

Gaussian Processes for Monitoring Air-Quality in Kampala

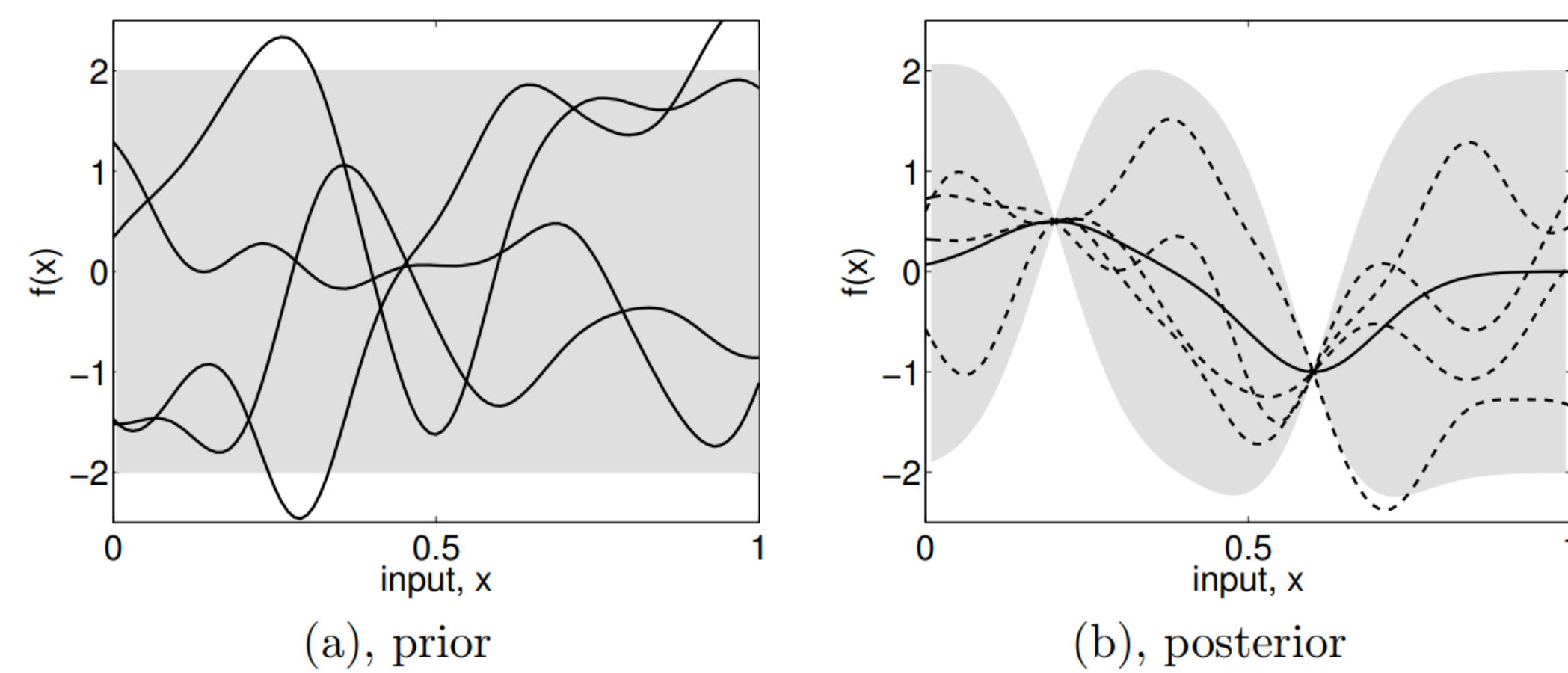
Imperial College
London



Clara Stoddart¹, Lauren Shrack², Richard Sserunjogi³, Usman Abdul-Ganiy³, Engineer Bainomugisha³, Deo Okure³, Ruth Misener¹, Jose Pablo Folch¹, Ruby Sedgwick¹

Imperial College London¹, Massachusetts Institute of Technology², AirQo / Makerere University³

