Multi-resolution Multi-task Gaussian Processes

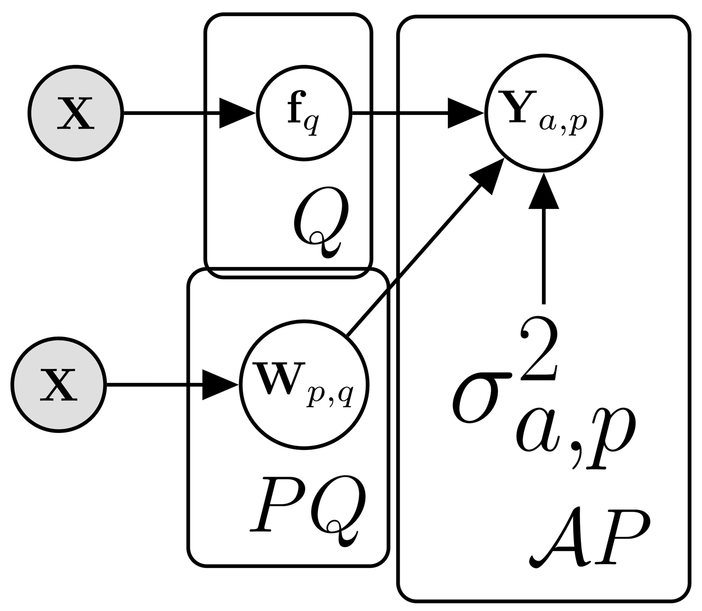
Oliver Hamelijnck[1,2] ohamelijnck@turing.ac.uk
Theodoros Damoulas[1,2,3], tdamoulas@turing.ac.uk
Kangrui Wang[1,2] kwang@turing.ac.uk
Mark Girolami[1,4] mgirolami@turing.ac.uk

[1]The Alan Turing Institute
[2]University of Warwick, Department of Computer Science
[3]University of Warwick, Department of Statistics
[4] Cambridge University

Multi-resolution Multi-task Learning

We consider evidence integration from potentially dependent observation pro- cesses under varying spatio-temporal sampling resolutions and noise levels. We develop a multi-resolution multi-task (MRGP) framework while allowing for both inter-task and intra-task multi-resolution and multi-fidelity. We develop shallow Gaussian Process (GP) mixtures that approximate the difficult to estimate joint likelihood with a composite one and deep GP constructions that naturally handle biases in the mean. By doing so, we generalize and outperform state of the art GP compositions and offer information-theoretic corrections and efficient variational approximations. We demonstrate the competitiveness of MRGPs on synthetic settings and on the challenging problem of hyper-local estimation of air pollution levels across London from multiple sensing modalities operating at disparate spatio-temporal resolutions.

