Spatially Adaptive $PM_{2.5}$ Estimation in Low-sensor Regions using Variational Gaussian Processes

Shrey Gupta
Boston College
shrey.gupta@bc.edu

Avani Wildani
Cloudflare
avani@cloudflare.com

Yang Liu
Emory University
yang.liu@emory.edu

***Abstract*—Air pollution, particularly particulate matter 2.5 ($PM_{2.5}$), poses a significant public health challenge in densely populated developing regions. Moreover, deploying an extensive ground sensor network to monitor $PM_{2.5}$ accurately is economically unfeasible in such regions. To address this problem, we utilize Sparse Variational Gaussian Process (SVGP) models to generate approximate data using the limited ground sensor data. Since SVGPs use computational approximators for Gaussian Process modeling, we hypothesize that their inducing points can be trained to adapt spatially, i.e., these points, when optimized, can spread over the region of interest. Hence, well-initialized inducing points allow SVGPs to model $PM_{2.5}$ data by capturing spatial variations of the region. We evaluate our hypothesis using $PM_{2.5}$ data from Lima, Peru, one of the most polluted cities in the Americas, and with very few $PM_{2.5}$ ground sensors. Our experiments qualitatively validate our hypothesis of spatial adaptation and provide a quantitative justification of improved performance over the baseline models.**

***Index Terms*—$PM_{2.5}$ Modeling, Sparse Variational Gaussian Processes**

I. Introduction

Air pollution poses a significant public health concern, with its impact most severe in densely populated regions [1]. Among various pollutants, the *fine particulate matter* ($PM_{2.5}$) constitutes the highest public health risk due to its small size ($<$ 2.5 microns), allowing it to penetrate the lungs as well as enter the bloodstream [2]. Therefore, accurate monitoring of $PM_{2.5}$ is crucial for effective and informed policy decisions for public health [3]. Such effective monitoring can be achieved by either installing ground sensors or using satellite-derived (remote sensing) measurements. However, proximity to the ground generally improves accuracy for sensing, thereby making the ground sensors a better option.

Installing a dense network of $PM_{2.5}$ ground sensors across a large geographical area is often economically unfeasible, especially in developing regions or countries that are unwilling to invest in such critical infrastructure [4], [5]. Hence, the little data that gets collected using the sparse network of sensors in such regions generally contains spatiotemporal irregularities [6] and distribution imbalance [7], [8]. *Spatiotemporal irregularities* refer to spatial and temporal variations in the data collected from a sensing network. Spatial irregularities arise due to the sparsity of sensor locations, while temporal irregularities occur due to unsynchronized sampling at these sensors. *Distribution imbalance* refers to data points being unevenly distributed across the target variable. For example, for a sparse network region, the data collected has mostly moderate $PM_{2.5}$ concentration levels, thereby having fewer data points with high $PM_{2.5}$ concentration levels.

These non-uniformities (irregularities and imbalance) affect the underlying data structure, resulting in non-IID (non-independent and identically distributed) characteristics, i.e., distortion of the relationships between features and the target variable. Hence, the efficacy of the machine learning (ML) models is reduced as they traditionally rely on big, context-rich datasets for robust learning and prediction.

We pose two questions here: **(i) Among the extensive solutions to overcome data scarcity issues for sparse networks, can we avoid leveraging auxiliary data or sensing technologies to improve the prediction of machine learning (ML) models (self-reliant models)? (ii) Are these self-reliant models affected by the placement of ground sensors in such scenarios? Can their performance be optimized if the placement of the sensors is optimized?**