Gaussian Processes



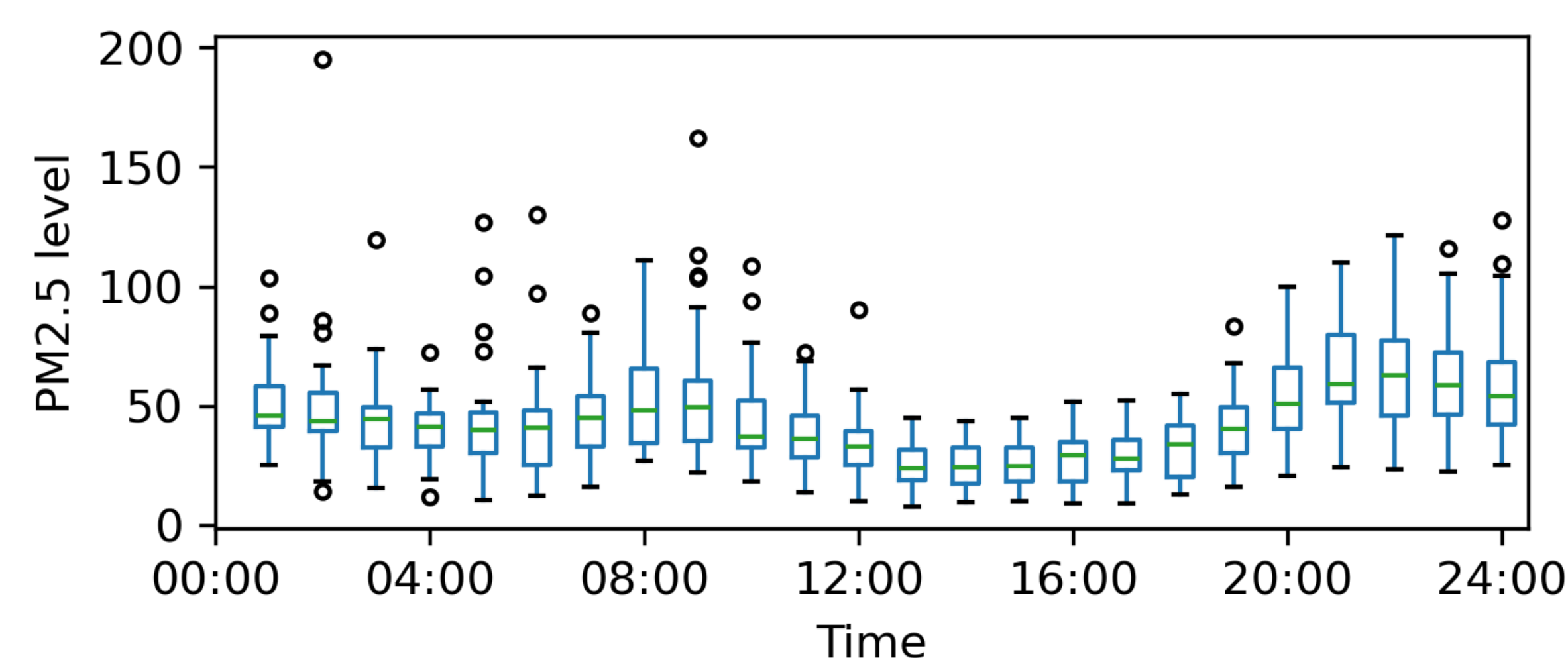
- Gaussian processes are a form of supervised machine learning where we seek to model a probability distribution over an infinite number of potential functions.
- A Gaussian process is specified by its **mean function**, which defines our predicted function value at any point of the input space, and a **covariance function**, which defines the covariance between any two points.

$$f(\mathbf{x}) \sim \mathcal{GP}(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')) \quad (1)$$

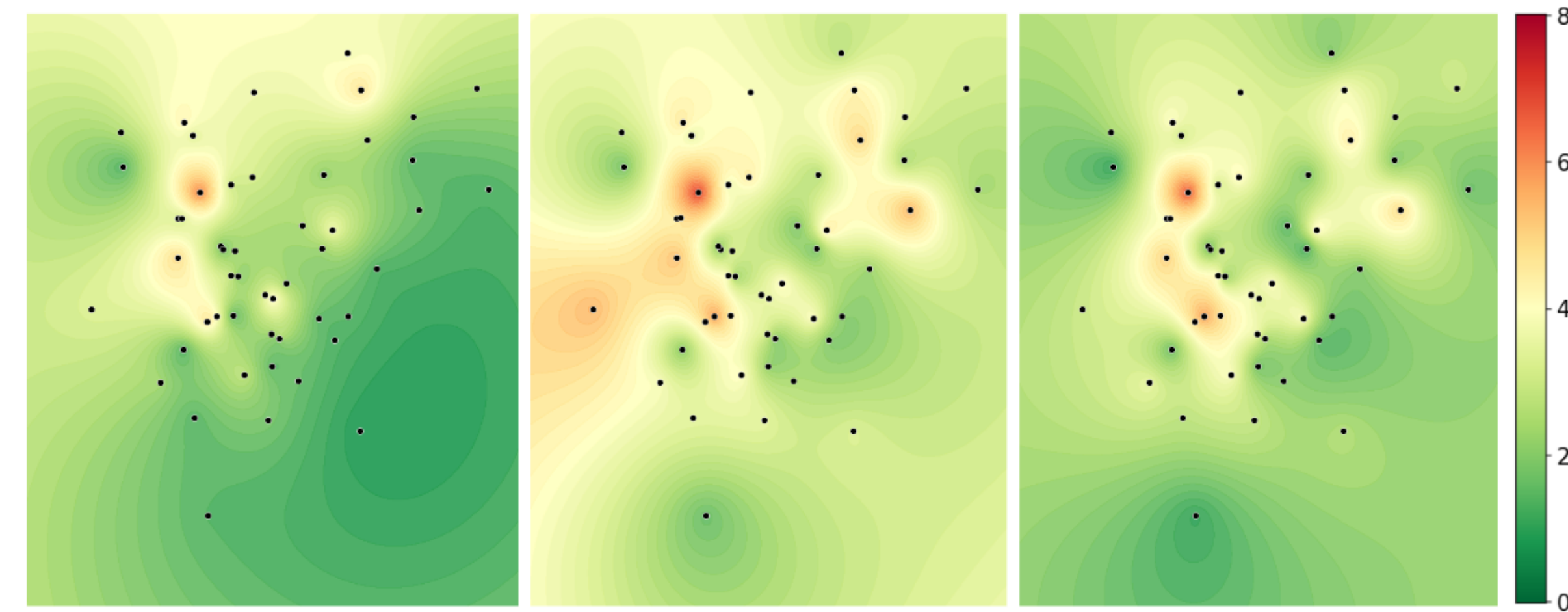
- Gaussian processes applied to a data set of size N , have a complexity of $O(N^3)$ due to matrix inversion, making it difficult for them to scale well.
- Sparse GPs are based on finding a small set of M inducing points on which to base computation. Using these inducing variables facilitates the reduction of the time complexity from $O(N^3)$ to $O(NM^2)$
- In our work, we compare the performance and efficiency of two sparse approximations, namely Stochastic Variational Gaussian Processes developed by Hensman et al. [2] and Spatio-Temporal Variational Gaussian Processes developed by Hamelijnck et al. [1].

