Locally Sparse Varying Coefficient Mixed Model

with Application to Longitudinal Differential Abundance Analysis

Simon Fontaine[1] **Nisha J D'Silva**[2] **Ji Zhu**[1] **Gen Li**[3]

[1]Department of Statistics, University of Michigan [2]Department of Periodontics and Oral Medicine, University of Michigan [3]Department of Biostatistics, University of Michigan

Joint Statistical Meeting 2023
Contributed Poster Presentations
Biometrics Section
August 8th, 2023

bit.ly/lsvcmm
fontaine618/LSVCMM
simfont@umich.edu

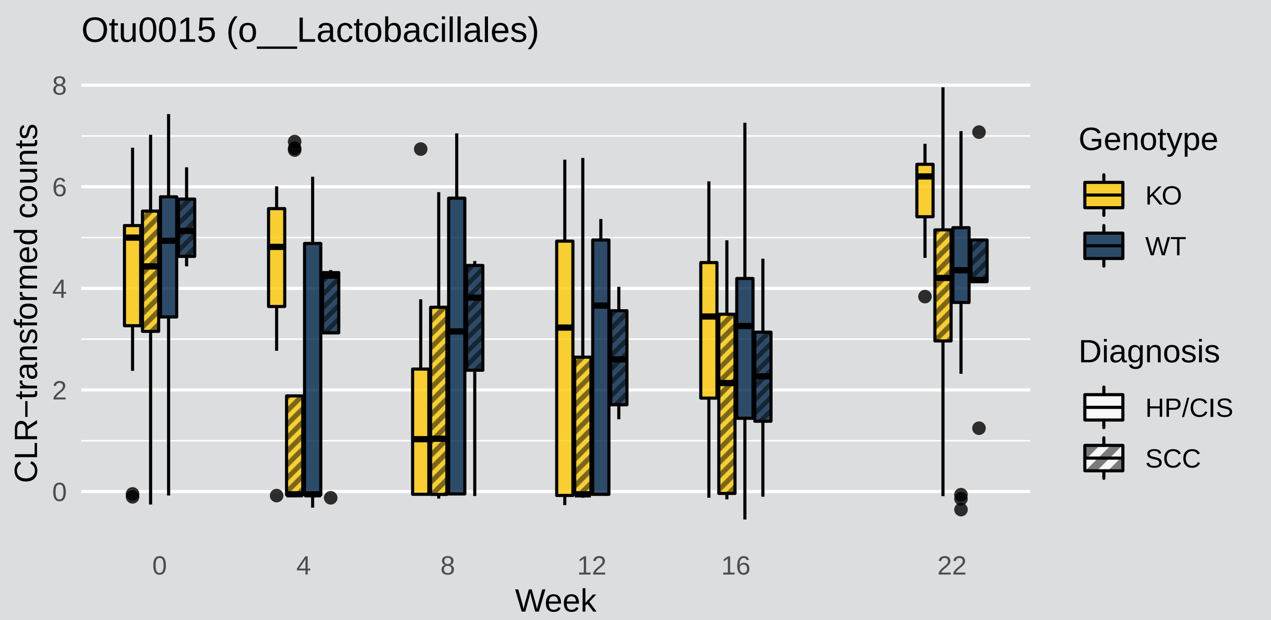
Motivation: Longitudinal Differential Abundance

Researchers are interested in the interaction between **oral squamous cell carcinoma** (OSCC) development, **oral microbiome**, and the **DMBT1 gene**.

To study this interaction, **mice** were bred with (**WT**) and without (**KO**) the DMBT1 gene and a **carcinogen** was introduced to induce OSCC development. **Microbial samples** were **collected over time** and **histopathology** was performed at the end of the study period.

