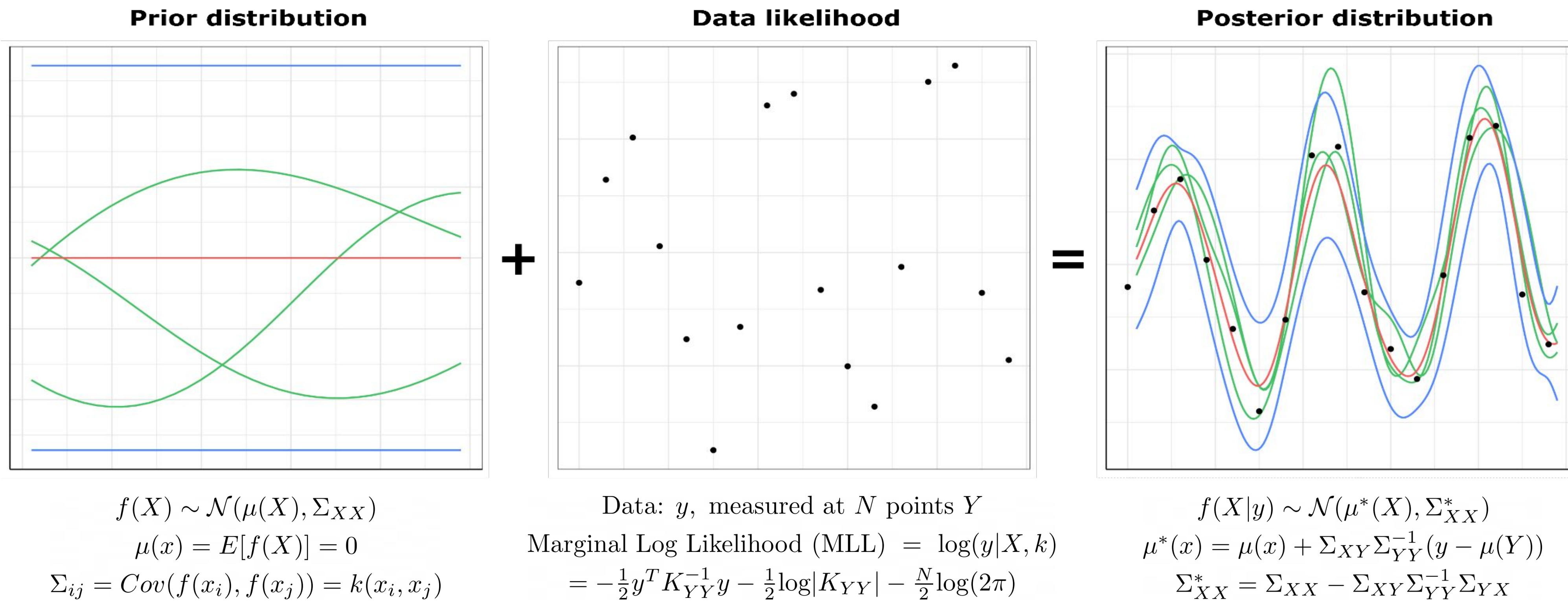


BACKGROUND

Quantitative inference of the presence of oscillations in time-series datasets is challenging, particularly the detection of circadian clock rhythms in dividing cells. In single cells, oscillations can be observed in lineage trajectories when the dividing cells contain reporters for circadian clock proteins. However, they have very low amplitude, are noisy, and non-stationary. Population level measurements like qPCR also exhibit these complex noisy patterns, making oscillation detection a difficult problem.

