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

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

Tilburg University

Author Note


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

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

Abstract

Here could your abstract be!

Keywords: Here could your keywords be!

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

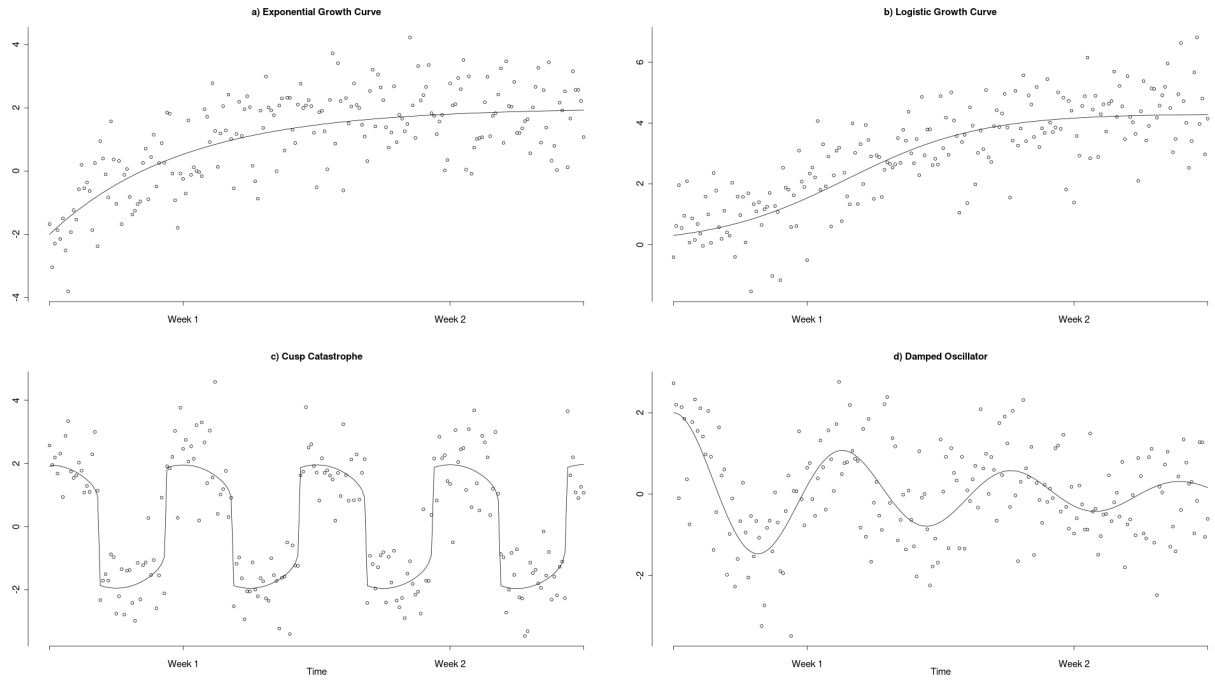
Psychological constructs are increasingly understood as components of complex dynamic systems (Nesselroade & Ram, 2004; Wang et al., 2012). This perspective emphasizes that these constructs fluctuate over time and within individuals. To study these variations and the underlying processes, researchers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

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

Another common non-linear phenomenon is that the construct of interest switches between distinct states, which often correspond to different mean levels. This occurs, for example, during the sudden perception of cognitive flow, where individuals abruptly switch from a 'normal' state to a flow state and back (Ceja & Navarro, 2012). Another example is alcohol use relapse, where patients suddenly switch from an abstinent state to a relapsed state (Witkiewitz &

Figure 1

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



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

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

As a final example, one may consider (self-) regulatory systems, which maintain a desired state by counteracting external perturbations. In these systems the regulatory force often depends on the distance between the current and the desired states. The common autoregressive model describes such a system in which the regulation strength depends linearly on this distance. However, this relationship may also be non-linear, such that the regulatory force changes disproportionately with larger mismatches. Such a (self-) regulatory model has been used to model, for example, emotion regulation (Chow et al., 2005) using a damped oscillator model.

This model is exemplified in Figure 1 (d).

Although initial evidence for non-linearity in psychological research exists, theories about the nature and form of non-linear psychological processes remain scarce (Tan et al., 2011). Frequently, psychological theories are too general (Oberauer & Lewandowsky, 2019) to predict specific non-linear dynamics and ILD studies could have the potential to refine these theories through a nuanced understanding of how the involved psychological variables interact over time. Such refined theories could, for instance, take the form of formal dynamic models, such as differential equation (Boker, 2012) or state-space models (Durbin & Koopman, 2012), describing how a given process changes over time. However, in order to develop these types of theories it is first necessary to identify phenomena in ILD, which are replicable and empirically observable features of the underlying processes, such as the described state switching or regulatory oscillations. Formal theories about the underlying process should then be able to explain these phenomena and different candidate theories can be compared on their success to do so (Borsboom et al., 2021). While the study of non-linear phenomena in ILD is receiving increasingly more attention in psychology and different statistical techniques are developed to explore these phenomena (Cui et al., 2023; Humberg et al., 2024), researchers are currently still limited in their ability to infer non-linear phenomena from ILD due to a lack of advanced statistical methods that are flexible enough to adequately capture and explore these processes, which hinders the development and evaluation of guiding theories. Due to this lack of adequate available statistical methods, non-linear trends are most often addressed in psychology through polynomial regression or 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, such as kernel regression, Gaussian processes, and smoothing splines are available outside of psychology. However, these methods have rarely been applied in psychology because they have not been reviewed for an applied audience, nor have their assumptions and inference possibilities been evaluated in the context of ILD. As a result, psychological researchers struggle to select the most suitable method for a specific context. This challenge is further complicated by the fact that the ideal statistical method may depend on the characteristics of the underlying non-linear process, which are generally unknown. Especially, since the assumed smooth processes for which many of these methods were originally developed are unlikely to occur in psychological research.

To address this important gap, this article reviews three advanced non-linear analysis methods and evaluates their applicability to typical ILD scenarios (Section 1). The methods reviewed in this article are different semi- and non-parametric regression techniques that are able infer non-linear functions from data while accommodating varying degrees of prior knowledge. We compare how well each method can recover different processes under common ILD conditions in a simulation study (Section 2). Lastly, we also demonstrate how some of the

introduced methods can by 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

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 process value that the LPR predicts at a specific time point, the data is first centered around that point (by shifting the data along the time axis so that the chosen time point is at zero), and a low-order polynomial is fitted around it. Additionally, to account for the fact that the polynomial approximation is more accurate for data points closer in time, a weighting function is applied during the polynomial estimation, which assigns weights to each data point based on its distance from the point of interest. The value that the LPR predicts for the chosen time point is then given by the intercept of the locally weighted polynomial at this point in time.

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^*[1]. 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^*.

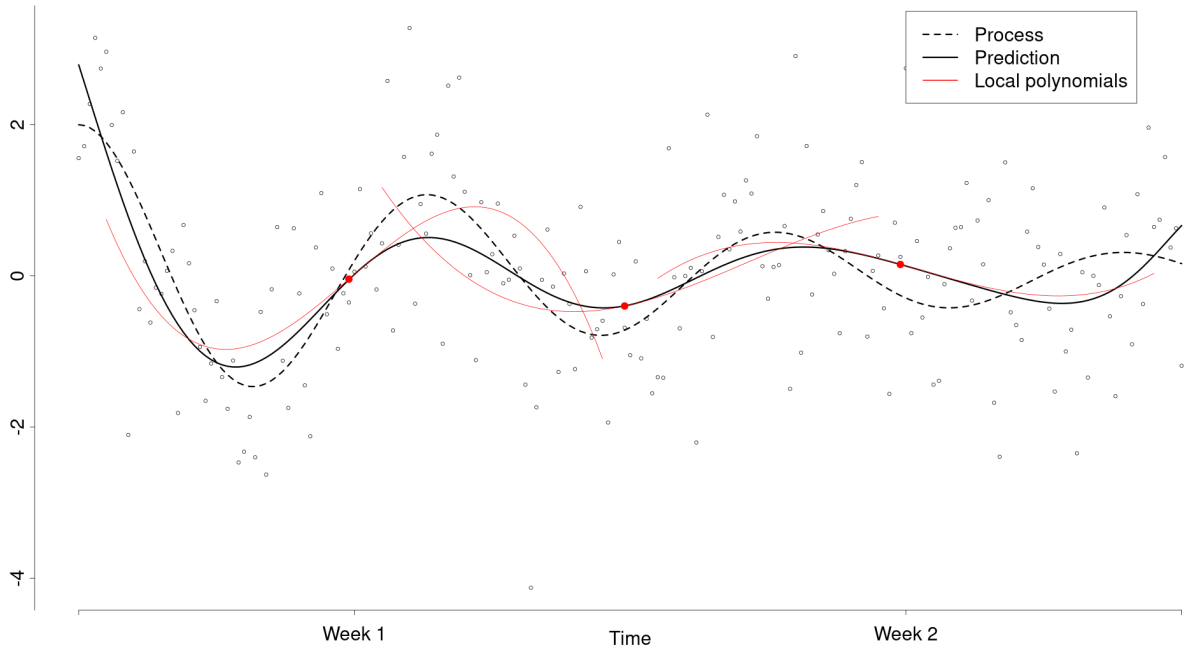
To find the process value that the LPR predicts for a different time point, the same procedure can be repeated, centering the data around the new point of interest. Since it is theoretically possible to repeat this process at infinitely many time points, LPR is a non-parametric technique. Figure 2 shows the estimated LPR for the damped oscillator example process introduced in Figure 1 (d). In this figure, three truncated examples of local cubic regressions are shown in red, which contribute to the overarching LPR for this process.

