A hierarchical Ornstein-Uhlenbeck model for stochastic time series analysis

Ville Laitinen[1 [0000−0002−2653−1673]] and Leo Lahti[1 [0000−0001−5537−637X]]

Department of Mathematics and Statistics, University of Turku, Finland
velait@utu.fi

Abstract. Longitudinal data is ubiquitous in research, and often complemented by broad collections of static background information. There is, however, a shortage of general-purpose statistical tools for studying the temporal dynamics of complex and stochastic dynamical systems especially when data is scarce, and the underlying mechanisms that generate the observation are poorly understood. Contemporary microbiome research provides a topical example, where vast cross-sectional and longitudinal collections of taxonomic profiling data from the human body and other environments are now being collected in various research laboratories world-wide. Many classical algorithms rely on long and densely sampled time series, whereas human microbiome studies typically have more limited sample sizes, short time spans, sparse sampling intervals, lack of replicates and high levels of unaccounted technical and biological variation. We demonstrate how non-parametric models can help to quantify key properties of a dynamical system when the actual data-generating mechanisms are largely unknown. Such properties include the locations of stable states, resilience of the system, and the levels of stochastic fluctuations. Moreover, we show how limited data availability can be compensated by pooling statistical evidence across multiple individuals or studies, and by incorporating prior information in the models. In particular, we derive and implement a hierarchical Bayesian variant of Ornstein-Uhlenbeck driven t-processes. This can be used to characterize universal dynamics in univariate, unimodal, and mean reversible systems based on multiple short time series. We validate the model with simulated data and investigate its applicability in characterizing temporal dynamics of human gut microbiome.

Keywords: Longitudinal analysis · Hierarchical models · Ornstein-Uhlenbeck process · Resilience · Stochastic processes

1 Introduction

Many natural and social systems are complex and cannot be studied in isolation. The underlying data-generating mechanisms are often largely unknown in such cases, and the observed dynamics can be characterized only indirectly [8]. Non-parameteric models that focus on characterizing observed data properties, rather than modeling the underlying mechanisms, can provide valuable information on

the system behavior. In the context of human microbiome dynamics, for instance, such non-parametric models have been used to describe and infer the presence of alternative ecosystem states [13], periodicity, stochasticity, and chaos [6, 5]. In many real applications, the data is scarce, and new methods are needed in order to derive maximal information from limited observations.

Our study is motivated by the analysis of temporal dynamics of human gut microbiome. This refers to the totality of microbial communities living on skin, gastrointestinal tract and other body sites. Contemporary human microbiome research has largely focused on cross-sectional cohorts with limited follow-ups, providing information on the composition and inter-individual variation of the microbiome. The dynamics of these systems are yet, however, not well understood despite their clinical importance [1, 10]. As understanding of these systems is accumulating, the research focus is beginning to shift from general descriptions towards actionable clinical applications and manipulation.

In this work, we show how key dynamical properties of poorly understood dynamical systems can be inferred from limited time series by pooling information can across multiple individuals. In the present work, we focus specifically on mean-reversible stochastic processes. Such dynamic behavior is frequently observed in the human gut microbiome. Many bacterial species in the human gut ecosystem have been reported to exhibit characteristic abundance levels around which they tend to fluctuate over time (see e.g. [13]). It has been reported that the average abundance levels of many gut bacteria remain relatively stable over long time periods but on a shorter (daily) time scale the abundances can exhibit considerable fluctuations [3]. Mean-reverting stochastic processes, in particular the Ornstein-Uhlenbeck (OU) process , provide well-established means to characterize key properties of such systems, including the location and resilience of the potential wells, speed of mean reversion, and volatility of abundance levels, even when the underlying mechanisms regulating those dynamics are unknown. Stochastic processes and generative probabilistic models provide a rigorous framework for the characterization of the observed dynamics in such cases, with wide applicability across different application domains [9, 11, 16, 18].