GAUSSIAN PROCESSES

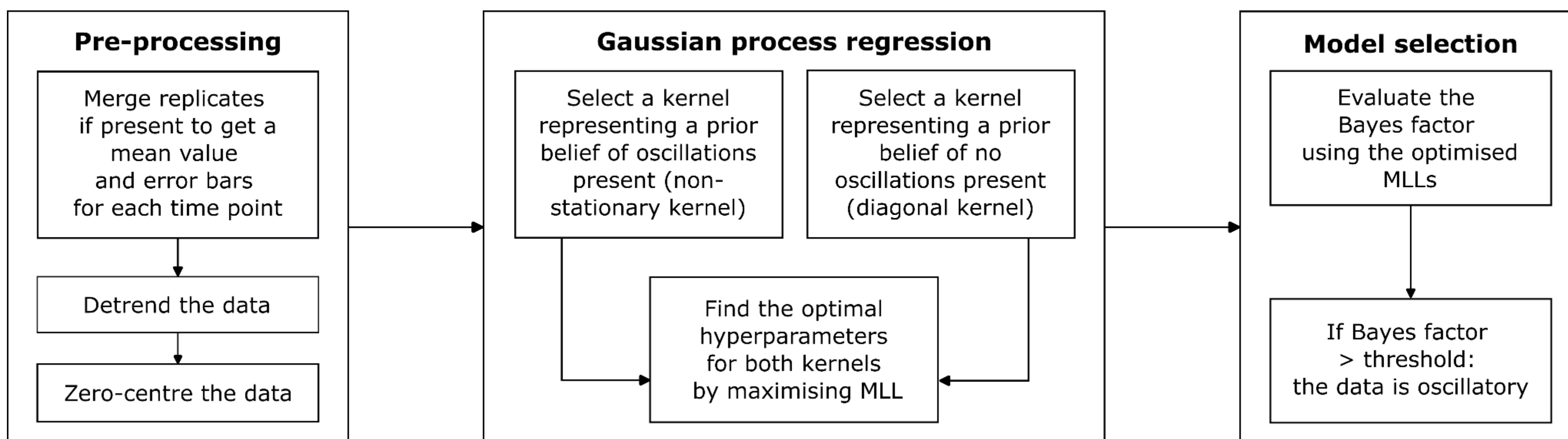
A Gaussian Process (GP) is a set of random variables, any finite combination of which has a multivariate normal distribution. When fitting data, the n th random