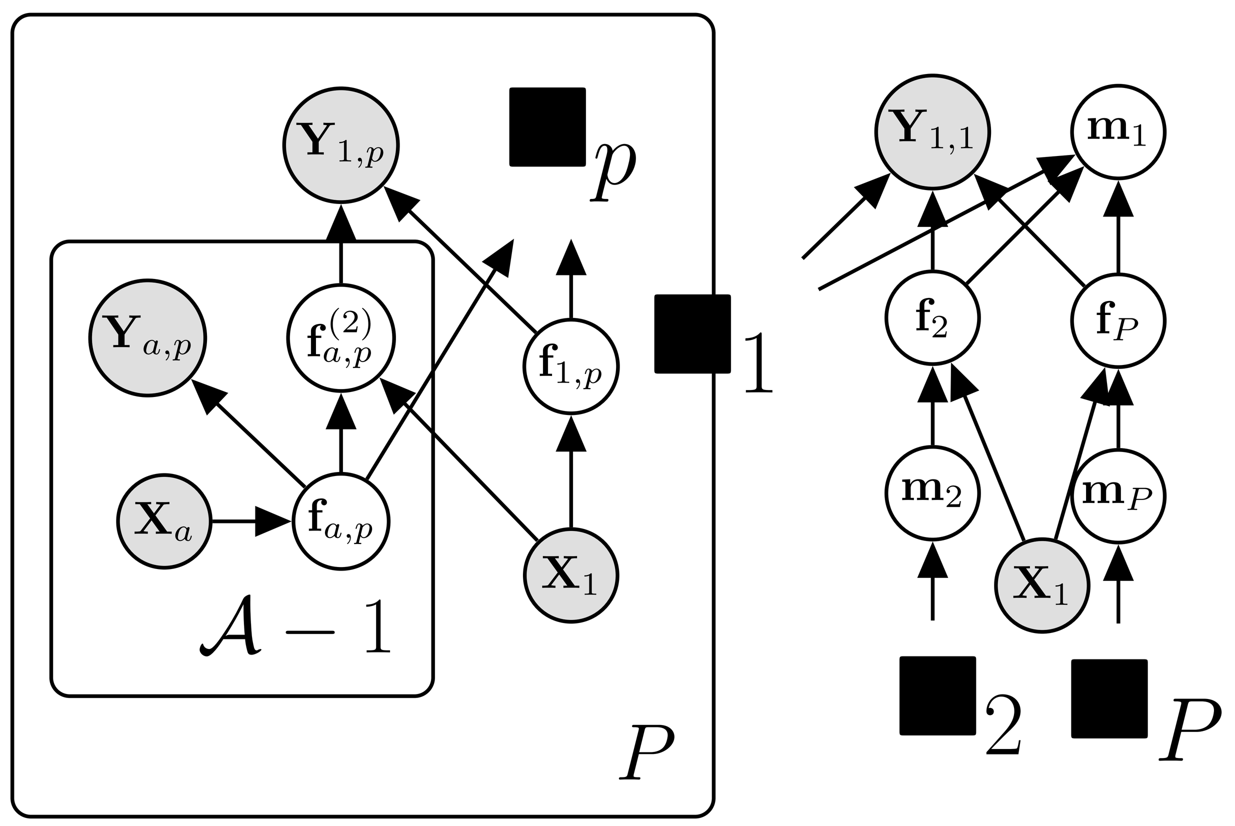
MR-GPRN



We construct sparse Gaussian processes and use variational inference to learn our models. Using a mixture of Gaussians variational posterior we have derived closed form expected log likelihood as $\mathrm{ELL}_{a,p,n,k}$,

$$\begin{aligned}\mathrm{ELL}_{a,p,n,k} &= \pi_k \log \mathcal{N}\left(Y_{a,p,n} \mid \frac{1}{|\mathcal{S}_{a,n}|}\sum_{\mathbf{x}\in\mathcal{S}_{a,n}}\sum_{q=1}^{Q}\boldsymbol{\mu}^{(w)}_{k,p,q}(\mathbf{x})\boldsymbol{\mu}^{(f)}_{k,q}(\mathbf{x}), \sigma^2_{a,p}\right) \\ &- \frac{\pi_k}{2\sigma^2_{a,p}}\frac{1}{|S_{a,n}|^2}\sum_{q=1}^{Q}\sum_{\mathbf{x}_1,\mathbf{x}_2}\boldsymbol{\Sigma}^{(w)}_{k,p,q}\boldsymbol{\Sigma}^{(f)}_{k,q} + \boldsymbol{\mu}^{(f)}_{k,q}(\mathbf{x}_1)\boldsymbol{\Sigma}^{(w)}_{k,p,q}\boldsymbol{\mu}^{(f)}_{k,q}(\mathbf{x}_2)\boldsymbol{\mu}^{(w)}_{k,p,q}(\mathbf{x}_1)\boldsymbol{\Sigma}^{(f)}_{k,q}\boldsymbol{\mu}^{(w)}_{k,p,q}(\mathbf{x}_2)\end{aligned} \tag{1}$$