We adapt and apply these techniques to model human gut microbiome dynamics. A key practical limitation of the existing methods in our application is that the available implementations of the OU process depend on the availability of long time series with dozens of time points or more. The currently available longitudinal data sets in typical human microbiome studies have more limited sample sizes and time series lengths, or sparse sampling intervals. Combined with high levels of variation and limited knowledge of the data-generating processes, these limitations form considerable challenges for the application of previously established stochastic models, such as the the OU process, in contemporary human microbiome research. In order to address these limitations, we derive, implement, and validate a hierarchical extension to the OU process. This can be used to recover key information of the system dynamics from limited data by aggregating information across short time series from multiple individuals. Further potential advantages of the probabilistic formulation include the

opportunities to model individual variation, and to incorporate of prior information from the cross-sectional background collections in the model. We validate the implementation with simulated data, investigate its robustness to varying modeling assumptions including the numbers, lengths and densities of the time series, and ranges of parameter values, and finally explore the applicability of this model on topical human gut microbiome data sets.

In order to maximize the flexibility we have constructed the implementation so that the number of observation per time series and the observation times do not have to be identical. Thus, our implementation of the OU process provides a rigorous and justified method for modeling dynamics of single potential wells.

2 Preliminaries

This section outlines the statistical model and the relevant technical derivations.

2.1 The Ornstein-Uhlenbeck Process

Many natural processes can be modeled by a combination of deterministic drift and stochastic fluctuations. These assumptions naturally lead to stochastic differential equations, which are commonly encountered in literature in the form:

$$dX_t = f(X,t)dt + L(X,t)dZ_t.$$

Here, X_t is the system state at time t, Z is a stochastic process and f, and L are called the *drift* and *dispersion* terms, respectively. The drift defines the deterministic behavior, whereas dispersion characterizes the stochastic component of the system. Unlike the solutions of ordinary differential equations, the solutions of the stochastic counterparts are non-unique and nowhere differentiable as they are different for different realizations of the noise term. The deterministic solution can be recovered by averaging over these solutions.

The Ornstein-Uhlenbeck (OU) process, also known as the Langevin equation in physics and Vasicek model in finance, is a stochastic process with a wide range of applications [12]. It is frequently used to model systems that have a steady state, and a tendency to recover from perturbations by gradually returning, or drifting, towards the long-term mean value. The OU process is the continuous-time extension of autoregressive AR(1) model and is defined as the solution to the stochastic differential equation with drift function $f(X,t) = \lambda(\mu - X)$ and constant dispersion $L(X,t) = \sigma$. The parameters $\lambda \in [0,1]$, $\mu \in \mathbb{R}$ and $\sigma \geq 0$ have natural interpretations as mean-reversion rate, long-term mean and size of stochastic fluctuations, respectively. The OU process can be formulated as a Gaussian process on the real line $\mathrm{GP}(\mu,\ K)$ with a covariance function $K = \mathrm{Cov}(X_{t_1}, X_{t_2}) = \frac{\sigma^2}{2\lambda} e^{-\lambda \Delta t}$, and as all diffusion processes, is also a Markov process [12].

2.2 The Ornstein-Uhlenbeck Driven t-process

We adopt the Student's t-process, instead of the traditionally used Wiener process as the driving process of the OU process. This choice is more robust to outliers and short term volatility, with little if any additional computational cost as the critical analytical equations are available in both cases.

Although the stochastic process in OU process is often modeled as white noise, requiring Z_t to have Gaussian transition density is often a too limiting assumption for practical purposes as it does not allow large enough fluctuations. Thus, robustness against outliers is compromised and a more general process would be preferred [15]. The Student's t-process is a non-Gaussian alternative to a prior over functions that allows more flexibility and room for outliers. Using t-processes is a convenient choice also in the sense that the Gaussian process can be obtained as a special case by taking the limit $\nu \to \infty$ [15]. Thus we will adopt t-processes as the driver of dispersion of the OU process. See Figure 1 for an example of simulated OU process time series and corresponding parameter estimates. The t-process has recently been studied in e.g. [14, 15] and the following definition can be found in these references.