When fitting an LPR, several decisions must be made regarding the degree of the local polynomials and the optimal weighting of the data. Typically, the degree of these local polynomials is kept low and odd. This choice reflects a bias-variance tradeoff, where higher-order polynomials reduce bias but increase variance only when transitioning from an odd to an even

[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 exemplary local cubic regressions (red) are shown that provide the values that the LPR predicts at the time points 50, 100, and 150. The complete LPR was plotted by evaluating such local cubic regressions at 200 points along the time axis.

power (Ruppert & Wand, 1994). The data weighting in an LPR is achieved through a kernel function of the form $w_{i,i} = K(\frac{t_i - t^*}{h})$, which is usually centered and symmetric to assign weights based on the distance from the origin. Common kernel choices include the Gaussian and Epanechnikov, with the latter being optimal in many applications and aspects (Fan et al., 1997). Each kernel is further defined by a bandwidth parameter h, which determines its width and effectively controls the influence of more distant data points. The bandwidth parameter represents the wiggliness of the estimated process in practice. Several methods are available to find the optimal bandwidth by optimizing a data-dependent criterion function, such as the cross-validation error or the mean integrated squared error (Debruyne et al., 2008; Köhler et al., 2014).

Due to its non-parametric nature, LPR makes minimal assumptions about the data.

However, it does require that the underlying process is p times differentiable, which is a necessary condition for local polynomial approximation (under Taylor's theorem). It is further noteworthy that unless the process follows a polynomial of at most degree p the approximation with local polynomials is biased. However, this bias is usually negligible and there are methods available to correct for it (Calonico et al., 2019). Another key assumption is that the process has constant wiggliness, represented by a single bandwidth parameter. However, this assumption may be relaxed by using a time-varying bandwidth (Fan & Gijbels, 1995b) or polynomial degree (Fan & Gijbels, 1995a), but this extension is beyond the scope of this paper.

1.2 Gaussian process regression

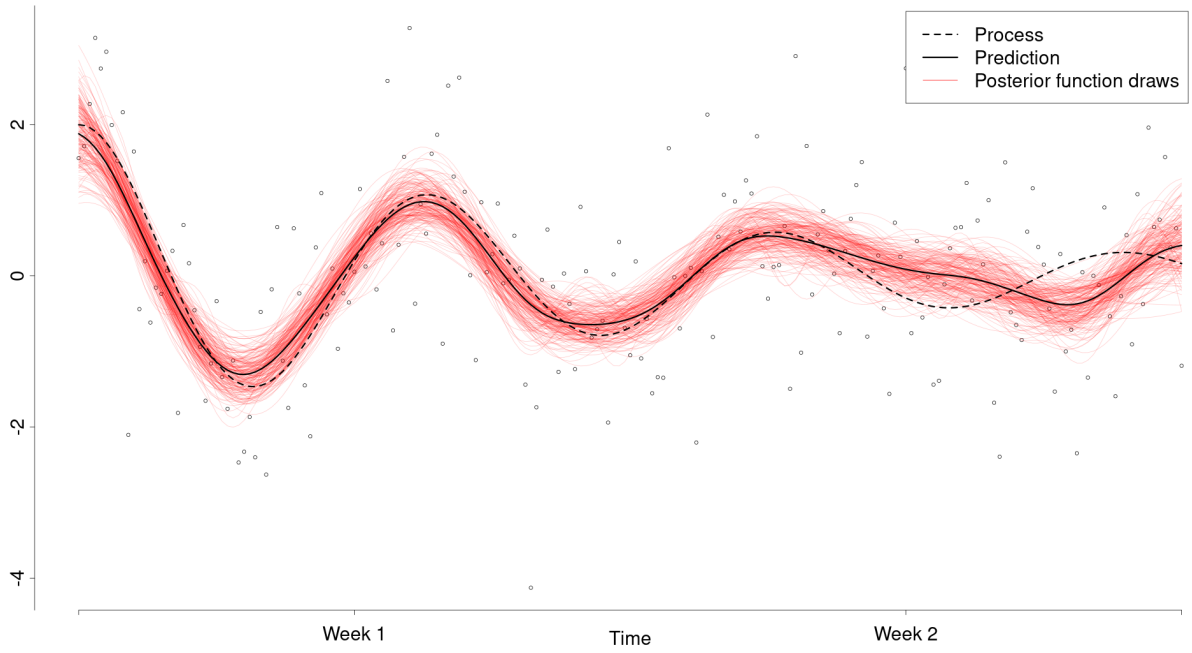
The second non-parametric technique is Gaussian process (GP) regression, a Bayesian approach that directly defines a probability distribution over an entire family of non-linear functions, which is flexible enough to capture many complex processes effectively (Betancourt, 2020; Rasmussen & Williams, 2006; Roberts et al., 2013). Unlike regular probability distributions (e.g., normal distribution) that specify the plausibility of single values, Gaussian processes determine the plausibility of entire (non-linear) functions. A GP is defined such that the functions it describes, take values at any finite set of time points that 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.

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

This posterior distribution represents an updated belief about which functions describe the latent process well (Kruschke, 2011), allowing one to draw inferences about the process itself. Figure 3 illustrates such a posterior distribution for the running example process. The red lines represent a sample of non-linear functions drawn from the posterior distribution, such that the pointwise average of these functions provides a mean estimate for the underlying process.

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.

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 time points depending on their distance (e.g., quadratic exponential, Matern class kernels), such that

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

For some combinations of GP priors and likelihoods, is possible to derive an analytic posterior distribution for a GP. One example of this is a GP prior with a zero mean function and a Gaussian likelihood with a zero mean and a standard deviation of σ. Then the posterior process values at a selected set of time points $\mathbf{t}^*$ are distributed according to a multivariate normal distribution with a mean vector and covariance matrix of:

$$\begin{aligned} f(\mathbf{t}^* \,|\, \mathbf{y}) &\sim MvN(\mu(\mathbf{t}^*), \Sigma(\mathbf{t}^*)) \\ \mu(\mathbf{t}^*) &= K(\mathbf{t}^*, \mathbf{t})[K(\mathbf{t}, \mathbf{t}) + \sigma^2 \mathbf{I}]^{-1} \mathbf{y} \\ \Sigma(\mathbf{t}^*) &= K(\mathbf{t}^*, \mathbf{t})[K(\mathbf{t}, \mathbf{t}) + \sigma^2 \mathbf{I}]^{-1} K(\mathbf{t}, \mathbf{t}^*) \end{aligned} \tag{4}$$

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

Lastly, the mean and covariance function typically contain parameters themselves, which can be treated as hyperparameters and can be estimated as such by specifying appropriate priors. These hyperpriors reflect prior beliefs about the hyperparameters and can be used to constrain them to sensible values. The corresponding posterior distributions then provide interpretable inferences for the hyperparameters. Most common covariance kernels build on a characteristic lengthscale and a marginal standard deviation parameter. The characteristic lengthscale effectively determines the wiggliness 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 model comparison.

Whereas LPR is a mainly data driven procedure GP regression is more model based. 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. One example of this could be to combine in one model a linear mean function with non-linear deviations captured by a GP. However, this also means that to accurately capture a process, it is important that the functional family generated by a GP is similar to the actual process. Most common choices for covariance kernels for example result in smooth and covariance-stationary GPs with constant wiggliness. Another difference between GP regression and LPR is that the Bayesian estimation underlying the GP regression provides a natural approach to uncertainty quantification. 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 mitigate the knot placement problem by providing enough flexibility to the spline 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. Here, using as many basis functions as there are data points is analogous to placing a knot at each data point. To prevent the resulting overfitting, smoothing splines add an additional penalty term, similar to those used in a lasso or ridge regression, to control the smooth term's flexibility during the estimation (Gu, 2013; Wahba, 1980). This penalty balances the flexibility and fit of the smooth term, ensuring the model captures the underlying process accurately without introducing unnecessary complexity. The optimal weight of the penalty is typically determined by minimizing a criterion function, such as the generalized cross-validation

criterion or thorough likelihood maximization (Golub & von Matt, 1997; Wood, 2006).

