Accurate and Scalable Gaussian Processes for Fine-grained Air Quality Inference

Zeel B Patel, Palak Purohit, Harsh Patel, Shivam Sahni, Nipun Batra

IIT Gandhinagar

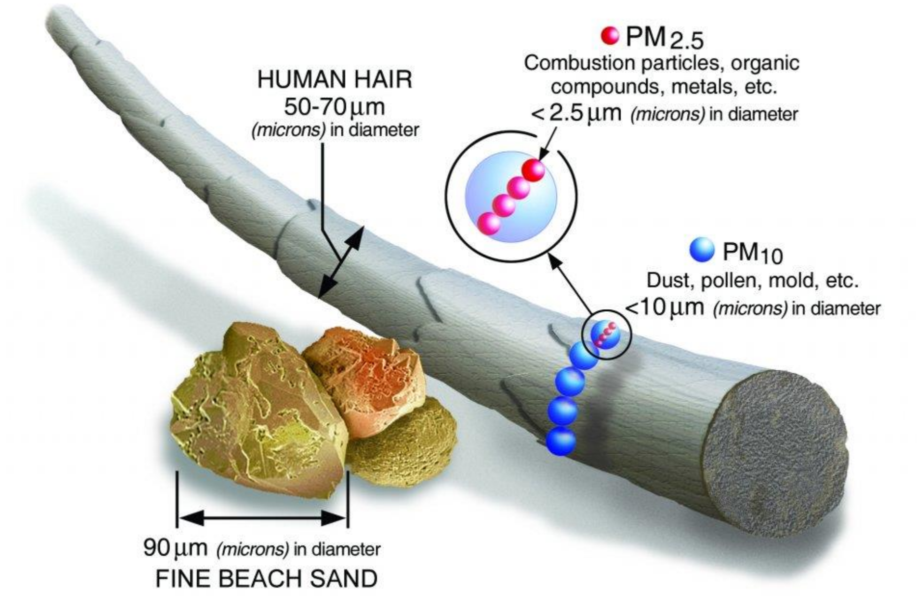
AAAI 2022

Motivation

- 8 M deaths per annum worldwide: WHO
- 90% of the world breathes
  low quality air
- Sparse and non-uniform AQ stations
  (India has 573 over the demand of 4K)
- Stations are costly

src:https://indianexpress.com/article/india/at-2-5-million-india-tops-list-of-pollution-linked-deaths-study-4898337/

Particulate Matter : 2.5 microns ($PM_{2.5}$)



src:https://www.epa.gov/pm-pollution/particulate-matter-pm-basics

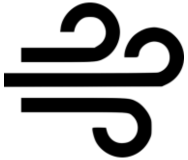
src:https://mypositiveparenting.org/2019/11/21/air-pollution-and-health/