Definition 1. *A vector $\bar{y} \in \mathbb{R}^n$ is multivariate Student-t distributed with ν degrees of freedom, mean parameter μ and shape matrix Σ, $\bar{y} \sim \mathcal{ST}_n(\nu, \mu, \Sigma)$, if it has density*

$$p(\bar{y}) = \frac{\Gamma(\frac{\nu+n}{2})}{((\nu-2)\pi)^{\frac{n}{2}}\Gamma(\frac{\nu}{2})}|\Sigma|^{-\frac{1}{2}} \times \Big(1 + \frac{(\bar{y}-\bar{\mu})^T\Sigma^{-1}(\bar{y}-\bar{\mu})}{\nu-2}\Big)^{-\frac{\nu+n}{2}} \quad (1)$$

Definition 2. *The process f is a Student-t process, $f \sim \mathcal{ST}(\nu, \mu, \Sigma)$, if any finite set of values is multivariate t-distributed.*

The covariance matrix K is related to the shape matrix via $\Sigma = \frac{\nu-2}{\nu}K$.

2.3 Hierarchical extension

The model outlined above describes the Ornstein-Uhlenbeck driven t-process as implemented in [9]. Our novel contribution that we present now is to equip the model with hierarchical structure and testing the robustness of the extended implementation. Let $\mathcal{X} = \{\bar{X}_i, i \in \{1, \ldots, N\}\}$ be a set of OU process values, with n_i observations in each, each i representing e.g. a different measurement site. We assume a hierarchical structure for the parameters λ, μ and σ,

$$dX_{j,t} = \lambda_j(\mu_j - X_{j,t})dt + \sigma_j dZ_t,$$

for all $j \in \{1, \ldots, n_i\}$. As the OU process is a Markov process the generative model for the data can be described as in 2. We have implemented the model using the multivariate t-distribution formulation but it is possible to implement the model using transition densities between consecutive observations.

Adding a level of hierarchy to the implementation for a single series can be obtained by modifying the model likelihood in the extended version so that it

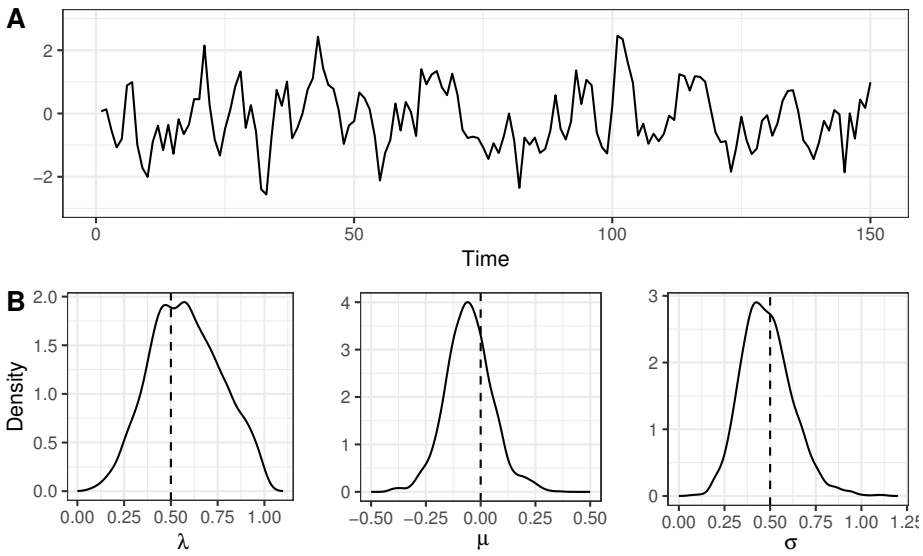

Fig. 1: **A** A simulated OU driven t-process time series with $\nu = 7$ and parameter values $\lambda, \sigma = 0.5$, $\mu = 0$. **B** Posterior estimates of the model parameters. Dashed lines mark the simulation values used to generate the data.

