Bayesian Optimisation for Active Monitoring of Air Pollution [1]

Sigrid Passano Hellan, Christopher G. Lucas and Nigel H. Goddard

School of Informatics, University of Edinburgh

s.p.hellan@ed.ac.uk

Aims

Ambient air pollution causes about 4 million deaths annually [2]. Recent progress in low-cost sensors [3] means that more and denser monitoring networks can be set up.

We are interested in this research question:

Q: Can Bayesian optimisation and low-cost sensors efficiently monitor urban air pollution?

Data

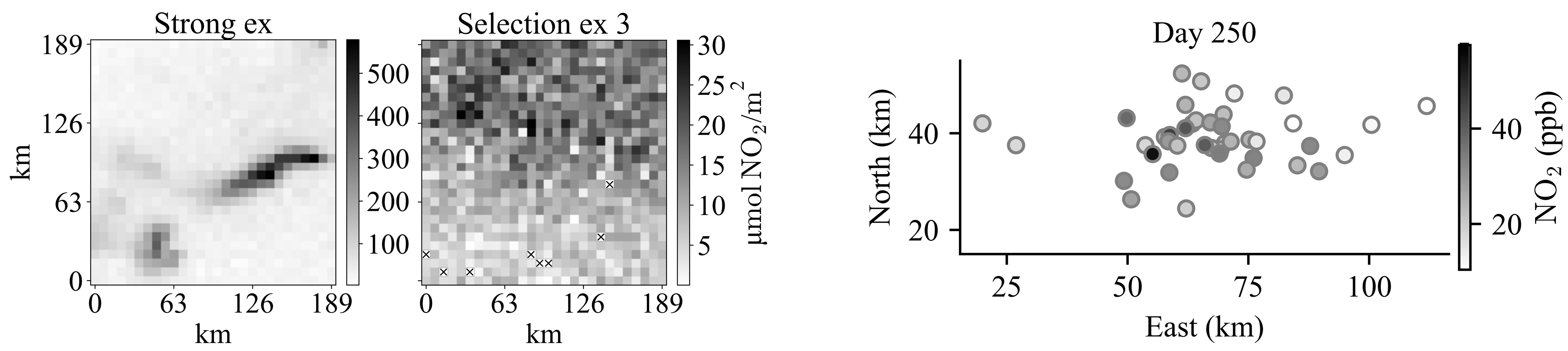

Figure 1: *Left and Middle*: Example data from the EU Copernicus project, from the Strong (high pollution) and Selection (assorted pollution levels) subsets. *Right*: Example data from the London Air Quality Network [4].