Problem: Air Quality Inference

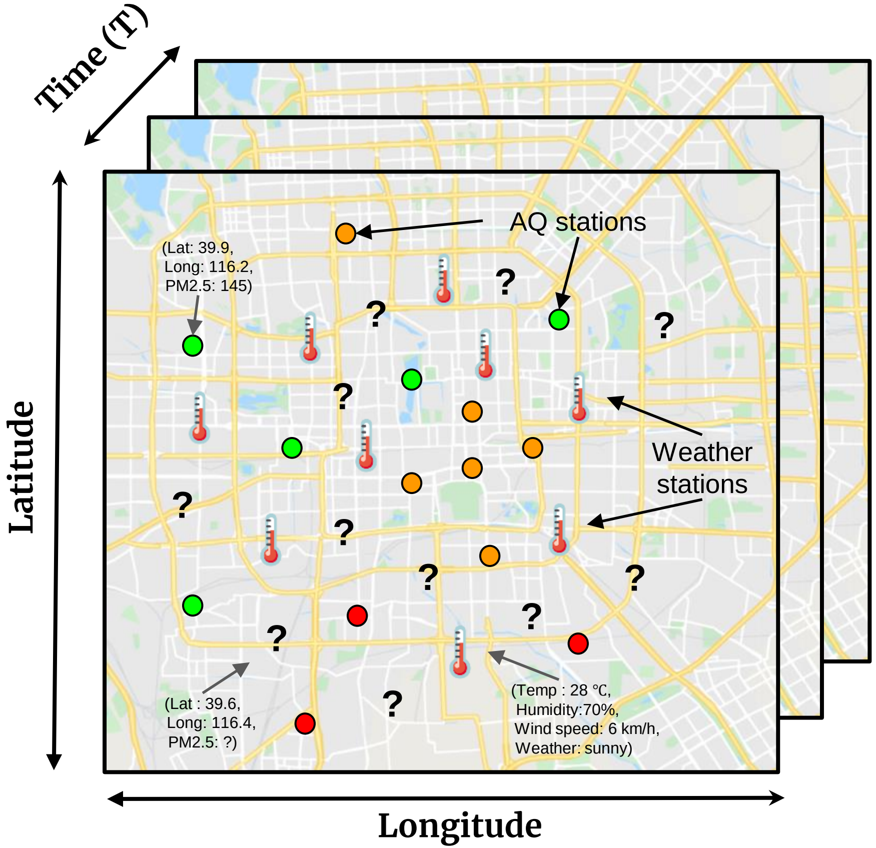
Temperature

Humidity

Wind

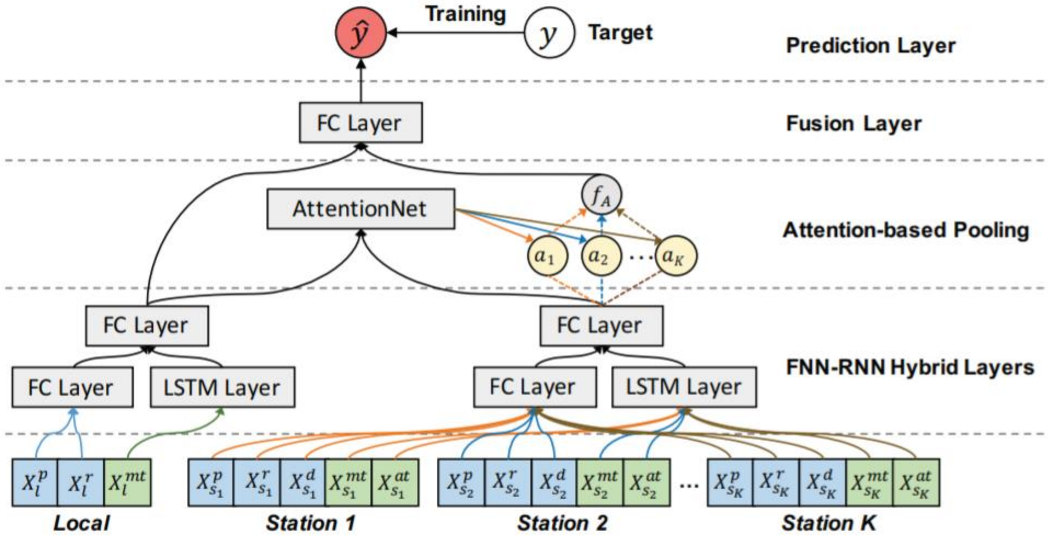


Weather



Related work

Attentional Deep Air Quality Inference Network

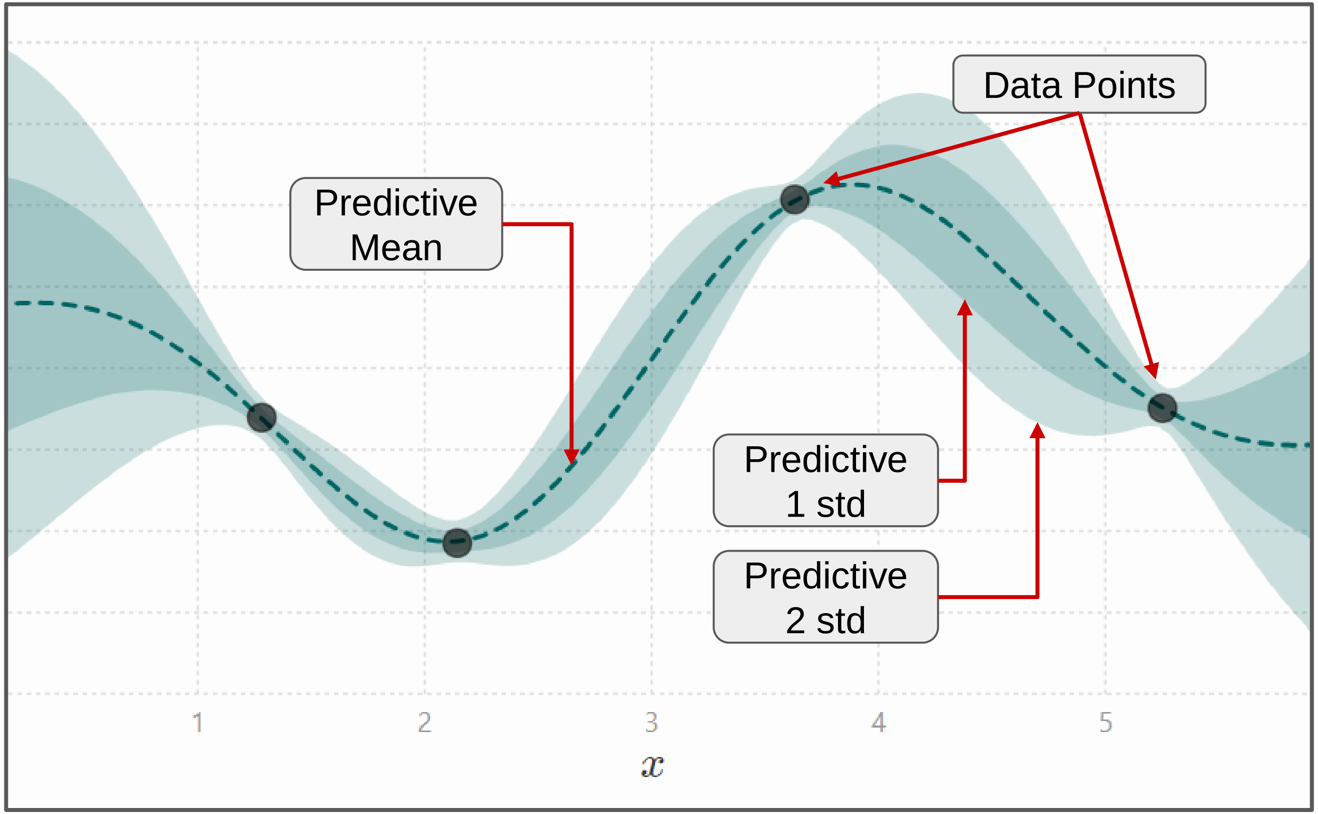


src: Cheng, W., Shen, Y., Zhu, Y., & Huang, L. (2018). A Neural Attention Model for Urban Air Quality Inference: Learning the Weights of Monitoring Stations. Proceedings of the AAAI Conference on Artificial Intelligence, 32(1).