Therefore, the proposed solution should be able to overcome non-iid characteristics present in the data and improve the ML model prediction by utilizing only the data collected from the ground sensors. We propose using non-parametric ML models that allow data augmentation to solve the non-iid and data scarcity issues. Sparse Variational Gaussian Processes (SVGP), a subset of Gaussian Process models, can be used to overcome the above issues [9], [10]. A Gaussian Process is a probabilistic model that places a distribution over all functions fitting the observed data, and quantifies how likely each function is given the data [11], [12]. Such behavior is termed non-parametric because it does not assume a fixed structure of the data and, therefore, does not assign a single function to the entire data distribution. This allows them to adapt to complexities in data, such as the non-iid characteristics of the data. In SVGPSs, the *sparse* aspect addresses scalability issues as Gaussian Processes are unfit for large datasets. SVGPs overcome this by selecting a set of data points called inducing points that are representative of the entire dataset. The *variational* aspect allows them to approximate a distribution (make predictions) using these inducing points.

Hence, in this work, we propose a new perspective on Sparse Variational Gaussian Processes (SVGPs), arguing that their inducing points can serve as synthetic training data points that generalize for sparse sensor networks when approximated using variational inference.

Hence, we answer the two posed questions by hypothesizing and experimentally validating two theories associated with inducing points of SVGPs: **(i) Well-initialized inducing points spread over the sparse sensing region and adapt to the underlying data distribution.** We term this phenomenon of spreading as *spatial adaptation*. **(ii) Strategic placement of the sensors such that they have a large region of cover allows for an improved spatial adaptation.**

We evaluate these hypotheses using the $PM_{2.5}$ data of Lima, Peru, which has a sparse sensor network [13] (only 10 sensors). We validate our hypotheses by visualizing the spread of inducing points at the end of the SVGP training. Moreover, we compare the efficacy of this spread by comparing SVGPs prediction accuracy to competitive regression models. Our experiments show that the SVGP model outperforms the second-best performing model by 84% and 10% for R2 and RMSE scores, repsectively.

II. Related Works

We review prior studies that focus on ML models for predicting $PM_{2.5}$ for sparse sensor networks and Gaussian Process models applied for $PM_{2.5}$ and data-scarce scenarios.

A. Estimating $PM_{2.5}$ for Sparse Sensor Network

To estimate $PM_{2.5}$ in regions with sparse sensor networks, two solutions can be incorporated – (i) Installing low-cost sensors in the region in addition to the existing ground sensors [4], (ii) Adopting transfer or meta-learning models where data from alternate regions that have a dense sensor network is incorporated to develop a robust ML model.

For the first case, deploying low-cost sensors can improve the spatial coverage of the sensing network [14], [15]; however, studies have demonstrated that these low-cost sensors, such as PurpleAir and more, often exhibit systematic biases due to fluctuating environmental conditions like temperature, humidity, and more. conditions [4], [16]. Hence, statistical calibration models that incorporate these environmental conditions as input are required for bias removal.

For the second case, transfer and meta-learning models use auxiliary data from *data-rich* regions to develop robust $PM_{2.5}$ prediction models [14]. For example, Fong et al. [17], Yao et al [18], Ma et al [19], and Yadav et al [20] use transfer models that incorporate deep learning methods such as Recurrent Neural Networks (RNN, Variational auto-encoders (VAE), and Long-Short Term Memory with RNN (LSTM-RNN) for time-series forecasting. They utilize ground sensor networks from neighboring cities as source data to forecast $PM_{2.5}$, incorporating both the source data and historical values of the target sensors for training.

In our work, we aim to use sparse-sensor networks without relying on auxiliary data or additional sensors and predict on unseen test data.

B. Gaussian Processes for $PM_{2.5}$ Estimation

Gaussian processes (GPs) are powerful models for predicting $PM_{2.5}$ concentrations, as their non-parametric, kernel-based framework effectively captures complex spatial and temporal dependencies [12], [21]. For example, Cheng et al. [22] propose a GP-based model that predicts $PM_{2.5}$ concentration at test locations via spatial interpolation. Patel et al [12] introduce a GP-based model that uses non-stationary spatiotemporal kernels to predict air quality ($PM_{2.5}$ and PM_{10} concentration levels). Jang et al. [23] demonstrate that Gaussian Process regression (GPR) outperforms linear and ARIMA models in forecasting $PM_{2.5}$ and PM_{10} concentrations over three-day periods. However, these studies utilize a dense sensor network for predicting $PM_{2.5}$ and are unsuitable for sparse sensor networks.