Predicting air quality in Kampala

- Using data provided by AirQo, our aim was to develop a GP model that could predict PM2.5 levels in Kampala based on time and location. Below we show how PM2.5 levels vary over November 2021 for the site Makerere University Weather Station:



We propose a Gaussian Process model that is able to predict air quality in Kampala based on temporal and spatial inputs



Examples of learnt models. From left to right, the actual PM2.5 levels, our SVGP model's predictions and the ST-SVGP model's.

Model adaptations

We investigated the effect of a range of adaptations to the model:

1. **Periodic versus non-periodic kernels:** We assessed the impact of using a periodic kernel for our temporal inputs, against using standard squared-exponential kernels for all inputs.
2. **Forecasting versus nowcasting:** We compared our model against forecasting future PM2.5 levels at locations where we had air quality sensors, and nowcasting PM2.5 levels at locations between sensors where we had data.
3. **Additional variables:** We evaluated the impact of additional meteorological information including but not limited to wind speed, precipitation and temperature.
4. **Sparse GPs:** We compared standard GPs, against two sparse variations including the Stochastic Variational Gaussian Process model [2], and the Spatio-Temporal Sparse Variational Gaussian Process model [1].

Evaluation method

Nowcasting: We treated each site as a hold out site and trained the model on the rest of the 65 sites across Kampala. This left us with 66 RMSEs for each site, which we took the mean of to assess overall model performance.

Forecasting: We trained the model on a subsample of all data up to but excluding that of the 30th of November, and then used readings from this day as our test data.

Results

Periodic	Outliers Removed	Additional Inputs	Sparse	Min MSE	Average RMSE	Max RMSE
✗	✗	✗	✗	17.12	27.65	86.36
✓	✗	✗	✗	11.03	25.49	87.26
✓	✓	✗	✗	15.49	25.06	86.01
✓	✓	✓	✗	12.98	24.94	86.50
✓	✓	✓	SVGP	11.25	23.63	85.86
✗	✓	✗	ST-SVGP	9.01	22.34	87.04

Table 1: Results for nowcasting task

Periodic	Outliers Removed	Additional Inputs	Sparse	Min MSE	Average RMSE	Max RMSE
✗	✗	✗	✗	10.52	22.99	141.77
✓	✗	✗	✗	9.65	17.75	44.27
✓	✓	✗	✗	5.31	15.66	27.18
✓	✓	✓	✗	5.89	15.17	30.15
✓	✓	✓	SVGP	2.86	14.57	27.06
✗	✓	✗	ST-SVGP	5.56	15.34	32.85

Table 2: Results for forecasting task

Tables 1 and 2 show the results of our investigations, showing how with each adaptation we improved our model.

- We found that removing outliers is very important for forecasting.
- The final iteration of the ST-SVGP model outperformed the SVGP model at nowcasting, but was slightly worse than the SVGP model at forecasting. We conjecture that improved nowcasting would require non-stationary kernels and more sensors.

References

- [1] Oliver Hamelijnck, William Wilkinson, Niki Loppi, Arno Solin, and Theodoros Damoulas. Spatio-temporal variational gaussian processes. *Advances in Neural Information Processing Systems*, 34:23621–23633, 2021.
- [2] Michalis Titsias. Variational Learning of Inducing Variables in Sparse Gaussian Processes. In *Proceedings of the Twelfth International Conference on Artificial Intelligence and Statistics*, volume 5, pages 567–574, 2009.

Acknowledgements:

Acknowledgements to the Wehubit programme implemented by Enabel and to Google LLC, the EPSRC (StatML) CDT (grant no. EP/S023151/1), BASF SE, Ludwigshafen am Rhein, the UKRI AI4Health CDT (grant no. EP/S023283/1) and the AI for Social Good Dagstuhl seminar (no. 22091).

Link to paper:

