preprint

DGP-LVM: Derivative Gaussian process latent variable models

Soham Mukherjee, Manfred Claassen, Paul-Christian Bürkner

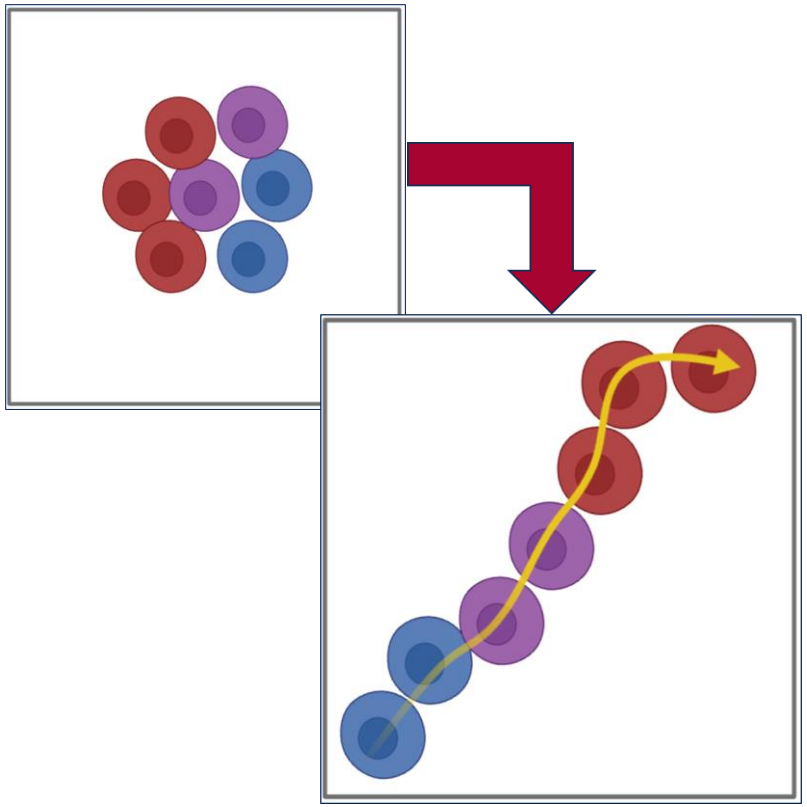
technische universität dortmund

EBERHARD KARLS UNIVERSITÄT TÜBINGEN

imprs-is

Motivation: single-cell biology

Single-cell RNA sequencing data



Ordered cells

y

Genes

Cells

y'