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

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

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 wiggliness of the estimate, measured by the integrated squared second derivative of the smooth term. Here, λ 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 α_k.

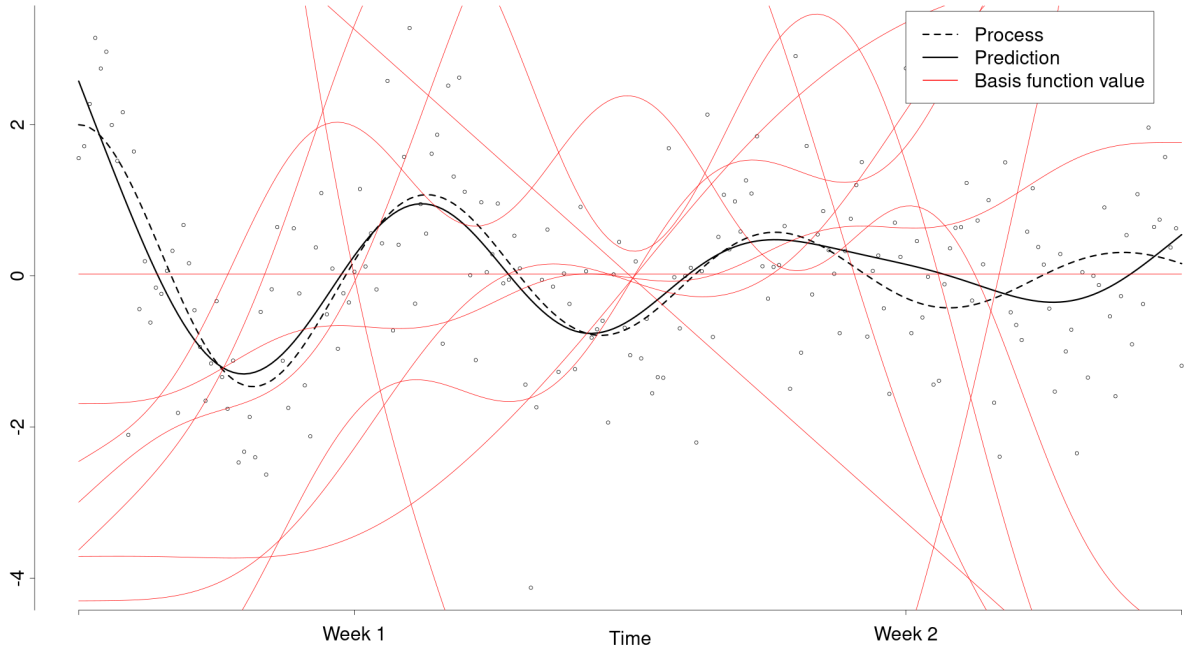
GAMs extend on the smoothing spline approach by making it possible to combine multiple smooth terms in an overarching additive regression model, where each smooth term essentially functions as a predictor within a regular regression analysis. In this model, smooth terms (with potentially different input variables) may be multiplied by covariates x_j and summed into a single overall non-linear function f, which estimates the process.

$$\begin{aligned} \hat{f}(t) &= \underset{\alpha}{\operatorname{argmin}}\, P(\mathbf{y} \mid f(t)) + \sum_{j=1}^{J} \lambda_j \int (\beta_j(t)'')^2 dt \\ f(t) &= \sum_{j=1}^{J} \beta_j x_j \\ \beta_j(t) &= \sum_{k=1}^{K} \alpha_{j,k} R_{j,k}(t) \end{aligned} \tag{6}$$

This approach makes it possible to formulate models such as a time-varying autoregressive model, where the intercept and autoregressive parameters are smooth terms of time (Bringmann et al., 2015; Bringmann et al., 2017). By integrating non-parametric smooth terms into a broader parametric model, GAMs become semi-parametric models that are well-suited for testing specific hypotheses while keeping the flexibility needed to accurately capture the latent process. Figure 4

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).

illustrates a simple GAM construction with a single smooth term for time, fitted to the example process. The first ten thin-plate smoothing spline basis functions of the nearly 200 basis functions that make up the 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 techniques, as each GAM can be understood as a (unconventional) GP with a partially improper prior[2]. These improper priors make the GAMs more flexible to model processes, which might be outside the functional family generated by a standard GP construction (Wahba, 1978). In addition to this and similar to GPs, GAMs offer an

[2] Any basis function regression with Gaussian priors for the coefficients is a GP in weight space. The smoothing spline underlying the GAMs is such a basis function regression in which the coefficient priors serve the same purpose as the penalty term. Due to the null space of the penalty term, some of these priors are generally going to be imporper.

intuitive way to combine partial theories with data driven 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 of the model. These are a measure of the models effective complexity, where an effective degree of freedom of one corresponds to a linear model. However, in contrast to LPR and the standard GP construction, GAMs do not assume constant wiggliness. Unfortunately, GAMs do not provide a measure of the uncertainty of the effective degrees of freedom. Lastly, it is possible to interpret the basis function coefficients in a GAM. The interpretation of these coefficients does however depend on the chosen basis functions and on the other terms in the model. 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 addition to that, the introduced methods were also compared to a polynomial regression, which is currently used most often to model non-linear trends in psychology. Lastly, to obtain a benchmark for how accurately non-linear processes can be recovered when their parametric form is correctly specified, the data-generating parametric models were also included in the simulation. 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.

Table 1

A comparison of LPR, GP regression and GAMs

	LPR	GP	GAM
Advantages	• Intuitive theory • Completely data driven	• Most interpretable parameters • Natural uncertainty quantification • Flexible modelling framework can incorporate prior theory	• Intuitive theory • Some interpretable parameters • Flexible modelling framework can incorporate prior theory
Disadvantages	• Least interpretable parameters • Biased for most processes	• Unintuitive theory • Difficult to specify in practice	• ?
Extensions	• Variable bandwidth • Variable polynomial degree	• Many in theory • Few implemented in accessible software	• Multivariate input data • Multilevel data
Key assumptions	• P-times differentiable process • Constant wiggliness	• Assumptions depend on chosen specifications	• Smooth process • Homoscedasticity
Estimation	• OLS	• Bayesian	• OLS, MLE, Bayesian
Key sources of information	• Fan and Gijbels (2018)	• Rasmussen and Williams (2006)	• Wood (2006)

2.2 Design and Hypotheses

For the LPR and the GP, we expect that both method, using the configurations in which they are most often applied and implemented in standard software, will most accurately infer processes that are (a) continuous (i.e., without sudden jumps), (b) have constant wiggliness, and (c) are smooth (i.e., differentiable). We expect this, because both methods by default produce continuous, smooth estimates with a single constant bandwidth or lengthscale. For the GAMs, we expect that only criteria (a) and (c) will influence the performance, as GAMs do not assume constant wiggliness. The parametric modeling approach is expected to provide the most accurate inferences, serving as a benchmark for comparison with the other methods. Additionally, we expect that the overall sample size will influence the accuracy of the inferences, with both (d) the overall length of the sampling period and (e) the sampling frequency being varied.