equals the product of likelihoods of individual series. In addition priors have to be assumed to follow some distribution. We have used normal distributions for μ and σ and inverse gamma distribution for λ. Hyperpriors for the hyperparameters ϕ were chosen to be uninformative but still strong enough to guide the parameter estimates to practically reasonable ranges. We can now write the generative model for the hierarchical OU process with partially pooled estimates

$$\begin{aligned} X_i &\sim \mathrm{MVT}_n(\nu, \mu_i, \Sigma_i) \\ \mu_i &\sim \mathcal{N}(\mu_{\mu,i}, \sigma_{\mu,i}) \\ \sigma_i &\sim \mathcal{N}(\mu_{\sigma,i}, \sigma_{\sigma,i}) \\ \lambda &\sim \Gamma^{-1}(\alpha_i, \beta_i) \\ \phi &\sim \mathcal{N}(\phi_\mu, \phi_\sigma), \end{aligned} \tag{2}$$

where $i = \{1, \ldots, N\}$ and ϕ represents all hyperpriors.

The model can also be specified so that both the prior shape and hyperparameters are fixed. This version corresponds to no pooling between distinct observations. It does not share information as it assumes that the differences between series are too large to be modeled together. The other extreme is complete pooling in which all data are assumed to be generated by identical parameters. Partial pooling assumes some, but not full, similarity between time series and

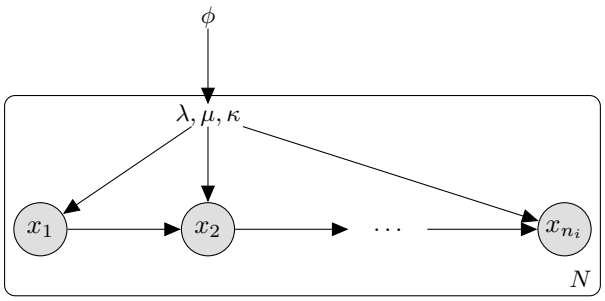

Fig. 2: Bayesian graph representation of the hierarchical OU process. Hyperparameters are denoted with ϕ.

thus represents a compromise between the other two alternatives. These models are compared in subsection 3.1 below.

For a general and simplified treatment of the OU process we assumed that our observations are directly generated from the OU process and use uniform time intervals in the following simulations. The model can, however, incorporate unequal time intervals and varying numbers of observations per time series. Alternative models of observation noise represent opportunities for further extension. In ecological studies that motivate the present work, the observation noise is often modeled with a Gaussian or Poisson distribution, where the rate parameter is obtained from the OU process by exponentiation. This so called stochastic Gompertz model is frequently used in ecological time-series analysis [4]. For OU process implementation of the Gompertz model in the context of a single time series, see [9].

3 Model validation

Next, we tested the implementation with simulation experiments. The simulations were motivated by recent human microbiome studies that are introduced in more detail in section 4. The data sets in these studies have considerable differences in sample sizes and in this respect represent the scope of the currently available human microbiome data.

In the simulations the values for λ, σ and μ were sampled separately for each series from priors $\Gamma^{-1}(6,4)$, $\mathcal{N}(0,1)$ and $\mathcal{N}(3,1)$ respectively. The degrees of freedom in the multivariate Student's t-distribution was set to 7. These distributions and parameter values were chosen as they generate values and variation resembling those encountered in (log-transformed and centered) human gut microbiome time series. Hyperprior distributions for the model parameters were chosen to be vague as no prior understanding of these parameters exists in this context. Normal distributions with relatively large variance were used.

Parameter estimates are obtained by coding the model in rstan [2]. Stan requires the user to specify data, parameters and model in the corresponding

code blocks and uses Hamiltonian Monte Carlo and No-U-Turn Sampler techniques to sample from the posterior distribution. To minimize the amount of divergent transitions in HMC sampling we have used a non-centered parameterization for μ and σ. This is in agreement with [17] where it is mentioned that hierarchical models often perform better with non-centered parameterizations, especially when the sample size is limited. Non-centering λ led to additional divergences so its parameterization was kept centered. We encountered no divergences of other pathologies in the MCMC diagnostics, which yields additional confirmation for the validity of our implementation. In principle the degrees of freedom of the multivariate t-distribution could be estimated in addition to the other model parameters. In our experiments we were, however, unable to reliably recover this parameter so the implementation assumes it to be fixed and input to the model. The source code for the Stan model is available at https://github.com/velait/OU_IDA.