Genes

Cells

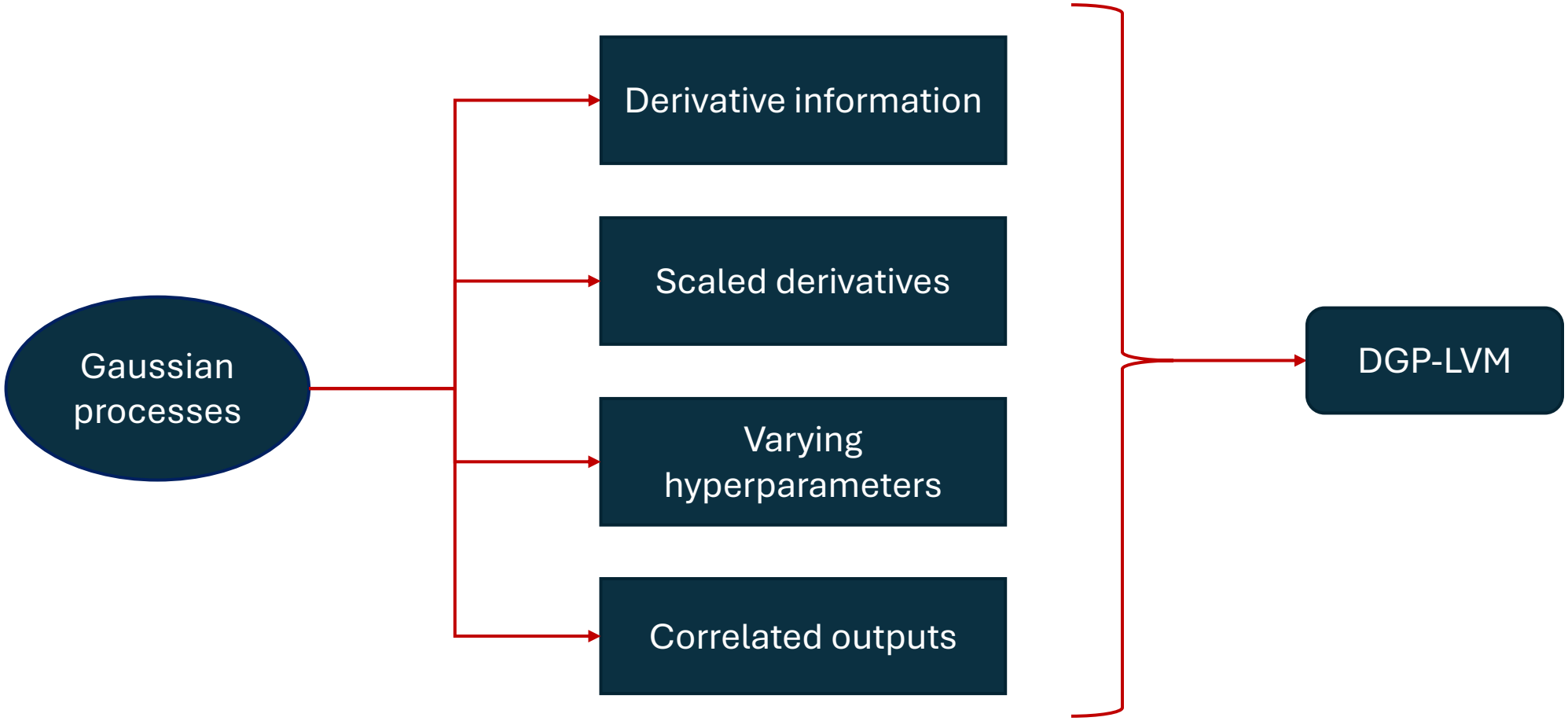
In real world…

$$y' \propto \frac{\delta}{\delta x} y$$

x

Ordering

Model framework



DGP-LVM

For a choice of covariance function K_d and output dimension d

$$\begin{pmatrix} f_d(x) \\ f'_d(x) \end{pmatrix} \sim \mathcal{GP}\left(\begin{pmatrix} m_{f_d} \\ m_{f'_d} \end{pmatrix}, \begin{pmatrix} K_d & K'_d \\ K'^T_d & K''_d \end{pmatrix} \right)$$

Considering derivative SE with length scale ρ_d and GP marginal SD α_d

$$K_d(x_i, x_j) = \alpha_d^2 \exp\left(-\frac{(x_i - x_j)^2}{2\rho_d^2} \right),$$

$$K'_d(x_i, x_j) = \alpha_d \alpha'_d \frac{(x_i - x_j)}{\rho_d^2} \exp\left(-\frac{(x_i - x_j)^2}{2\rho_d^2} \right),$$

$$K''_d(x_i, x_j) = \frac{\alpha'^2_d}{\rho_d^4}(\rho_d^2 - (x_i - x_j)^2) \exp\left(-\frac{(x_i - x_j)^2}{2\rho_d^2} \right).$$

