

Non-parametric modeling of microbiome dynamics

with continuous mean-reversible stochastic processes

Ville Laitinen and Leo Lahti

Department of Mathematics and Statistics, University of Turku, Finland

velait@utu.fi

Background & Motivation

- Cross-sectional cohorts dominate population-level studies of the human microbiome, but universal patterns of microbial community dynamics are less well understood.
- Longitudinal data sets are now rapidly accumulating and there is demand for robust statistical tools that can take into account the low sample sizes, sparse irregular observations, and high levels of unaccounted variation that pose challenges for modeling microbiome dynamics.
- Due to high complexity data generating mechanism are typically not known and inference with mechanistic models such as the generalized Lotka-Volterra equations is unreliable.

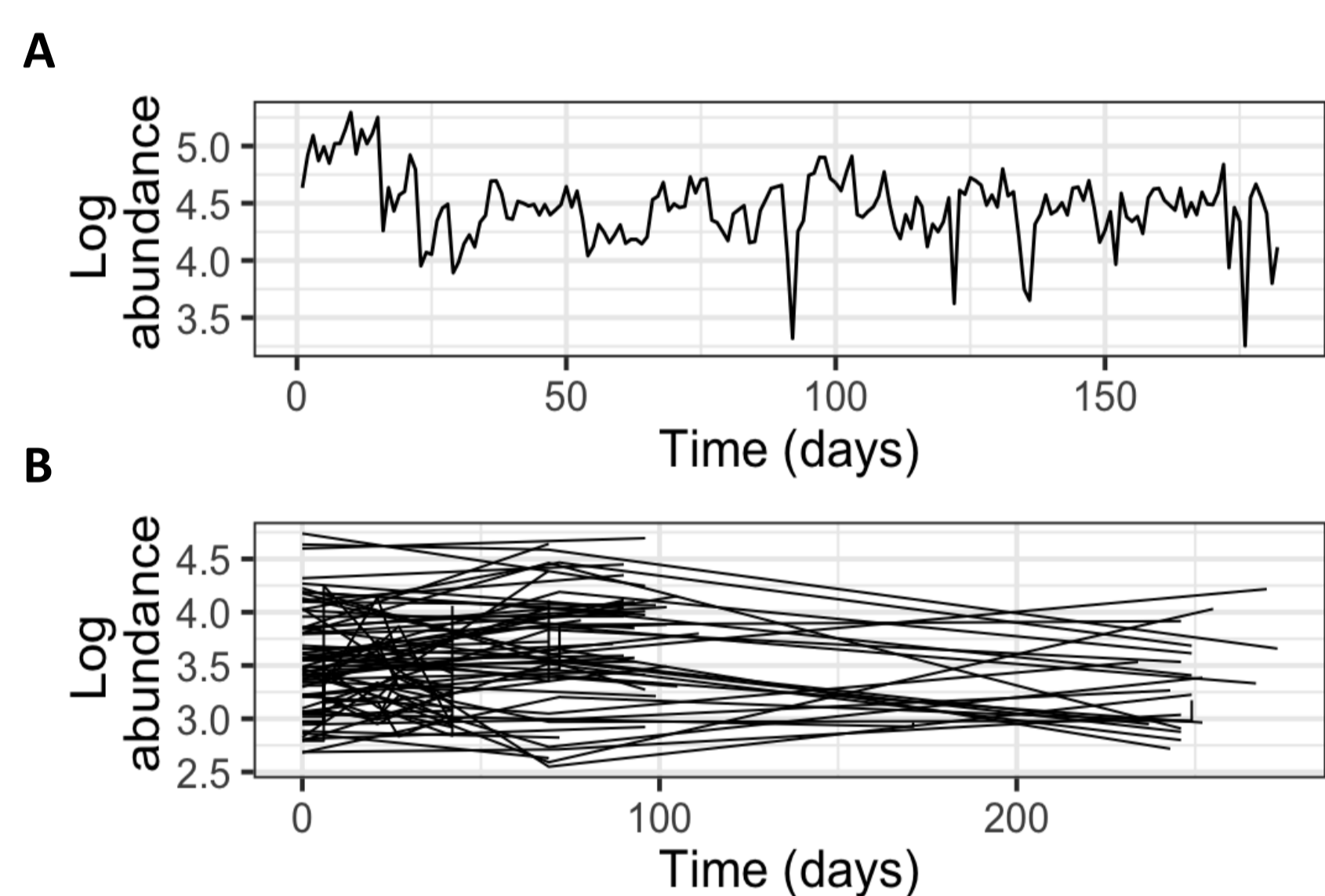


Figure 1. Log abundance levels of *Bacteroides* in **A** 188 samples from a single subject [1] and **B** 2-5 samples from 78 subjects [2]. Long time series are typically available only for a few subjects. Short or sparse series can be analyzed with multi-level models, if there is a large collection of them.