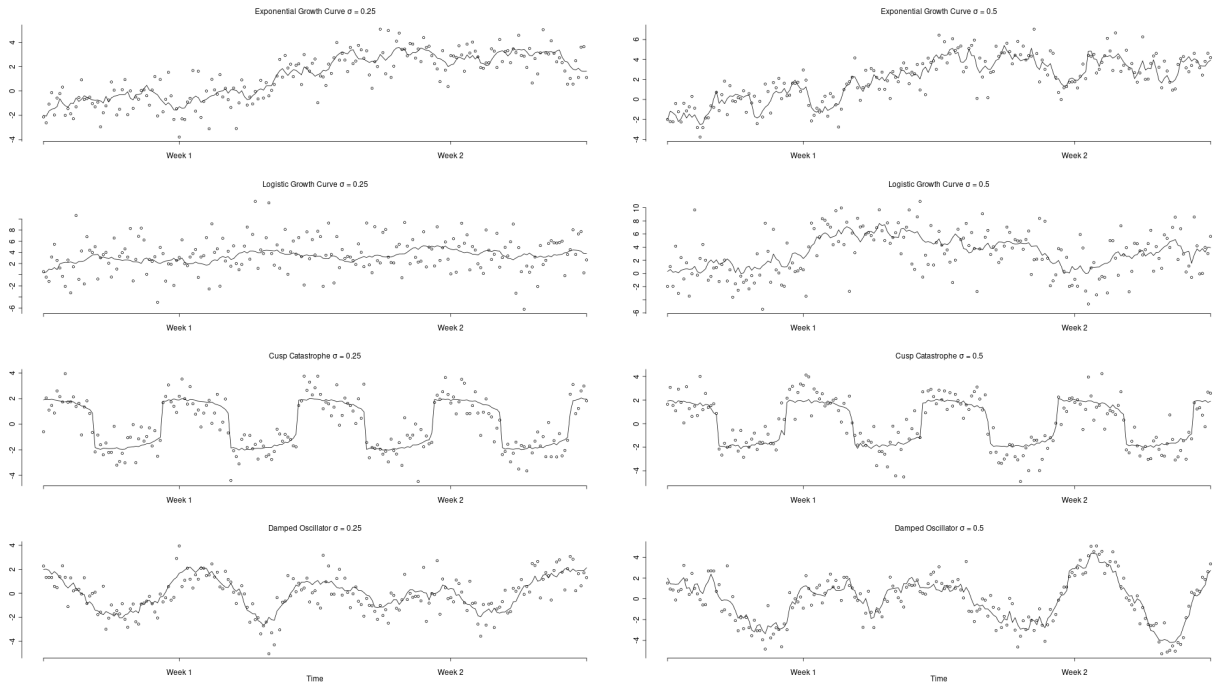
To conduct the simulation with processes that might be encountered in real EMA studies, we selected the exemplar processes illustrated in Figure 1 as a basis. These include two growth curves, modeled as an exponential and a logistic growth curve, a mean-level switching process, modeled as a cusp catastrophe, and a self-regulatory process, represented by a damped oscillator. These processes make it possible to test the impact of (a) sudden jumps and (b) changing wiggliness on the four methods. First, we hypothesize that the cusp catastrophe model, which is the only process featuring jumps, will be least accurately inferred by all methods. Second, all four processes exhibit changes in wiggliness (i.e., changes in the second derivative) over time. However, while the wiggliness of the exponential and logistic growth functions and the damped oscillator decreases monotonically, the cusp catastrophe's wiggliness changes cyclically (i.e., low wiggliness during the plateau phases and very high wiggliness during the jumps). Therefore, we hypothesize that longer sampling periods for the exponential and logistic growth curves and the damped oscillator will reduce the inference accuracy of the LPR and the GP, as the single bandwidth or lengthscale parameter becomes increasingly inadequate to capture the changing wiggliness over time. We do not expect this effect to occur for the cusp catastrophe process, or when using GAMs.

To manipulate the (c) smoothness of the processes a dynamic error component was added to each process. 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 one possible realization of the exemplar processes with dynamic errors. Importantly, we intentionally omitted a condition without dynamic noise from this simulation, as dynamic noise is reasonably expected to be present in all psychological intensive longitudinal data (ILD).

Additionally, the sample size was varied during the simulation by manipulating both (d) the sampling period and (e) the sampling frequency, as these distinct methodological choices are expected to impact the performance of the analysis methods differently. Specifically, for the LPR and the GP, which rely only on data in local neighborhoods during the estimation, we expected that extending the sampling period beyond this neighborhood will not increase the inference accuracy. In fact, if the process exhibits changing wiggliness over the extended period, as previously discussed, increasing the sampling period might even negatively affect the inference accuracy. In contrast to this, we expected the GAMs, which incorporate the entire dataset in their estimation, to perform better with a longer sampling period. Since within the simulation there is no inherent scaling to the time variable, we simulated either only the first half of each process or the entire process to represent different sampling periods. For the ease of reading and to correspond to typical EMA conditions, this will be referred to as sampling over either one or two weeks. However, this scaling is arbitrary and could be changed to any other time frame. Lastly, we expected that increasing the sampling frequency will generally improve the inference accuracy

Figure 5

Non-linear exemplar processes with dynamic errors



Note. This figure shows one possible realization of how the exemplar processes could unfold with dynamic errors. The processes have dynamic error standard deviations of 0.25 (left) and 0.5 (right).

across all methods, since there is more information about the latent process available. Relative to the introduced weekly scale, we tested sampling frequencies of three, six, and nine measurements per day, to cover typical EMA sample sizes (Wrzus & Neubauer, 2023).

2.3 Data generation

To simulate data for each process, they must first be represented as parametric generative models that replicate the structure of psychological time series data. In such a time series, any psychological construct follows a (potentially non-linear) function over time, as depicted by the lines in Figure 1. However, since these psychological processes are typically unobservable or latent, they are measured through observable indicators, such as questionnaire items. The observed values on these indicators (Figure 1, dots) differ from the true values of the latent process due to measurement error, which may come from an imperfect measurement instrument.

For this simulation, we assume that all time-point-specific measurement errors are independent and normally distributed. The model for the observations of a single indicator can then be expressed as follows:

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

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

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

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

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

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

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

The Wiener process is a continuous non-differentiable stochastic process, which describes normally distributed dynamic errors over any given discrete time interval. These errors have a mean of zero and a variance 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.

All four processes were modeled as stochastic differential equations by substituting their respective deterministic dynamics into Equation 8, as detailed in the online supplementary material. Latent process data were 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 from a standard normal distribution, generating the final sets of observations

To determine the number of data sets that should be simulated per condition, a power simulation was conducted based on an initial pilot sample of 30 generated data sets per condition. Based on the pilot sample, the outcome measures (e.g., MSE, GCV, and confidence interval coverage) and their corresponding standard deviations were calculated by condition. These standard deviations were then used to predict the Monte Carlo standard errors of the means of each outcome measure across increasing sample sizes and conditions (Siepe et al., 2023). These Monte Carlo standard errors reflect the expected variation in the outcome statistics due to random processes within the simulation. Based on this analysis, we concluded that 100 replications per cell were feasible while providing sufficiently low maximum Monte Carlo standard errors ($se_{MSE} \approx 0.05$, $se_{GCV} \approx 0.38$, $se_{CIC} \approx 0.03$)

2.4 Model estimation

After simulating the data, all introduced methods were applied to each data set using the statistical software R (R Core Team, 2024). First, the LPR was estimated using the nprobust package (Calonico et al., 2019), which allows to correction for the bias inherent in LPRs. Second, GPs were estimated in STAN (Gabry et al., 2024) with a zero mean and a squared exponential covariance function, following common practice. Third, GAMs with a single smooth term for time were fitted using the mgcv package (Wood, 2011). The polynomial regressions were estimated using base R, with correlated (i.e., standard) polynomial terms. Finally, the parametric differential equation models corresponding to the true data-generating models were estimated using the Dynr package (Ou et al., 2019). While the same non-parametric models were used across all conditions, the parametric models were tailored to each specific data-generating process. After fitting each model to the data, they were used to obtain point and interval estimates (i.e., 95% confidence and credible intervals) for the latent process at each time point. A detailed description of each model fitting procedure is provided in the online supplementary material.

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

2.5 Outcome measures

To evaluate and compare the performance of the different analysis methods, we focused on three outcome measures. The first two assessed each method's accuracy in predicting the process values at or between the observed time points. These predictive accuracy measures indicate how well each method captures the underlying non-linear process. The third outcome

measure evaluated the accuracy of the uncertainty estimates provided by each method. Specifically, whether the confidence or credible intervals produced by each method correctly included the true state value the expected proportion of times.

2.5.1 Capturing the non-linear process

To assess how effectively each method captured the non-linear process at the observed time points, we calculated the mean squared error (MSE) between the estimated and generated process values. Additionally, to evaluate how well each method would predict 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 of these metrics 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 leave-one-out cross-validation is calculated by removing one data point and refitting the model while keeping certain parameters fixed (e.g., the bandwidth of the LPR, the smoothing penalty weight of the GAMs, and the covariance function hyperparameters of the GP). The left out observation is then predicted by the refit model and the squared prediction error is obtained. This procedure is repeated for each data point and averaged. The GCV instead exploits the structure of the used models to obtain the cross-validation errors for optimally rotated data without refitting the models (Golub et al., 1979). Due to this computational efficiency it is optimal for the use in large scale simulation studies.

To analyze whether the simulated samples indicate true differences in the MSE and GCV values across the different processes, analysis methods, and simulation conditions (i.e., sampling period, frequency, and dynamic error variance) two separate ANOVAs were fit. Both ANOVAs included all possible main and interaction effects. However, the parametric models were excluded from the ANOVAs as they were expected to outperform and hinder the comparison of the other methods. The fitted ANOVAs were used to identify which effects were significant and have at

least a small effect size according to the partial η^2 (> 0.01). Since, two separate ANOVAs were fit, the resulting p-values were adjusted for multiple comparisons by multiplying them by two. 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.

