Running head: Gaussian Process Panel Modeling (GPPM)

Gaussian Process Panel Modeling – Kernel-Based Longitudinal Modeling

Julian D. Karch

Leiden University, The Netherlands

Max Planck Institute for Human Development, Germany

Andreas M. Brandmaier

Max Planck Institute for Human Development, Germany

Max Planck UCL Centre for Computational Psychiatry and Ageing Research, Germany

Manuel C. Voelkle

Humboldt University Berlin, Germany

Max Planck Institute for Human Development, Germany

Author Note

Julian D. Karch, Center for Lifespan Psychology, Max Planck Institute for Human Development; Andreas M. Brandmaier, Center for Lifespan Psychology, Max Planck Institute for Human Development; Manuel C. Voelkle, Department of Psychology, Humboldt University Berlin

Julian D. Karch is now at the Institute of Psychology, Leiden University.

Julia Delius provided highly valuable editorial assistance. This manuscript is based on Julian D. Karch's doctoral dissertation.

GPPM has already been presented at the following conferences: 7th European Congress of Methodology (2016), Conference of the German Society for Psychology (2016), 11th International Multilevel Conference (2017), and International Meeting of the Psychometric Society (2017).

Correspondence concerning this article should be addressed to Julian D. Karch, Institute of Psychology, Leiden University, P.O. Box 9555, 2300 RB Leiden, The Netherlands. Email: j.d.karch@fsw.leidenuniv.nl.

Abstract

Longitudinal panel data obtained from multiple individuals measured at multiple time points are crucial for psychological research. To analyze such data, a variety of modeling approaches such as hierarchical linear modeling or linear structural equation modeling are available. Such traditional parametric approaches are based on a relatively strong set of assumptions, which are often not met in practice. We present a flexible modeling approach for longitudinal data that is based on the Bayesian statistical learning method Gaussian Process Regression. We term this novel approach *Gaussian Process Panel Modeling* (GPPM). We show that GPPM subsumes most common modeling approaches for longitudinal data such as linear structural equation models and state-space models as special cases but also extends the space of expressible models beyond them. GPPM offers great flexibility in model specification, facilitates both parametric and nonparametric modeling in a single framework, enables continuous-time modeling as well as person-specific predictions, and offers a modular system that allows the user to piece together hypotheses about change by selecting from and combining predefined types of trajectories or dynamics. We demonstrate the utility of GPPM based on a selection of models and data sets.

Keywords: continuous-time modeling; longitudinal data analysis; statistical learning

Introduction

Longitudinal panel data obtained from multiple individuals measured at multiple time points are among the most common types of data collected for psychological research. To analyze longitudinal data, different modeling approaches are commonly used, and their choice largely depends on the researcher's discipline or statistical training. Most longitudinal analyses are performed using the general linear model (GLM; Cohen, 1968), hierarchical linear modeling (Raudenbush & Bryk, 2001), time series methods (Hamilton, 1994), or structural equation modeling (SEM; Bollen, 1989). These traditional parametric approaches have the advantage that specification, inference, and interpretation is relatively straightforward. However, this comes at the price of being restricted to a relatively strong set of assumptions. There is increasing appreciation of the fact that these strong assumptions may incur severe model misfit (and potentially wrong conclusions). More general approaches may provide more meaningful and correct models for the more complex situations that are often encountered in psychology (Lee & Zhu, 2002).

Here, we present a novel modeling approach for longitudinal data, which we call *Gaussian Process Panel Modeling* (GPPM). It offers great flexibility in model specification, facilitates both parametric and nonparametric modeling, appreciates the continuous-time nature of the data, and offers a modular system that allows the researcher to piece together hypotheses about change by selecting from predefined types of trajectories or dynamics.

We refer to our approach as GPPM because it is based on the flexible Bayesian non-parametric multivariate regression method Gaussian process regression (GPR; Rasmussen &

Williams, 2006). GPR is a popular statistical learning method that has been applied successfully for the analysis of time series data from fields such as astronomy (Damouras, 2008; Roberts et al., 2013), meteorology (Roberts et al., 2013), economics (Damouras, 2008; Roberts et al., 2013), biology (Saatçi, Turner, & Rasmussen, 2010), medicine (Brahim-Belhouari & Bermak, 2004; Liu, Wu, & Hauskrecht, 2013), and neuroimaging (Ziegler, Ridgway, Dahnke, & Gaser, 2014). We refer to GPR as applied in the analysis of time series as temporal GPR. Temporal GPR is popular in statistical learning, but has received little attention from most social and behavioral scientists. Temporal GPR is a single-subject technique, and its extension to multi-subject data requires augmentation with a between-person model. Here, we present a mechanism to extend temporal GPR in this way, resulting in GPPM.

GPPM is closely related to longitudinal SEM (lSEM). lSEM and GPPM are both means- and covariance-modeling techniques. The central mechanism for model specification is the translation of hypotheses into restricted covariance matrices and mean vectors. A main difference between these methods lies in the way the model covariance matrix is specified. In lSEM, structural equations that relate observed and latent variables to each other are at the heart of model specification. In contrast, GPPM uses a covariance-matrix generating and mean-vector generating function using the so-called *kernel language*. This allows more efficient expression of a broader class of models than linear SEM, as we will demonstrate later on. As a consequence, GPPM extends lSEM in two important ways: First, it is not the covariance matrix of the observations that is modeled but rather the infinite dimensional covariance matrix of the underlying continuous process from which the observations are sampled. Thus, it enables continuous-time modeling (Voelkle, Oud, Davidov, & Schmidt, 2012). Second, it allows specification of all lSEMs as part of an even broader model class.

GPPM is also closely related to continuous-time linear Gaussian multi-subject state-space modeling (Boker, 2007; Chow, Ho, Hamaker, & Dolan, 2010; Driver, Oud, & Voelkle, 2017; Oud & Singer, 2008), which we will abbreviate to state-space modeling in the remainder of this article. State-space modeling importantly contains the more popular autoregressive integrated moving average (ARIMA) models (Hamilton, 1994), and thereby vector autoregressive models as special cases. Like GPPM, state-space modeling also allows direct modeling of the underlying continuous-time process. Again, the core difference between GPPM and state-space modeling is the language for model specification. While in state-space modeling the model is specified via stochastic difference or differential equations, making it well suited for dynamic models, GPPM uses the kernel language for model specification, which is more closely related to model specification in lSEM. Thus, GPPM could be more suitable for formulation of trajectory-based continuous-time models. However, as we will show below, GPPM also includes state-space models as a special case (for comparison between temporal GPR and state-space modeling for time series analysis, see Grigorievskiy & Karhunen, 2016).

To our knowledge, temporal GPR has only rarely been applied to psychological data. For example, Ziegler and colleagues (2014) used temporal GPR to estimate a normative cross-sectional age gradient of volumetric changes in the brain. Two publications already brought up the notion of extending temporal GPR for the analysis of hierarchical data: Within the field of statistics, Hall, Müller, & Yao (2008) proposed an extension of temporal GPR to model non-Gaussian panel data, and within psychology, Cox, Kachergis, & Shiffrin (2012) suggested an hierarchical extension to model computer-mouse trajectories nested in trials and conditions across multiple participants. The present article proposes an extension of temporal

GPR for the analysis of psychological panel data. In the following, we provide a detailed comparison with conventional psychological panel data analysis methods and include a full set of corresponding frequentist inference procedures.

We provide proof-of-concept implementations of GPPM in both R (R Core Team, 2015) and Matlab (MATLAB, 2014), which enables readers to apply this approach to their own data. We showcase the usage of both packages by providing the code that was used to implement the demonstration analyses as well as the code for production of the figures below as a supplement. While the R software is easier to use, it is – at the time of writing – less mature than the Matlab package. Parameter estimation, for example, is substantially slower. For the novice in GPPM, we advise starting with the R software and eventually switching over to the Matlab package that provides more advanced features. However, this is likely to change, as there are plans to extend the R package substantially.

This article is structured as follows: First, we formally define temporal GPR. Our presentation of temporal GPR differs from previous treatments (e.g., Roberts et al., 2013) in that we do not assume that the reader has knowledge of GPR. Instead, we connect temporal GPR directly to the more familiar linear model. Second, we propose a generalization of temporal GPR for the analysis of longitudinal panel data, resulting in the more general framework, GPPM. Third, we demonstrate the advantages of GPPM by comparing it to existing longitudinal panel modeling methods based on a select set of example models and data sets. Specifically, we show how person-specific predictions, even on non-observed time points, can be obtained easily. We demonstrate this using an application of the latent growth curve model (LGCM; Duncan, Duncan, & Strycker, 2006) to data from the Berlin Aging Study (BASE; Baltes & Mayer, 2001; Delius, Düzel, Gerstorf, & Lindenberger, 2017), a longitudinal study tracking the development of older adults. We also demonstrate how the popular first-order autoregressive model (AR(1) model) can easily be translated into its continuous-time equivalent using a GPPM representation and how the resulting model can be combined with the LGCM to implement AR(1)-correlated errors for the LGCM. We proceed to show that GPPM can also represent the nonparametric technique of generalized additive modeling, which was recently introduced as a promising modeling approach for psychological data (Bringmann et al., 2017; Shadish, Zuur, & Sullivan, 2014; Sullivan, Shadish, & Steiner, 2015). Furthermore, GPPM can extend generalized additive modeling to become a panel modeling technique. Finally, we demonstrate the utility of the squared exponential model, one of the most commonly used GPR models in statistical learning, for the analysis of longitudinal panel data. We close this paper with a summary and a discussion of work to be done in the future.

Gaussian Process Time Series Modeling

To introduce the model specification mechanism of GPPM, the kernel language, we will first provide a short recap of the linear growth model, which most researchers will be familiar with, before demonstrating how the linear growth model can be expressed as a temporal GPR model. Then we introduce model specification using the kernel language in its general form.

Linear Growth Model

One of the simplest models for a time series is to assume that the repeated measures of some observable variable, like a person's height, follow a linear trend. Formally, we assume that each observation y_t is made at a time point t. The scale of time is arbitrary and could be the person's age or the time that has elapsed since the beginning of a study. For illustration, assume t encodes age in days. If the perfectly measured observations follow a linear trend, for every observation y_t, we obtain the regression equation with time as independent variable:

$$y_t = b_0 + b_1 t . \quad \textbf{(1)}$$

The parameters b_0 and b_1 represent the linear trend's intercept and slope (i.e., the rate of linear change) respectively. Both parameters are allowed to take on any continuous scalar value, $b_0, b_1 \in \mathbb{R}$.

Typically, we assume that all observed variables are subject to measurement error. To account for this, we add an explicit measurement model. In this simplest case, we assume uncorrelated, normally distributed measurement errors at each time point. Formally, we account for measurement error by adding a Gaussian random variable ϵ_t with a mean of zero and measurement error variance σ_ϵ^2 to Equation (1), resulting in

$$Y(t) = b_0 + b_1 t + \epsilon_t . \quad \textbf{(2)}$$

We have changed the notation of $Y(t)$ because it now represents a random rather than a fixed variable and no longer corresponds to the observations but to a stochastic representation of the underlying process. Each observation y_t is a realization of the random variable $Y(t)$.

Linear Growth Model as Gaussian Process Time Series Model

The linear growth model introduced above corresponds to linear regression with time as the only predictor applied within persons. In this section, we show how the linear growth model can be expressed equivalently as a temporal GPR using the kernel language.

First note that for each time point t, the regression model in Equation (2) defines a Gaussian random variable. Meanwhile, time is modeled as a continuous variable, that is, $t \in \mathbb{R}$. As a consequence, the set $\{Y(t): t \in \mathbb{R}\}$ contains an infinite number of random variables, one for each time point t. This type of set is called a *stochastic process*. If and only if every random variable Y_t has a Gaussian distribution, the set $\{Y(t): t \in \mathbb{R}\}$ is called a *Gaussian process*.