3.1 Model comparison

To demonstrate the differences between the three basic model variants available for multiple observation units (complete, partial and no pooling) we now compare the estimates they provide. We use a single simulated test set with sample size similar to [7]: 30 time series, 30 samples each with 3 time units between observations. The parameter values were sampled from prior distributions individually for each series and parameter.

Maximum a posteriori (MAP) estimates for the parameter λ from each model as well as their distance to simulation values and widths of the 50% interquartile ranges are displayed in 3. The MAP estimates of the partially pooled model are on average closer to the simulation values, although some individual estimates are farther as they get shrunk towards the estimate from completely pooled model (dashed line). The IQRs are shorter compared to the model with no pooling, which yields additional confirmation for the models improved accuracy. Similar results were obtained for the other parameters.

One of the advantages of a hierarchical model with partially pooled parameters is that the prior distributions can be estimated as well. This provides information on the parameters' population level variation. In Figure 4 the simulation and estimated priors are compared. Prior of μ is recovered best and lambda on a relatively satisfactory level as well. The prior of the variation parameter σ, however, is not very well estimated as the mode and variance are clearly different from the target. The reason for less than ideal estimates may lie in the low number of values simulated in the first place, as only 30 values are drawn from each distribution. Thus there is plenty of room for stochastic variation. Additional uncertainties may arise due to possibly challenging regions in the (λ, σ) space. As these parameters are intertwined it is possible that certain combinations (e.g. small λ, large σ) pose challenges beyond the capabilities of our implementation.

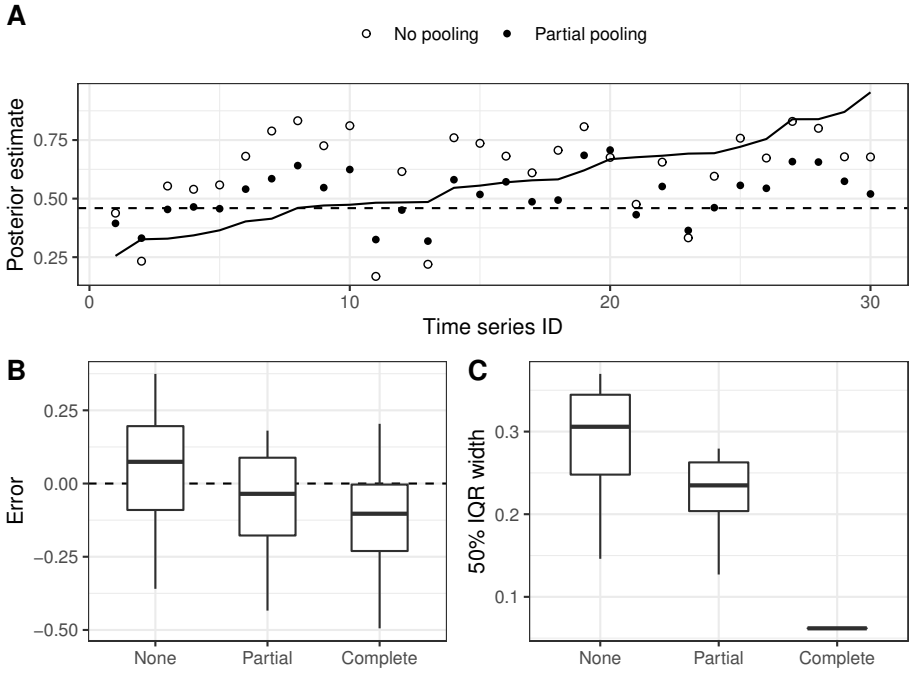

Fig. 3: **A** MAP estimates from different model variants. Dashed line marks the completely pooled estimate and solid line the simulation values, sorted in increasing order. **B** Distribution of estimates error, defined as difference between MAP and simulation value. **C** Distribution of 50% IQR widths.