2.5.2 Uncertainty quantification

To evaluate the uncertainty estimates provided by each method, we recorded whether the true generated process was located within the confidence or credible intervals at each time point. Subsequently, the average confidence interval coverage proportion for each method and data set was obtained, by averaging over all time points. Given that all confidence or credible intervals were set at a 95% confidence level, the expected coverage proportion should ideally be close to 95%. Due to Monte Carlo error in the simulation ($\max(se_{CIC}) \approx 0.03$) 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.

2.6 Results

In the simulation some of the 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 did 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

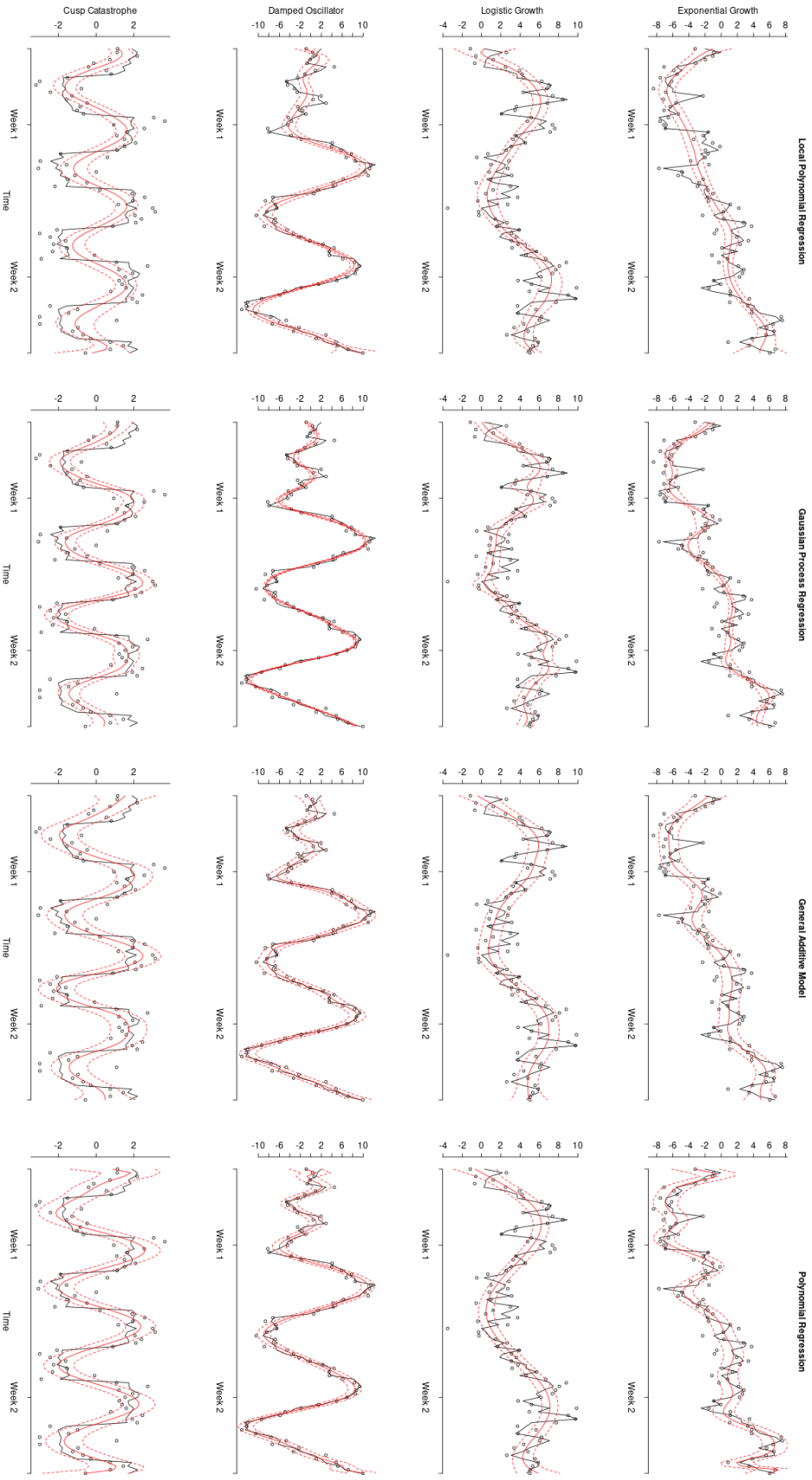
First, it is noteworthy that all considered methods visually performed well in mean predicting the simulated processes. Figure 6 illustrates an example from each process being inferred by all methods. The predicted means produced by each method closely followed the simulated processes, although the LPR appears to underfit for some data sets. Additionally, near the boundaries (i.e., the ends) of the simulated time series, the GP regression sometimes tends towards zero, which is a characteristic of the squared exponential kernel (this can 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 and objective analysis of the performance of each method.

To achieve this, 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-η^2, which gives the proportion of variance explained by an effect after partialling all other effects out, 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

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.

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 process, averaged across sampling periods, frequencies, and dynamic error variances. This figure highlights the main effect of the analysis method, as there are 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. However, this effect was least pronounced for the cusp catastrophe.

Figure 8 displays the corresponding effects for the mean GCV values. Similar to the MSE

Table 2

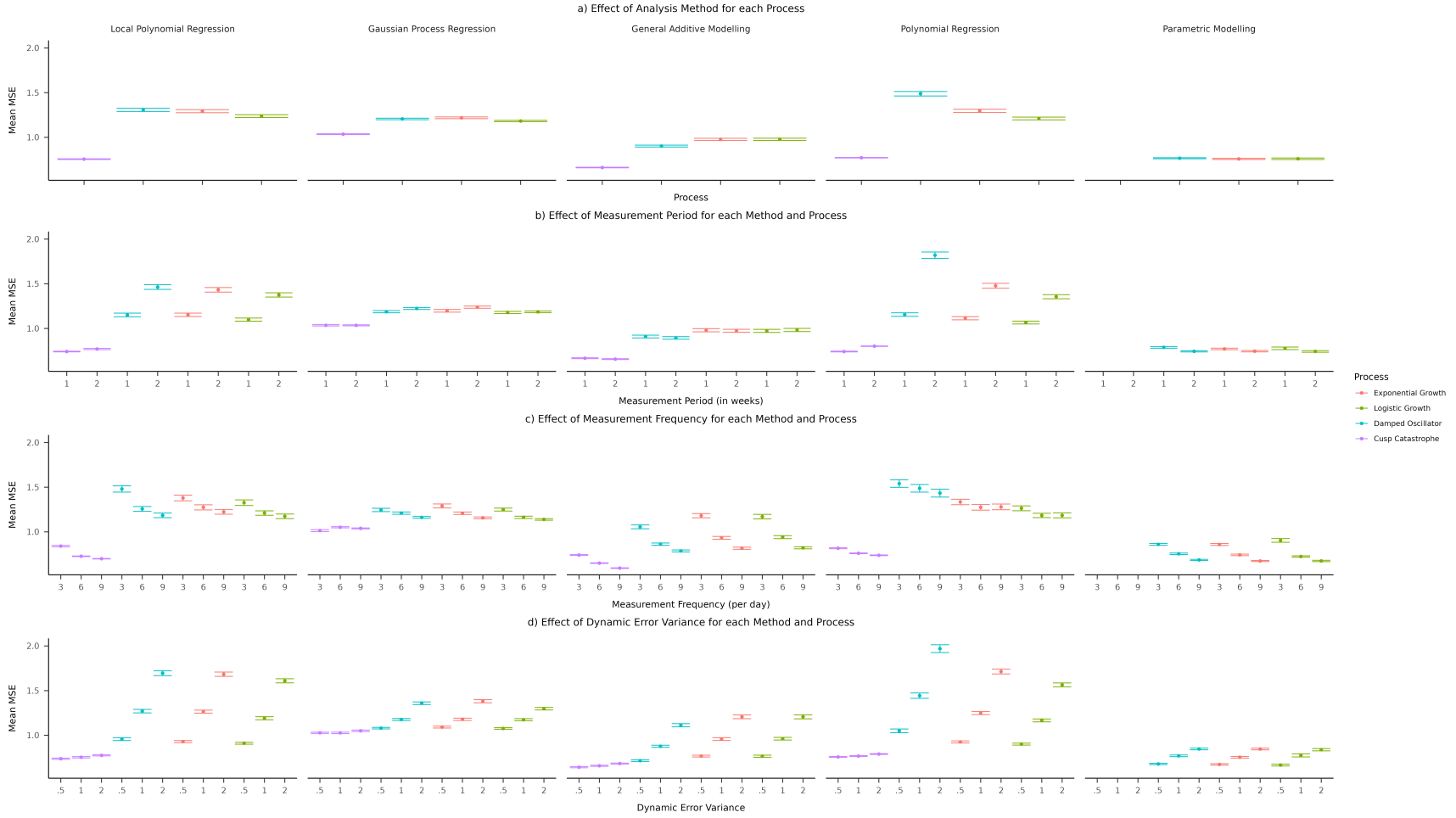
Effect sizes from the MSE and GCV ANOVAs

Effect	partial-η^2 MSE	partial-η^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-$\eta^2 >= 0.01$ on either outcome. SP: Sampling period; SF: Sampling frequency; DEV: Dynamic error variance.