Equation (2) implies a mean for every time point t and a covariance for each pair of time points s, t. They both depend on the model parameters $\theta = [b_0, b_1, \sigma_\epsilon^2]$:

$$\begin{aligned} \mathbb{E}_\theta[Y(t)] &= b_0 + b_1 t \\ \mathrm{Cov}_\theta[Y(s), Y(t)] &= \delta(s-t)\sigma_\epsilon^2 \end{aligned},$$

where $\delta(\cdot)$ is the Dirac delta function, that is, it yields 0 for all cases; unless its function argument is zero (here, when s equals t) when it yields 1. Following the above model specification, we can define a mean function $m(t)$ that returns the model-implied expectation for every time point and a kernel (also, covariance function) $k(s,t)$ that returns the model-implied covariance for every pair of time points s, t:

$$\begin{aligned} m(t;\theta) &= b_0 + b_1 t \\ k(s,t;\theta) &= \delta(s-t)\sigma_\epsilon^2 \end{aligned}. \tag{3}$$

We find that the simple linear regression model in Equation (2) and the mean and kernel functions in Equations (3) define the same model in the sense that the model-implied distribution is identical for both. In the supplement, we also demonstrate this practically by showing that the parameter estimates obtained by linear regression modeling software and the GPPM software are identical.

Representing a model using mean and kernel functions is referred to as kernel language representation. The kernel language is not only an alternative language for model specification, it is also much more expressive than the linear model, as we will show throughout this paper.

For readers familiar with the GPR literature, it may seem strange that the core definition of the linear model is done here through the mean and not the kernel function, as models are mostly defined by their kernel functions in the GPR literature. We explain this in Appendix A.

Model Specification Using the Kernel Language

In the previous section, we obtained a temporal GPR model by translating the linear growth model into its kernel language representation. In general, a GPR is defined by specifying parameterized mean and kernel functions. These functions can be derived through multiple avenues, such as translating a theory directly into mean and kernel functions or translating a model expressed in a different specification language (as carried out above for the linear model).

Another approach for specifying a temporal GPR model that facilitates the representation of complex hypotheses is to formulate a model by flexibly combining a set of predefined template models. To this end, one can make use of the fact that new mean and kernel functions can be created from sets of available template mean and kernel functions using a variety of operators (Duvenaud, 2014, Chapter 2; Rasmussen & Williams, 2006, Chapter 4.2.4; Roberts et al., 2013), for instance, by simply adding them together. In this paper, we will primarily capitalize on the fact that new functions can be obtained by multiplying functions by a scalar and that both the sum and the product of two functions again produce new valid functions. These simple combination rules allow us to efficiently specify a wide range of hypotheses in form of mean and kernel functions.

As an easy example of this compositional mechanism, we present an alternative avenue to obtain the kernel language representation of the linear model. The linear growth model can be specified by taking the sum of the constant mean function, $m(t;b_0)=b_0$, and the linear mean function, $m(t;b_1)=b_1 t$. We will present many more examples of the combination mechanisms throughout the text.

Gaussian Process Panel Modeling

In the following, we extend the time series modeling method, temporal GPR, to GPPM for the analysis of longitudinal panel data. The extension consists of a framework to formulate a between-person model as well as corresponding frequentist inference procedures.

Between-Person Model

The core difference between time series modeling and panel data modeling is that only a model for the within-person structure is required for the former, whereas a model for the within- and the between-person structure is needed for the latter. Thus, extending a time series method to a panel method requires augmenting the method by a way to account for the between-person structure. In this section, we discuss how we propose to do this when extending temporal GPR to GPPM.

Before we can proceed, we need to introduce some notation. Formally, a time series consists of multiple observations y_t at multiple time points $t \in \{t_1, t_2, \ldots, t_T\}$ originating from one person. A longitudinal panel data set consists of multiple observations y_{it} from multiple persons $i \in \{1, 2, \ldots, N\}$ and time points $t \in \{t_{i1}, t_{i2}, \ldots t_{iT_i}\}$. Note that the time points at which persons were observed do not have to be the same across persons when using this notation. With y_i, we denote the vector containing the complete time series of person i.

The easiest way to adapt temporal GPR for the analysis of longitudinal panel data is to treat each time series y_i as an independent data unit and consequently to perform an independent GPR analysis for each person. The implicit assumption underlying this approach is that the differences between people are so large that they should be treated as completely independent analysis problems. Therefore, no between-person model is specified. This approach corresponds to a no-pooling analysis in multilevel parlance (Gelman, 2006).

At the other end of the spectrum we need to assume that there are no between-person differences. This can be implemented by contending that the time series of each person y_i is a realization of the same Gaussian process. In multilevel models, this would conform to the complete-pooling approach. Formally,

$$Y_i(t) \sim GP(m^*(t), k^*(s,t)) \ ,$$

where the mean $m^*(t)$ and the kernel function $k^*(s,t)$ describe the true but unknown distribution of the Gaussian process. Additionally, the process $Y_i(t)$ for a person i is assumed to be independent of the processes of all other persons. In this case, the task of statistical inference reduces to recovering the true mean $m^*(t)$ and kernel function $k^*(s,t)$. Also, it is sufficient to define a statistical model on the level of the individual time series. The model for the panel data set follows directly. Essentially, the same mechanism as used for SEM can be employed when it is assumed that the data for each person is a realization of one true multivariate Gaussian distribution. Here, however, we exchange multivariate Gaussian distribution with an infinite-dimensional Gaussian process.

We would argue that neither assumption – no differences between persons or no relationship between them at all – is typically realistic. The middle ground is to specify a between-person model. This model could, for example, allow estimation of a sample average and individual deviations from that average. To elaborate: The person-specific analysis approach, mentioned at the beginning of this section, results in estimates for the person-specific parameters θ_i. No between-person distribution is imposed on the person-specific parameters θ_i. In the nonspecific analysis approach, the implicit assumption is that $\theta_1 = \theta_2 = \ldots$. The middle ground is to suppose that there is some between-person parameter distribution $\mathbb{P}(\theta)$ of which the person-specific parameters are realizations. This approach, referred to as partial pooling, is typically taken in hierarchical linear modeling, multilevel models, or random-effects models, in which it is assumed that the between-person distribution of regression coefficients is a multivariate normal distribution (Raudenbush & Bryk, 2001). This approach can also be taken in SEM. One way to see this is to acknowledge that multilevel models can be considered a special case of SEM (Curran, 2003). An extension of this approach is to assume that the between-person distribution itself depends on stable observable characteristics z_i of a person by formulating a model for their conditional distribution $\mathbb{P}(\theta|z_i)$. This is often done in SEM, for example, where person-specific characteristics z_i are employed to modify the mean of a latent variable for each person.

Somewhat counterintuitively, specifying one GPPM that is shared across all persons, also permits one to represent between-person distributions of parameters as just shown. This is best understood based on an example. Consider the following GPPM, which encodes the assumption that the observed measurements are constant across both persons and time

$$\begin{aligned} m(t;\theta) &= c \\ k(s,t;\theta) &= \delta(s-t)\sigma_\epsilon^2 \end{aligned}$$

We now add a between-distribution for the c parameter ($c \sim \mathcal{N}(\mu_c, {\sigma_c}^2)$). This changes the model to

$$\begin{aligned} m(t;\theta) &= \mu_c \\ k(s,t;\theta) &= \sigma_c^2 + \delta(s-t)\sigma_\epsilon^2 \end{aligned}.$$

Thus, by introducing a between-person model for a fixed parameter, only the mean and kernel function are changed.

Not every between-person model can be expressed using GPPM. Essentially, only those between-person models that can be expressed by hierarchical linear modeling and SEM can also be expressed by GPPM. More formally, a Gaussian between-person distribution can be specified for each linear parameter of the mean function. The reason for this is that the resulting person-level model is again a model on a Gaussian process. In the next section, we give a more elaborate example of this mechanism by demonstrating how the linear growth model can be extended by a Gaussian between-person model, resulting in the well-known LGCM that allows between-person differences in linear trajectories.

In order to implement conditional between-person distributions $\mathbb{P}(\theta|z_i)$, the mean and kernel functions must accept the stable characteristics z_i as input, in addition to time. This does not pose a problem since mean and kernel functions can be defined for arbitrary inputs. Instead of using only the time point of each observation as input for the mean and kernel function, one can use arbitrary predictors from each time point. This is achieved by the following straightforward change in formalism. Instead of the time point t, the mean function accepts a vector x_{it} as input, which contains arbitrary information about the measurement from person i at time point t. So, the vector x_{it} typically contains at least one marker of time but can also contain other predictors of interest. A special case of this is of course to assume that predictors are constant for each person, as is done for the stable characteristics z_i. The kernel function is extended in the same fashion. That is, its form changes to $k(x_{it},x_{is})$. This inclusion of stable characteristics allows for a broader class of between-person models than available in classical frameworks such as SEM and hierarchical linear modeling. Below, we demonstrate this extension of the GPPM framework in detail by showing how the GPPM representation of the LGCM can be augmented by a stable characteristics variable z_i that is related to the slope's mean as well as its variance.

Latent Growth Curve Model and Extensions as GPPM

The LGCM starts out from the linear growth model introduced for time series, repeated here for convenience; with an emphasis on the fact that this is a model for a specific person i:

$$\begin{aligned} m(t;\theta_i) &= b_{i0} + b_{i1}t \\ k(s,t;\theta_i) &= \delta(s-t)\sigma_\epsilon^2 \end{aligned} \tag{4}$$

In LGCM, the between-person model is introduced by assuming that the individual parameters describing growth are distributed according to a Gaussian distribution. Formally,

$$\begin{bmatrix} b_{i0} \\ b_{i1} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu_{b_0} \\ \mu_{b_1} \end{bmatrix}, \begin{bmatrix} \sigma_{b_0}^2 & \sigma_{b_0,b_1} \\ \sigma_{b_0,b_1} & \sigma_{b_1}^2 \end{bmatrix} \right).$$

In Equation (4), we treated b_{i0} and b_{i1} as parameters. With the inclusion of the between-person distribution they are no longer parameters but random variables. As a consequence, the mean function itself becomes a Gaussian process. To see this, note that

$$\begin{aligned} \mathbb{E}(m(t)) &= \mu_{b_0} + \mu_{b_1}t \\ \text{Cov}(m(s),m(t)) &= \sigma_{b_0}^2 + \sigma_{b_0,b_1}(s+t) + \sigma_{b_1}^2 st \end{aligned}.$$

To bring the measurement error back into this model, one simply needs to add its kernel representation. This shows that the GPPM formulation of the LGCM is

$$\begin{aligned} m(t;\theta) &= \mu_{b_0} + \mu_{b_1}t \\ k(s,t;\theta) &= \sigma_{b_0}^2 + \sigma_{b_0,b_1}(s+t) + \sigma_{b_1}^2 st + \delta(s-t)\sigma_\epsilon^2 \end{aligned}$$

In the supplement, we provide a practical demonstration of the equivalence of the LGCM represented as GPPM and as lSEM.

To showcase the ability of GPPM to represent conditional between-person models, we now assume that the between-person model additionally depends on two stable characteristics variables z_{i1}, z_{i2} in the following fashion

$$\begin{bmatrix} b_{i0} \\ b_{i1} \end{bmatrix} | z_i \sim \mathcal{N}\left(\begin{bmatrix} \mu_{b_0} \\ \mu_{b_1} + c_1 z_{i1} \end{bmatrix}, \begin{bmatrix} \sigma^2_{b_0} + & \sigma_{b_0,b_1} \\ \sigma_{b_0,b_1} & \sigma^2_{b_1} f(c_2, z_{i2}) \end{bmatrix} \right).$$

We include a stable characteristics variable z_{i1} that is linearly related to the mean of the slope. This corresponds to a regression of the latent slope on z_{i1}, and is often done in LGCM to investigate whether a proportion of the slope variance can be explained by a stable characteristics variable. We also included a stable characteristics variable z_{i2} that is related to the slope variance according to some generic and potentially nonlinear function $f(c_2, z_{i2})$ with the additional parameter c_2. This is not typically done in LGCM and also not possible in (standard) SEM or hierarchical linear modeling.