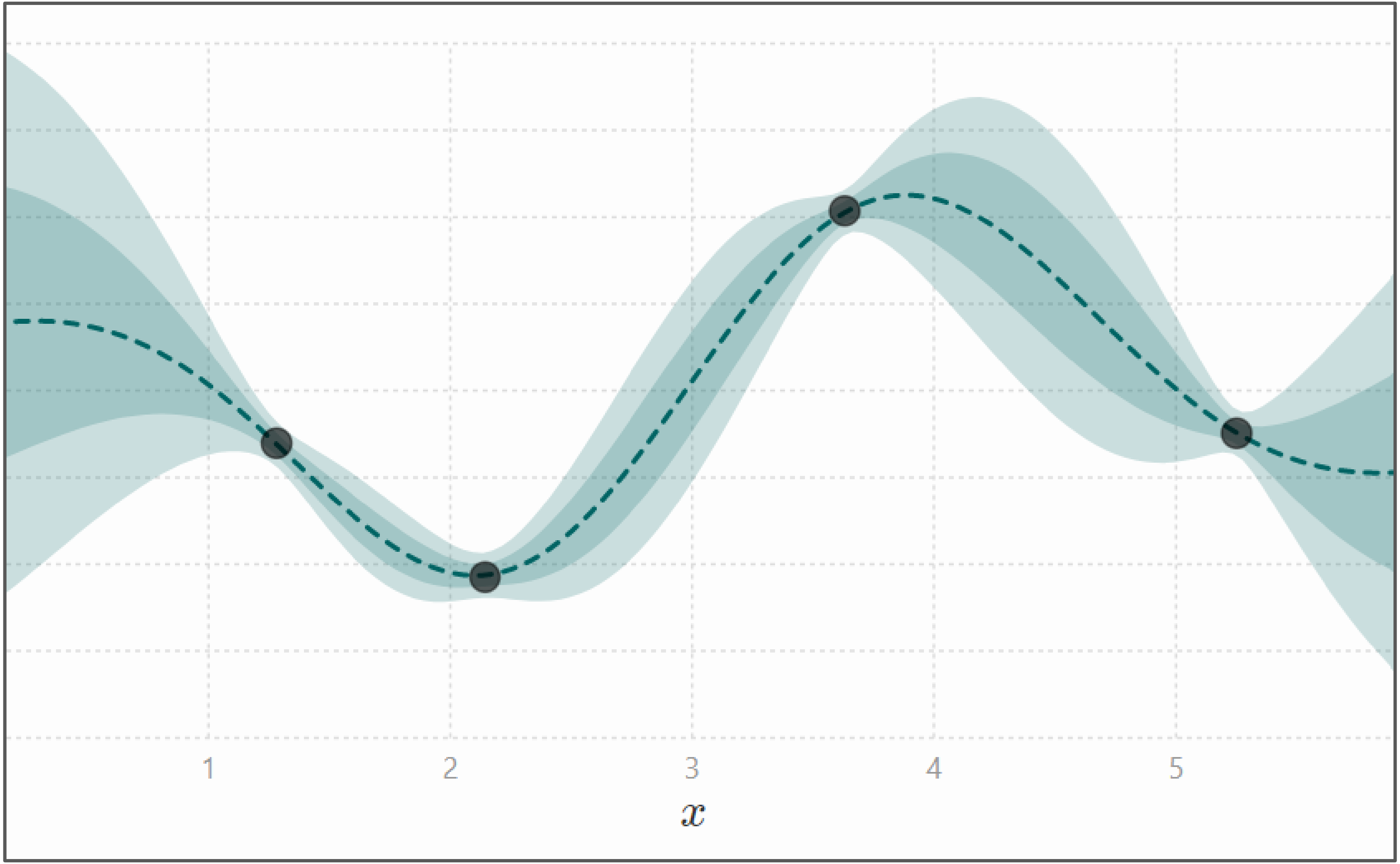
Shortcomings

- No scope for domain knowledge
- No uncertainty provision

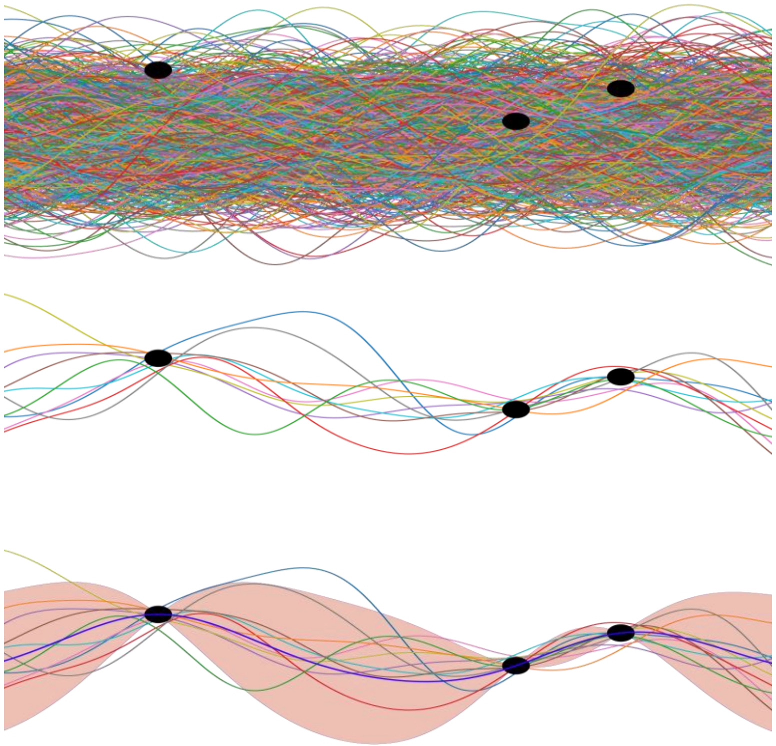
Our Approach: Gaussian Process (GP) Regression



src: http://www.infinitecuriosity.org/vizgp/


1
2
3
4
5
x

Gaussian Process Regression : Intuition



src: http://inverseprobability.com/talks/notes/gaussian-processes.html

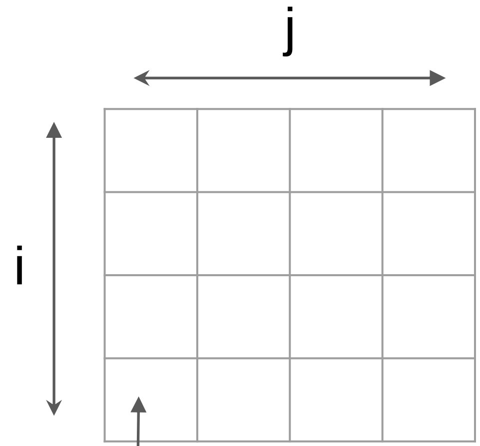
Gaussian Process Regression - Prior

$$k(X, X)$$

$$p(\mathbf{f}|X, \boldsymbol{\theta}) \sim \mathcal{N}(\mathbf{f}|\mathbf{0}, k(X, X))$$