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.

results, the GAMs show a mean GCV value closest to the benchmark parametric models. However, the GPs also display comparable performance to the GAMs in terms of the GCV. 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

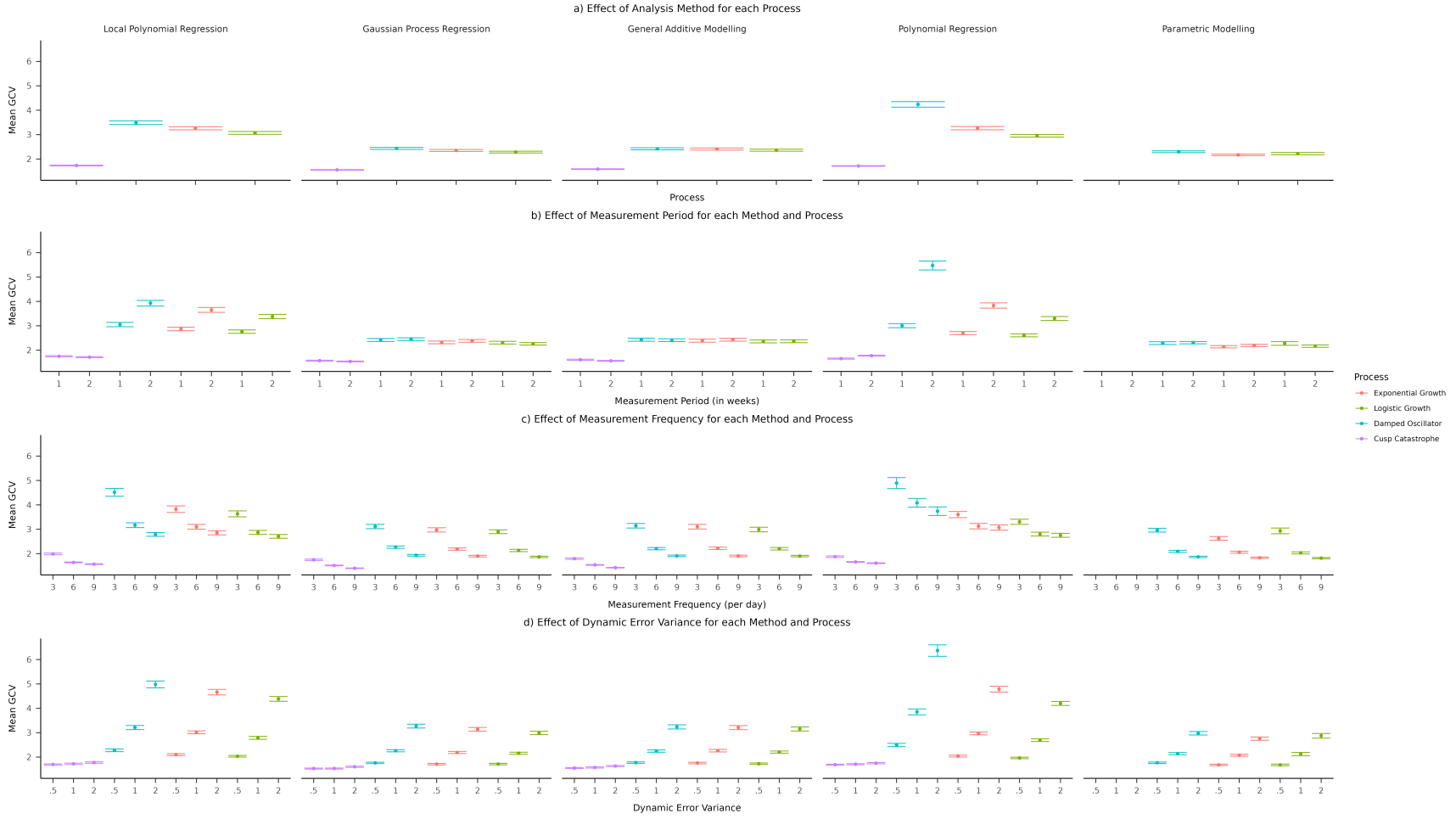
Figure 9 shows the average confidence interval coverage proportion for the conditions described above. The grey area represents an average confidence interval coverage between 89% and 100%, which indicates no considerable deviation from the ideal 95% given the Monte Carlo error of the simulation. Only the parametric models produced 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 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 MS, 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, since the GP had the lost 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 wiggliness, and over the GP, by effectively relying on improper priors, enabled the GAMs to infer the processes more accurately in the simulation. From this follows, that GAMs are an attractive starting point for modelling

Figure 8

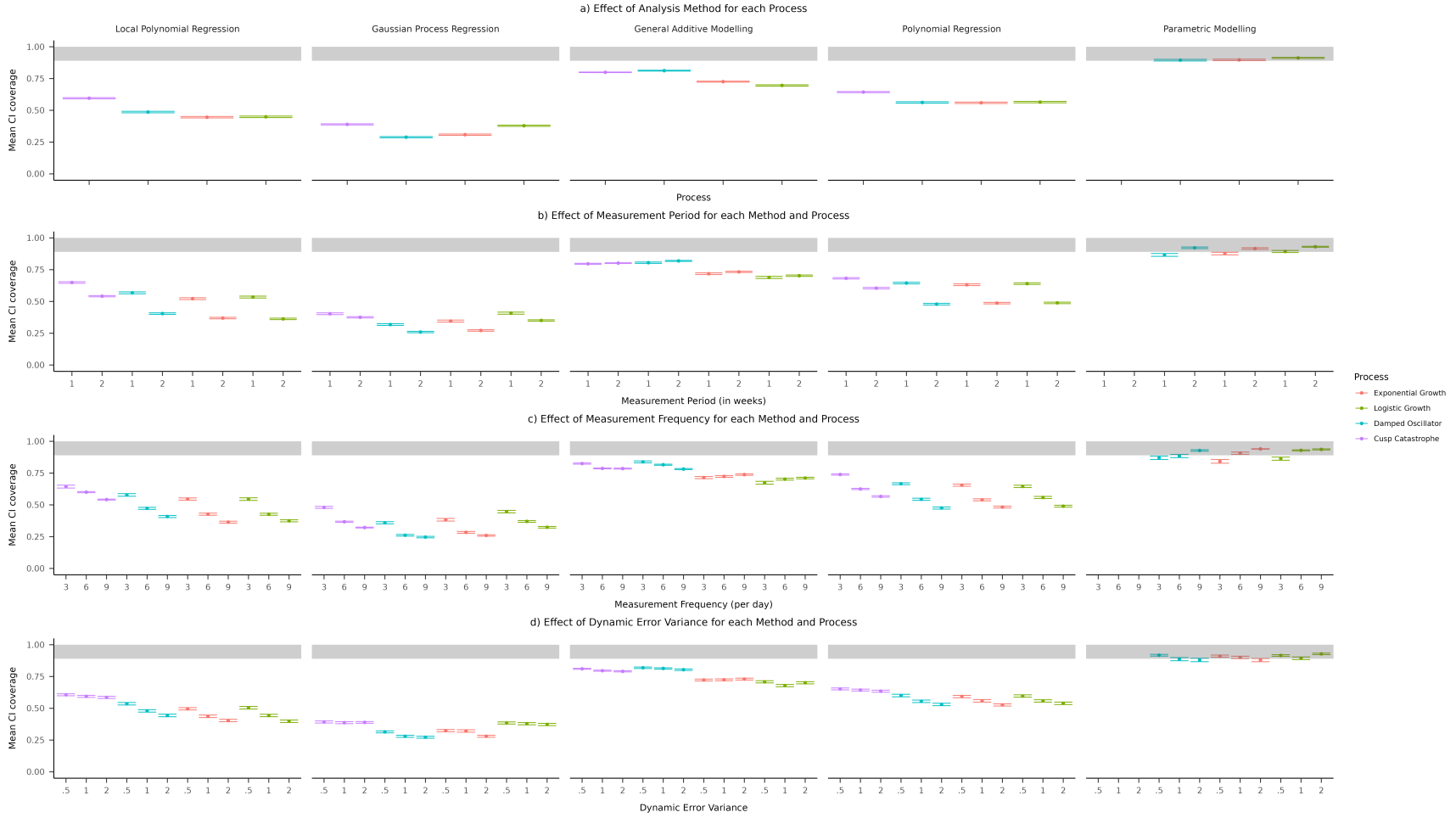
Mean GCV effects across all processes, analysis methods, and simulation conditions



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.