Latent inputs with measurement SD s

$$\tilde{x}_i \sim \mathcal{N}(x_i, s^2)$$

Specify the likelihood with error SD σ_d

$$y_{di} \sim \mathcal{N}(f_d(x_i), \sigma_d^2),$$

$$y'_{di} \sim \mathcal{N}(f'_d(x_i), \sigma'^2_d).$$

Simulation study

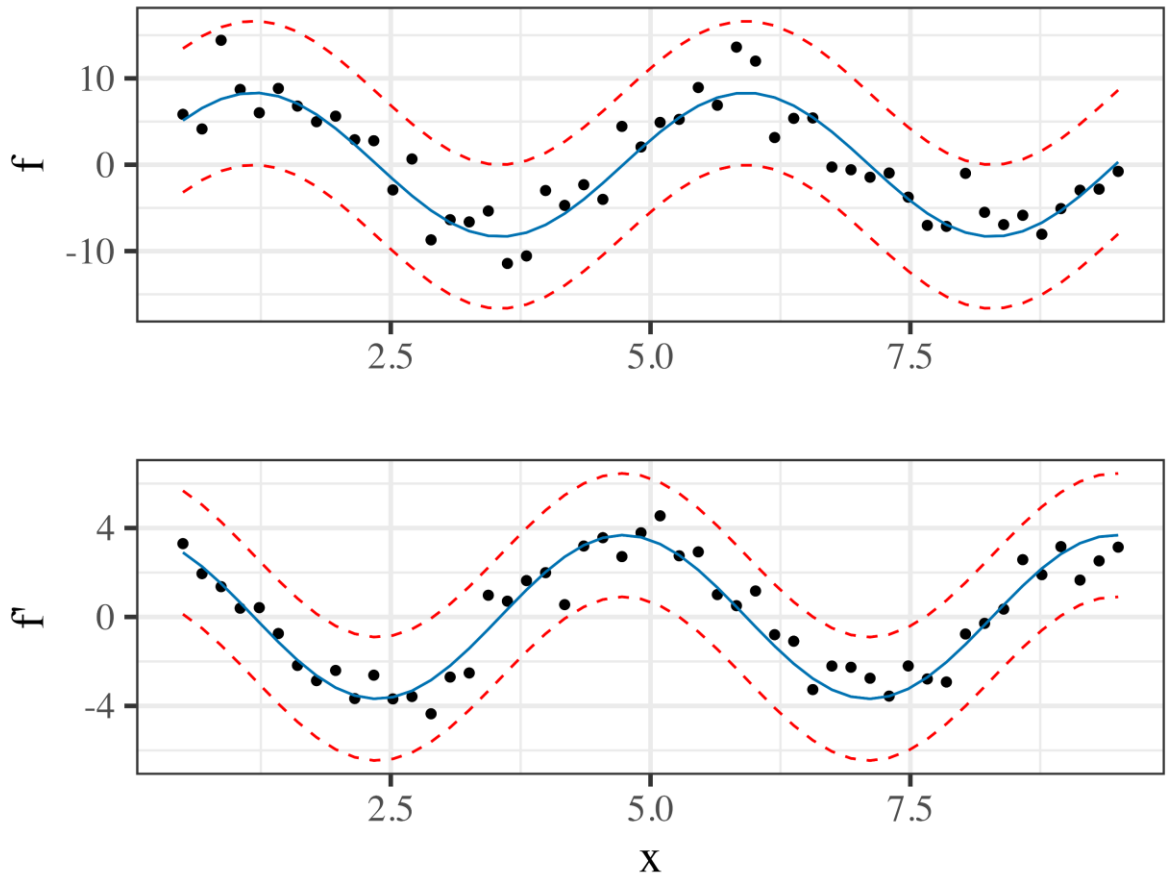
Consider a periodic data simulation scenario