MR-DGP



Modelling NO2 Across London

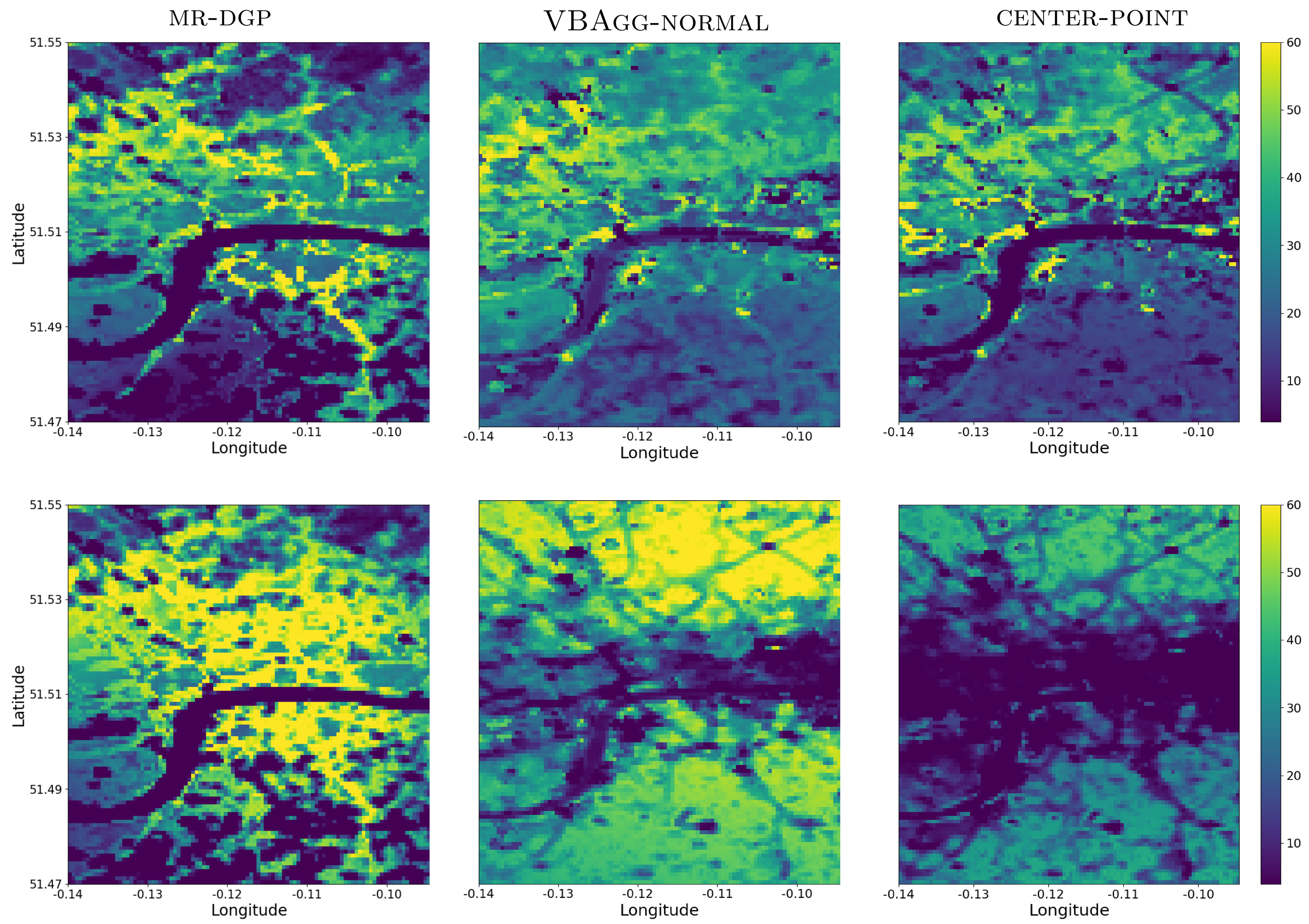


Figure 1: Spatio-temporal estimation and forecasting of NO_2 levels in London. **Top Row**: Spatial slices from MR-GPRN, VBAgg-normal and center-point respectively at 19/02/2019 11:00:00 using observations from both LAQN and the satellite model (low spatial resolution). **Bottom Row**: Spatial slices at the base resolution from the same models at 19/02/2019 17:00:00 where *only* observations from the satellite model are present.