Mean vector

Cov matrix



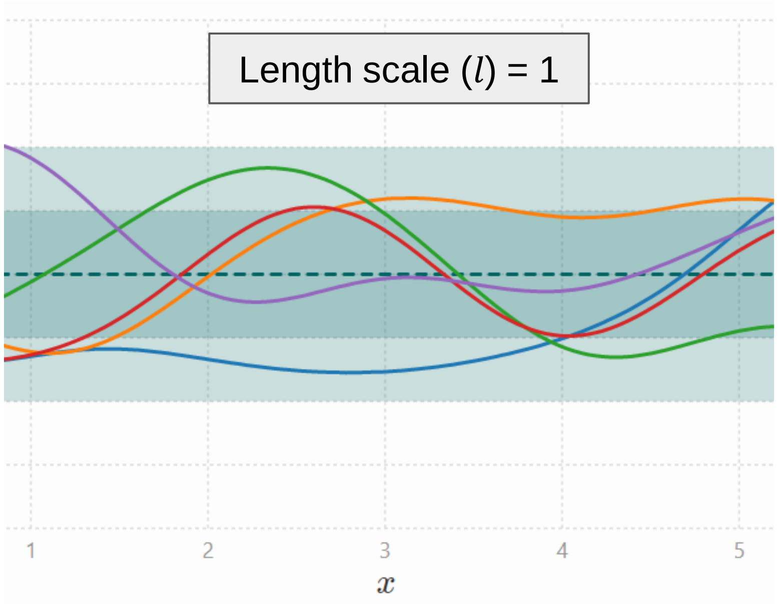
Cov function

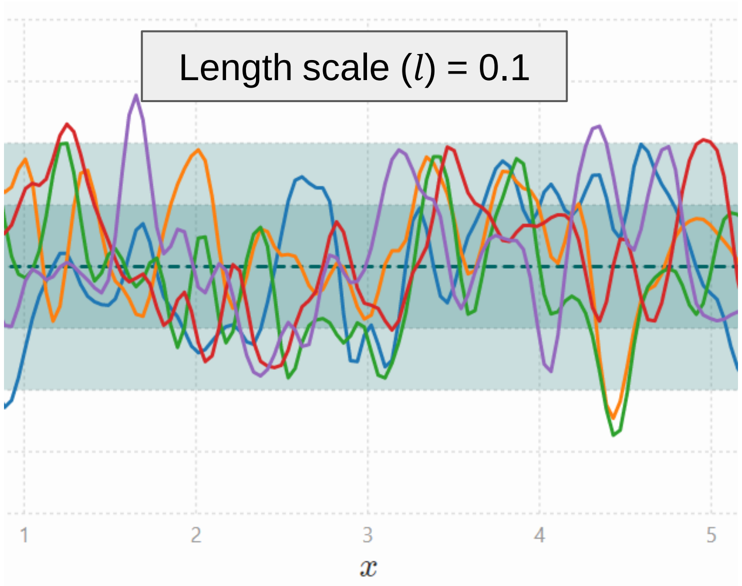
$$k(\mathbf{x}_i, \mathbf{x}_j)$$

$$\mathbf{x}_i = (Lat_i, Long_i, Temp_i, ...)$$

Domain Inspired Kernels (Covariance Functions)

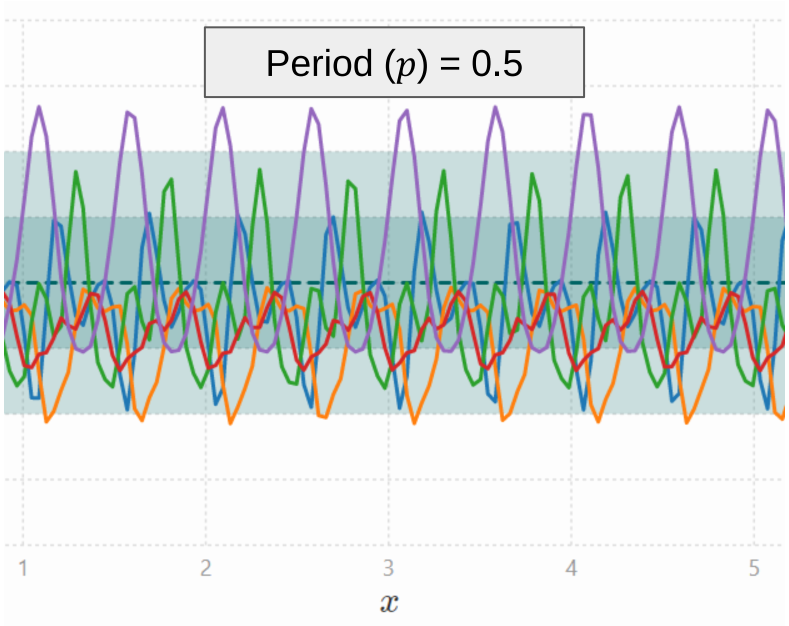
Squared Exponential Kernel $\longrightarrow K(\mathbf{x}, \mathbf{x}') = \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}'\|^2}{2l^2}\right)$

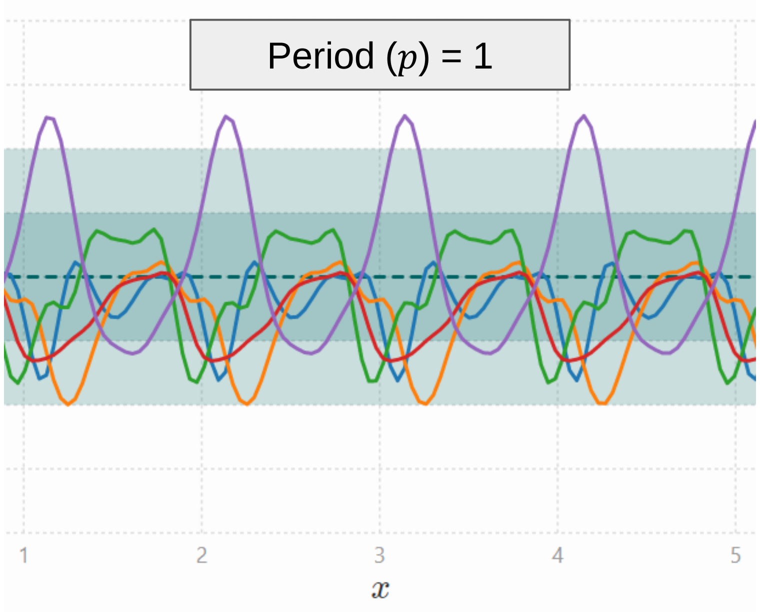




Domain Inspired Kernels (Covariance Functions)

Periodic kernel $\longrightarrow K(\mathbf{x}, \mathbf{x}') = \exp\left(-\frac{2\sin^2(\pi||\mathbf{x} - \mathbf{x}'||/p)}{l^2}\right)$