For the model-implied mean and kernel function, we obtain

$$\begin{aligned} \mathbb{E}(m(t)) &= \mu_{b_0} + (\mu_{b_1} + c_1 z_{i1}) &&= m(x_{it}; \theta) \\ \text{Cov}(m(s), m(t)) &= \sigma^2_{b_0} + \sigma_{b_0,b_1}(s+t) + \sigma^2_{b_1} f(c_2, z_{i2}) st &&= k(x_{it}, x_{is}; \theta) \end{aligned}, \quad \textbf{(5)}$$

with $x_{it} = [t, z_{i1}]$, $x_{it} = [s, z_{i1}]$. For the model to hold $f(c_2, z_{i2})$ must be such that the kernel function is still valid. This, however, still allows many different functional forms to be employed.

The kernel representation of the LGCM again illustrates the fact that GPPMs can typically be specified through combinations of template mean and kernel functions. The mean function consists of the sum of the constant $m(t;\mu_{b_0})=\mu_{b_0}$ and the linear $m(t;\mu_{b_1})=\mu_{b_1}t$ mean functions. The kernel function consists of the sum of the constant $k(s,t;\sigma_{b_0}^2)$, a scaled (times $\sigma_{b_1}^2$) version of the linear $k(s,t)=st$, the white noise $k(s,t;\sigma_\epsilon^2)=\delta(s-t)\sigma_\epsilon^2$, and a non-standard kernel function that represents the correlation between the intercept and the slope $k(s,t;\sigma_{b_0,b_1})=\sigma_{b_0,b_1}(s+t)$.

Inference for Gaussian Process Panel Models

Despite temporal GPR typically being used in conjunction with Bayesian inference, we focused on developing frequentist inference procedures for GPPM to make it better comparable to standard methods (specifically, such that parameter estimates would be identical to other specification languages that use maximum likelihood inference). In this section, we also show how GPPM makes it easy to obtain person-specific predictions after the population-level model has been fitted. We also provide suggestions for model selection when multiple hypotheses are to be compared on empirical data.

For maximum likelihood estimation, a likelihood function is required. Usually, the likelihood function can be obtained easily because the data are assumed to be a realization of the random vector on which the model is formulated. In GPPM however, the model is formulated on a stochastic process; and it is impossible to observe a complete realization of a stochastic process as this would equate to a time series with infinitely many time points. Instead, the stochastic process is partially observed at a finite set of time points. This problem however is easily solved as a GPPM also implies a model for every finite set of time points. This implied discrete model on a random vector can be used to perform maximum likelihood estimation as usual. It technically corresponds to a marginal likelihood, which is marginalized over all infinitely many unobserved time points. In Appendix B, we show this formally and also demonstrate that the likelihood ratio test for hypothesis testing is valid for GPPMs. Confidence intervals and regions can be calculated based on the likelihood ratio test (Pek & Wu, 2015).

A more informal proof that these inference methods are valid for GPPM can be obtained by acknowledging that GPPM can be implemented using extended SEM (Neale et al., 2016). In contrast to conventional SEM, extended SEM allows path coefficients and covariances to vary for each person, for instance, via definition variables. Additionally, path coefficients and covariances can be arbitrarily complex functions of parameters. In Figure 1, we present how the discrete model implied by a GPPM can be translated into an extended SEM. Essentially, the mean function becomes a path coefficient from the constant to the observations and the covariance function populates the values of the covariances between the observations. This translation might represent the most straightforward clarification of GPPM for researchers accustomed to extended SEM.

The features of GPPM lead to a natural approach to obtaining person-specific predictions. Assume that we have a panel data set and a corresponding GPPM. Using maximum likelihood estimation, we obtain a parameter estimate $\hat{\theta}$. Given this estimate and the observed data y_i for a person i at time points t_i, we can now ask: What are the predictions for the observations at unobserved time points t_i^*? For example, this question could be of interest to predict the development of a particular child or of a patient with regard to some score of their mental health.

Formally, the maximum likelihood estimate $\hat{\theta}$ represents the distribution of a Gaussian process. Thus, it especially expresses the joint distribution $p_{\hat{\theta}}(y_i, y_i^*)$ of the process at the observed and the unobserved time points, which is a multivariate Gaussian. Person-specific predictions can be obtained by conditioning on the observations. This results in the conditional distribution $p_{\hat{\theta}}(y_i \mid y_i^*)$, which is again a Gaussian and also known as the posterior predictive distribution in the Bayesian inference literature. If a point estimate is needed, different approaches to reduce the posterior predictive distribution to a point, like the mode or the expectation, are possible.

For model selection, the same ideas can be used as for SEM. Besides its application for hypothesis testing, the likelihood-ratio test can also be used to compare two nested models. The more restrictive model becomes the null hypothesis and is selected if the null hypothesis is not rejected. Fit indices like the Akaike information criterion (AIC) and the Bayesian information criterion (BIC) can also be used for general, non-nested model selection and are implemented in the provided packages.

Both the fit indices (Barrett, 2007) and the likelihood-ratio test (West, Taylor, & Wu, 2012) have been criticized in the context of SEM. The same criticisms also apply to their use in the context of GPPM. As a modern, statistical learning-inspired alternative we have therefore also developed cross-validation for GPPMs. We use the word cross-validation here in the same sense as it is used in statistical learning. That is, the procedure of leaving out data is repeated multiple times until every data point has been left out exactly once. The idea is as follows: The maximum likelihood estimate represents the distribution of a Gaussian process. With new, unseen data we can now estimate the out-of-sample likelihood of those data. We choose the model that leads to the highest out-of-sample likelihood. Thus, we choose the model that best approximates the true distribution in a likelihood sense. For the cross-validation, we can either choose to leave out persons in each fold, which would estimate how well the model generalizes across persons, or we leave out time points, which would measure how well the model generalizes over time for the observed persons.

Flexibility of Gaussian Process Panel Modeling

The main advantage of GPPM is its flexibility in formulating a model for the within-person structure. Using the kernel language, every model that can be expressed as a set of candidate distributions for a Gaussian process can be specified as a GPPM. Besides lSEM, and state-space modeling, this also includes generalized additive modeling, which has recently been proposed as a flexible, nonparametric method for the analysis of psychological time series (Bringmann et al., 2017; Shadish et al., 2014; Sullivan et al., 2015). Essentially, all these techniques employ different, less expressive languages to describe a model on a Gaussian process or a multivariate Gaussian, which can be considered the finite dimensional special case of a Gaussian process. Furthermore, in terms of statistical inference all aforementioned methods rely on frequentist inference, specifically, maximum likelihood estimation and the likelihood-ratio test. Formal proofs of the fact that GPPM generalizes these methods as well as a detailed discussion of how GPPM extends these methods can be found in Karch (2016; see Chapter 4.1.1 for lSEM, and Chapter 4.1.2 for multi-subject continuous-time state-space modeling). We do not present a formal proof for generalized additive modeling. However, Duvenaud, Nickisch, and Rasmussen (2011) present a GPR model that subsumes generalized additive modeling. Also, Rasmussen and Williams (2006, Chapter 2) show how any Bayesian kernel regression model can be implemented as a GPR model. Just by comparing the definition of Bayesian kernel regression (Rasmussen & Williams, 2006, p.12, Equation 2.10) and generalized additive modeling (Wood, 2006, p. 199, Equation 3.1), one can see that Bayesian kernel regression generalizes the former and consequently also GPR. For the purposes of this paper, we find it most instructive to demonstrate how GPPM can represent smoothing splines, the core technique used within generalized additive models.

As GPPM is an inherently continuous-time modeling approach, it shares the same advantages that other continuous-time modeling approaches (e.g., continuous-time state-space modeling) have over discrete-time modeling approaches (e.g., discrete-time state-space modeling; Voelkle et al., 2012): It adequately models different time intervals both between and within persons, results are comparable across studies with different measurement intervals, and missing data are treated “automatically” as they simply result in a different time interval between successive measurements.

In the following, we demonstrate the flexibility and utility of GPPM for psychological data analysis based on a selected set of example models. We start by presenting the GPPM formulation of popular longitudinal panel models. Additionally, we show that the GPPM representations of these models have certain advantages compared to their traditional representations. We close by presenting novel models for longitudinal panel data provided by GPPM that are not expressible as hierarchical linear models, lSEMs, state-space models, or generalized additive models.

Latent Growth Curve Model:
Person-Specific Predictions of Cognitive Change in the Berlin Aging Study

We mentioned earlier that one of the strengths of GPPM is that it allows us to obtain person-specific predictions after the population-level model has been fitted in a straightforward fashion. For instance, we might be interested in the cognitive development of older adults. In the Berlin Aging Study (BASE; Baltes & Mayer, 2001; Delius et al., 2017), cognitive functioning of initially 516 older adults (aged 70+ years) was measured up to 8 times over the course of roughly 20 years. After having performed a LGCM analysis on the BASE data to get at a population-level model, one might be interested in the development of one particular person, for example, in order to select individuals for an intervention.

A data analysis using LGCM results in parameter estimates that can be interpreted as describing a distribution of a Gaussian process. As parameter estimates for the LGCM we assume $\mu_{b_0} = 58$, $\mu_{b_1} = -1$, $\sigma^2_{b_0} = 258$, $\sigma^2_{b_1} = 0.4$, $\sigma^2_{b_0,b1} = 0$, $\sigma^2_{\epsilon} = 10$, which are adapted versions of the values that Ghisletta et al. (2014) reported based on a LGCM-type analysis of perceptual speed on the BASE data (cf. Stoel & Van Den Wittenboer, 2003, for the adaptation procedure). For simplicity, we set the covariance between the intercept and the slope to $\sigma_{b_0,b_1} = 0$, recoded age to be centered at 70, and rounded all values to 0 decimals. These parameters amount to an average mild decline of perceptual speed with no relationship between perceptual speed assessed at age 70 and the severity of the decline. There is, however, considerable between-person variation both with regard to perceptual speed at age 70 and the severity of the decline.

For our exemplary participant, we assume to have obtained two measurements: The value 40 at age 70 (recoded age 0) and the value 25 at age 80 (recoded age 10). We are now interested in predicting the participant's perceptual speed at age 85 (recoded age 15). The joint distribution of the observed $[Y(0), Y(10)]^\top$ and the unobserved perceptual speed values $Y(15)$ according to the fitted GPPM is obtained by plugging in the corresponding values into the mean and kernel functions. Thus, it is a multivariate Gaussian with mean vector

$$\mu = \begin{bmatrix} \mu_{b_0} + 0\mu_{b_1} \\ \mu_{b_0} + 10\mu_{b_1} \\ \mu_{b_0} + 15\mu_{b_1} \end{bmatrix} = \begin{bmatrix} 58 \\ 48 \\ 43 \end{bmatrix}$$

and covariance matrix

$$\begin{aligned} \Sigma &= \begin{bmatrix} \sigma_{b_0}{}^2 + \sigma_{b_1}{}^2 0\cdot 0 + \sigma_\epsilon^2 & \sigma_{b_0}{}^2 + \sigma_{b_1}{}^2 0\cdot 10 & \sigma_{b_0}{}^2 + \sigma_{b_1}{}^2 0\cdot 15 \\ \sigma_{b_0}{}^2 + \sigma_{b_1}{}^2 10\cdot 0 & \sigma_{b_0}{}^2 + \sigma_{b_1}{}^2 10\cdot 10 + \sigma_\epsilon^2 & \sigma_{b_0}{}^2 + \sigma_{b_1}{}^2 10\cdot 15 \\ \sigma_{b_0}{}^2 + \sigma_{b_1}{}^2 15\cdot 0 & \sigma_{b_0}{}^2 + \sigma_{b_1}{}^2 15\cdot 10 & \sigma_{b_0}{}^2 + \sigma_{b_1}{}^2 15\cdot 15 + \sigma_\epsilon^2 \end{bmatrix}. \\ &= \begin{bmatrix} 268 & 258 & 258 \\ 258 & 308 & 318 \\ 258 & 318 & 358 \end{bmatrix} \end{aligned}$$

To obtain the predictive distribution for the perceptual speed at age 85 (recoded as 15), one simply has to condition this joint distribution on the observations at age 70 and 80, which were recoded as $y_i = [\ 40 \quad 25\ ]^\top$. The conditional distribution is again a Gaussian with mean and variance as follows (see Bishop, 2006, Chapter 2.3.1, for the general formula):

$$\begin{aligned} \mathbb{E}\left[Y(15) \mid y_i\right] &= 43 + [\ 258 \quad 318\ ] \begin{bmatrix} 268 & 258 \\ 258 & 308 \end{bmatrix}^{-1} \left(\begin{bmatrix} 40 \\ 25 \end{bmatrix} - \begin{bmatrix} 58 \\ 48 \end{bmatrix} \right) \approx 25 \\ \mathrm{Var}\left[Y(15) \mid y_i\right] &= 358 - [\ 258 \quad 318\ ] \begin{bmatrix} 268 & 258 \\ 258 & 308 \end{bmatrix}^{-1} \begin{bmatrix} 258 \\ 318 \end{bmatrix} \approx 28 \end{aligned}$$

These results are in accordance with the following intuition. Without any measurements, the expectation for a particular person at age 85 is 43. The measurements taken at age 70 and 80 were considerably worse than to be expected given the model. Thus, the expectation for age 85 is also lowered compared to the unconditional expectation.