Biased and Dependent Observations

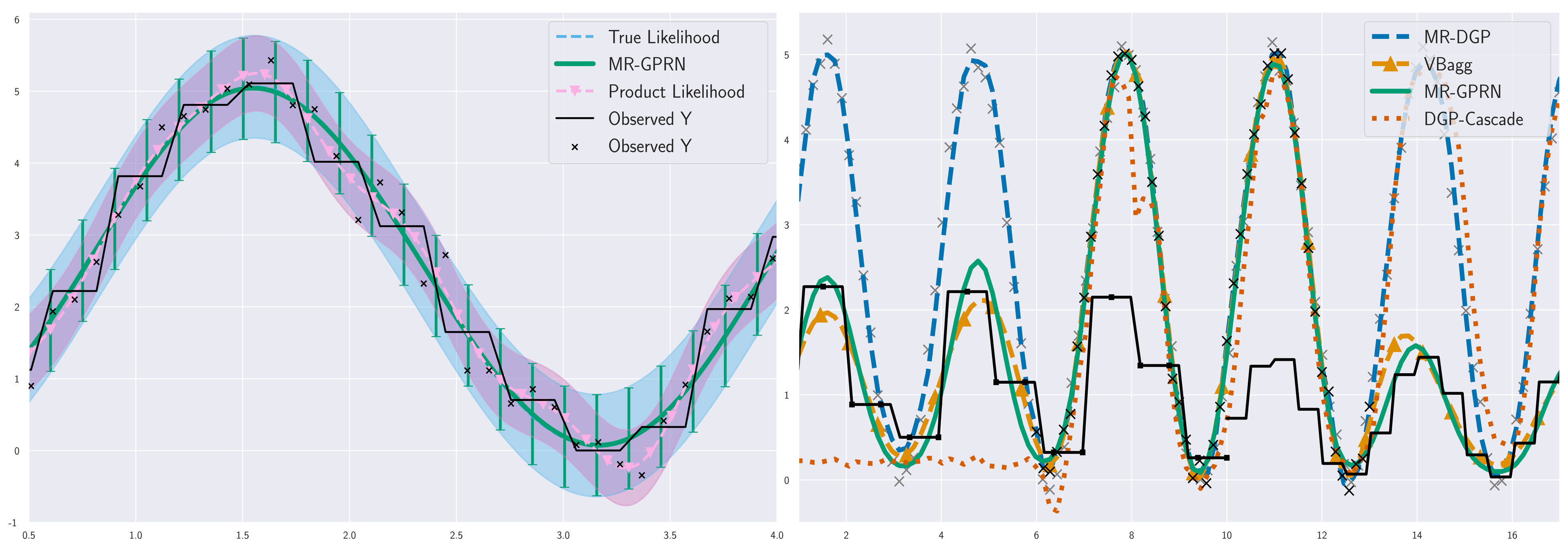


Figure 2: **Left**: MR-GPRN recovers the true predictive variance whereas assuming a product likelihood assumption leads to posterior contraction. **Right**: MR-DGP recovers the true predictive mean under a multi-resolution setting with scaling biases. Both VBAgg-normal and MR-GPRN fail as they propagate the bias. Black crosses and lines denote observed values. Grey crosses denote observations removed for testing.

Comparisons

Model	RMSE	MAPE
Single GP	20.55 ± 9.44	0.8 ± 0.16
center-point	18.74 ± 12.65	0.65 ± 0.21
VBAgg-normal	16.16 ± 9.44	0.69 ± 0.37
MR-GPRN w/o CL	12.97 ± 9.22	0.56 ± 0.32
MR-GPRN w CL	11.92 ± 6.8	0.45 ± 0.17
MR-DGP	$\mathbf{6.27 \pm 2.77}$	$\mathbf{0.38 \pm 0.32}$

Figure 3: We find that MR-DGP is able to substantially outperform both VBAgg-normal, MR-GPRN and the baselines as it is learning the forward mapping between the low resolution satellite observations and the high resolution LAQN sensors, while handling scaling biases. This is further highlighted in the bottom of Fig. 1 where MR-DGP is able to retain high resolution structure based only on satellite observations whereas VBAgg-normal and center-point over-smooth.

2*Model	PM_{10} Resolution			
	2 Hours	5 Hours	10 Hours	24 Hours
center-point	4.67 ± 0.74	5.04 ± 0.45	5.26 ± 0.91	5.72 ± 0.91
MR-GPRN	4.54 ± 0.93	5.09 ± 1.04	4.96 ± 1.07	5.32 ± 1.14
MR-DGP	5.14 ± 1.28	4.81 ± 1.06	4.61 ± 1.43	5.42 ± 1.15

Figure 4: *Inter*-task multi-resolution. Missing data predictive MSE on PM_{25} from MR-GPRN, MR-DGP and baseline center-point for 4 different aggregation levels of PM_{10}. VBAgg-normal is inapplicable in this experiment as it is a single-task approach..

Key References

[1] Varin, C., Reid, N., and Firth, D. An overview of composite likelihood methods. *Statist.Sinica*, 2011.
[2] Law, H. C. L., Sejdinovic, D., Cameron, E., Lucas, T. C., Flaxman, S., Battle, K., and Fukumizu,K. Variational learning on aggregate outputs with Gaussian processes. *Advances in NeuralInformation Processing Systems (NeurIPS)*,2018.
[3] Salimbeni, H. and Deisenroth, M. Doubly stochastic variational inference for Deep Gaussian processes. *Advances in Neural Information Processing Systems 30*, 2017..
[4] Wilson, A. G., Knowles, D. A., and Ghahramani, Z. Gaussian process Regression Networks, 2012.

Acknowledgments

O. H., T. D and K.W. are funded by the Lloyd's Register Foundation programme on Data CentricEngineering through the London Air Quality project. This work is supported by The Alan TuringInstitute for Data Science and AI under EPSRC grant EP/N510129/1 in collaboration with the Greater London Authority. We would like to thank the anonymous reviewers for their feedback and LibbyRogers, Patrick O'Hara and Daniel Tait for their help on multiple aspects of this work.

CODE

PAPER

Montréal 2018