C. Gaussian Processes for Low-data Scenarios

Multiple recent studies have explored Gaussian processes (GPs) and their variants for low-data scenarios, such that these GP models act as transfer learning models for robust predictions. Moreno-Muñoz et al. [24] introduce modular variational GPs, an ensemble model that incorporates uncertainty transfer without revisiting the source data. Tighineanu et al. [25] unified hierarchical GP models for transfer learning using Bayesian optimization, thereby improving the model's performance in low-data regimes. Urtasun et al. [26] proposed a shared latent space GP framework that learns with few target examples by transferring nonlinear representations. Karaletsos et al. [27] introduce a hierarchical GP-based prior over the neural network weights that can capture correlations among weights as well as allow flexible encoding of inductive biases. This improves uncertainty estimation and generalization in deep learning models. However, these models are limited by the need for auxiliary training data (source data for transfer).

III. Methodology

Gaussian Processes (GP) are non-parametric, kernel-based models that can model complex data with non-iid characteristics and also quantify uncertainty in the predictions. Uncertainty quantification means that these models don't just provide a single predicted value but also a measure of confidence in these values.

A regression problem can be represented by a function $f : \mathbb{R}^d \to \mathbb{R}$, where d is the dimensionality of the feature space. It is said to follow a Gaussian Process (GP), if for a finite set of input points $\{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$, the function values at these points, $\{f(\mathbf{x}_1), \ldots, f(\mathbf{x}_N)\}$, look like they were drawn together from a joint Gaussian distribution:

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

where $m(\mathbf{x}) = \mathbb{E}[f(\mathbf{x})]$ is the mean function (often set to zero), and $k(\mathbf{x}, \mathbf{x}') = \text{Cov}(f(\mathbf{x}), f(\mathbf{x}'))$ is the covariance function (also the kernel), describes how much two function values at different input points are related.

The prediction, y_i for the above setting for the dataset $\mathcal{D} = \{(\mathbf{x}_i, y_i)\}_{i=1}^N$ where N is the number of samples can be modeled using a latent function $f(\mathbf{x}_i)$ and noise ϵ_i as,

$$y_i = f(\mathbf{x}_i) + \epsilon_i \tag{2}$$

where $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$ and denotes a Gaussian distribution with mean 0 and variance σ^2.

We use equation 1, to model the observed values, $\mathbf{y}$ at input points $\mathbf{X}$ and the predicted values, $\mathbf{f}_*$ at test points $\mathbf{X}_*$ as:

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{f}_* \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mathbf{m} \\ \mathbf{m}_* \end{bmatrix}, \begin{bmatrix} \mathbf{K} + \sigma^2 \mathbf{I} & \mathbf{K}_{**}^T \\ \mathbf{K}_{**} & \mathbf{K}_* \end{bmatrix} \right), \tag{3}$$

Where $\mathbf{K}$, m, and $\mathbf{K}_*$, $\mathbf{m}_*$ are the kernel and mean matrices of train (input) and test points, respectively. $\mathbf{K}_{**}$ is the cross-covariance between them. Hence, the predictions at test points is represented as,

$$\mathbf{f}_* \mid \mathbf{X}, \mathbf{y}, \mathbf{X}_* \sim \mathcal{N}(\bar{\mathbf{f}}_*, \mathrm{Cov}[\mathbf{f}_*]), \tag{4}$$

Sparse Variational Gaussian Processes (SVGP) are useful in approximating Gaussian Processes by employing a set of M inducing points where $M \ll N$. These inducing points are learned and are denoted as, $\mathbf{Z} = \{\mathbf{z}_j\}_{j=1}^{M}$. The predicted values (latent function values) are represented as $\mathbf{u} = f(\mathbf{Z})$. Instead of calculating the full posterior $p(\mathbf{f} \mid \mathbf{y})$, we approximate it by using a simpler, and flexible distribution $q(\mathbf{u})$, using the inducing points. We approximate the posterior for $q(\mathbf{u})$ as,

$$q(\mathbf{f}) = \int p(\mathbf{f} \mid \mathbf{u}) q(\mathbf{u})\, d\mathbf{u}, \tag{5}$$

This expresses the posterior distribution in terms of a simpler (variational) distribution.