Results

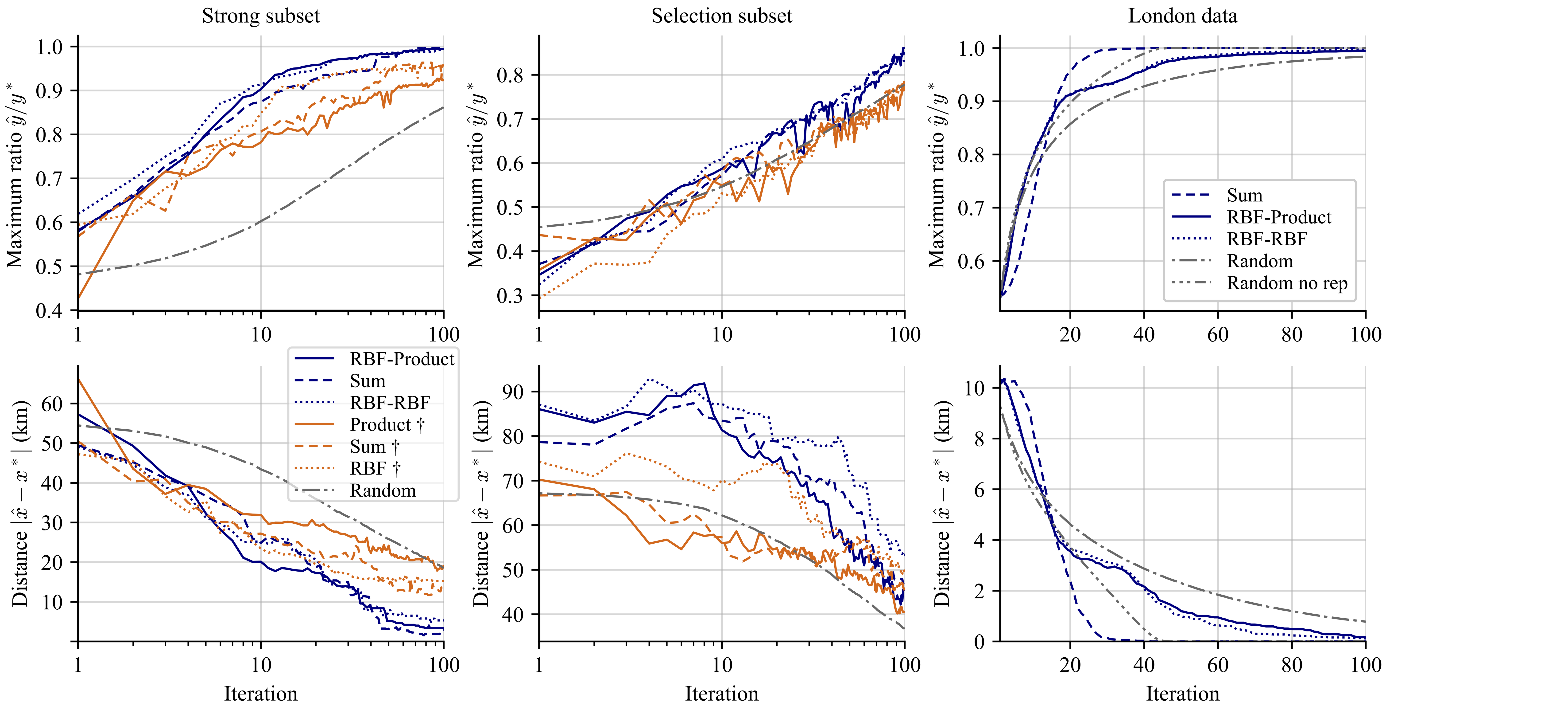

Figure 2: Hierarchical models in blue, simpler BO models in orange (†) and random sampling in grey. $\hat{x}$: Estimated maximiser, x^*: True maximiser, $\hat{y}$: True value at esimated maximiser, y^*: True value at true maximiser. Values given as means over test sets. *Left and middle*: Copernicus data. *Right*: London data.

The hierarchical models in blue outperformed random sampling in grey and simpler BO models in orange on the Copernicus data. Our Sum model performed best on the London data, finding the maxima fastest.