The predictive distribution can be obtained for and based on an arbitrary number of time points and for any type of kernel and mean function. In this case, we demonstrated how to obtain it for a single time point based on two measurement occasions for simplicity's sake. With more than one time point, the predictive distribution becomes a multivariate Gaussian, so there can be covariance between predictions. Also, predictions can be obtained for latent constructs, like the perceptual speed before being contaminated by measurement error. In this example, all that needs to be done is removal of the measurement error from the model-implied variance for $\Sigma(3,3)$, resulting in $\sigma_{b_0}^{\ 2}+\sigma_{b_1}^{\ 2}15\cdot 15$. In Figure 2, we visualize the predictive distribution for the latent (measurement-error free) perceptual speed for our exemplary individual for the ages 81 to 100. We also present the unconditional predictions, that is, those without conditioning on any observations. These can be interpreted as predictions on the group level, or alternatively, as the predictions for an individual for whom no data is available. As to be seen in the figure, conditioning on the observations not only shifts the predictions but also decreases their uncertainty.

AR(1) Model: Continuous-Time Representation

In this section, we will showcase the ability of GPPM to represent dynamic models like the AR(1) model. We also show how the GPPM representation of dynamic models facilitates the transformation of traditional discrete-time dynamic models to modern continuous-time dynamic models (Oud & Jansen, 2000; Voelkle et al., 2012).

Dynamic models are still predominantly employed in their discrete-time variant, that is, the model is typically formulated for sequential discrete time points (which would translate to integer-valued time indices in our framework). To reflect this fundamental difference in comparison to GPPM, we will denote time as an index instead of a functional argument in our notation and call it j.

The AR(1) model is arguably the most popular dynamic model employed in psychology. The basic idea is to use a linear model to predict the observation at the current time point based on the observation at the previous time point:

$$Y_j = b_0 + b_1 Y_{j-1} + \epsilon_j,$$

with b_0, b_1 being parameters and ϵ_j a Gaussian error term with mean zero and variance σ_ϵ^2. Typically, the wide-sense stationary variant of the AR(1) model is employed. The process Y_j is wide-sense stationary if and only if $|b_1| < 1$.

To convert the AR(1) model in its linear model formulation into a GPPM, we compute the mean and the kernel functions. They are as follows (for the derivation, see Appendix C):

$$\begin{aligned} \mathbb{E}\left(Y_j\right) &= \frac{b_0}{1-b_1^2} := m(j;\theta) \\ \text{Cov}\left(Y_j, Y_{j+r}\right) &= b_1^{\,r} \frac{\sigma_\epsilon^2}{1-b_1^2} := k(j, j+r;\theta) \end{aligned} \tag{5}$$

where r is the discrete-time interval between two not necessarily successive measurement occasions. Thus, the AR(1) model can also be expressed as a GPPM. In the supplement, we again demonstrate that the solutions obtained by SEM software and the GPPM software for this model are identical.

Recently, the advantage of using a continuous-time approach for dynamic models has become more and more appreciated within psychology (Asparouhov, Hamaker, & Muthén, 2017; Boker, 2007; Chow, Ram, Boker, Fujita, & Clore, 2005; Driver et al., 2017; Oud & Jansen, 2000; Voelkle et al., 2012). This is due to the fact that it allows the correct treatment of unequal measurement intervals, for example. Translating the discrete-time AR(1) model into its continuous-time variant has proven to be rather difficult. Using standard approaches to specify the continuous-time AR(1) model, such as SEMs, relies on non-standard extensions of the SEM framework (Oud & Jansen, 2000) or approximate solutions (Boker, 2007). In contrast, specifying the continuous-time AR(1) model as a GPPM is relatively effortless. We simply employ a continuous representation of time and slightly modify the mean and kernel functions in the following way:

$$\begin{aligned} m(t;\theta) &= \frac{b_0}{1-b_1^2} \\ k(t,t+r) &= \exp\left(b_1 r\right)\frac{\sigma_\epsilon^2}{1-b_1^2} \end{aligned} \tag{6}$$

We have exchanged b_1^r by $\exp\left(b_1 r\right)$. Importantly, the distance between two measurements r is allowed to be a continuous value here, while it needs to be discrete in Equation (5). For a more detailed treatment of this topic including a proof that these mean and kernel functions represent the continuous-time AR(1) model, see Karch (2016, Chapter 4.2.1).

The modular nature of GPPM can also be used to combine the AR(1) model (Equation (6)) and the LGCM (Equation (5)). One simply has to add the two models:

$$m(t;\theta) = \mu_{b_0} + \mu_{b_1} t$$
$$k(s,t) = \sigma_{b_0}^2 + \sigma_{b_0,b_1}(s+t) + \sigma_{b_1}^2 st + \exp\left(b_1 \,|\, s-t \,|\right)\frac{\sigma_\epsilon^2}{1-b_1^2}$$

This model can be interpreted as the LGCM with AR(1)-correlated error terms. It accounts for auto-correlations between error terms as they often occur in longitudinal panel studies (Sivo, Fan, & Witta, 2005). If the auto-correlations are not corrected for, they will bias parameter estimation (Sivo et al., 2005). For a full demonstration of this model based on positive affect data from the COGITO study (Schmiedek, Bauer, Lövdén, Brose, & Lindenberger, 2010), an intensive longitudinal data set, see Karch (2016, Chapter 4.2.2). On the COGITO study data, the LGCM with AR(1)-correlated errors was selected over the regular LGCM by all employed model-scoring methods, confirming that this model is indeed more appropriate for some longitudinal data sets.

Smoothing Spline: Extension to Panel Model

Recently, generalized additive modeling (Wood, 2006) has been introduced as a non-parametric alternative for modeling psychological time series data (Bringmann et al., 2017; McKeown & Sneddon, 2014; Shadish et al., 2014; Sullivan et al., 2015). The core technique utilized by generalized additive modeling is the smoothing spline (Hastie, Tibshirani, & Friedman, 2009, Chapter 5.4). In this section, we will demonstrate that GPPM is also able to implement this non-parametric technique. By formulating splines as GPPM, we translate it from a nonparametric function fitting approach into a parametric statistical model. This has the advantage that more inference methods can be leveraged (for example Bayesian inference). More importantly, the GPPM representation permits the application of smoothing splines for the analysis of panel data. The relationship between GPR and smoothing splines is discussed in detail by Seeger (1999).

The starting point for smoothing splines is again the linear growth model introduced at the beginning of this paper:

$$Y(t) = b_0 + b_1 t + \epsilon_t .$$

Here, from a least-squares perspective, we are trying to find a function $f(t)$ such that the residual sums of squares

$$\sum_{t \in \{t_1, \ldots, t_T\}} \left(y_t - f(t) \right)^2$$

is minimized. As candidates, we only consider functions of the form $f(t) = b_0 + b_1 t$, that is, linear functions.

For smoothing splines the residual sums of squares are also minimized. However, “all functions” (technically, only those functions that have two continuous derivatives) are considered as candidates, thus making it a non-parametric method. To avoid overfitting by always selecting a function that perfectly fits the data, an additional constraint favoring smooth functions is added: $\int f''(t)^2 dt$. Thus, smoothness is quantified here as the average-squared second derivative, as denoted by the double prime, that is, a monotonic function of the average-squared change in change. The complete term to minimize is then

$$\sum_{t \in \{t_1, \ldots, t_T\}} \left(y_t - f(t) \right)^2 + \lambda \int f''(t)^2 dt . \qquad \textbf{(7)}$$

The positive scalar λ controls the compromise between model fit and smoothness. $\lambda = \infty$ results in a flat line while $\lambda = 0$ yields any function that fits the data perfectly – and such a function always exists as the set of all functions is considered. The function $g(t)$ that minimizes Equation (7) is called a smoothing spline.

To find the smoothing spline $g(t)$, the following result can be employed, which transforms this hard non-parametric problem into a solvable penalized regression model. It can be shown that the smoothing spline $g(t)$ is a natural cubic spline (Hastie & Tibshirani, 1990, Chapter 2.10). That is, $g(t)$ can be written as

$$g(t) = \sum_{k=1}^{T} b_k h_k(t),$$

where $h_k(t)$ is a *T*-dimensional set of basis functions representing the natural cubic spline. It follows that the penalty term that punishes smoothness can be rewritten as follows (Wood, 2006, p. 140, Exercise 7):

$$\int f''(t)^2 dt = b^\top S b$$

with $b = \left[b_1, \ldots, b_T\right]^\top$ and $S = \int d(t) d(t)^\top dt$, where $d_k(t) = h_k''(t)$. For notational simplicity, we introduce the matrix $X \in \mathbb{R}^{T \times T}$ with $X_{lk} = h_k(t_l)$. Equation (7) simplifies to the following penalized regression problem

$$|| y - Xb || + \lambda b^\top S b,$$

which has the minimizer

$$\hat{b} = \left(X^\top X + \lambda S\right)^{-1} X^\top y. \quad \textbf{(8)}$$

As the first step of translating smoothing splines into a GPPM, we utilize the ability of Bayesian regression to implement penalized regression (Bishop, 2006, Chapter 3.3.1). Linear regression written in the Bayesian notation amounts to the likelihood function $p(y\,|\,X,b,\sigma_\epsilon^2)=\mathcal{N}(y;Xb,\sigma_\epsilon^2)$. As the prior for the regression coefficients weights, we use $p(b)=\mathcal{N}(b;0,\Sigma_p)$ here. The maximum *a-posteriori* estimate for the regression weights is

$$\hat{b}=\left(X^\top X+\sigma_\epsilon^2\Sigma_p^{-1}\right)X^\top y\,.$$

(for the formula, see Rasmussen & Williams, 2006, p. 9). By setting $\Sigma_p^{-1}=S$ and $\sigma_\epsilon^2=\lambda$, we obtain Equation (8).

As the next step, this Bayesian regression model needs to be translated into a GPR model. This is covered in depth by Rasmussen and Williams (2006, Chapter 2). In the supplement, we demonstrate how GPPM software is able to implement smoothing splines by providing a GPPM reimplementation of the example presented by Wood (2006, Chapter 3.2). By representing splines as GPPM, the optimization problem has been turned into a statistical model. Thus, instead of applying the traditional cross-validation to optimize the parameter λ, we can now also use other methods such as model evidence maximization, which is favored by some (cf. Bishop, 2006).

Besides being able to represent spline models as parametric statistical models instead of a nonparametric optimization problem, GPPM also allows their extension. Specifically, it is possible to formulate a panel spline model. In the context of time series modeling, the smoothing parameter λ is optimized using the respective time series. However, if the time series is embedded in a panel data set, it might be better to take all of the participants' data into account to find a smoothing parameter that is optimal for everyone, in order to counteract overfitting. This amounts to the hypothesis that all person-level time series share the same compromise between smoothness and model fit, which nevertheless allows for considerable heterogeneity in the form of the person-level time series and represents a much broader assumption than those implemented by LGCMs, for example.