Since the full posterior, $p(\mathbf{f} \mid \mathbf{y})$, does not have a closed form solution, the distributions $q(\mathbf{u})$, and $q(\mathbf{f})$ are introduced. Then, variational inference is used to make these approximated distributions fit the data. This is achieved by minimizing the KL divergence between the approximated and the full posterior, $\mathrm{KL}\big(q(\mathbf{f}) \parallel p(\mathbf{f} \mid \mathbf{y})\big)$, which is the same as maximizing the *Evidence Lower Bound* (ELBO) [28] as,

$$\mathrm{ELBO}(q) = \mathbb{E}_{q(\mathbf{f})}[\log p(\mathbf{y} \mid \mathbf{f})] - \mathrm{KL}\big(q(\mathbf{u}) \parallel p(\mathbf{u})\big). \tag{6}$$

The first term uses the variational posterior $q(\mathbf{f})$ to explain the data, while the second term acts as a regularizer by penalizing deviations of $q(\mathbf{u})$ from the GP prior $p(\mathbf{u})$. The inducing points $\mathbf{Z}$ and the parameters of $q(\mathbf{u})$ are optimized during training, which allows the model to adapt to the structure of the data.

In this work, we initialize $\mathbf{Z}$ by applying K-Means clustering on the training data points and selecting the cluster centroids as inducing points. The number of inducing points ranges between 200 and 800, depending on the dataset size; since Lima has fewer data points, we chose 200. We observe that the inducing points adapt spatially and spread across the Lima region, which demonstrates the effectiveness of this approach in capturing the spatial structure.

The kernel function used is a combination of a Radial Basis Function (RBF) kernel and a Matern kernel, $k(\mathbf{x}, \mathbf{x}') = k_{\text{RBF}}(\mathbf{x}, \mathbf{x}') + k_{\text{Matern}}(\mathbf{x}, \mathbf{x}')$, both employed Automatic Relevance Determination (ARD), allowing the model to learn individual length scales per features.

IV. Evaluation

A. Dataset

We used the dataset with daily averaged $PM_{2.5}$ values of the Lima region from the year 2016. There are 16 features in the dataset, which include 6 daily averaged meteorological features: *relative humidity*, *precipitation*, *temperature*, and more; 2 topographical features, such as *population density* and *elevation* for the sensor's location; a pollution indicator: aerosol optical depth (AOD) that represents aerosol's scatter; 2 geo-location features: latitude and longitude; 2 temporal features: day and month. The spatial irregularity was due to sparse sensor placement, and the temporal irregularity was due to gaps in the collected $PM_{2.5}$ data over the 365-day period.

Lima has a total of 10 ground sensors to monitor air quality, as shown in Figure 1(a). The sensors are primarily clustered in central Lima, a densely populated region, while the high-elevation areas to the eastern part, i.e., the *Andes mountain* ranges, are notably devoid of sensors. This highlights the need for synthetic data that spans the entire region and requires minimal sensors for generation.

Lima currently has 10 ground sensors for monitoring $PM_{2.5}$ as shown in Figure 1(a). These sensors are concentrated in central Lima, the most densely populated area, while the high-elevation regions to the east, i.e., the *Andes Mountains*, are devoid of any sensors. This imbalance highlights the importance of generating synthetic data that can cover the entire region while being dependent on a limited number of ground sensors.

To evaluate model performance, we generated 5 randomized train-test splits where 4 sensors were used for training and the remaining 6 for testing.

B. Models

Sparse Variational Gaussian Process (SVGP) model is compared to 3 other regression models which includes (1) Gradient Boosting regressor, parameter optimized to *lr = 0.05, estimators = 1000*, (2) Lasso Regressor with α = *0.5*, and (3) Gaussian Process Regressor (GPR) using a combined RBF, Constant and White Kernel, and 10 optimizer restarts for parameter tuning. For the SVGP model, we utilized 200 inducing points selected via the KMeans clustering, a combination of RBF and Matern kernels, with the variational strategy optimizing the inducing points over 1500 epochs.

All models are trained on the same standardized input features and evaluated using 20 train–test splits to ensure fair comparison. The model's performance is assessed using two metrics: the coefficient of determination (R^2) and the Root mean squared error (RMSE).

C. Results

We validate our hypothesis first using qualitative analysis, where we visualize how the inducing points grow over the region. Next, we compare the performance of SVGP to competitive regression models to determine if such spatial adaptation is useful.

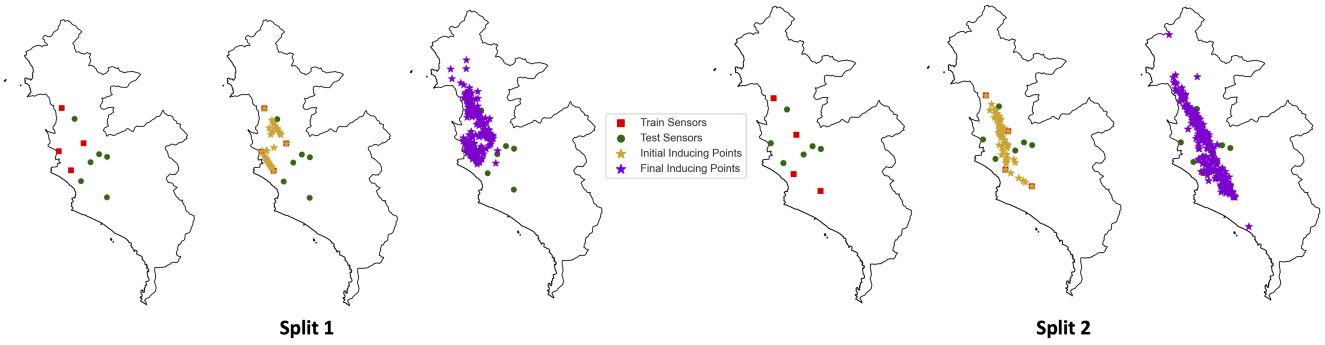

Fig. 1. Visualizing growth of inducing points when initialized (yellow) compared to when optimized (purple) over 2 splits.

1) Qualitative Analysis: In Figure 1, we analyze the evolution of the inducing points when they are initialized (Figure 1, yellow marker) and optimized (Figure 1, purple marker) for the first and second train-test split. Initially, the inducing points, selected via K-means clustering, are concentrated within the area around the training sensors (red marker). After optimization, the inducing points spread towards the northern region and cover a few test sensor locations (blue marker). This spatial expansion is indicative of potential generalization and the need to strategically place train sensors to improve the performance of the ML model.

We also observe the spread of inducing points in splits 1 and 2, which validates our hypothesis of sensor placement. In split 1, the sensors are sparse and we see the optimization growing northwards, whereas in split 2, the initial inducing points are almost linearly placed and therefore their spread has been limited around their ground locations. Hence, strategically placed sensors will affect the growth of inducing points, which will consequently impact model performance.

Model	R^2	RMSE
SVGP	**0.35**	**10.13**
Gaussian Process	0.14	11.24
Gradient Boosting	0.19	11.25
Lasso Regression	0.19	11.30

TABLE I
ACCURACY OF REGRESSION MODELS ON LIMA DATA.