- Many gut microbes have been reported to have persistent long-term abundance levels [1] (Fig. 1A). The Ornstein-Uhlenbeck process (OUP) can capture key characteristic of such dynamics: location of the stable abundance level, amplitude of stochastic variations and system resilience.

Modeling the stochastic dynamics

- The Ornstein-Uhlenbeck process is the continuous-time extension of the AR(1) model and is defined with the stochastic differential equation

$$dX = \lambda(\mu - X)dt + \sigma dS$$

Annotations: λ (mean-reversion rate), μ (long-term mean), X (abundance), σ (variation), dS (stochastic process)

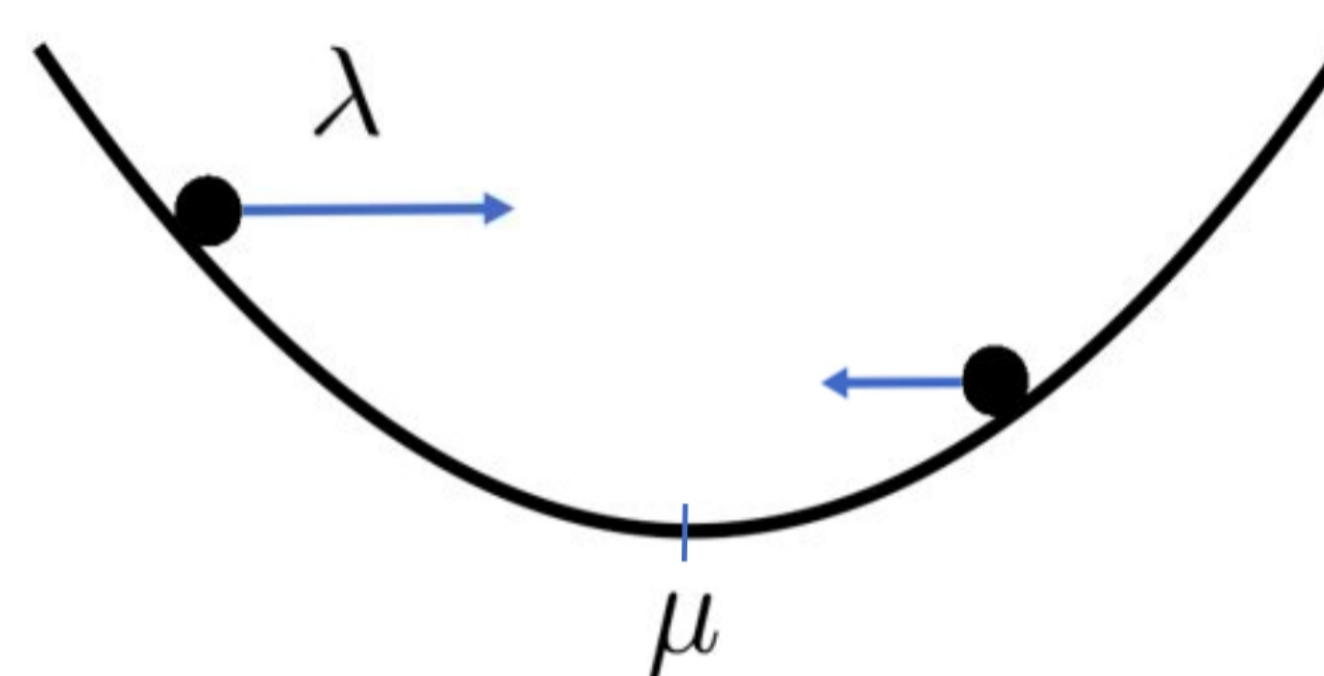


Figure 2. The OUP describes dynamics of a single potential well centered at μ . The mean-reversion rate is determined by λ .

- The OUP generates data resembling real gut microbiome data (Fig. 3A) and parameter inference with Stan [3] is reliable, as long as there are sufficiently many time points (Fig. 3B) [4].