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.

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. Especially, since the GAMs used in this simulation correspond to a GP with a linear mean function (with improper priors) and a non-stationary covariance function (Rasmussen & Williams, 2006; Wahba, 1978). This implies that there exists a GP configuration which performs at least as well as the GAMs. While such a configuration would heavily deviate from how GPs are most commonly applied, common changes like using a Matern class kernel, which makes less strict smoothness assumption about the data, could also be expected to improve the accuracy of the GPs. 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 p-times differentiable processes (Fan et al., 1997). It is however unclear whether these results generalize to the non-differentiable processes that we suspect can be found most often in psychology. Lastly, the tested polynomial regression inferred the underlying processes least accurately. This appeared to be due to numerical model instability that resulted in underfitting, since not enough higher-order terms could be included in the models, and due to divergence towards the ends of the timeseries. The numerical model instability can be 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 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, the three analysis methods previously introduced were applied to depression data from the Leuven clinical study. This study used experience sampling measures to study the dynamics of anhedonia in individuals with major depressive disorder (Heininga et al., 2019). This study was selected for its heterogeneous sample, which includes participants with major depressive disorder, borderline personality disorder, and healthy controls. This diversity increases the likelihood of the data exhibiting a range of (possibly non-linear) dynamics and processes. Specifically, Houben et al. (2015) found in their meta-analysis that individuals with lower psychological well-being tend to experience greater emotional variability, less emotional stability, and higher emotional inertia. Although, this finding did not replicate in an analysis of positive affect within the Leuven clinical study (Heininga et al., 2019). Further, emotional inertia, the extent to which an emotional state carries over across time points, has been shown to vary within individuals over time (Koval & Kuppens, 2012), which makes it likely that the processes underlying this data are non-stationary.

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

3.1 Sample and data description

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

see Heininga et al. (2019). The final data set used for this analysis contained 77 participants in the clinical sample and 40 participants in the control sample, who were matched to the clinical sample by gender and age, resulting in a total sample size of 117 [3].

During the study, all participants completed a baseline assessment, followed by seven days of semi-random EMA assessments, with 10 equidistant assessments per day. However, the starting date of the EMA measures varied between people. During each assessment, participants responded to 27 questions covering emotions, social expectancies, emotion regulation, context, and psychiatric symptoms. This analysis focused on the item assessing momentary depressive mood (i.e., 'How depressed do you feel at the moment?') rated on a scale from 0 to 100.

The published data set was obtained from the EMOTE database (Kalokerinos et al., n.d.). The initial study procedure was approved by the KU Leuven Social and Societal Ethics Committee and the KU Leuven Medical Ethics Committee. This secondary data analysis was approved by the Ethics Review Board of the Tilburg School of Social and Behavioral Sciences (TSB RP FT16).

3.2 Analysis Plan

The LPR, GP, and GAM were applied to explore the idiographic latent processes underlying the data. Each method was applied separately to the time-series of each participant, using the same specifications as in the simulation study (Appendix B). However, for the LPR, only local cubic polynomials were considered to keep the interpretation of the bandwidth consistent across participants. Since all participants were assessed over seven days, but not during the same period, the time series for each participant was centered so that the first measurement time point served as the zero point. The LPR bandwidth, GP lengthscale, and GAM smoothing parameter were then analyzed to assess the wiggliness of the idiographic processes. Additionally, the GCV values produced by each method were evaluated to determine which method provided

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

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

3.3 Results

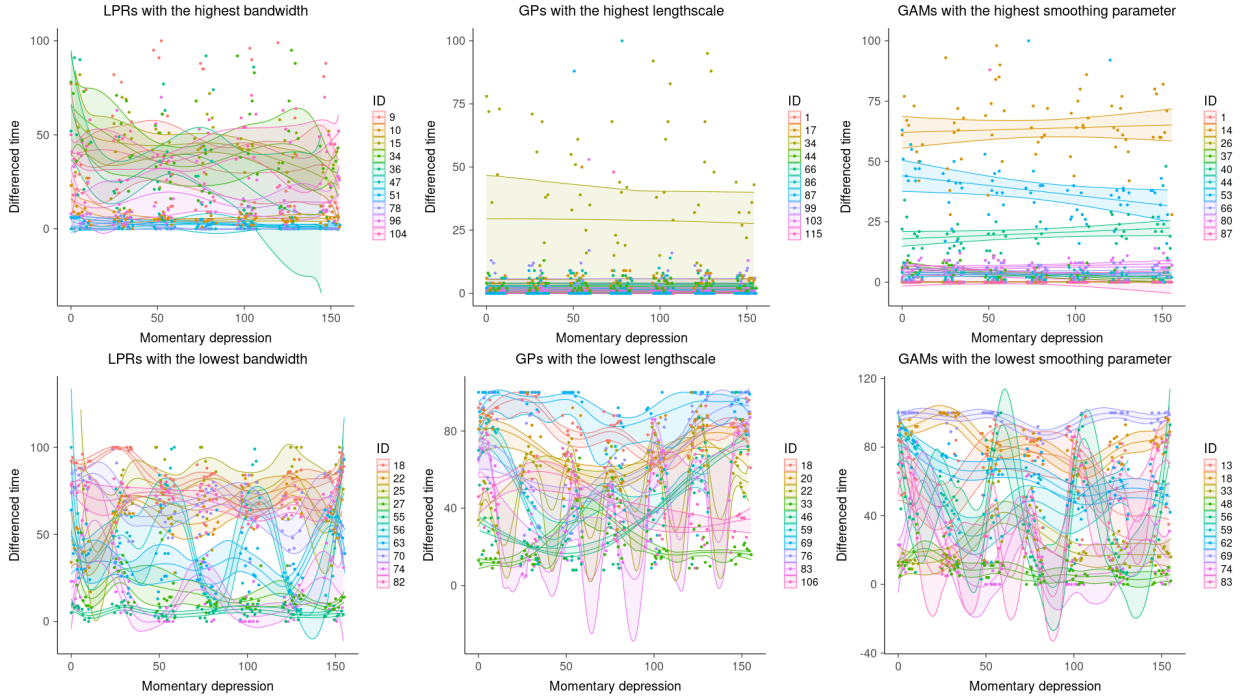
The LPR, GP, and GAM were used to estimate the individual latent depression processes. For the local cubic regression, the median optimal bandwidth was 21.28 hours (*IQR*: 5.52). For the GP, the median optimal length scale was 22.57 standard deviations (*IQR*: 15.09). Lastly, for the GAMs, the median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

Because of this, there is not much value in interpreting the absolute values of these parameters. Instead, we explored the range of functional behaviors inferred by the most extreme values of each parameters. Figure 10 shows the ten least and most wiggly processes inferred by each method. This figure reveals considerably heterogeneity in the functional behavior inferred by each method. Most interestingly, the least wiggly processes inferred by both the GPs and the GAMs are linear trends, indicating the absence of any dynamic errors for these individuals. In contrast to this, the processes with the highest inferred wiggliness, display either large dynamic errors around their respective person means or in addition to this a different non-linear dynamic.

Lastly, a cross-validation was conducted using the generalized cross-validation criterion to investigate which method predicted the latent processes most accurately. The median GCV for the

Figure 10

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



GAMs was 125.29 (*IQR*: 201.04), for the GP it was 248.27 (*IQR*: 578.00), and for the LPR it was 131.57 (*IQR*: 213.73). In addition to this, the mean squared error was calculated between the predictions generated by each method and the data. Here the median MSE of the GAM was 114.04 (*IQR*: 171.59), for the GP it was 197.54 (*IQR*: 388.40), and for the LPR it was 101.74 (*IQR*: 153.77). Together with the GCV this indicates that the GAM inferred the latent processes most accurately, whereas the LPR slightly overfit the data and the GPs tended to underfit the data.

