participants, this linear trend can be attributed to floor effects, but not for all. Moreover, the linear estimates suggest the absence of dynamic errors, which would otherwise result in deviations from the linear trend. This could, for instance, indicate that these participants are not reactive to external influences.

For the remaining participants, the GAMs estimated processes that deviated from a linear

Figure 10
Estimated depression processes



trend. However, this is not definitive evidence of non-linearity, as the observed deviations could also be explained by stronger dynamic errors in these individuals. This suggests considerable heterogeneity within the sample, either in terms of the linearity of the idiographic trends or in individual reactivity to external influences. Finally, the processes of certain participants exhibited interesting behaviors, such as oscillations and potential mean-level switching. However, these behaviors did not generalize to the rest of the participants.

4 Discussion

In this paper, we introduced three advanced semi- and non-parametric regression techniques for estimating non-linear processes in psychological ILD (Section 1). These methods address many of the limitations inherent in polynomial and piecewise spline regression, which are currently the most common approaches for modeling non-linearity in psychology. A simulation study (Section 2) further showed that the introduced methods inferred the types of non-linear

processes that could be found in psychological ILD more accurately than a polynomial regression. Particularly, GAMs closely approximated the underlying processes, performing almost as well as the data-generating parametric models. Finally, we showed how GAMs can be used to explore idiographic processes and identify potential phenomena from data (Section 3). This comprehensive analysis empowers psychological researchers to accurately model non-linear processes and to select analysis methods that align with their research goals and data characteristics.

The results obtained in the simulation aligned with almost all of our prior hypotheses. The first hypothesis that was falsified is that the cusp-catastrophe process was inferred most accurately by all methods. We had anticipated that the smooth, continuous estimates produced by all methods would struggle to adapt to the jumps exhibited by this process. However, this effect seems to have been overshadowed by the cusp-catastrophe's 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 the cusp-catastrophe⁴. The second hypothesis that was in partial misalignment with the data is that we expected the GP to perform worse for longer sampling periods (for the two growth curves and the damped oscillator). This effect was only pronounced for the confidence interval coverage and (nearly) absent for the mean MSE and GCV.

Regarding the presented simulation results, it is important to note that the observed differences in performance may be mainly caused by the specific configurations used for each method, rather than the general methods themselves. The specific configurations of each method were chosen to reflect how each method is most commonly applied in practice and not to

⁴ Further, rerunning the simulation with considerably smaller dynamic error variances resulted in the cusp-catastrophe being inferred least accurately.

optimally infer the simulated processes. Consequently, different configurations and extensions are likely to improve the performance of the LPR and GP⁵. For example, using a GP with a Matern class kernel, which makes less strict smoothness assumption about the process, could be expected to improve the accuracy for the rough processes considered in this simulation. Therefore, if there is prior knowledge about the form of the process available in practice, GPs are a very interesting modelling approach, due to their flexibility and interpretable parameters. 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). However, it is unclear whether these types of processes can be found in psychology. Lastly, the tested polynomial regression inferred the underlying processes least accurately, highlighting the limitations of this approach. The numerical model instability of large polynomial regressions can be partially mitigated by using orthogonal polynomials, which are more difficult to interpret. Based on the results presented in this paper, there appears to be no reason to use a global polynomial regression instead of one of the other presented methods in any situation in which prior theory does not strongly suggest that the process follows a low order polynomial trajectory.

The results presented in this article are accompanied by several important limitations. First, only a limited number of processes was tested in the simulation and it is possible that the presented results do not generalize to other processes. While the theory underlying each method guarantees a good performance for processes that satisfy their respective assumptions, it is generally not known how each method performs for other processes that violate their assumptions. Second, while it is necessary to use the GCV instead of the classical leave-one-out cross validation criterion to reduce the computational load of the simulation, the GCV may inadvertently favor the GP regression and the parametric models. This is because the analytic

⁵ Especially, since the GAMs used in this simulation can be expressed as a (unconventional) GP with a linear mean function and a non-stationary covariance function (Rasmussen & Williams, 2006; Wahba, 1978), which implies that there exists a GP configuration which performs at least as well as the GAMs.

form of the GCV implicitly requires certain parameters (such as the hyperparameters of the GP covariance function) to be estimated on the entire sample and then be fixed (to these values) during the cross-validation. This likely improves the ability of the GP and the parametric models, for which this occurs, to predict the omitted data points.