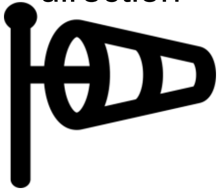




Domain Inspired Kernels (Covariance Functions)

Categorical features

Wind direction



Weather

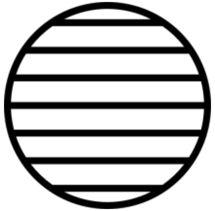
Temporal Features

Time

Continuous Features

Longitude

Latitude



Temperature

Humidity

Domain Inspired Kernels (Covariance Functions)

- Type: Categorical

Sample	Wind direction
1	North (N)
2	East (E)
3	South-West (SW)

One Hot Encoding

Sample	N	SW	E
1	1	0	0
2	0	0	1
3	0	1	0

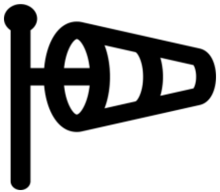
Hamming distance kernel[1] $\longrightarrow k(x, x') = \exp\left(-\frac{\mathbb{I}_{x \neq x'}}{2\ell^2}\right)$

[1]Hutter, F.; Xu, L.; Hoos, H. H.; and Leyton-Brown, K. 2014. Algorithm runtime prediction: Methods & evaluation. Artificial Intelligence, 206: 79–111

Domain Inspired Kernels (Covariance Functions)

Categorical Features

Wind speed



Weather

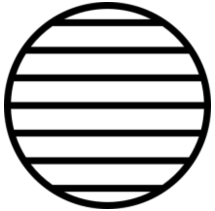
Temporal Features

Time

Continuous Features

Longitude

Latitude



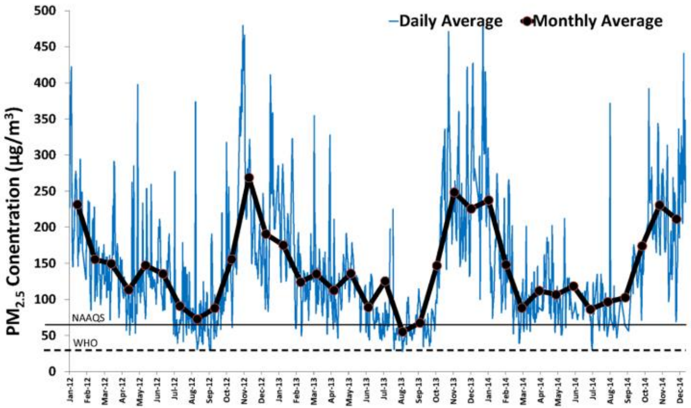
Temperature

Humidity

src: Flaticon.com

Domain Inspired Kernels (Covariance Functions)

- Features: Time



src:https://urbanemissions.info/delhi-india/delhi-ambient-monitoring-data-timeseries/

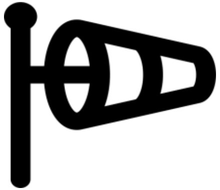
- Kernel: Periodic?

- Kernel: (Sq. Exp.) ✕ (Periodic)

Domain Inspired Kernels (Covariance Functions)

Categorical Features

Wind speed



Weather

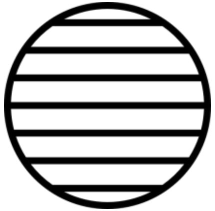
Temporal Features

Time

Continuous Features

Longitude

Latitude



Temperature

Humidity

src: Flaticon.com

Domain Inspired Kernels (Covariance Functions)

- Features: Longitude, Latitude, Temperature, Humidity

Squared Exponential Kernel $\longrightarrow K(\mathbf{x}, \mathbf{x}') = \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}'\|^2}{2l^2}\right)$

ARD (Automatic Relevance Determination)

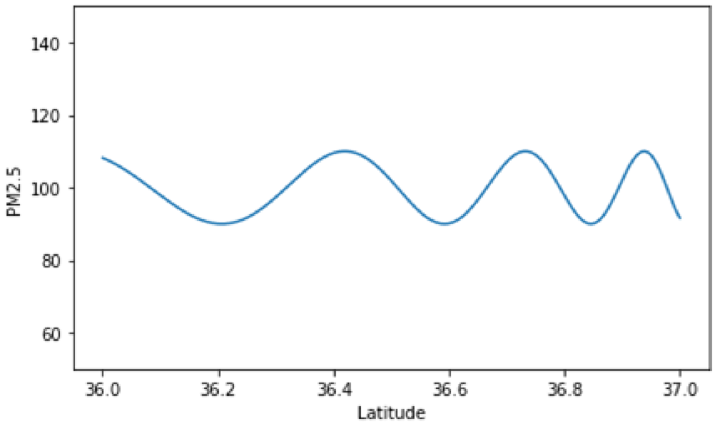
Non-ARD Kernel → $$K(\mathbf{x}, \mathbf{x}') = \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}'\|^2}{2l^2}\right)$$

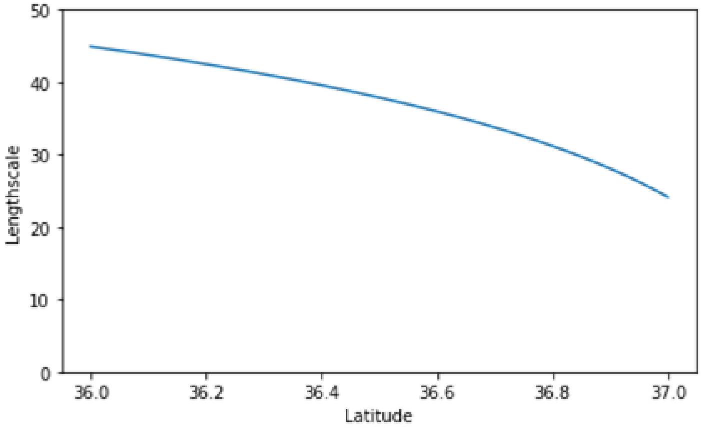
ARD Kernel → $$K(\mathbf{x}, \mathbf{x}') = \exp\left(-\sum_{i=1}^{|\mathbf{x}|} \frac{(x_i - x_i')^2}{2l_i^2}\right)$$

Domain Inspired Kernels (Covariance Functions)