$$f_{ij} = \alpha_j \sin\left(\frac{x_i}{\rho_j}\right)$$

$$f'_{ij} = \frac{\alpha'_j}{\rho_j} \cos\left(\frac{x_i}{\rho_j}\right)$$

$$\alpha \propto \alpha'$$

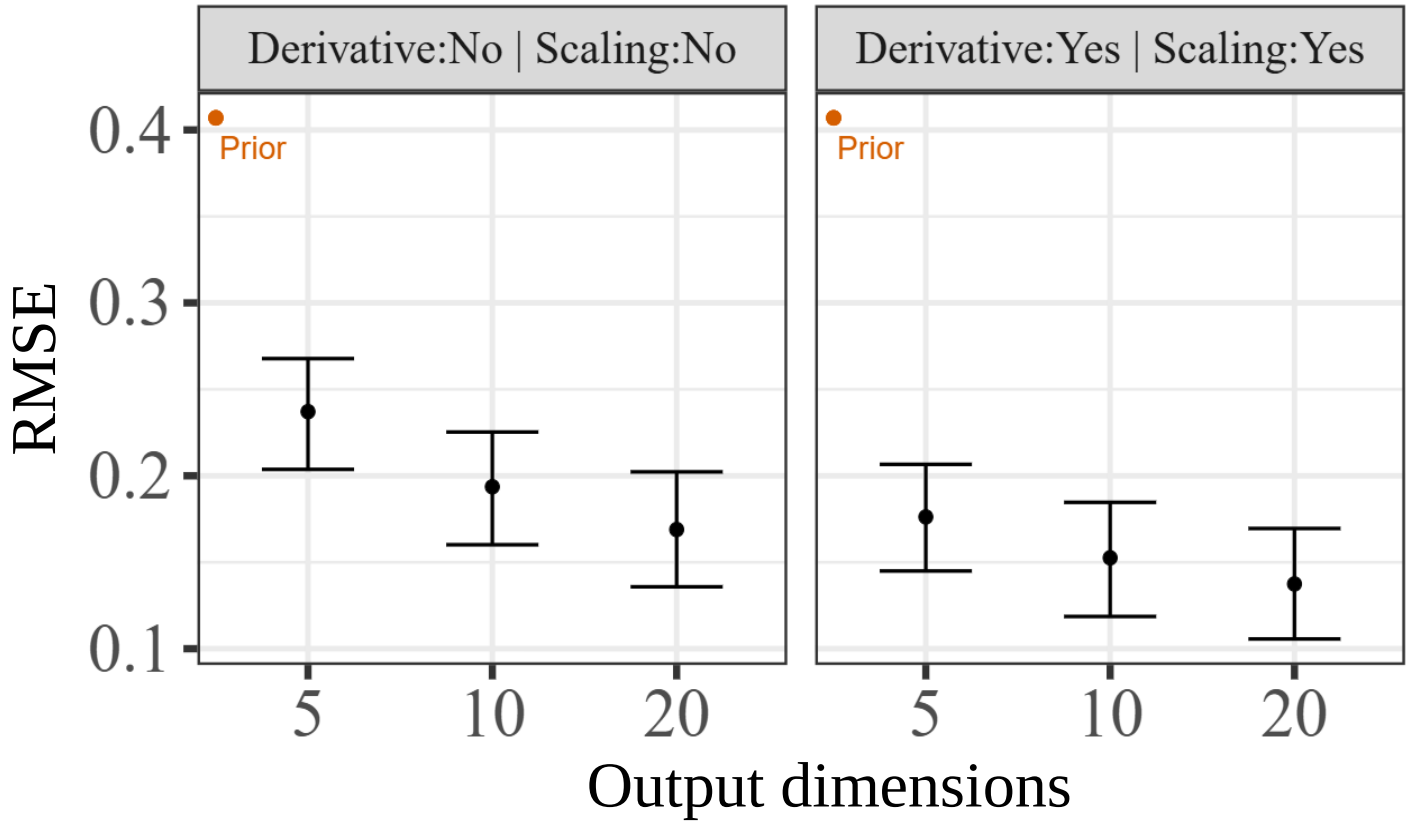
Periodic data



Effects of derivatives

We compare the effects of adding derivative information along with scaling customizations on estimating latent x

$$\mathrm{RMSE}(x_{post}) = \sqrt{\mathbb{E}(x_{post} - x_{true})^2}$$



Note: The prior RMSE is the error between the true x and prior $\tilde{x}$

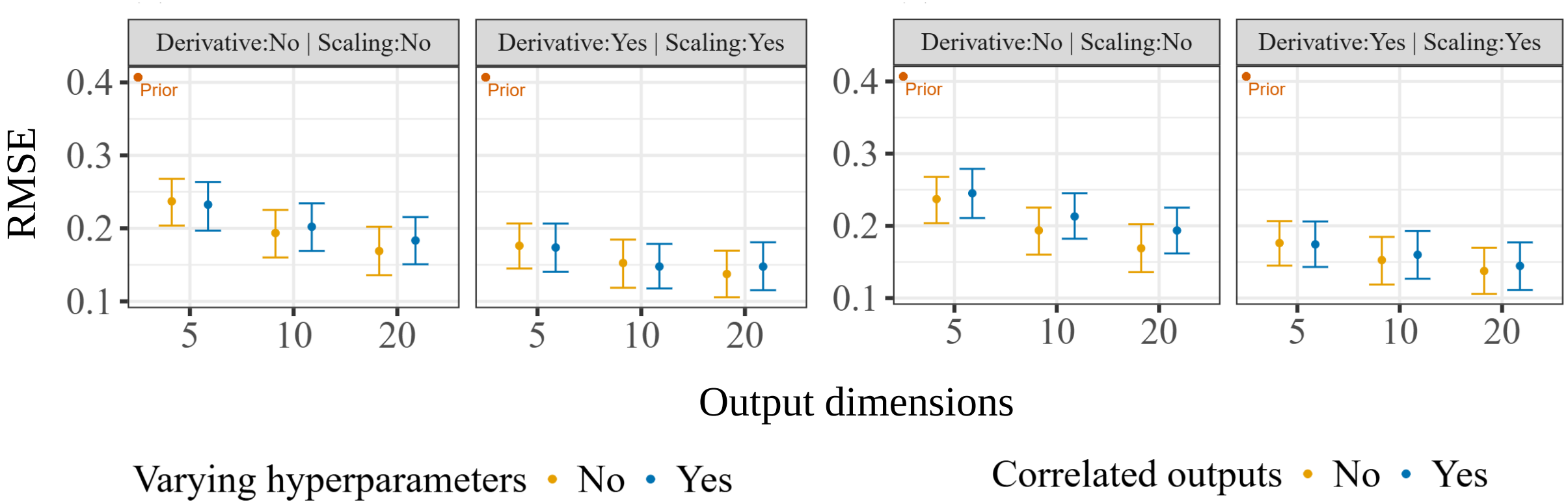
Summary

- GPs for scaled derivative multi-correlated-output data.
- Uncertainty estimates for latent samples using MCMC.
- Applications to single-cell biology and other fields dealing with derivative data.