To illustrate that this model is still able to fit relatively complex population models, we generated data according to a quadratic model:

$$Y_i(t) = b_{0i} + b_{2i}t^2 + \epsilon_{it}, \quad [b_0, b_2]^\top \sim \mathcal{N}(0, I), \quad \epsilon_{it} \sim \mathcal{N}(0, .05) \qquad \textbf{(8)}$$

Thus, both the intercept and the coefficient of the quadratic function varied between persons. We fitted the panel spline model to these data and explored the fit of the model visually based on three randomly selected persons. Figure 3 shows that the panel spline model indeed fits well for all three persons.

Modeling Smoothness with Squared Exponential Kernels

Probably the most popular model in GPR is the so-called squared exponential model (Rasmussen & Williams, 2006, p. 83)

$$\begin{aligned} m(t;\theta) &= 0 \\ k(s,t;\theta) &= \sigma^2 \exp\left(-\frac{(s-t)^2}{\rho}\right), \end{aligned}$$

with the length scale $\rho > 0$ being a parameter that controls how rapidly the auto-covariance declines and the signal variance σ^2 representing the variance at each time point. The squared exponential model serves as a local smoother, for which the locality is defined via the length scale. The smaller the length scale is, the greater the covariance between two equidistant time points, and the greater the trajectory's smoothness.

Both the squared exponential model and the smoothing spline model are smoothers. However, they differ in the ways they formalize smoothness. Which method is better will always depend on the actual data set (Duda, Hart, & Stork, 2001, Chapter 9.2).

Interestingly, the GPPM representation of the AR(1) model also reveals a similarity between the squared exponential and the autoregressive model. The AR(1) model kernel can be reparametrized as follows

$$k(s,t;\theta)=\sigma^2\exp\left(-\frac{|s-t|}{\rho}\right).$$

Thus, the kernel function representing the AR(1) model and the squared exponential model only differ in the distance function employed. The squared exponential function employs the exponential squared distance function $\exp\left(-\frac{(s-t)^2}{\rho}\right)$ whereas the AR(1) model employs the exponential distance function $\exp\left(-\frac{|s-t|}{\rho}\right)$. Karch (2016, Chapter 4.2.1) covered this topic in depth and showed that the squared exponential model may yield a better fit to a dataset that was previously thought to adhere to an AR(1) model, suggesting that it might not only be a worthwhile alternative to the recently introduced method of smoothing splines but also to the established AR(1) model.

In Figure 4, we compare the predictive means obtained by the panel extensions of the squared exponential model, the autoregressive model, and the spline model. As a basis for comparison, we again used the population model introduced in Equation (8) but added more noise ($\sigma_{\epsilon}^{2} = .5$) to induce differences between the methods. The predictions for the three individuals clearly differ by model. Both the spline and the squared exponential model essentially fall back to fitting a line per person. For the spline model, this is relatively easy to understand. The amount of noise in combination with the relatively few measurements per person does not justify a more complex model than a line in terms of compromise between smoothness and model fit. In contrast, the predictive mean obtained by the AR(1) model is considerably more complex and not smooth.

Summary and Discussion

Summary

In this paper, we introduced the novel longitudinal panel data modeling method GPPM. The main advantage of GPPM is its flexibility with regard to the specification of the within-person model, which makes it possible to formulate models appropriate to complex situations for which traditional methods might not work adequately. We have shown some of the possibilities that this flexibility opens up.

Throughout this manuscript, we have demonstrated how the combination rules introduced for temporal GPRs can be utilized to form novel GPPMs based on existing models that are popular in psychology. Another promising avenue in this context is to construct novel models based on (a combination of) existing temporal GPR models (Duvenaud, 2014; Roberts et al., 2013). We have only scratched the surface of this subject by discussing the utility of the squared exponential model for psychological research.

We have also emphasized the capability of GPPM to obtain person-specific predictions easily. This makes it possible to estimate person-specific trajectories that also take into account the data of all other study participants. Person-specific predictions can, for example, be used as screening tools for interventions by identifying those people whose trajectory is at risk of moving into an unsatisfactory direction. Of course, GPPM is not the only method that yields person-specific predictions. However, formulating the model on a Gaussian process in terms of mean and kernel functions results in an easy and natural method. All that needs to be done is to compute the conditional distribution of a multivariate Gaussian. In contrast, lSEM requires introduction of a latent variable into the model for every time point for which a prediction is desired. Different methods can be applied to obtain a prediction for this latent variable (Estabrook & Neale, 2013). Technically, the GPPM predictive distribution is equivalent to the expected posterior method in SEM (Estabrook & Neale, 2013).

In the context of the AR(1) model, we discussed that GPPM is inherently a continuous-time modeling technique and thus takes differences in measurement intervals into account adequately. We also demonstrated that GPPM permits extension of the LGCM with AR(1)-correlated errors. The same mechanism can be used to augment any GPPM with AR(1)-correlated, or alternative error structures, such as the squared exponential kernel or the spline kernel. Modeling smooth residuals in this way may be a good nonparametric approach to take into account the unmodeled complexity of the data.

We have shown that GPPM can also represent exploratory generalized additive models and demonstrated that GPPM representation affords the extension of smoothing splines to a panel modeling technique.

We proceeded with the squared exponential kernel, which is arguably the most used kernel in statistical learning. It is typically used as a universal smoothing method like smoothing splines. We also showed that the squared exponential kernel can be regarded as an alternative to smoothing splines but also to AR(1) models.

Besides the squared exponential model, many other models only expressible by GPPM could provide new modeling approaches for longitudinal psychology data. Roberts et al. (2013) provides an overview of models used for temporal GPR and Duvenaud (2014) reviews those for regular GPR. Among the most promising candidates are periodic models. In the field of psychology these might be useful for modeling oscillations, for example, in brain signals measured with electroencephalography (EEG) or in emotion regulation data (Chow et al., 2005).

Essentially, as a result of its flexibility, GPPM frees researchers from the boundaries set by specific methods. Although we consider this additional freedom as an important step forward, it also highlights the importance of an appropriate model specification. Ideally, the model is formulated based on a careful translation of substantive theory into mathematical formalism and then checked for correctness. The kernel language is beneficial for this task, as it is able to express more models than traditional methods. At the same time, GPPM also provides the means for exploratory analysis, for example, by applying the spline model or the squared exponential model.

GPPM subsumes and generalizes a variety of traditional longitudinal panel modeling methods. Thus, GPPM can be regarded as a unification of existing panel modeling methods in one consistent formal framework, the kernel language. Together with the combination rules we introduced, GPPM allows researchers to create hybrid models that consist of a mixture of models from different traditions, such as the LGCM with AR(1)-correlated errors.

The fact that GPPM can represent many different models in one common formalism can also be utilized for didactical purposes. Comparisons between different methods can be performed more easily as they are not hindered by differences in the modeling language. Fundamentally, GPPM can be used to understand the connections between apparently different methods. For example, the GPPM representations of the smoothing spline and LGCM presented above could be used to investigate the communalities between these two seemingly different methods in more depth. Although we emphasized the connection between GPPM and SEM, there is also work connecting GPR to many other methods stemming from supervised statistical learning such as support vector machines or artificial neural networks (Rasmussen & Williams, 2006, Chapter 6). One interesting result that follows from the identification of SEM as a special case of GPPM, and GPPM's close relation to Bayesian kernel regression, is that every conventional SEM is equivalent to linear regression in some high-dimensional space. Thus, GPPM provides a good starting point to better understand the differences and commonalities among many methods such as SEM and nonlinear regression, both popular in psychology, but also support vector machines, as used in statistical learning.

Limitations and Future Work

GPPM also has disadvantages. In contrast to lSEM, GPPM lacks a graphical model representation. SEMs can be represented by graphical representations as they describe linear relationships between a finite number of variables. GPPM, on the other hand, describes arbitrary relationships between an infinite number of variables within a stochastic process. Thus, it is not straightforward to come up with a graphical representation for GPPMs. At the same time, describing a model on the level of a stochastic process is one of the core advantages of GPPM that contributes to its favorable properties.

While being very flexible with regard to the formulation of the within-person model and also more flexible with regard to the formulation of the between-person model than classical methods such as lSEM and hierarchical linear model, GPPM is still rather limited in terms of the specification of the between-person model. A true between-person model, that is, a set of candidate probability distributions, can only be specified using Gaussian distributions, and only for linear parameters of the mean function. However, deterministic between-person models, that is, a set of functions that describe how a person-specific parameter changes depending on some person-level variables, can be formulated for every parameter. Nevertheless, we would like to be able to specify true between-person models for all parameters. It seems, for example, implausible and inconsistent to assume that parameters such as the autoregressive parameter b_1 in the AR(1) model exhibits no between-person variation, while also assuming that all parameters of the linear growth model vary between persons, as is the case when using LGCM. The multi-subject state-space modeling framework, which is less flexible with regard to the within-person model than GPPM, has recently been extended to allow specification of a between-person model for all parameters (Asparouhov et al., 2017; Driver & Voelkle, in press). This includes the equivalents of non-linear parameters of the mean function and parameters of the kernel function. We plan to extend the flexibility of GPPM with regard to the between-person model in the near future.

Of course, GPPM does not include *every* modeling approach for longitudinal data as a special case. For example, nonlinear relationships between latent and observed variables (Lee & Zhu, 2002) are not possible. If a nonlinear SEM is used to formulate a longitudinal model, the resulting stochastic process is typically no longer a Gaussian process. This results from the fact that transforming a Gaussian variable using a nonlinear function typically does not result in a Gaussian random variable. Given the general assumption that a Gaussian process generates the trajectories is fulfilled, however, GPPM is the most flexible approach possible by definition.

As introduced here, GPPM is only suitable for continuous but not for nominal and ordinal data. In GPR, in equivalence to generalized linear models, so-called link functions (Rasmussen & Williams, 2006, Chapters 3 and 9.3) are used to accommodate non-Gaussian data. This complicates parameter estimation. However, the appropriate algorithms have been developed (Rasmussen & Williams, 2006). Adapting them for use in GPPM remains to be done in the future.

For didactical reasons, we have focused on univariate GPPMs here. However, in practice the joint development of multiple variables is often of interest and requires multivariate models. The specification of multivariate GPPMs is possible. Beyond the auto-kernels, which restrict the form of the auto-covariance for each variable pair, this requires definition of a cross-kernel, which restricts the form of the cross-covariance.

Technological progress enables us to obtain unprecedented amounts of measurements per person, and therefore, speeding up model-fitting algorithms for longitudinal models is becoming increasingly important. GPPM has much potential in this regard. Karch (2016, Chapter 4.3.1) performed a theoretical analysis, which revealed that the standard GPPM fitting algorithm already avoids certain costly computations necessary for SEM fitting algorithms. However, the overall running time is about equal for SEM and GPPM, as the number of measurements T contributes cubically $O(T^3)$ in both frameworks. Various approximations for faster model fitting have been proposed for Gaussian process regression (Hartikainen & Särkkä, 2010; Lawrence, Seeger, & Herbrich, 2003; Leithead & Zhang, 2007; Quiñonero-Candela & Rasmussen, 2005; Zhang & Leithead, 2007) since it is typically applied to substantially larger data sets than SEM. Transferring these approximation algorithms to make them applicable to GPPM could substantially decrease the amount of time needed to fit GPPMs, and consequently SEMs. Whether the resulting approximation error is in an acceptable range needs to be investigated.

GPPM also provides many opportunities for exploratory data analysis. As we have discussed, the squared exponential model can be considered an alternative to the nonparametric regression technique smoothing splines and is able to fit any continuous trajectory. In the context of GPR models, exploratory analysis has been taken one step further by an algorithm that automatically learns the kernel function from data and then describes the model in natural language (Lloyd, Duvenaud, Grosse, Tenenbaum, & Ghahramani, 2014). Extending this algorithm for use in GPPM would allow the appropriate model to be learned on the basis of a longitudinal panel data set, including a natural language description of the model, and thus facilitate exploratory data analysis.

Conclusion

In this paper, we have presented GPPM, a novel flexible method for modeling longitudinal data based on the Bayesian non-parametric method of Gaussian process regression. The overarching advantage of this new method is its flexibility in specifying a within-person model. Due to lack of space, we could only cover a fraction of the possibilities that GPPM offers but hope to spawn interest and further research into kernel-based longitudinal modeling of psychological panel data.