- Features: Longitude, Latitude, Temperature, Humidity

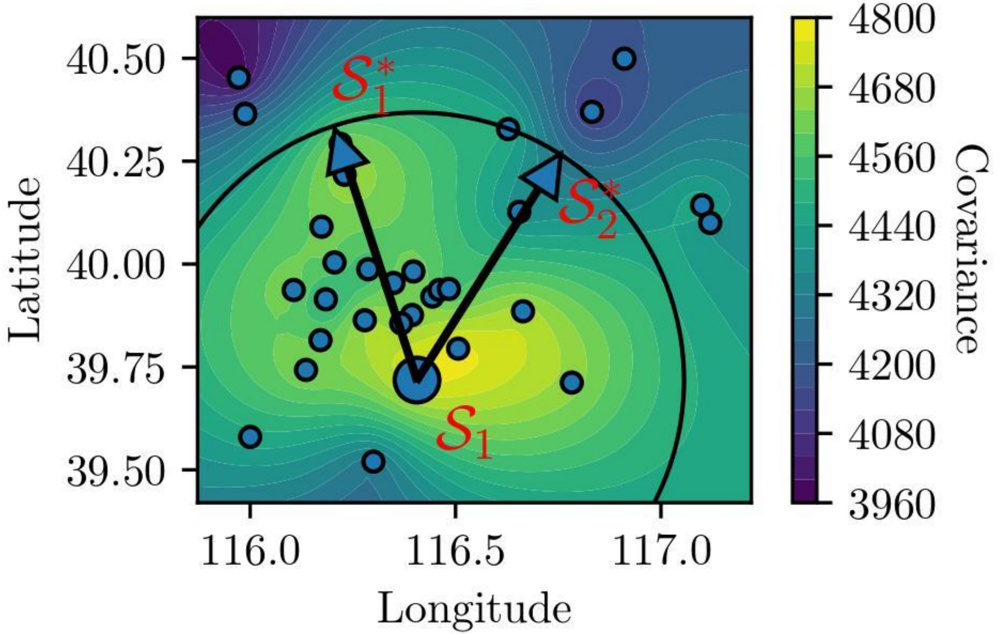
Squared Exponential Kernel → $K(\mathbf{x}, \mathbf{x}') = \exp\left(-\sum_{i=1}^{|\mathbf{x}|} \frac{(x_i - x_i')^2}{2l_i^2}\right)$





Checking Covariance Non-stationarity

Cov($\boldsymbol{s}_1$, $\cdot$)



Non-stationary GP

- Plagemann et al.[1]

$$K_{NS}(\mathbf{x},\mathbf{x}') = \sqrt{\frac{2 l_{\mathbf{x}} l_{\mathbf{x}'}}{l_{\mathbf{x}}^2 + l_{\mathbf{x}'}^2}} \cdot \exp\left(-\frac{||\mathbf{x}-\mathbf{x}'||^2}{l_{\mathbf{x}}^2 + l_{\mathbf{x}'}^2}\right)$$

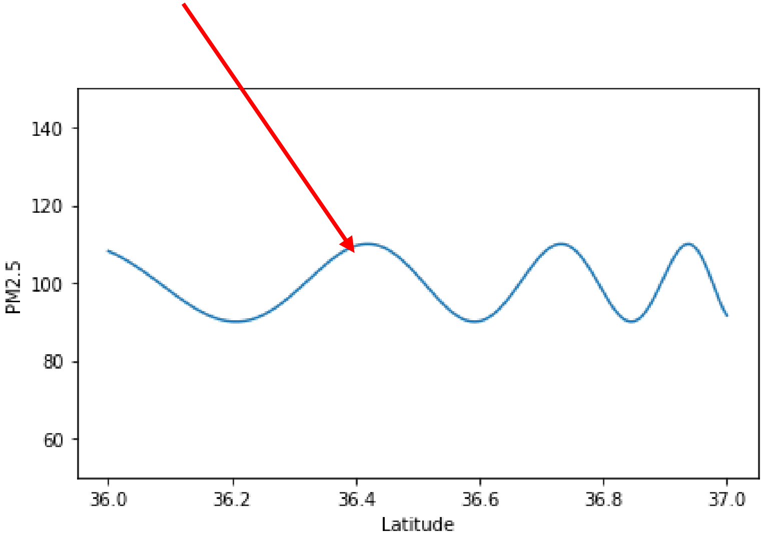
Input depended length scales

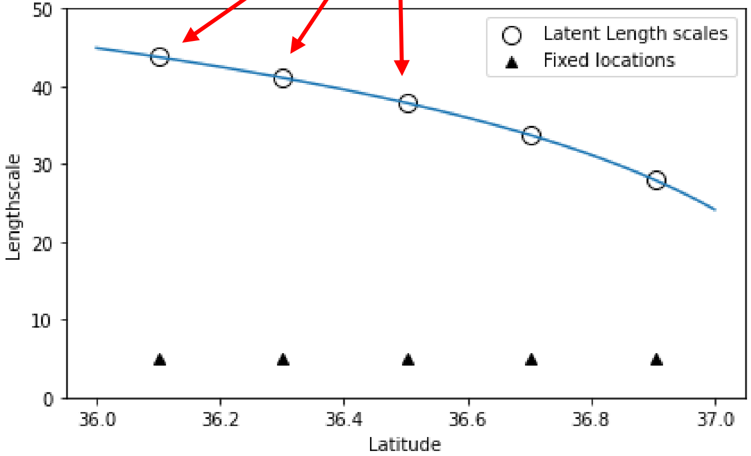
1. Plagemann, Christian, Kristian Kersting, and Wolfram Burgard. "Nonstationary Gaussian process regression using point estimates of local smoothness." Joint European Conference on Machine Learning and Knowledge Discovery in Databases. Springer, Berlin, Heidelberg, 2008.

Non-stationary GP

$$\mathcal{GP}_y \sim \mathcal{N}(\mathbf{0}, K_{NS})$$

$$f(\mathbf{x}) = l_{\mathbf{x}}, \mathcal{GP}_l \sim \mathcal{N}(\mathbf{0}, K_S)$$





Scalability

- Cost of GP training: O(n^3), memory required: O(n^2)
- Cost of Batch GP training: O(nm^2), memory required: O(m^2), m: batch size

Chen, Hao, et al. "Stochastic gradient descent in correlated settings: A study on gaussian processes." Advances in Neural Information Processing Systems 33 (2020).