2) Quantitative Analysis: Table I presents the comparative performance of the regression models, where we observe that SVGP achieves the best performance, outperforming the second-best-performing model, Gradient Boosting regression, by approximately 84% for R^2 score and 10% for RMSE. This can be attributed to the optimal initialization of inducing points (in our case, using KMeans clustering) and SVGP's ability to adapt spatially over the region. Conversely, while the Gaussian Process model employs the same training process to fit over the observed data points, it doesn't share the same advantage of using approximated data points that have been spatially adapted. The source code is available at: https://github.com/shrey-gupta/svgps-for-low-sensor

V. DISCUSSION AND FUTURE WORK

The optimized inducing points can be used as synthetic data points with any of the comparative models we have used. This would require approximating $PM_{2.5}$ values at these locations using standard interpolation strategies such as Kriging and more. We discuss how this strategy can be used in conjunction with other state-of-the-art techniques to design an improved prediction model, as well as its applications in other fields.

A. Generative Modeling to Reduce Spatial Irregularity

By utilizing the spatially adaptable behavior of SVGP, we propose that the optimized inducing points can be leveraged in conjunction with a generative architecture [29] to synthesize data that reduces spatial irregularities in the dataset. Such an approach could enhance predictive performance in regions with sparse sensor coverage by generating robust training data.

B. Low-Cost Sensors for Extreme Event Modeling

To improve coverage in regions with high $PM_{2.5}$ levels, we propose first deploying a few low-cost sensors in these hotspots. These sensors can help capture extreme values that are underrepresented in the dataset. The observations from these sensors can be used to collect inducing points that focus on modeling tail-end distributions of $PM_{2.5}$ [30]. Subsequently, a generative model can be applied to these inducing points to simulate high-pollution scenarios. This pipeline can be utilized to improve the robustness and generalization of the $PM_{2.5}$ model for forecasting extreme air quality events.

C. Reinforcement Learning for Inducing Point Selection

The optimized inducing points capture the locations where the model is most sensitive or uncertain, i.e., areas of high variance as inferred by the SVGP model. This configuration can also serve as an input to a reinforcement learning (RL) agent. The RL agent can utilize the inducing points as part of its state to generate new labels in the sparse sensor regions. The reward function can be tied to performance gains such as improved predictive accuracy in these regions.

Moreover, the inducing points can act as learned signals that affect decisions in active learning (where new data will help the most) [31] or adaptive sensing (where new sensors will help the most) [32].

VI. Conclusion

In this work, we investigate the problem of predicting $PM_{2.5}$ concentrations in regions with sparse sensor networks, where data scarcity and non-IID characteristics reduce the prediction accuracy of traditional machine learning models. We propose using Sparse Variational Gaussian Processes (SVGPs) as a solution that requires no auxiliary data sources and addresses limitations such as irregularities in sensor placement and distribution imbalance. Our findings demonstrate that inducing points of SVGPs exhibit spatial adaptation, i.e., they spread across the region to capture the underlying data distribution. We also show that strategic sensor placement improves the spatial adaptation. We validate these hypotheses using $PM_{2.5}$ data from Lima, Peru. We found that SVGPs outperform competitive regression models by 84% and 10% for R2 and RMSE scores, respectively. These results suggest that SVGPs provide a scalable and flexible framework for predicting $PM_{2.5}$ in sparse sensor regions. Future work can extend this approach to larger-scale sensor deployments and explore strategies for sensor placement to improve monitoring in developing regions.

References

[1] R. Burnett, H. Chen, M. Szyszkowicz, N. Fann, B. Hubbell, C. A. Pope III, J. S. Apte, M. Brauer, A. Cohen, S. Weichenthal *et al.*, "Global estimates of mortality associated with long-term exposure to outdoor fine particulate matter," *Proceedings of the National Academy of Sciences*, vol. 115, no. 38, pp. 9592–9597, 2018.