References

Asparouhov, T., Hamaker, E. L., & Muthén, B. (2017). Dynamic latent class analysis. *Structural Equation Modeling: A Multidisciplinary Journal*, *24*(2), 257–269. https://doi.org/10.1080/10705511.2016.1253479

Baltes, P. B., & Mayer, K. U. (2001). *The Berlin Aging Study: Aging from 70 to 100*. Cambridge, UK: Cambridge University Press.

Barrett, P. (2007). Structural equation modelling: Adjudging model fit. *Personality and Individual Differences*, *42*(5), 815–824. https://doi.org/10.1016/j.paid.2006.09.018

Bishop, C. M. (2006). *Pattern recognition and machine learning* (1st ed.). New York, NY: Springer.

Boker, S. M. (2007). Specifying latent differential equations models. In S. M. Boker & M. J. Wenger (Eds.), *Data analytic techniques for dynamical systems in the social and behavioral sciences* (pp. 131–160). Mahwah, NJ: Lawrence Erlbaum.

Bollen, K. A. (1989). *Structural equations with latent variables* (1st ed.). New York, NY: Wiley.

Brahim-Belhouari, S., & Bermak, A. (2004). Gaussian process for nonstationary time series prediction. *Computational Statistics & Data Analysis*, *47*(4), 705–712. https://doi.org/10.1016/j.csda.2004.02.006

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

Chow, S.-M., Ho, M. R., Hamaker, E. L., & Dolan, C. V. (2010). Equivalence and differences between structural equation modeling and state-space modeling techniques. *Structural Equation Modeling: A Multidisciplinary Journal*, *17*(2), 303–332. https://doi.org/10.1080/10705511003661553

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

Cohen, J. (1968). Multiple regression as a general data-analytic system. *Psychological Bulletin*, *70*(6, Pt.1), 426–443. https://doi.org/10.1037/h0026714

Cox, G. E., Kachergis, G., & Shiffrin, R. M. (2012). Gaussian process regression for trajectory analysis. In *Proceedings of the 34th Annual Conference of the Cognitive Science Society* (pp. 1440–1445). Retrieved from http://palm.mindmodeling.org/cogsci2012/papers/0256/paper0256.pdf

Curran, P. J. (2003). Have multilevel models been structural equation models all along? *Multivariate Behavioral Research*, *38*(4), 529–569. https://doi.org/10.1207/s15327906mbr3804_5

Damouras, S. (2008). *Nonparametric time series analysis using Gaussian processes* (Doctoral dissertation). Carnegie Mellon University. Retrieved from http://www.utsc.utoronto.ca/~sdamouras/files/thesis.pdf

Delius, J. A. M., Düzel, S., Gerstorf, D., & Lindenberger, U. (2017). Berlin Aging Studies (BASE and BASE-II). In N. A. Pachana (Ed.), *Encyclopedia of geropsychology* (pp. 386–395). New York: Springer. https://doi.org/10.1007/978-981-287-082-7_44

Driver, C. C., Oud, J. H. L., & Voelkle, M. C. (2017). Continuous time structural equation modeling with R package ctsem. *Journal of Statistical Software*, *77*(5). https://doi.org/10.18637/jss.v077.i05

Driver, C. C., & Voelkle, M. C. (in press). Hierarchical Bayesian continuous time dynamic modeling. *Psychological Methods*.

Duda, R. O., Hart, P. E., & Stork, D. G. (2001). *Pattern classification* (2nd ed.). New York: Wiley-Interscience.

Duncan, T. E., Duncan, S. C., & Strycker, L. A. (2006). *An introduction to latent variable growth curve modeling: Concepts, issues and applications* (2nd rev. ed.). Mahwah, N.J: Lawrence Erlbaum Associates Inc.

Duvenaud, D. K. (2014). *Automatic model construction with Gaussian processes* (Doctoral dissertation). University of Cambridge. Retrieved from https://www.repository.cam.ac.uk/handle/1810/247281

Duvenaud, D. K., Nickisch, H., & Rasmussen, C. E. (2011). *Additive Gaussian processes*. (J. Shawe-Taylor, R. S. Zemel, J. C. Bartlett, F. Pereira, & K. Q. Weinberger, Eds.). Red Hook, NY: Curran Associates. Retrieved from http://papers.nips.cc/paper/4221-additive-gaussian-processes

Estabrook, R., & Neale, M. C. (2013). A comparison of factor score estimation methods in the presence of missing data: Reliability and an application to nicotine dependence.

Multivariate Behavioral Research, *48*(1), 1–27. https://doi.org/10.1080/00273171.2012.730072

Gelman, A. (2006). Multilevel (hierarchical) modeling: What it can and cannot do. *Technometrics*, *48*(3), 432–435. https://doi.org/10.1198/004017005000000661

Ghisletta, P., Bäckman, L., Bertram, L., Brandmaier, A. M., Gerstorf, D., Liu, T., & Lindenberger, U. (2014). The Val/Met polymorphism of the brain-derived neurotrophic factor (BDNF) gene predicts decline in perceptual speed in older adults. *Psychology and Aging*, *29*(2), 384–392. https://doi.org/10.1037/a0035201

Grigorievskiy, A., & Karhunen, J. (2016). Gaussian process kernels for popular state-space time series models. In *2016 International Joint Conference on Neural Networks (IJCNN)* (pp. 3354–3363). Red Hook, NY: Curran Associates. https://doi.org/10.1109/IJCNN.2016.7727628

Hall, P., Müller, H.-G., & Yao, F. (2008). Modelling sparse generalized longitudinal observations with latent Gaussian processes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *70*(4), 703–723. https://doi.org/10.1111/j.1467-9868.2008.00656.x

Hamilton, J. D. (1994). *Time series analysis* (1st ed.). Princeton, N.J: Princeton University Press.

Hartikainen, J., & Särkkä, S. (2010). Kalman filtering and smoothing solutions to temporal Gaussian process regression models. In *2010 IEEE International Workshop on Machine Learning for Signal Processing (MLSP)* (pp. 379–384). Red Hook, NY: Curran Associates. https://doi.org/10.1109/MLSP.2010.5589113

Hastie, T., & Tibshirani, R. (1990). *Generalized additive models*. Boca Raton, FL: Chapman and Hall/CRC.

Hastie, T., Tibshirani, R., & Friedman, J. H. (2009). *The elements of statistical learning* (2nd ed., Vol. 1). New York: Springer.

Karch, J. D. (2016). *A machine learning perspective on repeated measures.* (Doctoral dissertation). Humboldt-Universität zu Berlin, Lebenswissenschaftliche Fakultät. Retrieved from https://edoc.hu-berlin.de/docviews/abstract.php?id=43051

Lawrence, N. D., Seeger, M., & Herbrich, R. (2003). Fast sparse Gaussian process methods: The informative vector machine. In *Advances in neural information processing Systems* (Vol. 15, pp. 609–616). Cambridge, MA: MIT Press.

Lee, S.-Y., & Zhu, H.-T. (2002). Maximum likelihood estimation of nonlinear structural equation models. *Psychometrika*, *67*(2), 189–210. https://doi.org/10.1007/BF02294842

Leithead, W. E., & Zhang, Y. (2007). O(N 2)-operation approximation of covariance matrix inverse in Gaussian process regression based on quasi-Newton BFGS method. *Communications in Statistics – Simulation and Computation*, *36*(2), 367–380. https://doi.org/10.1080/03610910601161298

Liu, Z., Wu, L., & Hauskrecht, M. (2013). Modeling clinical time series using Gaussian process sequences. In *Proceedings of the 2013 SIAM International Conference on Data Mining* (pp. 623–631). https://doi.org/10.1137/1.9781611972832.69

Lloyd, J. R., Duvenaud, D. K., Grosse, R., Tenenbaum, J. B., & Ghahramani, Z. (2014). Automatic construction and natural-language description of nonparametric regression models. In *Proceedings of the 28th AAAI Conference on Artificial Intelligence*. Retrieved from http://www.aaai.org/ocs/index.php/AAAI/AAAI14/paper/download/8240/8564

MATLAB. (2014). *Version 8.4.0 (R2014b)*. Natick, MA: The MathWorks Inc.

McKeown, G. J., & Sneddon, I. (2014). Modeling continuous self-report measures of perceived emotion using generalized additive mixed models. *Psychological Methods*, *19*(1), 155–174. https://doi.org/10.1037/a0034282

Neale, M. C., Hunter, M. D., Pritikin, J. N., Zahery, M., Brick, T. R., Kirkpatrick, R. M., … Boker, S. M. (2016). OpenMx 2.0: Extended structural equation and statistical modeling. *Psychometrika*, *81*, 535–549. https://doi.org/10.1007/s11336-014-9435-8

Oud, J. H. L., & Jansen, R. A. R. G. (2000). Continuous time state space modeling of panel data by means of sem. *Psychometrika*, *65*(2), 199–215. https://doi.org/10.1007/BF02294374

Oud, J. H. L., & Singer, H. (2008). Continuous time modeling of panel data: SEM versus filter techniques. *Statistica Neerlandica*, *62*(1), 4–28. https://doi.org/10.1111/j.1467-9574.2007.00376.x

Pek, J., & Wu, H. (2015). Profile likelihood-based confidence intervals and regions for structural equation models. *Psychometrika*, *80*(4), 1123–1145. https://doi.org/10.1007/s11336-015-9461-1

Quiñonero-Candela, J., & Rasmussen, C. E. (2005). A unifying view of sparse approximate Gaussian process regression. *Journal of Machine Learning Research*, *6*, 1939–1959.

R Core Team. (2015). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org/

Rasmussen, C. E., & Williams, C. K. I. (2006). *Gaussian processes for machine learning*. Cambridge, MA: MIT Press.

Raudenbush, S. W., & Bryk, A. S. (2001). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.

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

Saatçi, Y., Turner, R. D., & Rasmussen, C. E. (2010). Gaussian process change point models. In *Proceedings of the 27th International Conference on Machine Learning* (pp. 927–934). Retrieved from http://machinelearning.wustl.edu/mlpapers/paper_files/icml2010_SaatciTR10.pdf

Schmiedek, F., Bauer, C., Lövdén, M., Brose, A., & Lindenberger, U. (2010). Cognitive enrichment in old age: Web-based training programs. *GeroPsych*, *23*(2), 59–67. https://doi.org/10.1024/1662-9647/a000013

Seeger, M. (2002). *Relationships between Gaussian processes, support vector machines and smoothing splines* (Technical report). Institute for Adaptive and Neural Computation (ANC), University of Edinburgh, UK.

Shadish, W. R., Zuur, A. F., & Sullivan, K. J. (2014). Using generalized additive (mixed) models to analyze single case designs. *Journal of School Psychology*, *52*(2), 149–178. https://doi.org/10.1016/j.jsp.2013.11.004

Sivo, S., Fan, X., & Witta, L. (2005). The biasing effects of unmodeled ARMA time series processes on latent growth curve model estimates. *Structural Equation Modeling: A Multidisciplinary Journal*, *12*(2), 215–231. https://doi.org/10.1207/s15328007sem1202_2

Stoel, R. D., & Van Den Wittenboer, G. (2003). Time dependence of growth parameters in latent growth curve models with time invariant covariates. *Methods of Psychological Research*, *8*(1), 21–41.

Sullivan, K. J., Shadish, W. R., & Steiner, P. M. (2015). An introduction to modeling longitudinal data with generalized additive models: Applications to single-case designs. *Psychological Methods*, *20*(1), 26–42. https://doi.org/10.1037/met0000020

Taboga, M. (2012a). Likelihood ratio test. In *Lectures on Probability Theory and Mathematical Statistics* (2nd ed.). Lexington, KY: CreateSpace Independent Publishing Platform. Retrieved from http://www.statlect.com/likelihood_ratio_test.htm

Taboga, M. (2012b). Maximum likelihood. In *Lectures on Probability Theory and Mathematical Statistics* (2nd ed.). Lexington, KY: CreateSpace Independent Publishing Platform. Retrieved from http://www.statlect.com/fundamentals-of-statistics/maximum-likelihood

Voelkle, M. C., Oud, J. H. L., Davidov, E., & Schmidt, P. (2012). An SEM approach to continuous time modeling of panel data: Relating authoritarianism and anomia. *Psychological Methods*, *17*(2), 176–192. https://doi.org/10.1037/a0027543

West, S. G., Taylor, A. B., & Wu, W. (2012). Model fit and model selection in structural equation modeling. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 209–231). New York: Guilford Press.

