# Non-linear Methods for Intensive Longitudinal Data

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

Tilburg University

## Author Note

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

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

**Abstract**

Here could your abstract be!

*Keywords:* Here could your keywords be!

# Non-linear Methods for Intensive Longitudinal Data

## Simulation Study

### *Problem*

We conducted a simulation study to evaluate how well each of the introduced parametric and non-parametric statistical methods recovers different non-linear processes, which researchers are likely to encounter in ESM research. To facilitate an accessible introduction to these statistical methods and apply them under conditions for which software implementations are currently available, we employed a univariate single-subject design, such that the simulated data corresponds to repeated measurements of one variable for one individual. The theory underlying the non-linear analysis methods (i.e., generalized additive models, local polynomial regression, Gaussian processes) suggests, that each method in its default configuration should infer processes most accurately, if they (a) are without sudden jumps (i.e., continuous), (b) have constant wigglyness (i.e., constant second derivative), and (c) if they are smooth (i.e., differentiable). Thus, we specifically considered processes in this simulation that varied with regards to these characteristics. Additionally, it is expected that the inference is more accurate for larger sample sizes. Here we varied both (d) the overall length of the sampling period and (e) the sampling frequency, as both of these factors were expected to affect the inference differently for different processes and methods.

### *Design*

To investigate their impact on the perfomance of the different analysis methods we manipulated the continuity, wigglyness, and smoothness, of the underlying process. To test (a) the continuity and (b) wigglyness we chose four exemplar processes that vary on these two characteristics. Each of these exemplar processes represents a theoretically feasible and empirically demonstrated psychological process which might be interesting to study using ESM. Depictions of each of the processes can be seen in Figure 1, in which the points

represent the noisy observations and the line shows the latent process value at each time point.

The first two processes depict common growth curves in which the latent construct approaches a fixed value over time. Here, the exponential (1) and logistic (2) growth mainly differ in their respective growth velocity during the start of the process. Growth curves like these have been found to appear in cognitive and intellectual development (Kunnen, 2012; McArdle et al., 2002), motor learning (Newell et al., 2001), and second language acquisition (De Bot et al., 2007). Both of these growth curves are continuous and change their wigglyness over time.

The third process has been generated from a cusp catastrophe model. This dynamic model exhibits natural jumps or discontinuities when varying one of its parameters over time. This has made it popular for describing sudden changes in attitudes (van der Maas et al., 2003), flow (Ceja & Navarro, 2009), or alcohol use relapse (Witkiewitz & Marlatt, 2007). In addition to that, the cusp model also describes the average behavior of dichotomous network models (Finnemann et al., 2021), which are a popular way of modelling clinical symptoms (Borsboom & Cramer, 2013). This makes the cusp an attractive candidate for modelling therapy effects under these network theories.

The last depicted process is a damped oscillator. As such it describes the return to baseline after a perturbation, while oscillating around this baseline. Because of this behavior, the damped oscillator has been proposed as a model for affect regulation (Chow et al., 2005; Waugh & Kuppens, 2021). This model shows a clear decrease in wigglyness over time. All of the considered processes were generated from state-space models, such that the value of the process at a given time point is determined by a (non-) linear function of the previous process value.

To manipulate the (c) smoothness of the process we additionally added a dynamic or process noise component to the respective state-space models. This means, that at each time point the process value does not only depend on the previous process value but also

on a random error, that is usually attributed to external shocks or perturbations. This results in processes that are not smooth (i.e., non-differentiable), where the degree of the roughness depends on the variance of the dynamic error component. The dynamic error standard deviations where chosen in relation to the range of the process. Here error standard deviations of **0.25** and **0.5** were considered reasonable. Figure 2 shows possible realizations of the exemplar processes with dynamic noise. Importantly, a condition lacking dynamic noise was intentionally omitted from our study, as dynamic noise is reasonably expected to be present in all ESM measures.

Additionally, we altered the sample size during the simulation. Here we manipulated either (d) the sampling period or (e) the sampling frequency, since these are distinct methodological choices that are expected to affect the performance of the analysis methods differently. For example, since the local polynomial estimator (Fan et al., 1997) only uses data in the local neighborhood, extending the sampling period outside the neighborhood might not improve the estimation. On the contrary, if the process changes its optimal bandwidth over time, extending the sampling period might even negatively affect the inference. However, increasing the sampling frequency would provide more data within the local neighborhood and hence improve the inference. If one instead considers a method that uses all available data and does not assume constant wigglyness, such as the generalized additive models, both ways of increasing the sample size should facilitate the inference similarly.