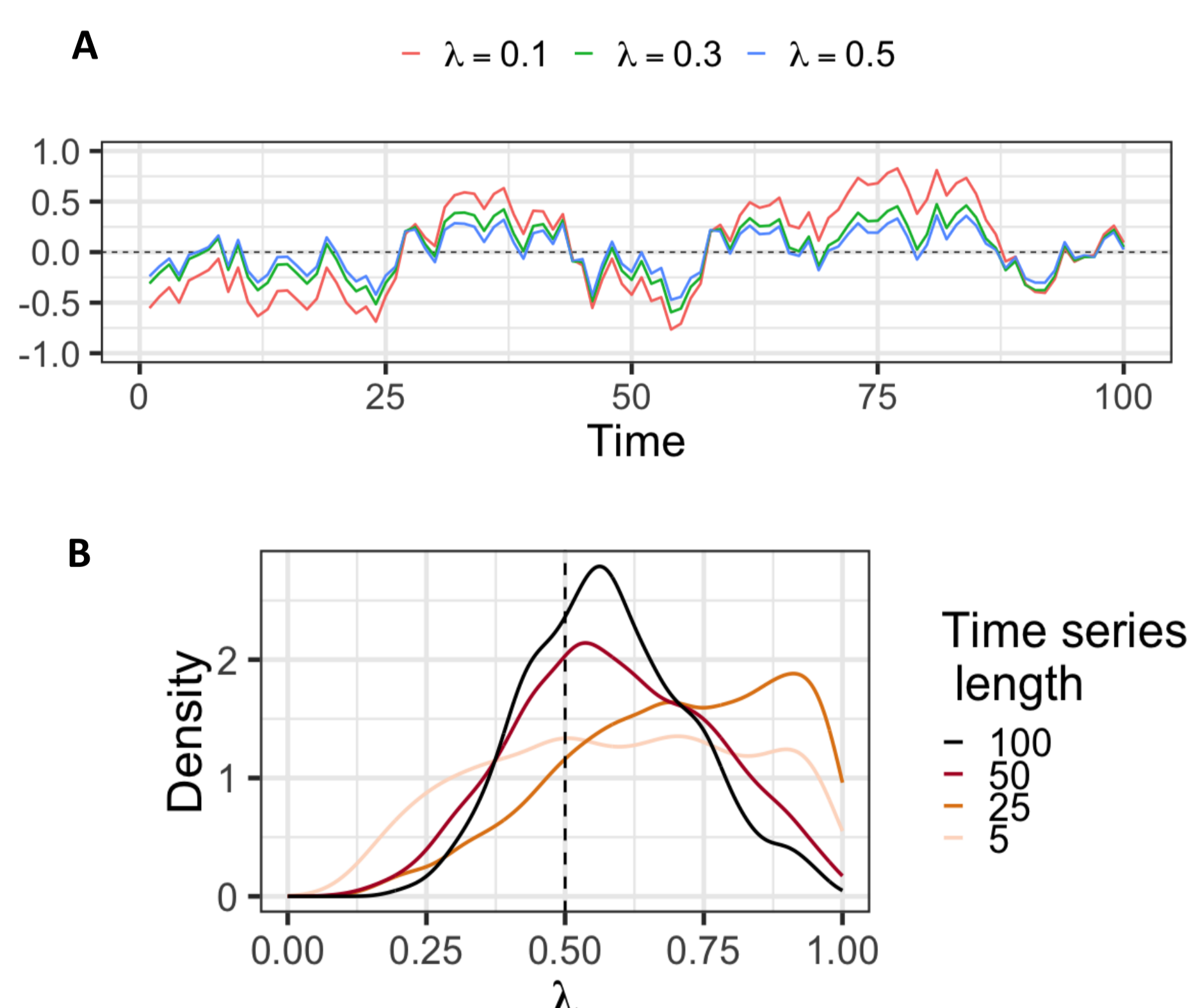


Figure 3. **A** Simulated OUP time series with $\mu = 0$, $\sigma = 0.5$ and various λ values. **B** Posterior estimates for λ are accurate with sufficient amounts of time points. Similar results are obtained for μ and σ .

- The OUP can equivalently be formulated as a stochastic process with covariance

$$K = \frac{\sigma^2}{\lambda} e^{-\lambda \Delta t}$$

- With multiple time series we can use a hierarchically extended model with individual parameters $\theta = \lambda_i, \mu_i, \sigma_i$ for each series and hyperparameters ϕ and -priors [5]. The posterior density is

$$p(\theta, \phi | X) \propto \prod_{i=1}^N [\text{MVT}_n(X_i | \nu, \mu_i, K_i)] p(\theta | \phi) p(\phi)$$

Results & Conclusions

- Simulations show that our implementation of the hierarchical model with partially pooled parameters is able to recover more accurate estimates for all parameters compared to the non-pooled model (Fig. 4).