Wood, S. (2006). *Generalized additive models: An introduction with R*. Boca Raton, FL: CRC Press.

Zhang, Y., & Leithead, W. E. (2007). Approximate implementation of the logarithm of the matrix determinant in Gaussian process regression. *Journal of Statistical Computation and Simulation*, *77*(4), 329–348. https://doi.org/10.1080/10629360600569279

Ziegler, G., Ridgway, G. R., Dahnke, R., & Gaser, C. (2014). Individualized Gaussian process-based prediction and detection of local and global gray matter abnormalities

in elderly subjects. *NeuroImage*, *97*, 333–348.
https://doi.org/10.1016/j.neuroimage.2014.04.018

Figure Captions

Figure 1. Extended SEM representation of generic GPPM. Here, t_{ij} refers to the time point of the j-th observation of person i.

Figure 2. Visualization of the person-specific predictive distribution obtained using a LGCM and an exemplary individual. On the left, the predictive distribution without taking the person's data into account is shown. On the right, the data are taken into account. The observed data are marked as squares (left) and circles (right) and are identical. The red line describes the mean of the predictive distribution and the shaded area delineates the variance. The black lines demarcate twice the point-wise standard deviation, that is, the shaded area is a 95% credibility region.

Figure 3. Visualization of the success of the spline model. The fitting result for one person is visualized in each of the panels. The mean of the predictive distribution is shown in brown. The true underlying latent process, which is corrupted by measurement error, is a quadratic and shown in purple. The circles represent the observations.

Figure 4. Visualization of the differences between the squared exponential, the spline, and the AR(1) model. The fitting results for one person are shown in each of the panels. The mean of the predictive distribution is shown in brown for the spline model, in red for the squared exponential model, and in green for the AR(1) model. The true underlying latent process, which is corrupted by measurement error, is a quadratic and shown in purple. The circles represent the observations.

Figure 1

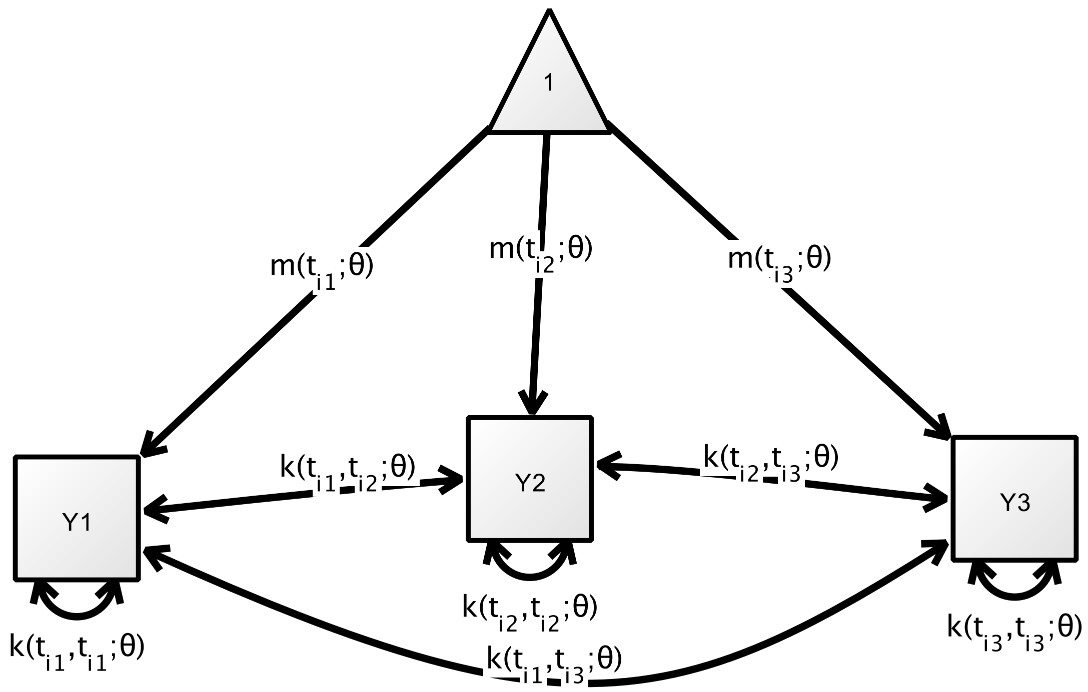

1
m(t_i1;θ)
m(t_i2;θ)
m(t_i3;θ)
Y1
Y2
Y3
k(t_i1,t_i2;θ)
k(t_i2,t_i3;θ)
k(t_i1,t_i1;θ)
k(t_i2,t_i2;θ)
k(t_i3,t_i3;θ)
k(t_i1,t_i3;θ)

Figure 2

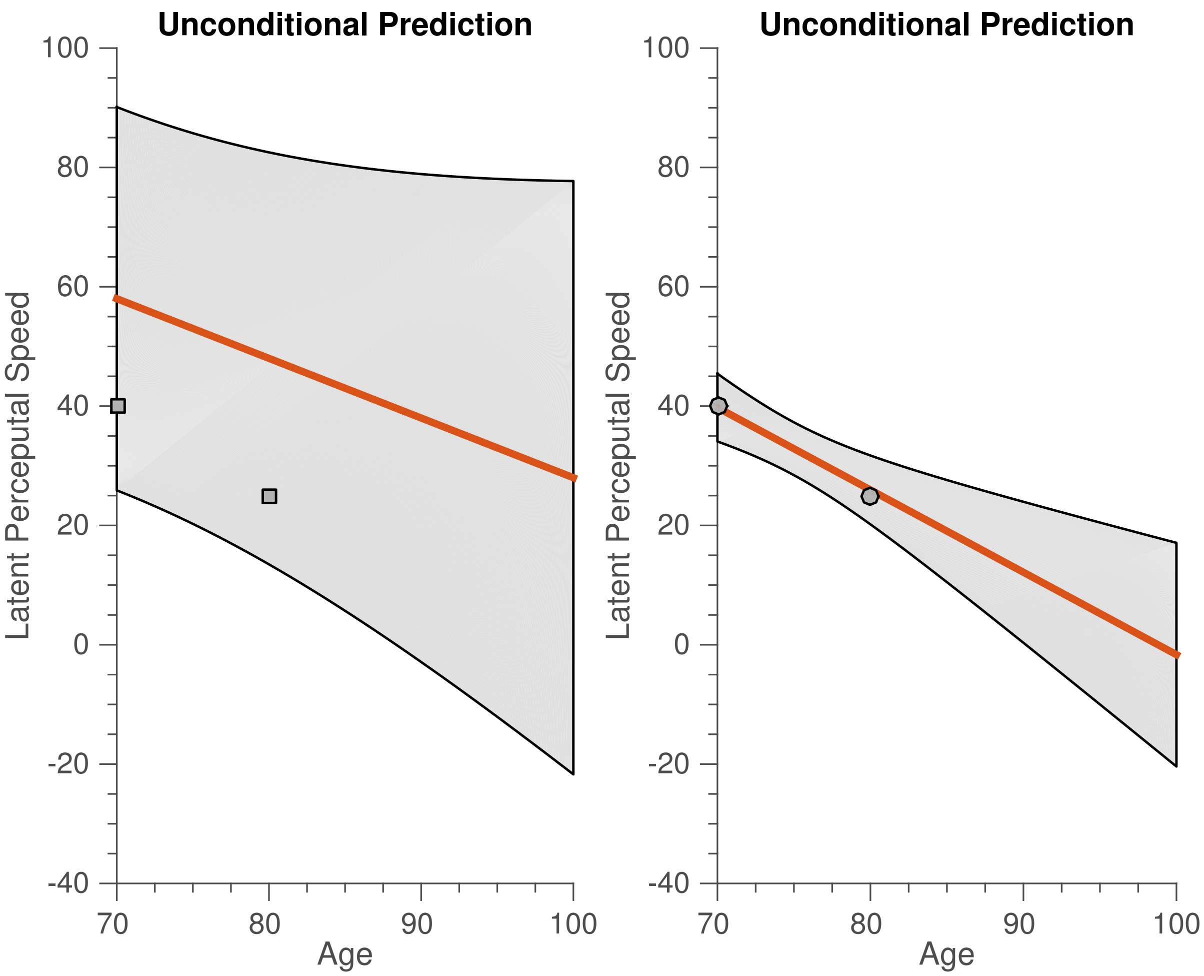

Unconditional Prediction
Latent Perceputal Speed
Age
100
80
60
40
20
0
-20
-40
70
80
90
100
Unconditional Prediction
Latent Perceputal Speed
Age
100
80
60
40
20
0
-20
-40
70
80
90
100