Therefore, the simulation consisted of four factors that were crossed in a full factorial design:

1. Exemplar processes (combining continuity and wigglyness)
2. Smoothness ($\sigma = 0.25$; $\sigma = 0.5$;)
3. Sample size: Frequency (20 - 225; Wrzus and Neubauer (2023))
4. Sample size: Period length (20 - 225; Wrzus and Neubauer (2023))

## Procedure

For simulating the data, each of the exemplar processes was represented as a generative state-space model, such that at each time-point the value of the process corresponds to a (non-) linear function of the previous value and the dynamic noise component. In each condition we simulated data from each of the exemplar processes with the respective dynamic error variance, sampling frequency, and sampling period length. The parameters of the exemplar processes were chosen, so that they vary over a comparable range. Lastly, measurement errors from a standard normal distribution were added to the process values at each time point to generate noisy observations. For a detailed and technical explanation of the data generation process, please see **Appendix A**.

In order to determine the required amount of data sets per condition, we conducted a power simulation based on a pilot sample of 30 generated data sets per cell. Based on these data sets, we calculated the outcome measures (i.e., one MSE and one GCV value per data set) to which we fit the appropriate MANOVA model. This MANOVA model was used to obtain initial effect size estimates for the power simulation, by retaining all significant parameter estimates and setting all insignificant parameters to zero. After simulating from this MANOVA model, we employed an exhaustive AIC and BIC weight based model search, to recover the correct MANOVA model from the data. In addition to finding evidence in favor of the effects that are likely present this approach can also find evidence to indicate which effects are likely to be absent in the data (Wagenmakers & Farrell, 2004). Here, **N** observations per condition were required to recover the correct model with at least **95%** probability.

## Model estimation

After simulating the data sets we fit each of the non-parametric models, specifically a generalized additive model (Wood, 2006), a local polynomial regression (Fan et al., 1997), and a Gaussian process (Rasmussen & Williams, 2006), as well as a parametric state-space

model (Durbin & Koopman, 2012) to each of the data sets. For a detailed description of the model fitting procedure, see **Appendix A**. Importantly, the model fitting procedure for each of the models was validated on a subsample of **5** data sets per condition, in order to ensure good model fit. If a model did not converge for a given data set, the respective data set was removed from the simulation and another data set was simulated instead. Lastly, it is important to note that the functional form of the state-space model in each condition was adjusted to the data generating model, while the same non-parametric models were used across conditions. After each of the models was fit to the data, they were used to obtain point and interval estimates (i.e., 95% confidence or credible intervals) for the latent process value at each time point.

## Outcome measures

To analyze and compare the performance of the different analysis methods, we focused on three different outcome measures. The first two measured the accuracy of each methods in predicting the value of the process at or between the time points at which the observations were taken. These measures of predictive accuracy provide insight into how well the non-linear process is captured by each of the methods. The last outcome measure, concerned the accuracy of the uncertainty estimates provided by each of the methods. Here we investigated whether the confidence intervals provided by each method indeed include the true state value the correct proportion of times.

### *Capturing the non-linear process*

In order to test how well each of the methods captured the non-linear process in each condition, the root mean squared error (RMSE) between the estimated and the generated values was calculated. Further, to estimate how well each of the methods predicts the process between two observed time-points, we calculated the generalized cross-validation (GCV; Golub et al. (1979)) criterion for each method and data set. The GCV is a computationally more efficient and rotation invariant version of the ordinary

leave-one-out cross-validation criterion with the same interpretation. The latter is calculated by removing one data point and refitting the model while keeping certain parameters fixed. Afterwards, the left out observation is predicted and the squared prediction error calculated. epeating this procedure for each data point and averaging the squared errors provides an estimate of how accurately the model predicts unobserved values within the observed range.

Subsequently, a MANOVA was conducted to explain differences in the RMSE and GCV values between the simulation conditions and analysis methods. In order to also find evidence in favor of which factors likely do not affect the RMSE and GCV values, we conducted an exhaustive model search. Here the AIC and BIC based model weights were used as a criterion to identify which model is closest to the true data generating model. The AIC weights can directly be interpreted as the conditional model probabilities (Wagenmakers & Farrell, 2004). Thus, if the true data generating model is among the tested models, the AIC weights indicate the probability that each tested model is the correct model given the data. This makes it possible, to quantify evidence both in favor and against the presence of any effect that could be included in the model. After establishing the most likely MANOVA model, we probed the included effects further by fitting univariate ANOVAs with the same predictors to the RMSE and the GCV values and followed up with Tukey's Honest Significant Difference test for the post-hoc comparisons.

### *Uncertainty quantification*

To investigate the uncertainty estimates that each method provides around the predicted process values, we recorded at each time point whether the true generated process value was within the confidence intervals created by each analysis method. Subsequently, by averaging over all time-points the average confidence interval coverage proportion for each method and simulation condition was obtained. Since, all obtained confidence or credible intervals had a confidence level of 95% the expected coverage proportion should

ideally be close to 95%. Due to the Monte Carlo error of the simulation coverage proportions between 93% and 97% were considered sufficiently close to the expected 95%. Conversely, coverage proportions larger than 97% indicated overestimated standard errors, whereas coverage proportions below 93% were interpreted to indicate either a poor approximation of the underlying process or a underestimation of the standard errors.