Available data: for each of the 65 mice

- Genotype (DMBT1: KO or WT)
- OTU counts at **up to** 6 times points (0, 4, 8, 12, 16, 22 weeks)
- Diagnosis at week 22 (hyperplasia, carcinoma in situ, SCC)
- Sex (known to be associated with microbiota and diagnosis)



Problem Formulation & Desirata

Function-on-scalar regression: for subject i

- t_{ij}: sampling time of the jth measurement
- $y_{ij} = y_i(t_{ij})$: response at time t_{ij}
- $\mathbf{x}_i$: covariates assumed to have a time-varying effect
- $\mathbf{u}_i$: covariates assumed to have a time-constant effect

Desired properties:

- **Local sparsity** to identify time points of differential effects
- **Smoothness** to borrow strength from neighboring time points
- Account for **dependency** to improve efficiency
- Allow **multiple covariates**, more than a dichotomous group identifier
- Allow **irregular designs** (missing or irregularly-sampled data)

Existing Solutions & their Limitations

Cross-sectional differential abundance (ALDEx2, DESeq2, ANCOM, etc.)

- Does not account for time dependence and cannot borrow strength from neighboring time points

Locally sparse kernel/spline methods (e.g., Kong, 2015)

- Does not account for time dependence

Dependent kernel/spline methods (e.g., Wang, 2003)

- Does not yield a sparse estimate

SPFDA: a locally sparse dependent spline model (Wang, 2023)

- Requires regular sampling times

Varying Coefficient Mixed Model

Mean model: **semi-varying coefficient model**

$$\mathbb{E}\{y_{ij}\} = \mu_{ij} = \mathbf{x}_i^\top \boldsymbol{\beta}(t_{ij}) + \mathbf{u}_i^\top \boldsymbol{\alpha}$$

Covariance model:

$$\mathrm{Cov}\left(\mathbf{y}_i\right) = \mathbf{V}_i, \qquad [\mathbf{V}_i]_{jk} = \kappa\left(t_{ij}, t_{ik}\right)$$

where $\kappa(\cdot,\cdot)$ is the **covariance kernel**.

- **We do not need the true covariance** structure to get **substantial efficiency gains** compared to an independence assumption.
- Instead of estimating the covariance kernel κ, we rather specify a **working covariance model**, where κ is **parameterized** by a few scalar quantities.
- For example, the **compound symmetry** structure (eq., a random intercept model or the exchangeable structure) assumes $\kappa(t_{ij}, t_{ik}) = \rho$.

Kernel Smoothing

To obtain a **smooth** estimate of the time-varying coefficients $\beta(t)$, we proceed to a **locally constant approximation**.

For each time point of interest t, and given a kernel function $k_h(s) = k(s/h)/h$, we consider the **seemingly unrelated kernel estimating equation** (Wang, 2003)

$$\sum_{i=1}^N \mathbf{X}_i^\top \mathbf{W}_i(t) \mathbf{V}_i^{-1} [\mathbf{y}_i - \boldsymbol{\mu}_i] = \mathbf{0}$$

where $\mathbf{W}_i(t) = \mathrm{diag}\left(k_h(t, t_{ij}), j = 1, \dots, n_i\right)$ and where $\mathbf{X}_i = \mathbf{1}\mathbf{x}_i^\top$.

Why not splines?

- Kernel smoothing is **more straightforward** to induce local sparsity. To obtain $\beta_j(t) = 0$, we simply need to use **shrinkage at** t; with **splines**, we need consecutive spline weights to be 0, which requires **overlapping group shrinkage** (see, e.g., Wang, 2023.)

Local & Global Sparsity

We impose a **sparse group Lasso** penalty (Simon et al., 2013) over the evaluation of $\beta(t)$ over a range of pre-specified time points t, namely $\mathbf{b}_j = \left(\beta_j(t^{(1)}), \dots, \beta_j(t^{(T)})\right)$, grouped across covariates $j = 1, \dots, p_x$:

$$\lambda \sum_{j=1}^p \alpha \left\|\mathbf{b}_j\right\|_2 + (1-\alpha) \left\|\mathbf{b}_j\right\|_1 .$$

- The group Lasso penalty encourages **global sparsity** ($\beta_j(\cdot) = 0$.)
- The Lasso penalty encourages **local sparsity** ($\beta_j(t) = 0$ for some t.)
- We further consider the **adaptive** sparse group Lasso (Poignard, 2020) to reduce shrinkage bias.

Estimation

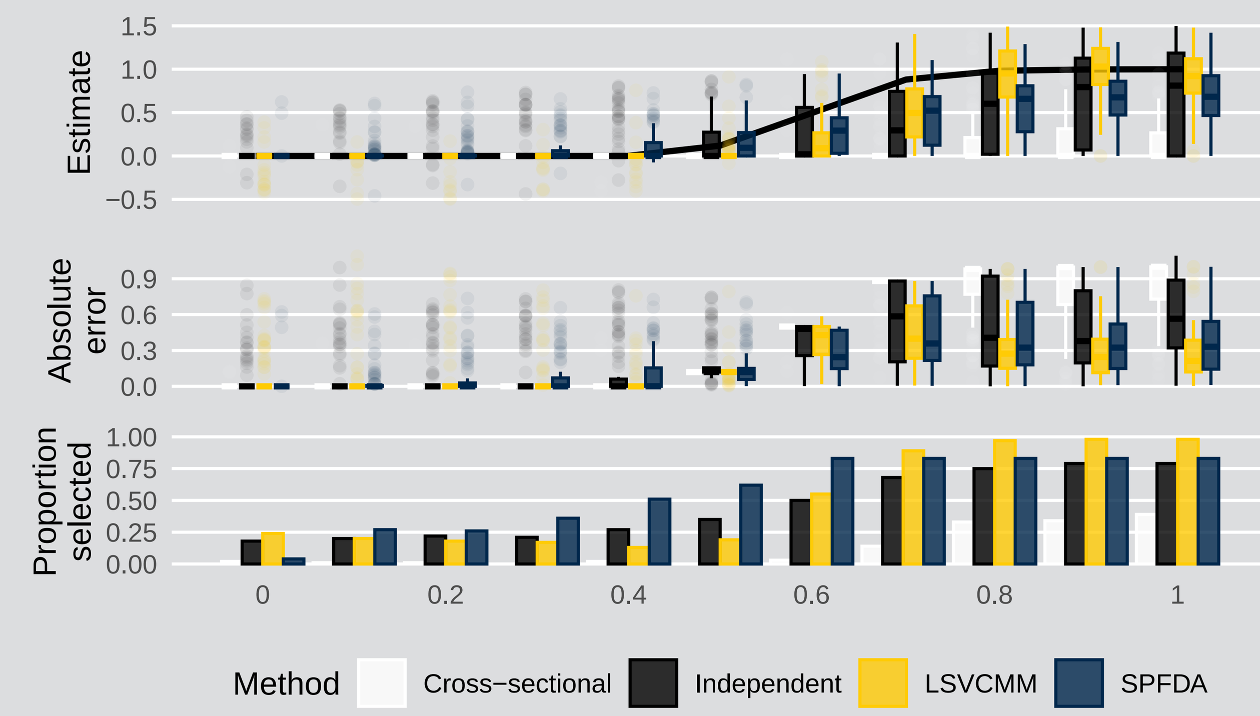
We **alternate** between estimation of mean and covariance parameters:

- **Mean parameters** updates are obtained by **proximal gradient descent**
- **Covariance parameters** updates are obtained by **Newton-Raphson steps**

Selection of tuning parameters (regularization parameter λ and kernel scale h) is performed through an **information criterion** (EBIC).

Simulation Experiments

- 100 datasets, each with 100 subjects split into two groups
- Group difference on the second half of the range
- Longitudinal effect: per-subject random intercept
- 11 sampling points per subject, half of which are discarded

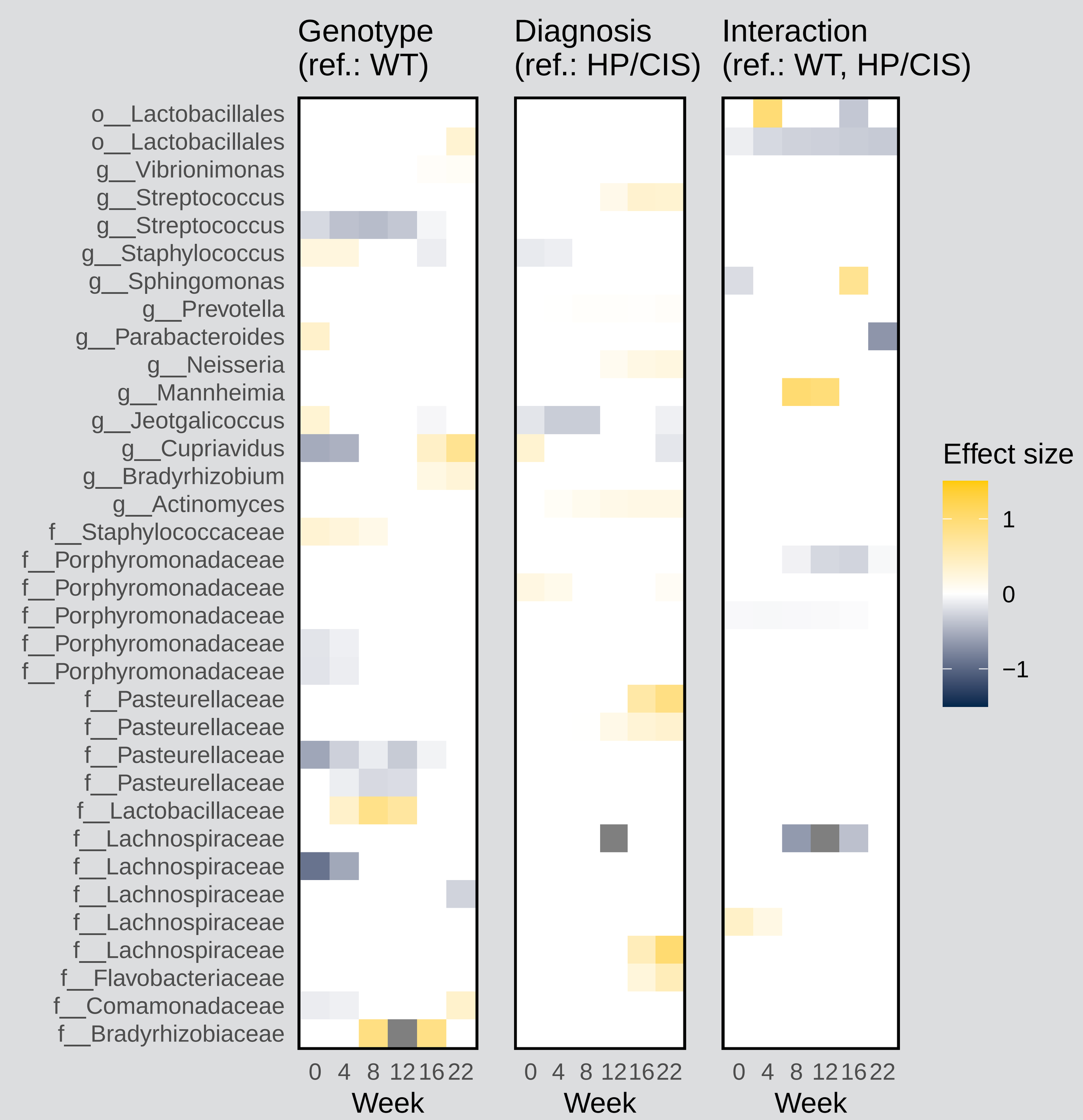


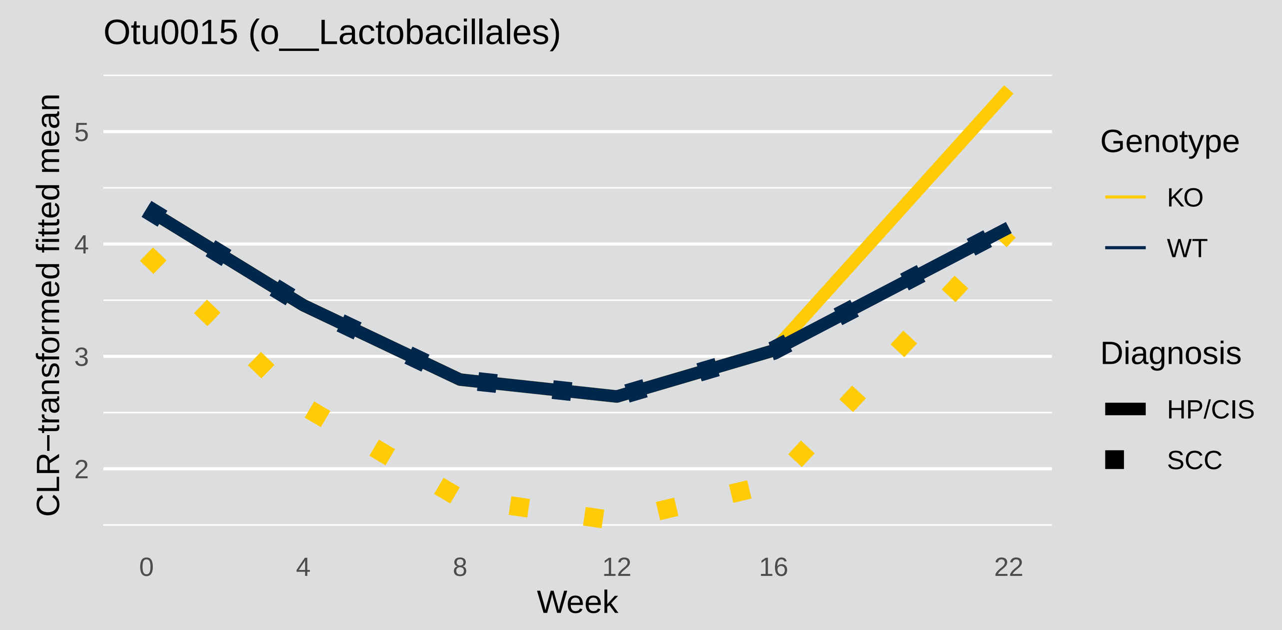
Real Data Application

For each clr-transformed OTU, we consider the following mean model:

```
1(Week) + Sex(Week) + Genotype(Week) * Diagnosis(Week)
```

Repeat for each of 187 OTUs, report those with Genotype or Diagnosis effects.





Extensions, Variants & Future Work

The current approach (Gaussian model on CLR-transformed response) may not be the most appropriate for microbial counts:

- **GLM with other families** (e.g., negative binomial)
- **Multivariate model** to account for correlation between taxa and compositional effects
- **Zero-inflated model** to account for excess zeros

Local linear approximation (instead of constant) could improve fit (& sparse derivatives could be of interest)

Additional working covariance structures, especially AR(1)

Alternative regularization, such as the SCAD penalty

Uncertainty quantification, such as bootstrap or asymptotic confidence intervals, or probability of inclusion

Summary

We propose a novel approach to function-on-scalar regression based on kernel smoothing producing locally sparse estimates of the functional coefficients with no requirements on the sampling design.

Our method improves both estimation accuracy and identification of time points with differential effects, compared to the independence covariance assumption and the SPFDA method.

Acknowledgements

Gen Li's research was partially supported by the National Institutes of Health grant R03DE031296.

We are grateful to Prof. **Grace Chen** (UM Hematology and Oncology) and Dr. **Marcell Costa de Medeiros** (UM Periodontics and Oral Medicine) for their useful comments.

References

Kong, D., Bondell, H. D., & Wu, Y. (2015). **Domain selection for the varying coefficient model via local polynomial regression**. *Computational statistics & data analysis*, 83, 236-250.

Poignard, B. (2020). **Asymptotic theory of the adaptive Sparse Group Lasso**. *Annals of the Institute of Statistical Mathematics*, 72, 297-328.

Simon, N., Friedman, J., Hastie, T., & Tibshirani, R. (2013). **A sparse-group lasso**. *Journal of computational and graphical statistics*, 22(2), 231-245.

Wang, N. (2003). **Marginal nonparametric kernel regression accounting for within-subject correlation**. *Biometrika*, 90(1), 43-52.

Wang, Z., Magnotti, J., Beauchamp, M. S., & Li, M. (2023). **Functional group bridge for simultaneous regression and support estimation**. *Biometrics*, 79(2), 1226-1238.