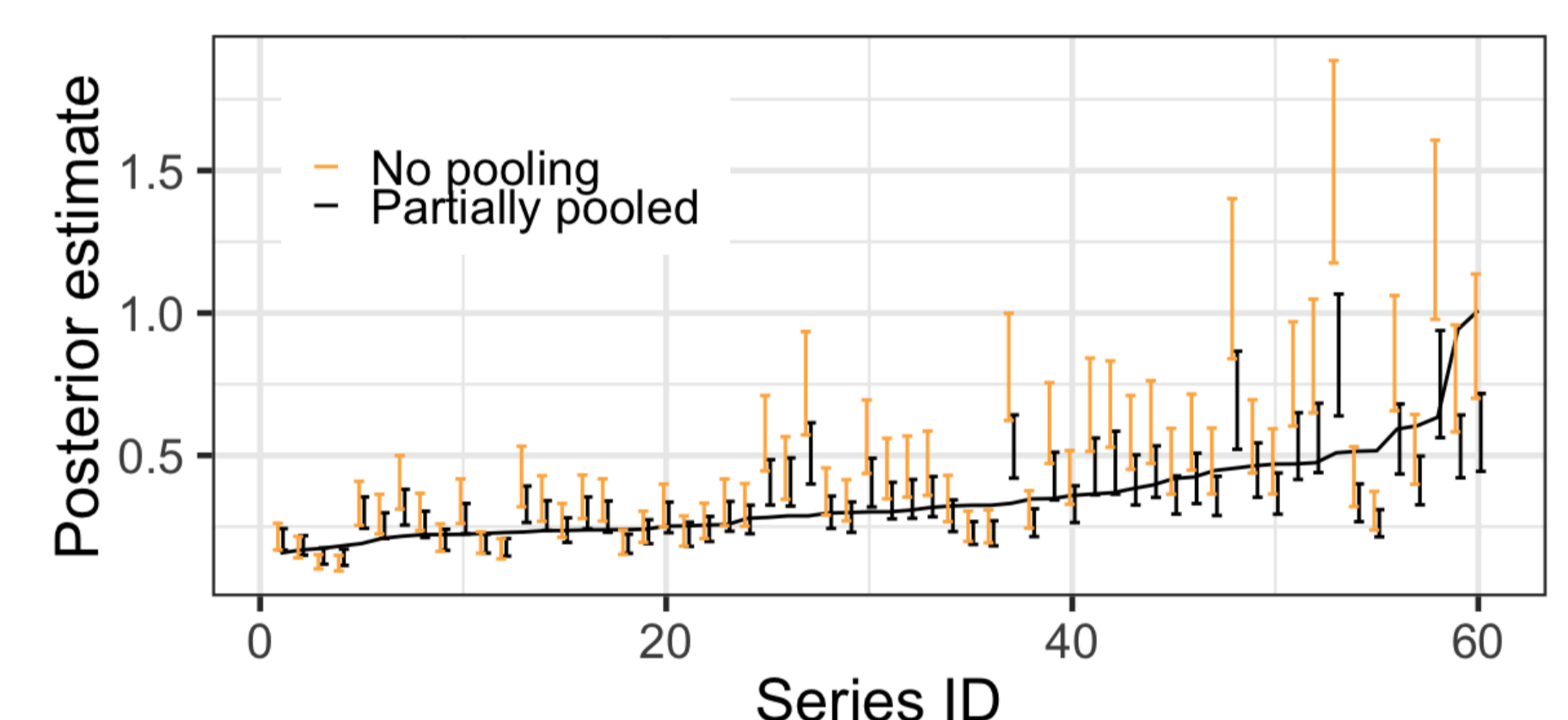


Figure 4. Simulation values of σ (solid line) and posterior 50% interquartile ranges from partially pooled (black bars) and unpooled (orange bars) models.

Ongoing research

- The standard OUP-type processes can be extended to multivariate models of microbiome composition, potentially including multiple stable states, to ultimately characterize the entire dynamic landscape of the gut ecosystem.