3.4 Conclusion

WIP

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 Twelth International Conference on Artificial Intelligence and Statistics*, 9–16. Retrieved September 11, 2023, from https://proceedings.mlr.press/v5/alvarez09a.html

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

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

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

Boker, S. M. (2012). Dynamical systems and differential equation models of change. In H. Cooper, P. M. Camic, D. L. Long, A. T. Panter, D. Rindskopf, & K. J. Sher (Eds.), *APA handbook of research methods in psychology, Vol 3: Data analysis and research publication.* (pp. 323–333). American Psychological Association. https://doi.org/10.1037/13621-016

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

Boyd, J. P., & Xu, F. (2009). Divergence (Runge Phenomenon) for least-squares polynomial approximation on an equispaced grid and Mock–Chebyshev subset interpolation. *Applied*

Mathematics and Computation, *210*(1), 158–168. https://doi.org/10.1016/j.amc.2008.12.087

Bringmann, L. F., Ferrer, E., Hamaker, E., Borsboom, D., & Tuerlinckx, F. (2015). Modeling Nonstationary Emotion Dynamics in Dyads Using a Semiparametric Time-Varying Vector Autoregressive Model. *Multivariate Behavioral Research*, *50*(6), 730–731. https://doi.org/10.1080/00273171.2015.1120182

Bringmann, L. F., Hamaker, E. L., Vigo, D. E., Aubert, A., Borsboom, D., & Tuerlinckx, F. (2017). Changing dynamics: Time-varying autoregressive models using generalized additive modeling. *Psychological Methods*, *22*(3), 409–425. https://doi.org/10.1037/met0000085

Calonico, S., Cattaneo, M. D., & Farrell, M. H. (2019). nprobust: Nonparametric kernel-based estimation and robust bias-corrected inference. *Journal of Statistical Software*, *91*(8), 1–33. https://doi.org/10.18637/jss.v091.i08

Ceja, L., & Navarro, J. (2012). 'Suddenly I get into the zone': Examining discontinuities and nonlinear changes in flow experiences at work. *Human Relations*, *65*(9), 1101–1127. https://doi.org/10.1177/0018726712447116

Chow, S.-M., Ram, N., Boker, S. M., Fujita, F., & Clore, G. (2005). Emotion as a Thermostat: Representing Emotion Regulation Using a Damped Oscillator Model. *Emotion*, *5*(2), 208–225. https://doi.org/10.1037/1528-3542.5.2.208

Chow, S.-M., Witkiewitz, K., Grasman, R., & Maisto, S. A. (2015). The cusp catastrophe model as cross-sectional and longitudinal mixture structural equation models. *Psychological Methods*, *20*(1), 142–164. https://doi.org/10.1037/a0038962

Clark, N. J., & Wells, K. (2023). Dynamic generalised additive models (DGAMs) for forecasting discrete ecological time series. *Methods in Ecology and Evolution*, *14*(3), 771–784. https://doi.org/10.1111/2041-210X.13974

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

De Bot, K., Lowie, W., & Verspoor, M. (2007). A Dynamic Systems Theory approach to second language acquisition. *Bilingualism: Language and Cognition*, *10*(01), 7. https://doi.org/10.1017/S1366728906002732

Debruyne, M., Hubert, M., & Suykens, J. A. K. (2008). Model Selection in Kernel Based Regression using the Influence Function. *Journal of Machine Learning Research*, *9*(78), 2377–2400. http://jmlr.org/papers/v9/debruyne08a.html

Durbin, J., & Koopman, S. J. (2012). *Time series analysis by state space methods* (Second edition). Oxford University Press

Hier auch später erschienene, unveränderte Nachdrucke Literaturverzeichnis: Seite [326]-339.

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

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

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

Fan, J., & Gijbels, I. (1995a). Adaptive Order Polynomial Fitting: Bandwidth Robustification and Bias Reduction. *Journal of Computational and Graphical Statistics*, *4*(3), 213–227. https://doi.org/10.1080/10618600.1995.10474678

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

B: Statistical Methodology, *57*(2), 371–394. https://doi.org/10.1111/j.2517-6161.1995.tb02034.x

Fritz, J., Piccirillo, M., Cohen, Z. D., Frumkin, M., Kirtley, O. J., Moeller, J., Neubauer, A. B., Norris, L., Schuurman, N. K., Snippe, E., & Bringmann, L. F. (2023). So you want to do ESM? Ten Essential Topics for Implementing the Experience Sampling Method (ESM). https://doi.org/10.31219/osf.io/fverx

Gabry, J., Češnovar, R., Johnson, A., & Bronder, S. (2024). *Cmdstanr: R interface to 'cmdstan'* [R package version 0.8.1, 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

Hastie, T., & Tibshirani, R. (1999). *Generalized additive models*. Chapman & Hall/CRC Originally published: London ; New York : Chapman and Hall, 1990.

Heininga, V. E., Dejonckheere, E., Houben, M., Obbels, J., Sienaert, P., Leroy, B., Van Roy, J., & Kuppens, P. (2019). The dynamical signature of anhedonia in major depressive disorder: Positive emotion dynamics, reactivity, and recovery. *BMC Psychiatry*, *19*(1), 59. https://doi.org/10.1186/s12888-018-1983-5

Houben, M., Van Den Noortgate, W., & Kuppens, P. (2015). The relation between short-term emotion dynamics and psychological well-being: A meta-analysis. *Psychological Bulletin*, *141*(4), 901–930. https://doi.org/10.1037/a0038822

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

Jebb, A. T., Tay, L., Wang, W., & Huang, Q. (2015). Time series analysis for psychological research: Examining and forecasting change. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.00727

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

Koval, P., & Kuppens, P. (2012). Changing emotion dynamics: Individual differences in the effect of anticipatory social stress on emotional inertia. *Emotion*, *12*(2), 256–267. https://doi.org/10.1037/a0024756

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

Kunnen, S. E. (2012). *A Dynamic Systems Approach of Adolescent Development* [OCLC: 823389367]. Taylor; Francis.

McArdle, J. J., Ferrer-Caja, E., Hamagami, F., & Woodcock, R. W. (2002). Comparative longitudinal structural analyses of the growth and decline of multiple intellectual abilities over the life span. *Developmental Psychology*, *38*(1), 115–142. https://doi.org/10.1037/0012-1649.38.1.115

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

Nesselroade, J., & Ram, N. (2004). Studying Intraindividual Variability: What We Have Learned That Will Help Us Understand Lives in Context. *Research in Human Development*, *1*(1), 9–29. https://doi.org/10.1207/s15427617rhd0101&2_3

Newell, K. M., Liu, Y.-T., & Mayer-Kress, G. (2001). Time scales in motor learning and development. *Psychological Review*, *108*(1), 57–82. https://doi.org/10.1037/0033-295X.108.1.57

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

Ou, L., Hunter, M. D., & Chow, S.-M. (2019). What's for dynr: A package for linear and nonlinear dynamic modeling in R. *The R Journal*, *11*, 1–20.

R Core Team. (2024). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria. https://www.R-project.org/

Rasmussen, C. E., & Williams, C. K. I. (2006). *Gaussian processes for machine learning* [OCLC: ocm61285753]. MIT Press.

Richardson, R. R., Osborne, M. A., & Howey, D. A. (2017). Gaussian process regression for forecasting battery state of health. *Journal of Power Sources*, *357*, 209–219. https://doi.org/10.1016/j.jpowsour.2017.05.004

Roberts, S., Osborne, M., Ebden, M., Reece, S., Gibson, N., & Aigrain, S. (2013). Gaussian processes for time-series modelling [Publisher: Royal Society]. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *371*(1984), 20110550. https://doi.org/10.1098/rsta.2011.0550

Ruppert, D., & Wand, M. P. (1994). Multivariate Locally Weighted Least Squares Regression. *The Annals of Statistics*, *22*(3). https://doi.org/10.1214/aos/1176325632

Siepe, B. S., Bartoš, F., Morris, T. P., Boulesteix, A.-L., Heck, D. W., & Pawel, S. (2023). Simulation Studies for Methodological Research in Psychology: A Standardized Template for Planning, Preregistration, and Reporting. https://doi.org/10.31234/osf.io/ufgy6

Tan, X., Shiyko, M., Li, R., Li, Y., & Dierker, L. (2011). A Time-Varying Effect Model for Intensive Longitudinal Data. *Psychological methods*, *17*, 61–77. https://doi.org/10.1037/a0025814

Tsay, R. S., & Chen, R. (2019). *Nonlinear time series analysis*. John Wiley & Sons Includes index.

van der Maas, H. L. J., Kolstein, R., & van der Pligt, J. (2003). Sudden Transitions in Attitudes. *Sociological Methods & Research*, *32*(2), 125–152. https://doi.org/10.1177/0049124103253773

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