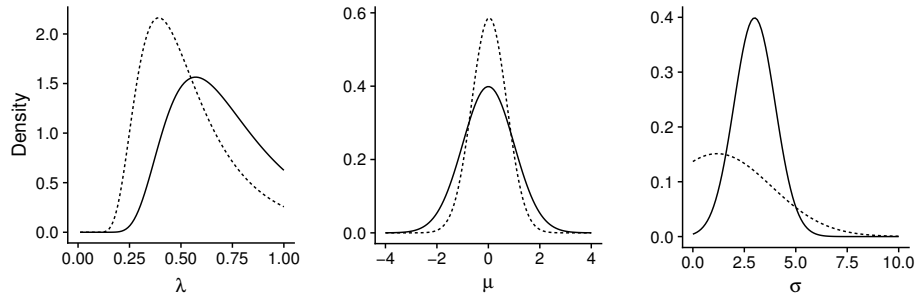

Fig. 4: Prior distributions used for data simulations (solid line) and posterior estimates of these distributions (dashed line) based on MAP estimates of the hyperparameters.

4 Application to human microbiome time series

Next, we demonstrate the use of these models in analyzing the dynamics of microbial ecosystems in the human body.

In the first case study, two healthy volunteers were followed over a year and provided hundreds of stool samples [3]. During the study the gut ecosystem of one of the individuals experienced a dramatic change in composition due to a *Salmonella* infection. This perturbation is beyond the capacities of the OU process model and for this reason we have limited our analysis to the samples prior to the infection, leaving 125 samples covering 4.5 moths for a closer analysis. The sample size in this study is large in the human gut microbiome context, consisting of nearly daily samples from two individuals over several months. In total 387 different genus level taxonomic units were observed out of which we chose to focus on the symmetric and unimodal *abundance types* as their observed dynamics roughly corresponds to the model assumptions. For demonstration purposes, we limit the analysis to a single genus-level taxonomic unit, *Bacteroides*, which is highly abundant and prevalent in human gut at least in the Western populations. We explored the estimates given by our implementation with the first 120 samples and subsets of these to assess how many samples are required for estimates close to the full sample size. Figure 5A displays the MAP estimates for various samples sizes, where values on the x-axis correspond the first n samples of the full 120 time points. The estimates level after sufficient amounts of time points suggesting that a there is little increase in accuracy after enough samples. We also experimented with randomly removing observations and discovered that roughly half of the samples can be removed without significant loss of accuracy compared to the full sample size, see B.

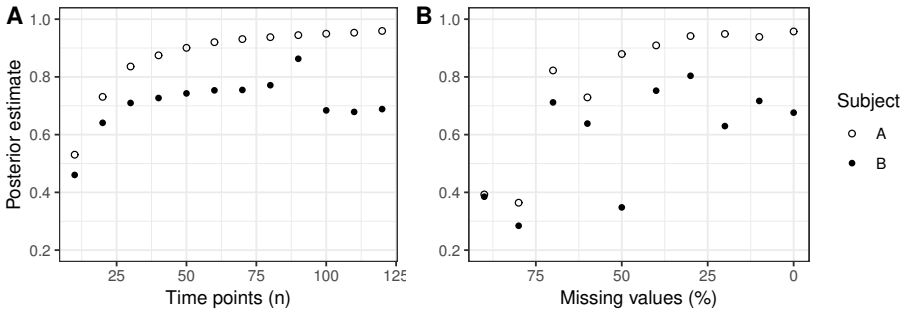

Fig. 5: MAP estimates for parameter λ against length of time series **A** and proportion of randomly removed samples **B**.