Finally, the scope of this paper was restricted to univariate single-subject data with independent normally distributed measurement errors. This somewhat idealized scenario was chosen to introduce all methods accessibly within the constraints of available software. However, this does not address many of the goals and challenges researchers are facing when working with ILD, which frequently contains complex measurements on several psychological variables for multiple individuals. Fortunately, there are many extensions available for the introduced methods, adapting them closer to the needs of ILD. For instance, GAMs and GPs can be naturally extended to multilevel data (Karch et al., 2020; Wood, 2006). Similarly, GP regression can be used to model different relationships between the observations and the latent process by changing their associated likelihood to a psychometric model, such as a factor model (Clark & Wells, 2023; Yu et al., 2009).

There are also many ways in which the present methods can be used to study multivariate data and structural relations between different psychological variables. This is because, although all methods were introduced in the context of inferring non-linear processes, they can theoretically be used to estimate a wide range of non-linear functions. This makes it possible to infer, for example, non-linear cross- and autoregressive relationships from data in discrete time (Bringmann et al., 2015; Eleftheriadis et al., 2017; Rasmussen & Williams, 2006; Wood, 2006) and non-linear differential equation models in continuous time (Yildiz et al., 2018). These models also present the exciting possibility to combine partial parametric models with non-linear data driven functions, to test incomplete theories.

While these extensions and possibilities exist in theory and are individually implemented in software, most are not yet available in a form that allows researchers to flexibly and accessibly adapt these methods to the specific characteristics of ILD. Therefore, it is essential that future

research unifies these methods and their extensions in software designed for the applied modeling of ILD.

References

- Betancourt, M. (2020). Robust Gaussian Process Modeling. Retrieved June 9, 2023, from
https://betanalpha.github.io/assets/case_studies/gaussian_processes.html
- Boker, S. M. (2012). Dynamical systems and differential equation models of change. In H. Cooper, P. M. Camic, D. L. Long, A. T. Panter, D. Rindskopf, & K. J. Sher (Eds.), *APA handbook of research methods in psychology, Vol 3: Data analysis and research publication.* (pp. 323–333). American Psychological Association.
<https://doi.org/10.1037/13621-016>
- 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>
- 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 (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>
- 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., 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).

- Karch, J. D., Brandmaier, A. M., & Voelkle, M. C. (2020). Gaussian Process Panel Modeling—Machine Learning Inspired Analysis of Longitudinal Panel Data. *Frontiers in Psychology*, 11, 351. <https://doi.org/10.3389/fpsyg.2020.00351>
- 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>
- 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.
- Roberts, S., Osborne, M., Ebdon, M., Reece, S., Gibson, N., & Aigrain, S. (2013). Gaussian processes for time-series modelling [Publisher: Royal Society]. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 371(1984), 20110550. <https://doi.org/10.1098/rsta.2011.0550>
- Ruppert, D., & Wand, M. P. (1994). Multivariate Locally Weighted Least Squares Regression. *The Annals of Statistics*, 22(3). <https://doi.org/10.1214/aos/1176325632>
- 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. (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
- Yu, B. M., Cunningham, J. P., Santhanam, G., Ryu, S. I., Shenoy, K. V., & Sahani, M. (2009). Gaussian-Process Factor Analysis for Low-Dimensional Single-Trial Analysis of Neural Population Activity. *Journal of Neurophysiology*, 102(1), 614–635.
<https://doi.org/10.1152/jn.90941.2008>