## References

Borsboom, D., & Cramer, A. O. (2013). Network Analysis: An Integrative Approach to the Structure of Psychopathology. *Annual Review of Clinical Psychology*, *9*(1), 91–121. https://doi.org/10.1146/annurev-clinpsy-050212-185608

Ceja, L., & Navarro, J. (2009). Dynamics of Flow: A Nonlinear Perspective. *Journal of Happiness Studies*, *10*(6), 665–684. https://doi.org/10.1007/s10902-008-9113-6

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

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

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

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

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

Finnemann, A., Borsboom, D., Epskamp, S., & van der Maas, H. L. J. (2021). The Theoretical and Statistical Ising Model: A Practical Guide in R. *Psych*, *3*(4), 594–618. https://doi.org/10.3390/psych3040039

Golub, G. H., Heath, M., & Wahba, G. (1979). Generalized Cross-Validation as a Method for Choosing a Good Ridge Parameter. *Technometrics*, *21*(2), 215–223. https://doi.org/10.1080/00401706.1979.10489751

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

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

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

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

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

Wagenmakers, E.-J., & Farrell, S. (2004). AIC model selection using Akaike weights. *Psychonomic Bulletin & Review*, *11*(1), 192–196. https://doi.org/10.3758/BF03206482

Waugh, C. E., & Kuppens, P. (Eds.). (2021). *Affect dynamics*. Springer. https://doi.org/10.1007/978-3-030-82965-0

Literaturangaben

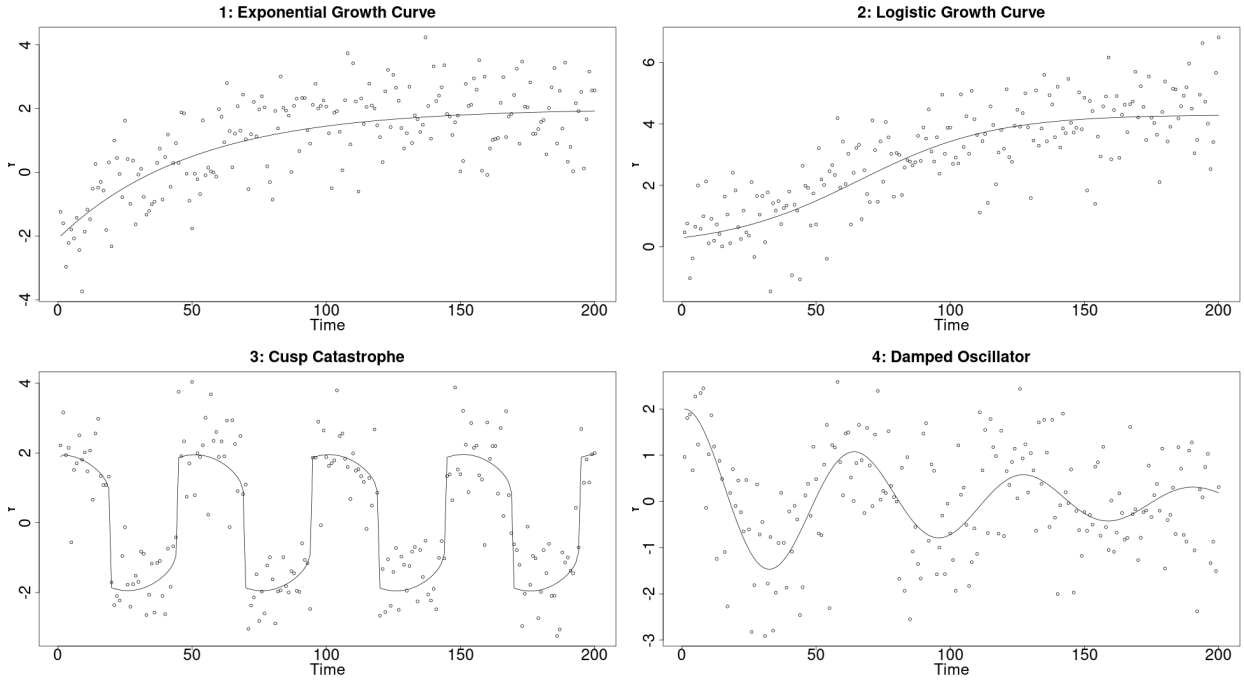
Witkiewitz, K., & Marlatt, G. A. (2007). Modeling the complexity of post-treatment drinking: It's a rocky road to relapse. *Clinical Psychology Review*, *27*(6), 724–738. https://doi.org/10.1016/j.cpr.2007.01.002

Wood, S. N. (2006). *Generalized additive models: An introduction with R* [OCLC: ocm64084887]. Chapman & Hall/CRC.

Wrzus, C., & Neubauer, A. B. (2023). Ecological Momentary Assessment: A Meta-Analysis on Designs, Samples, and Compliance Across Research Fields. *Assessment*, *30*(3), 825–846. https://doi.org/10.1177/10731911211067538

## Figure 1

*Non-linear exemplar processes*



## Figure 2

*Non-linear exemplar processes*