variable models the target function's value at n th domain point.



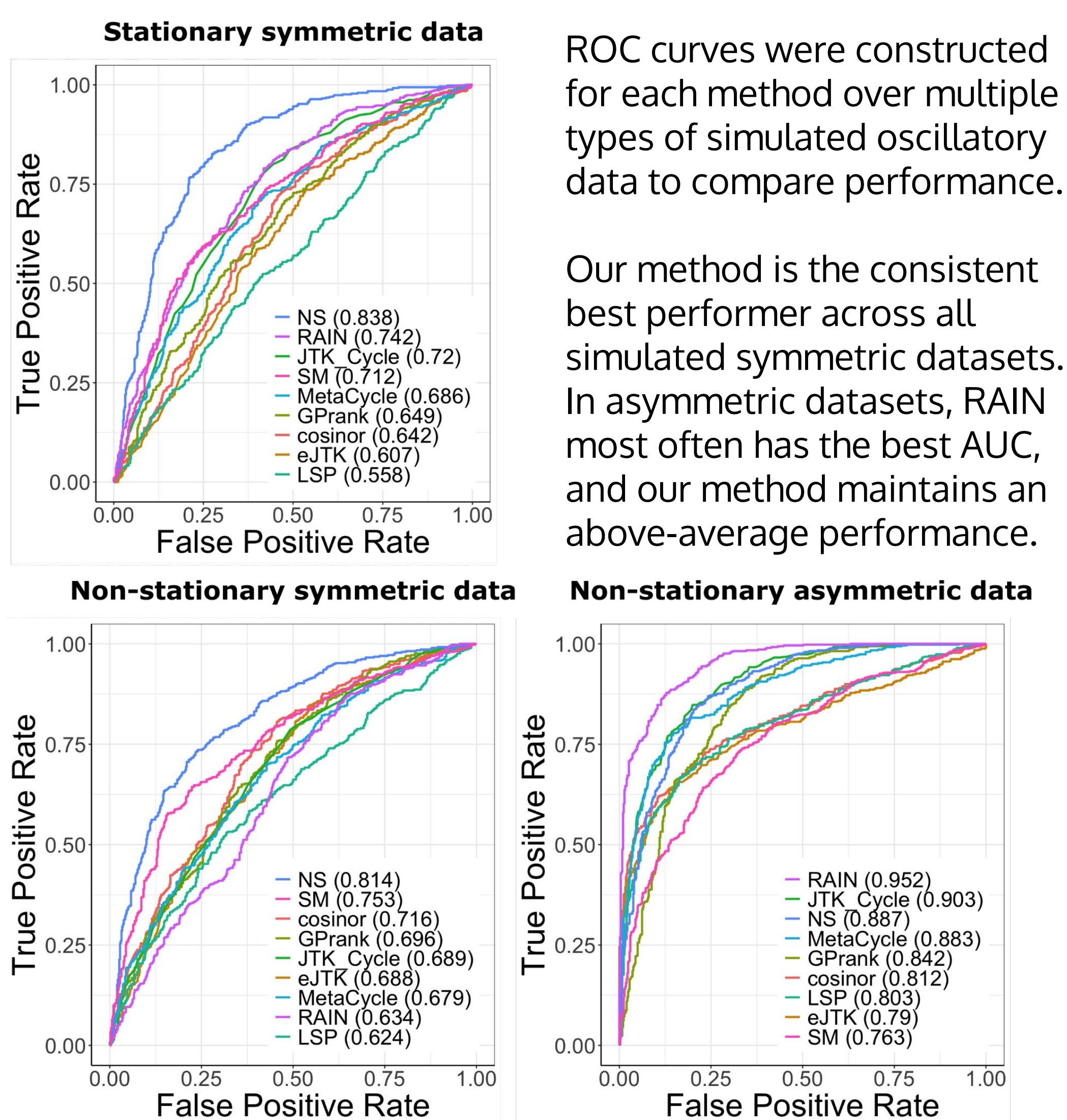
k describes the similarity between function values, and thus controls the shapes that a fitted function can adopt.

Diagonal kernel: $K_D = \sigma^2 \cdot \delta_{xx'}$ Non-stationary kernel: $K_{NS} = \sum_{i=1}^Q w_i(x)w_i(x')k_{gibbs,i}(x, x')\cos(2\pi(\mu_i(x) \cdot x - \mu_i(x') \cdot x')) + \sigma^2 \cdot \delta_{xx'}$

WORKFLOW OF THE METHOD



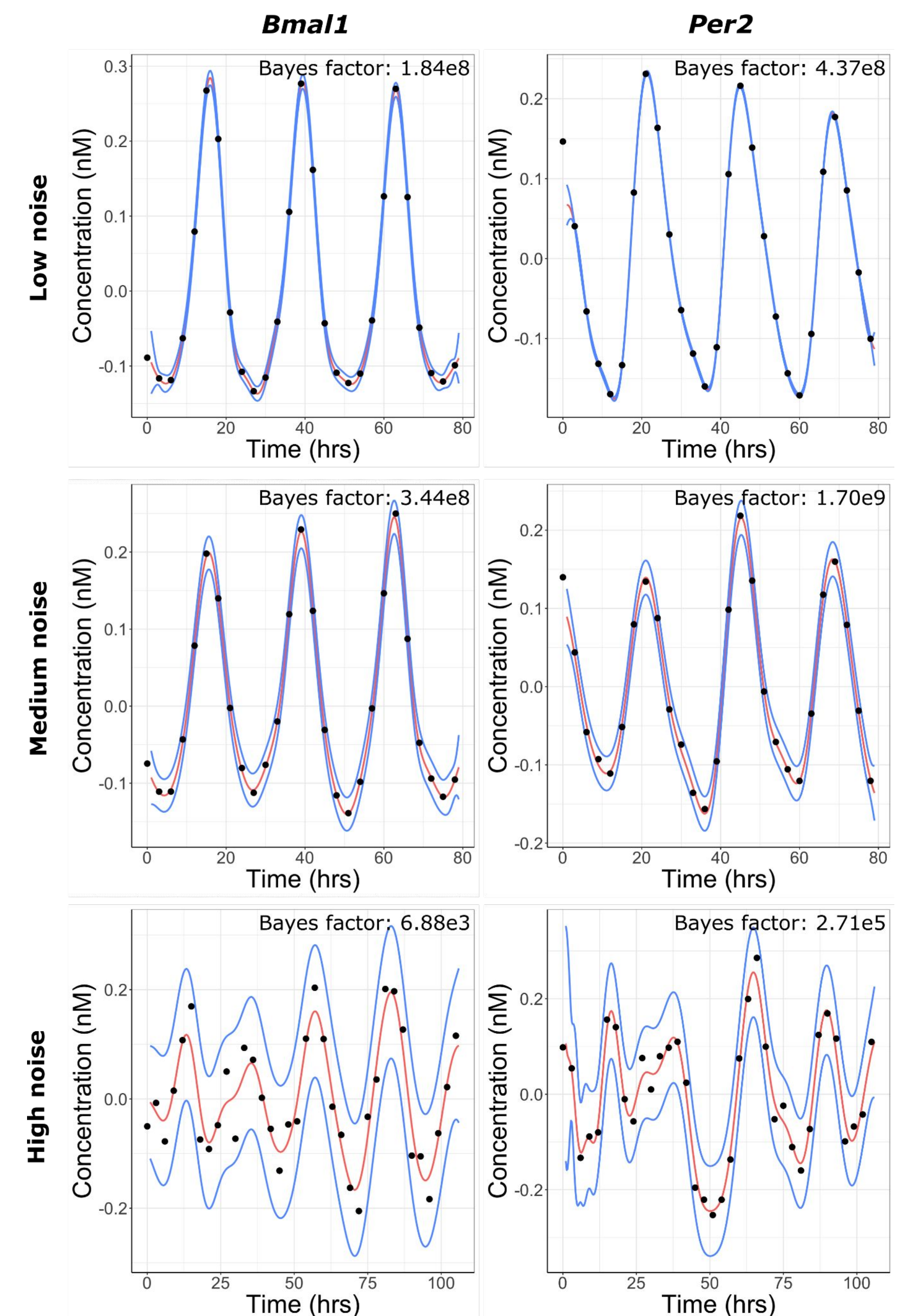
COMPARISON AGAINST EXISTING METHODS



Method	Times among best 3	Times among worst 3
Stationary symmetric datasets (Total: 9)		
NS Kernel	9	0
Cosinor	5	3
eJTK	5	3
Non-stationary symmetric datasets (Total: 15)		
NS Kernel	15	0
Cosinor	9	2
eJTK	7	4
Stationary asymmetric datasets (Total: 8)		
RAIN	7	0
JTK_Cycle	3	0
NS Kernel	2	1
Non-stationary asymmetric datasets (Total: 12)		
RAIN	10	0
JTK_Cycle	8	0
NS Kernel	5	0

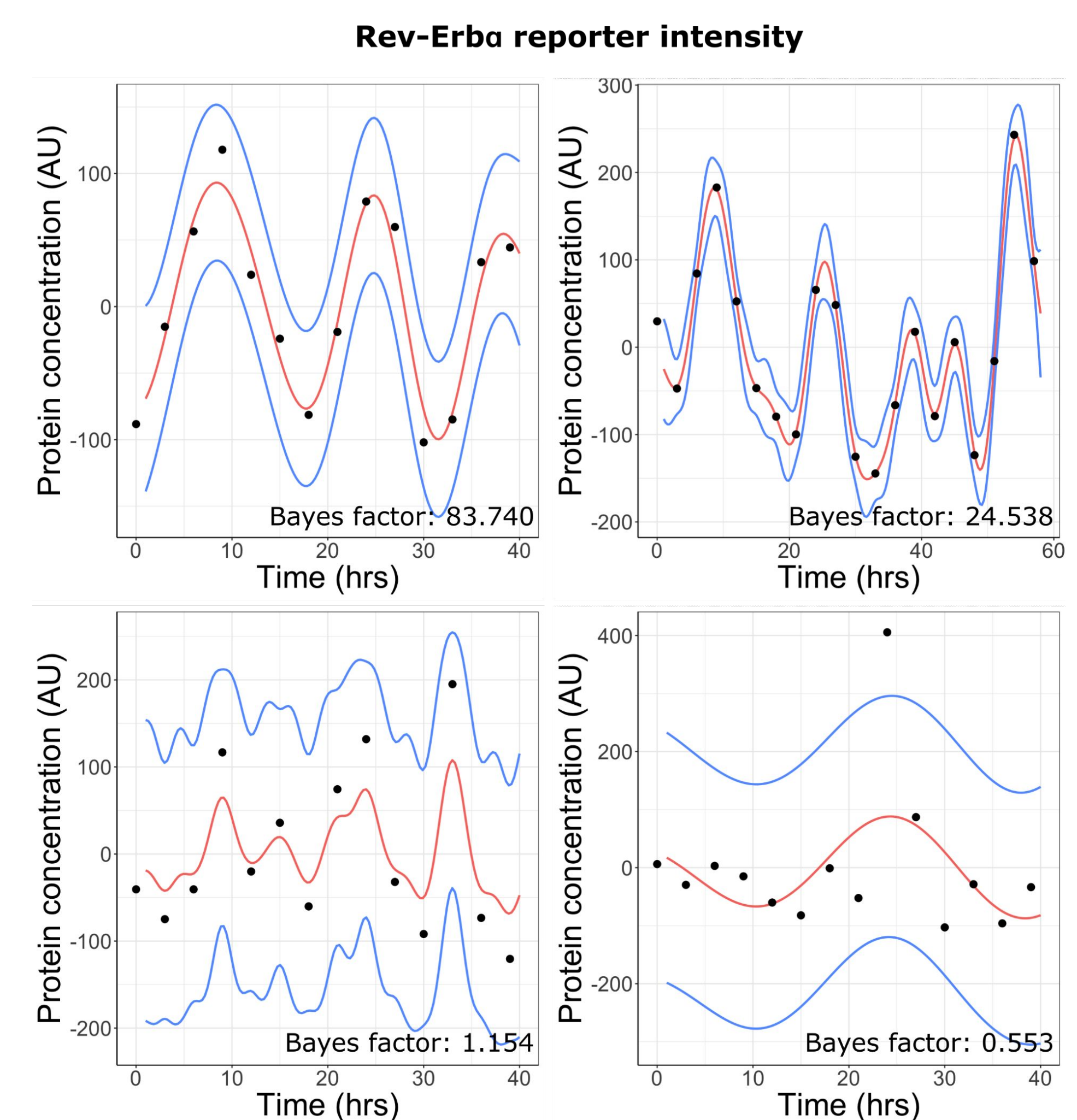
RESULTS: SIMULATED LINEAGES

A coupled circadian clock - cell cycle gene network was simulated with ODEs, and used to generate cell lineages. Concentration of clock gene products over time was extracted from individual lineages.



RESULTS: EXPERIMENTAL DATA

Time lapse microscopy was carried out over 72 hours on live U2OS (osteosarcoma) cells containing a reporter for circadian protein Rev-Erba. The corresponding channel intensity was extracted from all possible lineages via image analysis. (Data courtesy Granada Lab at Charité.)



CONCLUSIONS

Our method demonstrates the power and flexibility of GPs in modelling of noisy non-stationary time-series data, and provides a way of classifying oscillatory versus non-oscillatory datasets without using p-values. It outperforms existing methods for oscillatory/non-oscillatory classification of simulated symmetric data. It can effectively detect oscillations in noisy circadian clock protein time-series datasets extracted from both simulated and experimentally generated lineage trajectories.

REFERENCES

- [1] Remes, S., Heinonen, M., Kaski, S. (2017). Non-stationary spectral kernels. arXiv:1705.08736 [2] Heinonen, M. et. al. (2015) Detecting time periods of differential gene expression using gaussian processes: an application to endothelial cells exposed to radiotherapy dose fraction. Bioinformatics, 31(5):728–735. [3] El Cheikh, et. al.. (2014). Modeling circadian clock–cell cycle interaction effects on cell population growth rates. Journal of Theoretical Biology, 363, 318-331.