[2] S. Sharma, M. Chandra, and S. H. Kota, "Health effects associated with pm2. 5: A systematic review," *Current Pollution Reports*, vol. 6, no. 4, pp. 345–367, 2020.

[3] X. Wang and colleagues, "First close insight into global daily gapless 1 km pm2.5 pollution maps," *Nature Communications*, 2023.

[4] J. Bi, A. Wildani, H. H. Chang, and Y. Liu, "Incorporating low-cost sensor measurements into high-resolution pm2. 5 modeling at a large spatial scale," *Environmental Science & Technology*, 2020.

[5] T. Saeed, M. M. Khaliq, M. H. Bergin, P. V. Bhave, N. Khaleel, E. Mool, M. Senarathna, S. U. Zaman, S. Roy, A. Salam *et al.*, "Sustaining low-cost pm2. 5 monitoring networks in south asia: Technical challenges and solutions," *EGUsphere*, vol. 2024, pp. 1–24, 2024.

[6] D. Ganesan, S. Ratnasamy, H. Wang, and D. Estrin, "Coping with irregular spatio-temporal sampling in sensor networks," in *ACM SIGCOMM Computer Communication Review*. ACM, 2004.

[7] A. Baruah, D. Bousiotis, S. Damayanti, A. Bigi, G. Ghermandi, O. Ghaffarpasand, R. M. Harrison, and F. D. Pope, "A novel spatiotemporal prediction approach to fill air pollution data gaps using mobile sensors, machine learning and citizen science techniques," *npj Climate and Atmospheric Science*, vol. 7, no. 1, p. 310, 2024.

[8] J. Rentschler and N. Leonova, "Global air pollution exposure and poverty," *Nature Communications*, vol. 14, no. 1, p. 4432, 2023.

[9] H. Liu, J. Cai, Y.-S. Ong, and Y. Wang, "Understanding and comparing scalable gaussian process regression for big data," *Knowledge-Based Systems*, vol. 164, pp. 324–335, 2019.

[10] H. J. Cunningham, D. A. de Souza, S. Takao, M. van der Wilk, and M. P. Deisenroth, "Actually sparse variational gaussian processes," in *Proceedings of the International Conference on Artificial Intelligence and Statistics*, 2023, pp. 10 395–10 408.

[11] W. T. Stephenson, S. Ghosh, T. D. Nguyen, M. Yurochkin, S. K. Deshpande, and T. Broderick, "Measuring the robustness of gaussian processes to kernel choice," *arXiv preprint arXiv:2106.06510*, 2021.

[12] Z. B. Patel, P. Purohit, H. M. Patel, S. Sahni, and N. Batra, "Accurate and scalable gaussian processes for fine-grained air quality inference," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 36, no. 11. AAAI Press, 2022, pp. 12 080–12 088.

[13] B. N. Vu, O. Sánchez, J. Bi, Q. Xiao, N. N. Hansel, W. Checkley, G. F. Gonzales, K. Steenland, and Y. Liu, "Developing an advanced $pm_{2.5}$ exposure model in lima, peru," *Remote Sensing*, 2019.

[14] S. Gupta, Y. Park, J. Bi, S. Gupta, A. Züfle, A. Wildani, and Y. Liu, "Spatial transfer learning for estimating $pm_{2.5}$ in data-poor regions," in *Joint European Conference on Machine Learning and Knowledge Discovery in Databases (ECML PKDD)*. Springer, 2024, pp. 385–400.

[15] S. Gupta, J. Bi, Y. Liu, and A. Wildani, "Boosting for regression transfer via importance sampling," *International Journal of Data Science and Analytics*, pp. 1–12, 2023.

[16] P. Attey-Yeboah, C. Afful, K. Yeboah, C. H. Korkpoe, E. S. Coker, R. Subramanian, and A. K. Amegah, "Utility of low-cost sensor measurement for predicting ambient $pm_{2.5}$ concentrations: evidence from a monitoring network in accra, ghana," *Environmental Science: Atmospheres*, vol. 5, no. 4, pp. 517–529, 2025.