Algorithm 1: Minibatch SGD

Input: $\boldsymbol{\theta}^{(0)} \in \mathbb{R}^2$, initial step size $\alpha_1 > 0$.
for $k = 1, 2, \ldots, K$ **do**
  Randomly sample a subset of indices ξ_k of size m;
  Compute the stochastic gradient $g(\boldsymbol{\theta}^{(k)}; \mathbf{X}_{\xi_k}, \mathbf{y}_{\xi_k})$;
  $\alpha_k \leftarrow \frac{\alpha_1}{k}$;
  $\boldsymbol{\theta}^{(k)} \leftarrow \boldsymbol{\theta}^{(k-1)} - \alpha_k g(\boldsymbol{\theta}^{(k-1)}; \mathbf{X}_{\xi_k}, \mathbf{y}_{\xi_k})$;
end for

Data & Experimental setup

- Hourly granularity
- Datasets
  - Beijing[1] : 36 stations : March 2015
  - London[2] (KDD Cup 18) : 24 stations : May 2017
- Missing data is filled temporally
- Data scaled to mean 0, std 1
- K-Fold cross-validation

[1]Cheng, W., Shen, Y., Zhu, Y., & Huang, L. (2018). A Neural Attention Model for Urban Air Quality Inference: Learning the Weights of Monitoring Stations. Proceedings of the AAAI Conference on Artificial Intelligence, 32(1).

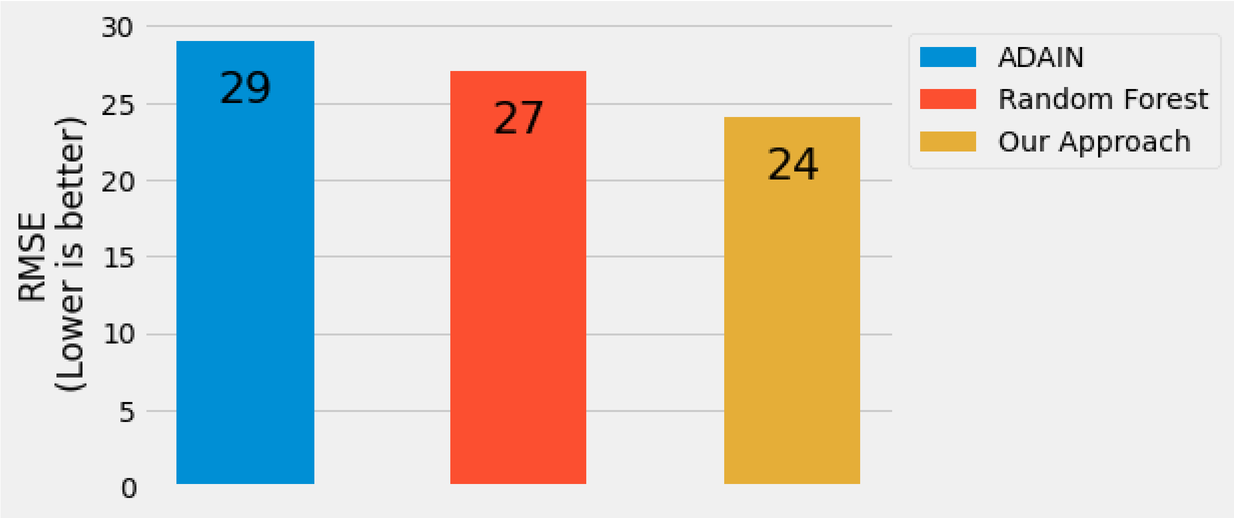
[2]https://www.kdd.org/kdd2018/kdd-cup

Baselines

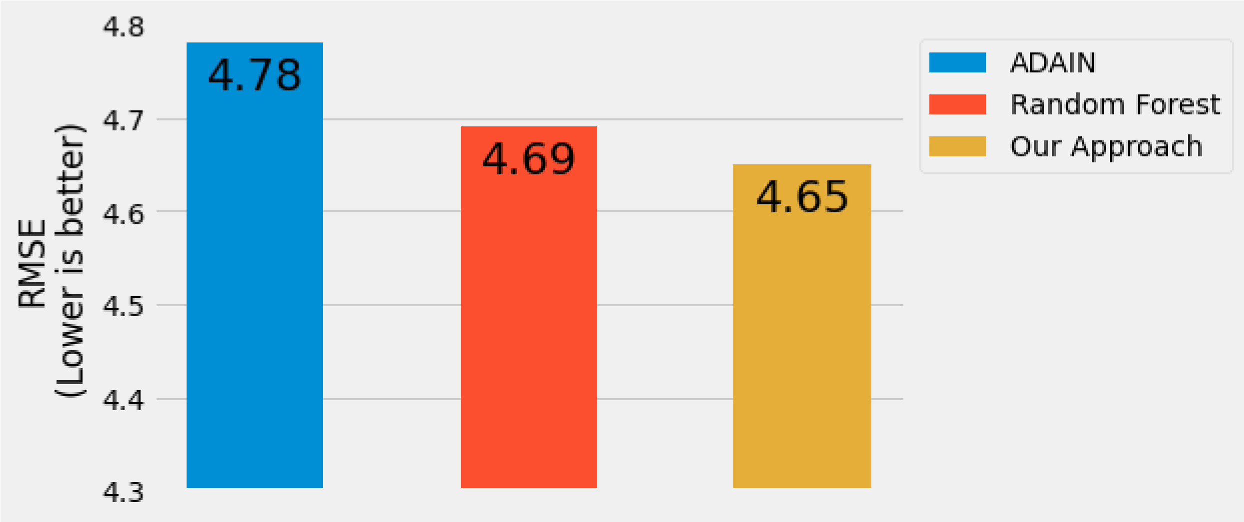
- Random Forest Regressor
- IDW (Inverse Distance Weighting)
- XGBoost
- K-Nearest Neighbors
- ADAIN (Attentional Deep Air Quality Inference Network)[1] - AAAI ‘18

[1]Cheng, W., Shen, Y., Zhu, Y., & Huang, L. (2018). A Neural Attention Model for Urban Air Quality Inference: Learning the Weights of Monitoring Stations. Proceedings of the AAAI Conference on Artificial Intelligence, 32(1)