We also carried out preliminary analyses on the HITChip Atlas data set [13], which has considerably shorter time series but from a larger number of individuals. The HITChip Atlas data set consists of stool samples from 1006 healthy western adults. Multiple (2-5) follow-up samples were available from 78 subjects with weeks or months between samples. We performed preliminary experiments on this data and data simulated with similar abundance, variation and sparsity profiles but recovered only unreliable and inaccurate results. The posterior esti-

mates had high levels of uncertainty, the model inference had convergence issues, and the results were sensitive to initialization and changes in the data, indicating that the 2-5 sparse time points will not suffice to distinguish the effects of the mean reversion and stochastic drift parameters in the hierarchical OU model.

5 Discussion

The main objective of this work has been to propose new general-purpose methods to characterize key properties of poorly understood dynamical systems based on scarce longitudinal data, and demonstrate their applicability in the topical research area of human microbiome studies.

We have extended the previously proposed Ornstein-Uhlenbeck (OU) driven t-process by deriving hierarchical version, which allows the pooling of information across multiple time series and parameter inference of the shared stochastic process. This is specifically motivated by topical problems in human microbiome research, where time series are often as short as 2-3 time points per individual, but available for a potentially large number of individuals. In such case, the traditional variants of the OU process are not applicable, and a hierarchical extension can potentially help to aggregate information across multiple experiments. We have implemented this model by adding a new level of hierarchy to the standard OU driven t-process [9]. Importantly, we designed the model so that the number of samples and observation times in each time series is flexible, allowing efficient utilization of real time series where the number and timings of the observations may differ across the available time series. This removes the need to impute missing values, or force synchronized observation times, thus facilitating application in many real-life scenarios. Following the work by [9], our model takes advantage of the Student's t-process based version of the OU process, rather than the Wiener process which is more common in the OU process literature, in order to increase robustness for outliers. This comes with little additional computational cost.

In simulation experiments we have demonstrated the advantage of partially pooling the parameters over the variants with complete and no pooling. In addition to increased accuracy the hierarchical OU process model offers information on the population level variation of the model parameters as it learns the prior distribution by estimating the hyperpriors. The model performance was satisfactory when tested on a simulated data set with moderate amounts of samples and series but failed to produce reliable estimates for very sparse and short time series. We anticipate that this failure could be explained by the narrow width of observation intervals compared to the simulated dynamics. Naturally observations need to be sufficiently dense and cover a large enough interval to be able to capture dynamics at of a particular scale. We also demonstrated the use of this model on longitudinal time series from human gut microbiome [3]. These experiments clearly demonstrate how the model parameters converge towards a saturation point with increasing time series lengths and densities. Regarding technicalities of he model fitting, the MCMC sampling converged well, also sup-

porting the validity of the implementation. For a more complete view on the robustness of parameter inference, a more extensive probing of the parameter ranges, alternative priors, observation noise and data with uneven and sparse sampling intervals should be undertaken. Alternative parameterizations should be tested to see if some perform better with higher sample sizes.

The hierarchical OU process provides several promising opportunities for future extensions that are directly applicable to microbial ecological time series. In particular, the standard OU process, which assumes unimodal and symmetrically distributed data, could be generalized to model other abundance types [13] of the human-associated microbial taxa abundance distributions. In particular, the analysis of alternative community states of dynamical systems, frequently observed the human vaginal microbiome[7], for instance, provides interesting challenges for further research and model extensions. Our current implementation of the hierarchical OU process currently only handles time series with unimodal density profiles. Moreover, generalizations of the hierarchical model to the multivariate setting would be valuable. These depend on the development of computationally more efficient implementations, for instance based on variational learning of simulation-based methods. Apart from [9] we are not aware of applications of these models, in particular its hierarchical extension that we develop here, in the context of human microbiome studies.

Whereas the focus in our current analysis is limited to investigating the applicability of the model to readily available real observations from a single taxonomic group, further studies could provide a systematic comparison of the stochastic, mean and drift parameters across different taxonomic units in order to characterize differences in the dynamical variation in the abundance level of various gut bacteria. By classifying the individuals to larger groups based on health status, life style factors, age or other meta data, clinically and biologically interesting connection could be learned. The methodology and the challenges of overcoming the limitations of scarce, noisy, and poorly understood observations that these models help to solve are very generic, and the potential applications naturally reach beyond population dynamics.