[17] I. H. Fong, T. Li, S. Fong, R. K. Wong, and A. J. Tallon-Ballesteros, "Predicting concentration levels of air pollutants by transfer learning and recurrent neural network," *Knowledge-Based Systems*, vol. 192, p. 105622, 2020.

[18] B. Yao, G. Ling, F. Liu, and M.-F. Ge, "Multi-source variational mode transfer learning for enhanced $pm_{2.5}$ concentration forecasting at data-limited monitoring stations," *Expert Systems with Applications*, 2024.

[19] J. Ma, Z. Li, J. C. P. Cheng, Y. Ding, C. Lin, and Z. Xu, "Air quality prediction at new stations using spatially transferred bi-directional long short-term memory network," *Science of The Total Environment*, vol. 705, p. 135771, 2020.

[20] K. Yadav, V. Arora, M. Kumar, S. Tripathi, V. Motghare, and K. Rajput, "Few-shot calibration of low-cost air pollution ($pm_{2.5}$) sensors using meta learning," *IEEE Sensors Letters*, vol. 6, no. 5, pp. 1–4, 2022.

[21] W. Yu, Y. Liu, Z. Ma, and J. Bi, "Improving satellite-based pm2. 5 estimates in china using gaussian processes modeling in a bayesian hierarchical setting," *Scientific reports*, vol. 7, no. 1, p. 7048, 2017.

[22] Y. Cheng, X. Li, Z. Li, S. Jiang, and X. Jiang, "Fine-grained air quality monitoring based on gaussian process regression," in *Neural Information Processing: Proceedings of the 21st International Conference, ICONIP 2014, Kuching, Malaysia, Part II*. Springer, 2014, pp. 126–134.

[23] J. Jang, S. Shin, H. Lee, and I.-C. Moon, "Forecasting the concentration of particulate matter in the seoul metropolitan area using a gaussian process model," *Sensors*, vol. 20, no. 14, p. 3845, 2020.

[24] P. Moreno-Muñoz, A. Artés, and M. Alvarez, "Modular gaussian processes for transfer learning," *Advances in Neural Information Processing Systems*, vol. 34, pp. 24 730–24 740, 2021.

[25] P. Tighineanu, K. Skubch, P. Baireuther, A. Reiss, F. Berkenkamp, and J. Vinogradska, "Transfer learning with gaussian processes for bayesian optimization," in *Proceedings of the International Conference on Artificial Intelligence and Statistics*, 2022, pp. 6152–6181.

[26] R. Urtasun, A. Quattoni, N. Lawrence, and T. Darrell, "Transferring nonlinear representations using gaussian processes with a shared latent space," *arXiv preprint arXiv:0808.0033*, 2008.

[27] T. Karaletsos and T. D. Bui, "Hierarchical gaussian process priors for bayesian neural network weights," in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 33, 2020, pp. 17 141–17 152.

[28] M. Titsias, "Variational learning of inducing variables in sparse gaussian processes," in *Artificial Intelligence and Statistics*. PMLR, 2009.

[29] L. Xu, M. Skoularidou, A. Cuesta-Infante, and K. Veeramachaneni, "Modeling tabular data using conditional gan," *arXiv preprint arXiv:1907.00503*, 2019.

[30] A. Chang, M. Wang, and G. I. Allen, "Sparse regression for extreme values," *Electronic Journal of Statistics*, 2021.

[31] T. Wan, K. Xu, L. Lan, Z. Gao, D. Feng, B. Ding, and H. Wang, "Tracing training progress: Dynamic influence based selection for active learning," in *Proceedings of the 32nd ACM International Conference on Multimedia*, 2024, pp. 9417–9425.

[32] H. B. Moss, S. W. Ober, and V. Picheny, "Inducing point allocation for sparse gaussian processes in high-throughput bayesian optimisation," in *Proceedings of the International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2023, pp. 5213–5230.