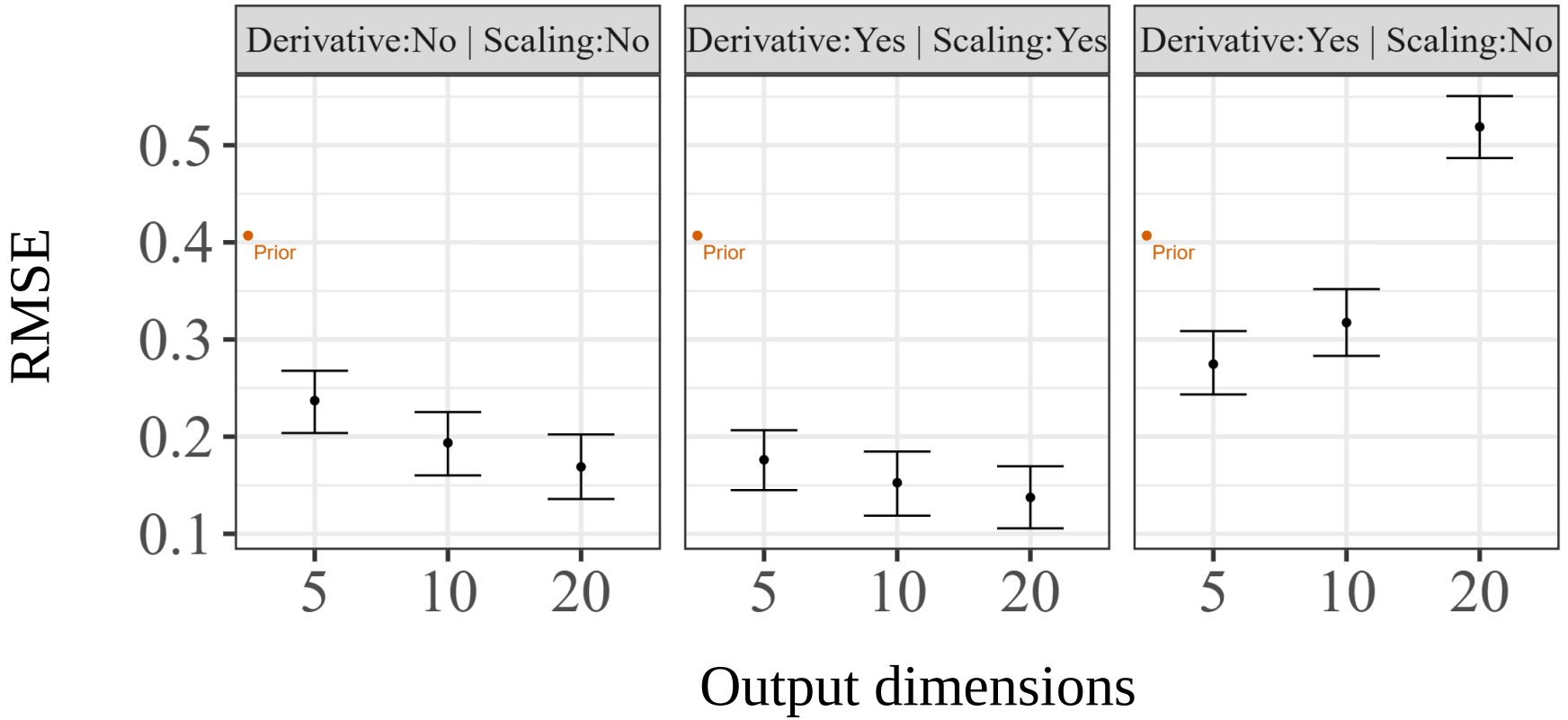
Mukherjee, S., Claassen, M., & Bürkner, P. C. (2024). DGP-LVM: Derivative Gaussian process latent variable models (*in review*). *arXiv preprint arXiv:2404.04074*.

https://arxiv.org/abs/2404.04074

Appendix


Derivative:No | Scaling:No
Derivative:Yes | Scaling:Yes
Derivative:No | Scaling:No
Derivative:Yes | Scaling:Yes
Prior
RMSE
0.4
0.3
0.2
0.1
5
10
20
Output dimensions
Varying hyperparameters • No • Yes
Correlated outputs • No • Yes

Appendix


Derivative:No | Scaling:No
Derivative:Yes | Scaling:Yes
Derivative:Yes | Scaling:No
RMSE
0.5
0.4
0.3
0.2
0.1
Prior
Prior
Prior
5
10
20
5
10
20
5
10
20
Output dimensions