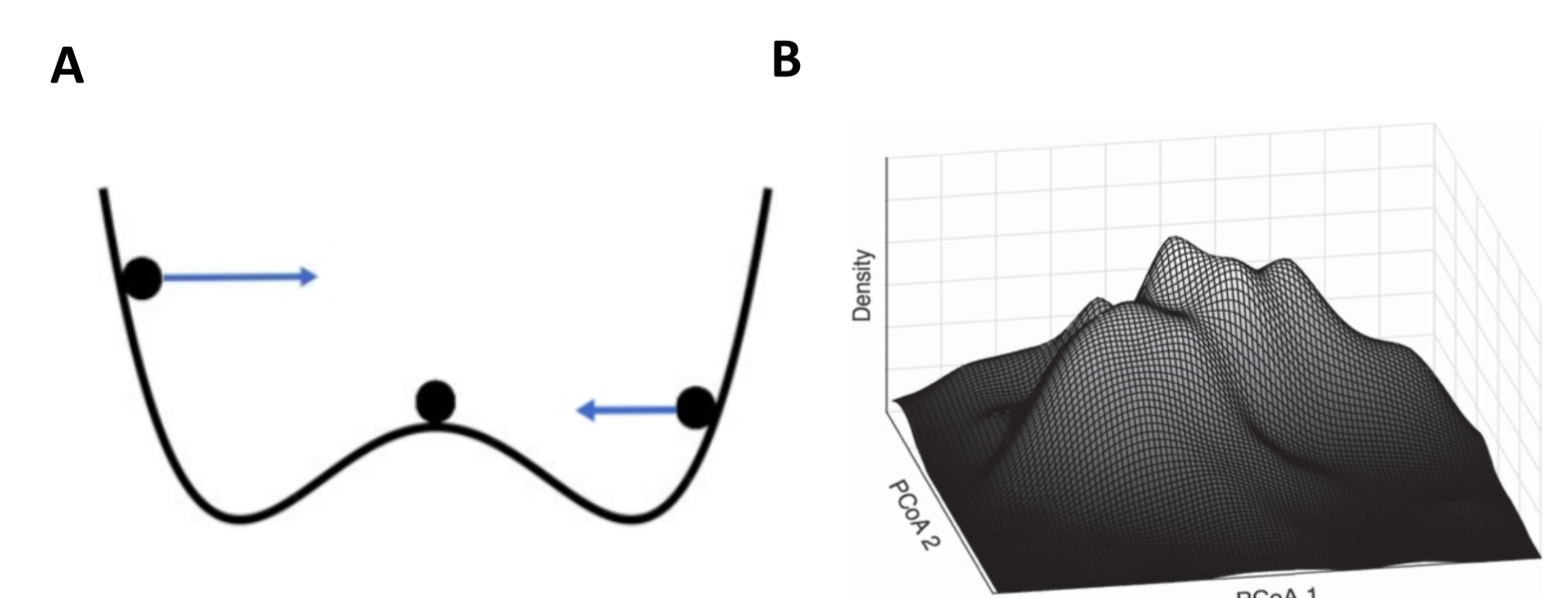


Figure 5. **A** Double well potential that could be used to model transitions between alternative stable states. **B** Variation in population frequency across alternative community types [6].

For preprint [7] and source code, check our research homepage:

<https://openresearchlabs.github.io>

References

- [1] David *et al.* (2014): Host lifestyle affect human microbiota on daily time scales, *Genome Biology*, DOI:10.1186/gb-2014-15-7-r89
- [2] L. Lahti *et al.* (2014): Tipping elements in the human intestinal ecosystem, *Nature Communications*, DOI:10.1038/ncomms5344
- [3] B Carpenter *et al.* (2017), Stan: A probabilistic programming language, *Journal of Statistical Software* 76(1), DOI: 10.18637/jss.v076.i01
- [4] A. Goodman (2018): Fitting ornstein-uhlenbeck-type student's t-processes in stan with applications for population dynamics data, DOI: 10.5281/zenodo.1284346
- [5] C. Driver, M. Voelkle: Hierarchical Bayesian continuous time dynamic modeling, *Psychol Methods*, 2018, DOI: 10.1037/met0000168. In Press.
- [6] L. Lahti *et al.* (2017): Multi-stability and the origin of microbial community types, *The ISME Journal*, DOI: 10.1038/ismej.2017.60.
- [7] V. Laitinen, L. Lahti (2018), A hierarchical Ornstein-Uhlenbeck model for stochastic time series analysis. *Proceedings of the Intelligent Data Analysis Conference (IDA)*. s'Hertogenbosch, The Netherlands. In Press.