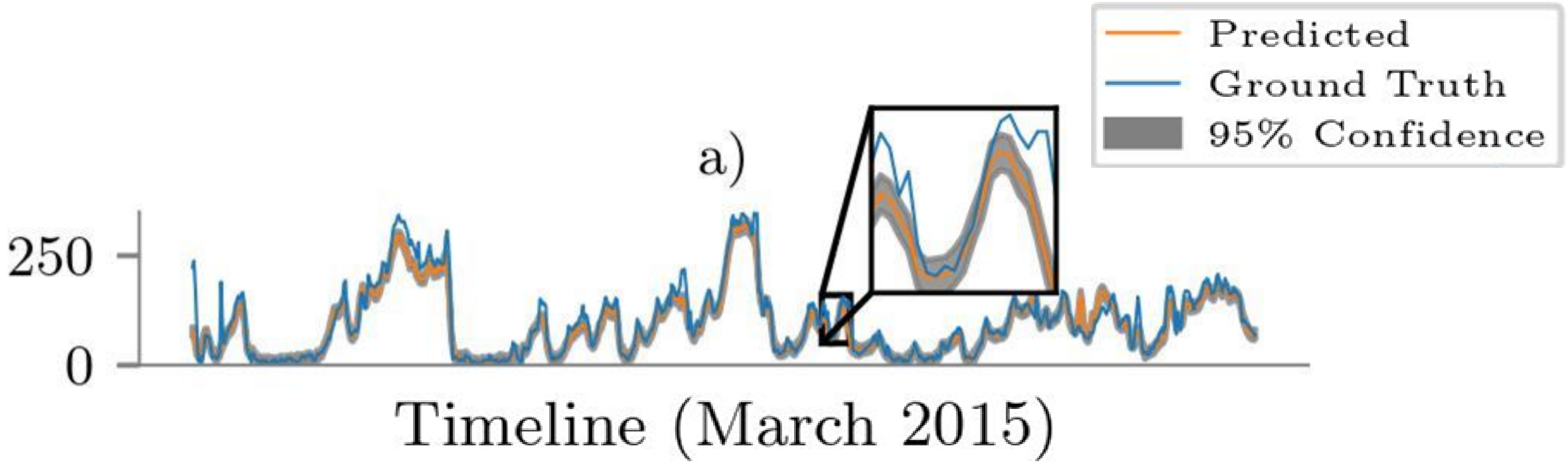
Results - Beijing dataset



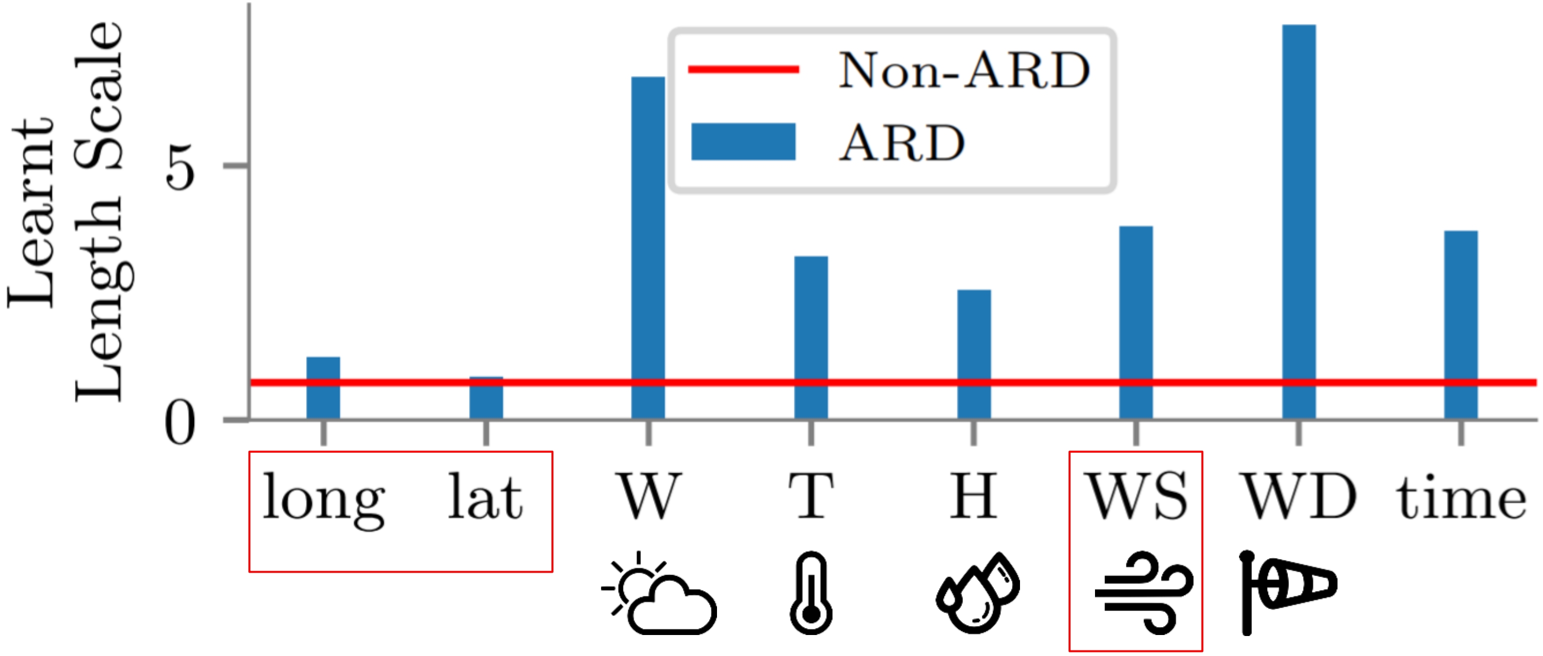
Results - London dataset



Results - Predictions



Results - Effect of Automatic Relevance Determination



Future work

src:https://ecolink.com/info/six-common-air-pollutants/

Future work

src:https://www.google.com/earth/outreach/special-projects/air-quality/

Summary

- Domain inspired and uncertainty aware Gaussian process model for air quality inference.
- Domain inspired kernels
- Non-stationarity
- Scalable training
- 17% improvement over state-of-the-art