Model

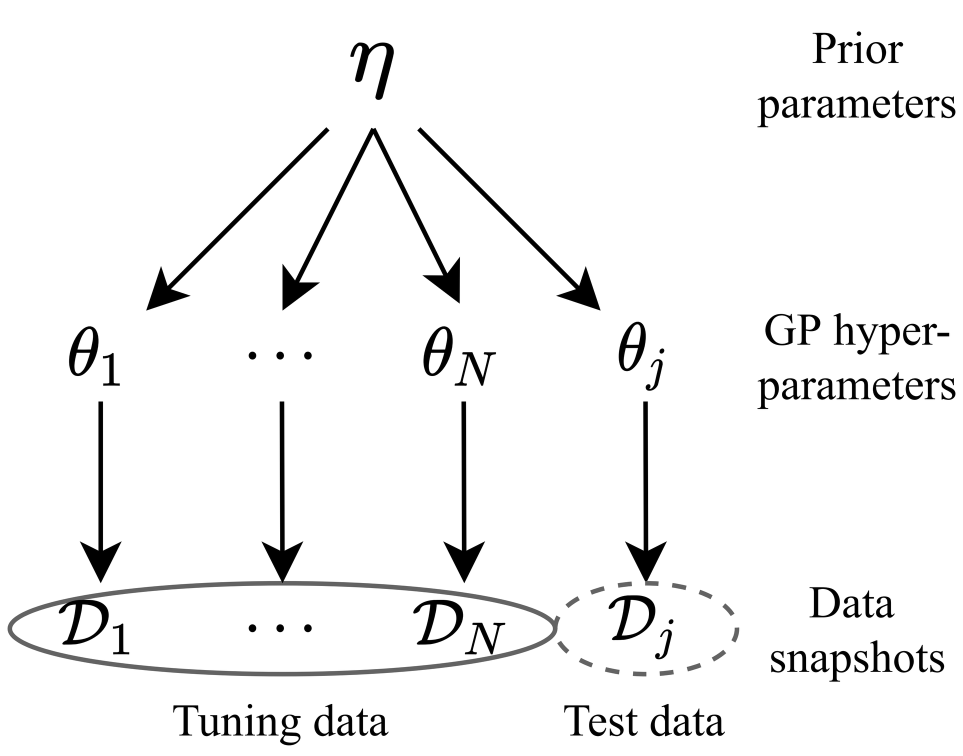

Figure 3: Hierarchical model used on the data.

Conclusion

Bayesian optimisation can successfully solve the problem, outperforming basic baselines.

Method

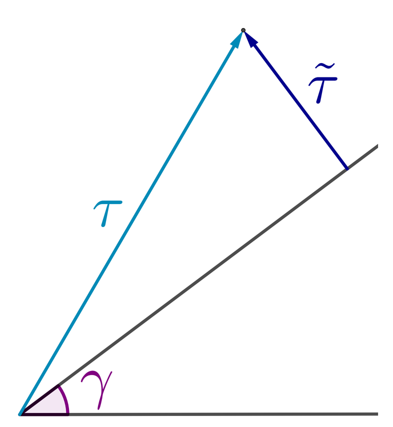

Figure 4: τ is the distance between two locations. $\tilde{\tau}$ is the part of τ orthogonal to wind direction γ.

GAUSSIAN PROCESSES:

We compare three covariance functions (yellow box), built from two base ones: RBF, $k_{\mathrm{R},i}(\tau)$, and Wind, $k_{\mathrm{W},i}(\tau) = k_{\mathrm{R},i}(\tilde{\tau})$ [5].

$$k_{\text{RBF-RBF}}(\tau) = k_{\mathrm{R},1}(\tau) + k_{\mathrm{R},2}(\tau)$$
$$k_{\text{RBF-Product}}(\tau) = k_{\mathrm{R},1}(\tau) + k_{\mathrm{R},2}(\tau)k_{\mathrm{W},3}(\tau)$$
$$k_{\text{Sum}}(\tau) = k_{\mathrm{R},1}(\tau) + k_{\mathrm{W},2}(\tau)$$

HYPERPARAMETERS:

Hyperparameter samples are generated using MCMC, and importance weighting is used to fit to the test data.

References

[1] Sigrid Passano Hellan, Christopher G. Lucas, and Nigel H. Goddard. Bayesian Optimisation for Active Monitoring of Air Pollution. In *Proceedings of AAAI*, 2022.

[2] Aaron J Cohen et al. Estimates and 25-year trends of the global burden of disease attributable to ambient air pollution. *The Lancet*, 389(10082):1907--1918, May 2017.

[3] Michael R. Giordano et al. From low-cost sensors to high-quality data. *Journal of Aerosol Science*, 158:105833, November 2021.

[4] Imperial College London, Environmental Research Group. London Air Quality Network, 1993.

[5] Sigrid Passano Hellan, Christopher G. Lucas, and Nigel H. Goddard. Optimising Placement of Pollution Sensors in Windy Environments. In *arXiv:2012.10770*, December 2020.

Resources

DATASCIENCE EPSRC CENTRE FOR DOCTORAL TRAINING

THE UNIVERSITY of EDINBURGH informatics

UK Research and Innovation