Figure 3

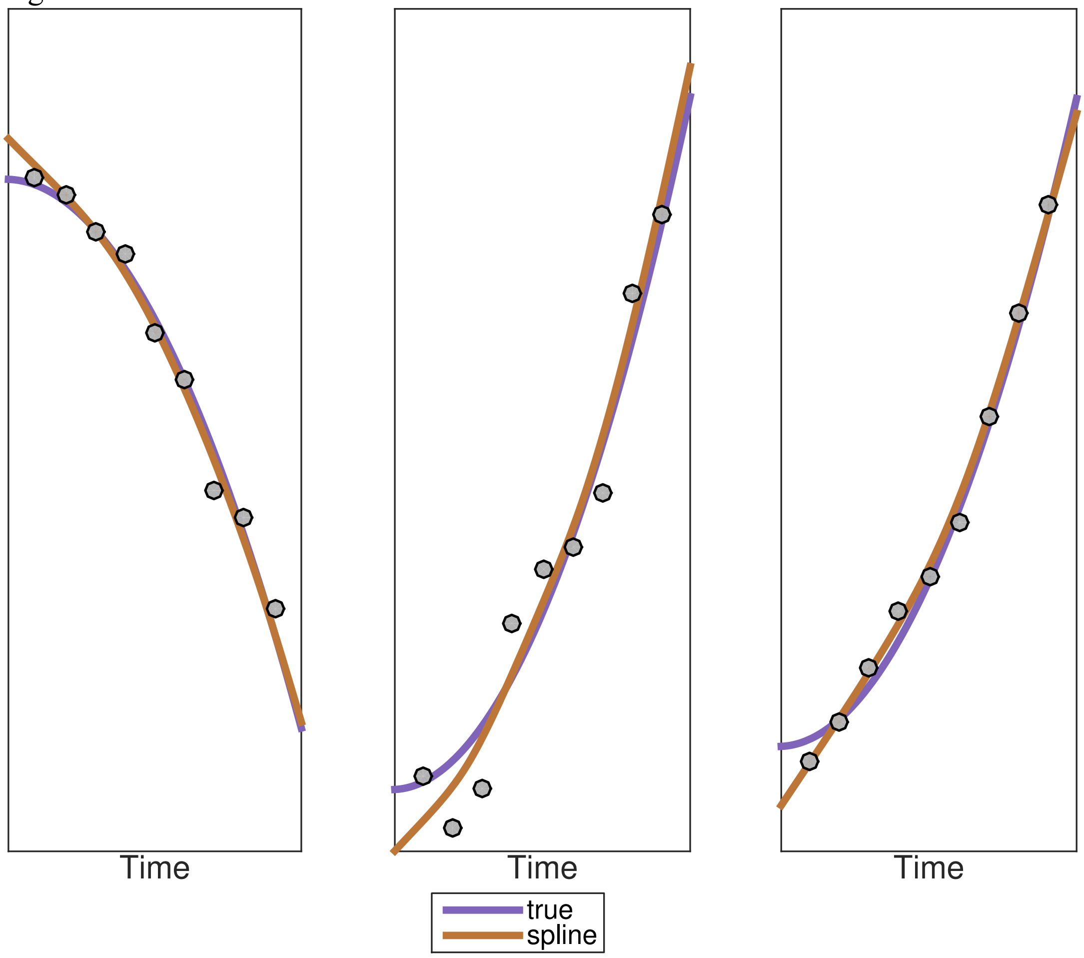
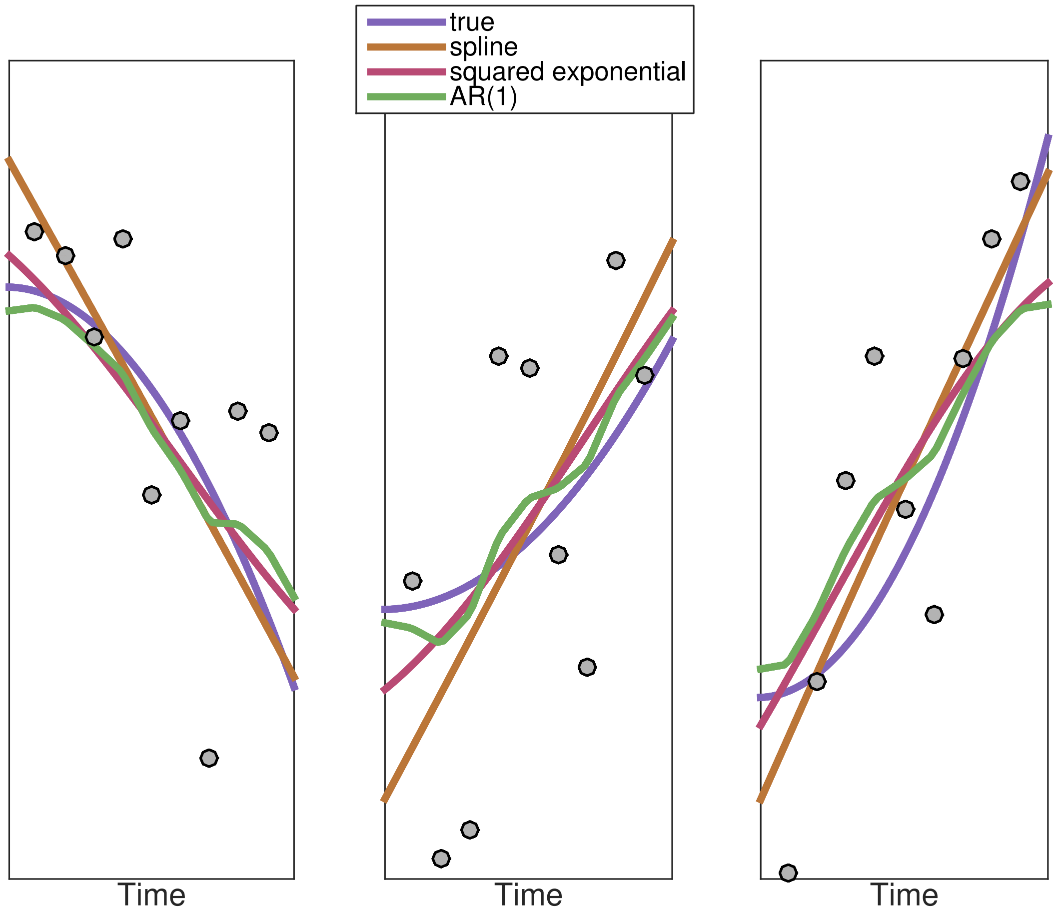


Figure 4



Appendix A

Temporal GPR is typically used within the Bayesian inference framework whereas we have introduced it within a frequentist framework. When using temporal GPR in conjunction with frequentist inference, the model is mostly specified in the mean function, and the kernel function only represents the measurement error. In contrast to that, when using temporal GPR in conjunction with Bayesian inference, the model is mostly specified in the kernel function. We will explain why this is based on an example.

If we take the linear model without measurement error introduced at the beginning

$$Y(t) = b_0 + b_1 t$$

and augment it with the prior $[b_0, b_1]^\top \sim \mathcal{N}(0, \sigma_b^2 I)$ the temporal GPR representation changes from

$$\begin{aligned} m(t;\theta) &= b_0 + b_1 t \\ k(s,t;\theta) &= 0 \end{aligned}$$

to

$$\begin{aligned} m(t;\theta) &= 0 \\ k(s,t;\theta) &= \sigma_b^2 + \sigma_b^2 st \end{aligned}$$

Thus, the core part of the model has moved from the mean function to the kernel function. This was a direct result of assuming a zero mean Gaussian prior on the regression weights. This is also the reason why most GPR models have constant or even zero mean functions, the assumptions are mostly encoded in the kernel function, and why different models are mostly characterized by their kernel function. Interestingly, both the mean and the kernel function are typically important for model specification in GPPM.

Appendix B

Let X_i be the predictor matrix for person i. That, is X_i has the predictor vector $x_{i1}^{\top},\ldots,x_{iT_i}^{\top}$ as rows. The vector y_i contains the observed time series for person i. For notational convenience, we define

$$
\begin{aligned}
M(X_i;\theta) &= [m(x_{i1};\theta),m(x_{i2};\theta),\ldots,m(x_{iT_i};\theta)]^{\top} \\
K(X_i) &= \begin{bmatrix} k(x_{i1},x_{i1}) & k(x_{i1},x_{i2}) & \cdots & k(x_{i1}x_{iT_i}) \\ k(x_{i2},x_{i1}) & k(x_{i2},x_{i2}) & & \vdots \\ \vdots & & \ddots & \\ k(x_{ij},x_{iT_i}) & \cdots & & k(x_{ij},x_{iT_i}) \end{bmatrix}.
\end{aligned}
$$

Maximum likelihood estimation

For a given data set $D=(X,\mathbf{y})$, with $X=(X_1,\ldots,X_N)$ and $\mathbf{y}=(y_1,\ldots,y_N)$ the likelihood function for a GPPM is

$$L(\theta|D)=p(\mathbf{y}|X,\theta)=\prod_{i=1}^{N}\mathcal{N}(y_i;M(X_i;\theta),K(X_i,X_i;\theta)).$$

The maximum likelihood estimate is thus $\hat{\theta}=\text{argmax}_{\theta\in\Theta}L(\theta|D)$.

For most GPPMs, the maximum of the likelihood function cannot be derived analytically. As a remedy, numerical optimization methods are used.

Likelihood-Ratio Test

One important condition for the likelihood-ratio test is that the null hypothesis can be expressed by a constraint g(θ) = 0. A GPPM can be restricted in such a way. For the likelihood-ratio test to be valid, i.e., for the test statistic to asymptotically converge to a Chi-squared distribution when the null hypothesis is true, both the restricted maximum likelihood (ML) estimator $\hat{\theta}_R(Y)$ and the unrestricted estimator $\hat{\theta}_R(Y)$ need to be asymptotically Gaussian (Taboga, 2012a).

A complete proof of the asymptotical normality of the estimators is not within the scope of this manuscript. However, we will provide the intuition for the proof: We assume that the number of time points observed for each person is equal. It is known that the ML estimator and the restricted ML estimator are asymptotically Gaussian for SEMs. Both a SEM and the statistical model implied by GPPM for a particular data set $D = (X, \mathbf{y})$ can be written in the form

$$p(\mathbf{y} \mid X) \in \left\{ \prod_{i=1}^{N} \mathcal{N}(y_i; \mu(X_i;\theta), \Sigma(X_i;\theta) : \theta \in \Theta \right\}$$

For a SEM, the implied mean $\mu(X_i;\theta)$ and covariance matrix $\Sigma(X_i;\theta)$ are the same for every person. Furthermore, the mean $\mu(X_i;\theta)$ and the covariance matrix $\Sigma(X_i;\theta)$ must correspond to a set of linear structural equations with Gaussian noise. In contrast to that, the implied mean $\mu(X_i;\theta)$ and covariance matrix $\Sigma(X_i;\theta)$ for GPPMs may be different for every person and can in principle have any form (as long as $\Sigma(X_i;\theta)$ is a valid covariance matrix for each parameter value θ).

Thus, GPPM extends SEM in two ways: first, there is no restriction on the parameterization of the mean vector and the covariance matrix, and second, the mean and the covariance matrix may be different for every person. The first extension does not violate any of the assumptions for asymptotic normality (Taboga, 2012b). The second extension violates the assumption of an independent and identical distribution. However, since the person-specific differences in the mean and the covariance matrix are produced by entering the predictors into template mean and covariance matrices, the conditional independent and identically distributed assumption still holds. This assumption is sufficient for the restricted as well as the unrestricted (conditional) maximum likelihood estimator to be Gaussian.

Confidence Regions

The validity of likelihood-based confidence intervals and regions follows directly from the validity of the likelihood-ratio test (Pek & Wu, 2015).

Appendix C

Proof that the GPPM representation of the discrete-time AR(1) model is

$$\mathbb{E}\left(Y_j\right) = \frac{b_0}{1-b_1^2} := m(j;\theta)$$
$$\mathrm{Cov}\left(Y_j, Y_{j+r}\right) = b_1{}^r \frac{\sigma_\epsilon^2}{1-b_1^2} := k(j, j+r;\theta)$$

Because of the stationarity it follows that $\mathbb{E}\left(Y_j\right) = \mathbb{E}\left(Y_{j-1}\right) =: \mu$.

$$\begin{aligned}
&\mathbb{E}\left(Y_j\right) = \mathbb{E}\left(b_0 + b_1 Y_{j-1} + \epsilon_j\right) \\
&\Rightarrow \mu = b_0 + b_1\mu + 0 \\
&\Leftrightarrow (1-b_1)\mu = b_0 \\
&\Leftrightarrow \mu = \frac{b_0}{1-b_1}
\end{aligned}$$

The calculation of the auto-covariance is slightly more complex. First, we calculate the variance. We use the same strategy as for the expectation. From stationarity, it follows that $\mathrm{Var}\left(Y_j\right) = \mathrm{Var}\left(Y_{j-1}\right) =: \sigma^2$.

$$\begin{aligned}
&\mathrm{Var}\left(Y_j\right) = \mathrm{Var}\left(b_0 + b_1 Y_{j-1} + \epsilon_j\right) \\
&\Rightarrow \sigma^2 = b_1^2\sigma^2 + \sigma_\epsilon^2 \\
&\Leftrightarrow (1-b_1^2)\sigma^2 = \sigma_\epsilon^2 \\
&\Leftrightarrow \sigma^2 = \frac{\sigma_\epsilon^2}{(1-b_1^2)}
\end{aligned}$$

Now, we are equipped to calculate the auto-covariance. First note that only the mean depends on the parameter b_0; not the variance. Thus, to calculate the auto-covariance we can set $b_0 = 0$.

Consequently,

$$
\begin{aligned}
Y_{j+1} &= b_1 Y_j + \epsilon_{j+1} \\
Y_{j+2} &= b_1{}^2 Y_j + b_1 \epsilon_{j+1} + \epsilon_{j+2} \\
Y_{j+3} &= b_1{}^3 Y_j + b_1{}^2 \epsilon_{j+1} + b_1 \epsilon_{j+2} + \epsilon_{j+3} \\
Y_{j+r} &= b_1{}^r Y_j + \sum_{i=0}^{r-1} b_1^i \epsilon_{j+r-i}
\end{aligned}
$$

Thus,

$$\mathrm{Cov}\left(Y_j, Y_{j+r}\right) = \mathrm{Cov}\left(Y_j, b_1{}^r Y_j + \sum_{i=0}^{r-1} b_1^i \epsilon_{j+r-i}\right)$$

However, since all ϵ_{j+r-i} are independent of Y_j this simplifies to:

$$\mathrm{Cov}\left(Y_j, Y_{j+r}\right) = \mathrm{Cov}\left(Y_j, b_1{}^r Y_j\right) = b_1{}^r \mathrm{Cov}\left(Y_j, Y_j\right) = b_1{}^r \sigma^2 = b_1{}^r \frac{\sigma_\epsilon^2}{(1-b_1^2)}$$