Acknowledgements

The work has been partially funded by Academy of Finland (grants 295741, 307127).

References

1. Bashan, A., Gibson, T.E., Friedman, J., et al.: Universality of human microbial dynamics. Nature **534**(7606), 259–262 (Jun 2016), doi: 10.1038/nature18301
2. Carpenter, B., Gelman, A., Hoffman, M.D., et al.: Stan: A probabilistic programming language. Journal of Statistical Software **76** (2017), doi: 10.18637/jss.v076.i01
3. David, L.A., Materna, A.C., Friedman, J., et al.: Host lifestyle affects human microbiota on daily timescales. Genome Biology **15**, R89 (Jul 2014), doi: 10.1186/gb-2014-15-7-r89

4. Dennis, B., Ponciano, J.M.: Density dependent state space model for population abundance data with unequal time intervals. Ecology **95**(8) (Aug 2014). https://doi.org/10.1890/13-1486.1
5. Faust, K., Bauchinger, F., Laroche, B., de Buyl, S., Lahti, L.o.: Signatures of ecological processes in microbial community time series. Microbiome **6**, doi: 10.1186/s40168-018-0496-2
6. Faust, K., Lahti, L., Gonze, D., de Vos, W.M., Raes, J.: Metagenomics meets time series analysis: unraveling microbial community dynamics. Current Opinion in Microbiology **25**(Supplement C), 56–66 (Jun 2015), doi: 10.1016/j.mib.2015.04.004
7. Gajer, P., Brotman, R.M., Bai, G., et al.: Temporal dynamics of the human vaginal microbiota. Science Translational Medicine **4**(132) (2012). https://doi.org/10.1126/scitranslmed.3003605
8. Gonze, D., Coyte, K.Z., Lahti, L., Faust, K.: Microbial communities as dynamical systems. Current Opinion in Microbiology **44**, 41–49 (Aug 2018), doi: 10.1016/j.mib.2018.07.004
9. Goodman, A.: Fitting ornstein-uhlenbeck-type student's t-processes in stan with applications for population dynamics data (2018), doi: 10.5281/zenodo.1284346
10. Halfvarson, J., Brislawn, C.J., Lamendella, R., et al.: Dynamics of the human gut microbiome in inflammatory bowel disease. Nature Microbiology **2**(5), 17004 (May 2017), doi: 10.1038/nmicrobiol.2017.4
11. Heinonen, M., Yildiz, C., Mannerström, H., et al.: Learning unknown ODE models with Gaussian processes (Mar 2018), http://arxiv.org/abs/1803.04303
12. Iacus, S.M.: Simulation and Inference for Stochastic Differential Equations: With R Examples (Springer Series in Statistics). Springer Publishing Company, Incorporated, 1 edn. (2008)
13. Lahti, L., Salojärvi, J., Salonen, A., Scheffer, M., de Vos, W.M.: Tipping elements in the human intestinal ecosystem. Nature Communications **5** (Jul 2014), doi: 10.1038/ncomms5344
14. Shah, A., Wilson, A.G., Ghahramani, Z.: Student-t processes as alternatives to gaussian processes. The Seventeenth International Conference on Artificial Intelligence and Statistics (AISTATS) (2014)
15. Solin, A., Särkka, S.: State space methods for efficient inference in student-t process regression. Proceedings of the 18th International Conference on Artificial Intelligence and Statistics (AISTATS) **38** (2015)
16. Srokowski, T.: Multiplicative levy noise in bistable systems. The European Physical Journal B **85**(2) (Feb 2012), doi: 10.1140/epjb/e2012-30003-9
17. Stan Development Team: Modeling language user's guide and reference manual, version 2.17.0 (2017), http://mc-stan.org
18. Yildiz, C., Heinonen, M., Intosalmi, J., et al.: Learning stochastic differential equations with gaussian processes without gradient matching (Jul 2018), http://arxiv.org/abs/1807.05748