Enhanced sensor environment graph based deep learning approach for air quality anomaly detection

Sahar Masmoudi, Christelle Garnier, Anne Savard, Vincent Itier, Stephane Sauvage, Florentin Bulot, Pascal Kaluzny

Enhanced sensor environment graph based deep learning approach for air quality anomaly detection

Sahar Masmoudi*†, Christelle Garnier*, Anne Savard*, Vincent Itier*, Stéphane Sauvage†, Florentin Bulot‡and Pascal Kaluzny‡

* IMT Nord Europe, Institut Mines Télécom, Centre for Digital Systems, F-59653 Villeneuve d'Ascq, France
† IMT Nord Europe, Institut Mines Télécom, Centre for Energy and Environment, F-59000, Lille, France
‡ TERA Sensor, 13790 Rousset, France
Email: {firstname.lastname}@imt-nord-europe.fr, {firstname.lastname}@groupe-tera.com,

***Abstract*—Air pollution is among the major threats to human well-being, highlighting the critical need for air quality monitoring, especially in urban areas. Whereas the development of low-cost pollution sensors has facilitated a widespread monitoring, a reliable anomaly detection system is required to properly characterize data for the end-users. In this paper, we propose an enhanced deep learning approach based on the A3T-GCN (Attention Temporal Graph Convolutional Network) model that accurately forecasts particulate matter $PM_{2.5}$ concentrations using real past measurements from a deployed sensor network. Our proposed Enhanced-A3T-GCN embeds all the available spatial and temporal correlations within the sensor network, along with additional information regarding the sensor environment in a graph. It is shown to achieve significant performance improvement with respect to other deep learning forecasting methods, emphasizing the importance of exploiting the sensor environment-based information. Further, the achieved accurate forecasting makes it possible to detect anomalies injected at both single and multiple sensor levels.**

***Index Terms*—Pollutant forecasting, Graph neural network, A3T-GCN, Anomaly detection.**

I. Introduction

Air pollution induces serious effects on human health worldwide. In particular, exposure to particulate matter (PM), is associated with worsening of respiratory and cardiovascular diseases that can lead to premature death. For this reason, monitoring pollutant concentrations is of prime importance. In recent years, the development of low-cost pollution sensors has completed the air quality monitoring strategy in order to enhance the spatial coverage. Their cost, size and ease of use enable to deploy large networks [1]–[3]. This new approach generates large sets of data for which characterization methods are indispensable. A reliable anomaly detection system is particularly needed to set up information and warning for the population and to trigger appropriate measures to reduce air pollution when needed.

More specifically, anomaly detection consists in identifying data that deviate from their expected behavior [4], [5]. In a pollution sensor network, the main sources of anomalies include:
(1) an unusual event leading to a significant increase in the pollutant concentration, e.g. fire, waste incineration, specific weather conditions [6], [7]. In this case, the data should be tagged and examined more thoroughly to discover the underlying event [5];
(2) a malfunction or failure of a low-cost sensor and/or the wireless communication system, more likely to lead to data erasure or erroneous data. In this case, noises or time drifts should be eliminated to improve the data quality and missing values should be filled.

A pollution sensor network provides air quality data in the form of several multivariate discrete-time time series. To take advantage of the temporal correlation, many anomaly detection approaches are based on time series forecasting. Then any observed value is considered as an anomaly if the distance between this value and the predicted value is higher than a predefined threshold. Time series models used for forecasting include autoregressive models [2], [8], e.g. Auto Regressive Integrated Moving Average (ARIMA), decomposition models, e.g. three components (trend, seasonality and holidays) in Facebook Prophet [9], and more recently deep learning models based on LSTM (Long Short Term Memory) or GRU (Gated Recurrent Unit) architectures specialized for processing time-sequential data [10]. The latter models have shown significant progress in multivariate time series forecasting, but they only focus on the temporal correlation. Nevertheless, the multiple time series from a sensor network are also spatially correlated. One promising solution for capturing the spatial correlation inherent to a sensor network is to add a GCN (Graph Convolution Network) [10]–[13], which is more general than the well-known CNN (Convolutional Neural Network) restricted to grid-type data structures, such as images.

In this context, one of the most advanced methods is A3T-GCN (Attention Temporal Graph Convolutional Network), originally proposed for traffic forecasting [14], that was then shown to outperform other baseline methods in several forecasting contexts [15], [16]. This method combines GCN to capture spatial dependencies, GRU to learn temporal dependencies, and introduces an attention mechanism to ensure a focus on the most relevant parts of the time series. To take full advantage of A3T-GCN, it needs to be fed with a graph embedding all the available spatio-temporal correlation within the considered sensor network. Exploiting graph theory, the latter is obtained as an adjacency matrix whose entries quantify

the similarity between each pair of sensors. Recently, in the context of air quality prediction [16] and anomaly detection in air quality data [12], this adjacency matrix was built from spatial distances between sensor locations and from temporal correlation between the time series provided by sensors.

In this paper, the objective is to enhance the A3T-GCN deep learning approach by enriching the spatio-temporal information supplied as input by exploiting features describing the environment in which the deployed sensors are installed, such as street width, height of nearby buildings, presence of trees... Indeed, the environment has a strong impact on the pollution sensor responses because atmospheric turbulences, caused by physical obstructions such as buildings and trees, are among the key factors influencing the dispersion of pollutants in the air. At a city scale, closely deployed sensors in contrasted street configurations can provide pollutant concentrations which are much less correlated than their proximity would suggest. On the other hand, sensors in similar streets can show a higher level of interdependence despite a greater distance. Thus, a novel adjacency matrix is defined from distances between sensors in the environment feature space such that its entries quantify the degree of similarity between sensor environments. It is then combined with the common adjacency matrix described previously to construct an enhanced spatio-temporal graph. As such, our main contributions are summarized as follows:

i) We first propose a novel graph construction, combining both spatial and temporal correlations across the deployed sensor network as well as additional information based on the environment of the deployed sensors.

ii) This graph is then fed to the A3T-GCN, which is referred to as *Enhanced-A3T-GCN* in the remaining. The forecasting accuracy of our proposed Enhanced-A3T-GCN is compared to the one of several other deep learning based forecasting approaches from the literature. Numerical results show that on our dataset, the Enhanced-A3T-GCN performs better.

iii) Finally, we investigate the ability of our Enhanced-A3T-GCN approach to detect anomalies. To this end, simulated anomalies are injected at a local level, i.e. at a sensor level, and at a global level, i.e. at multiple sensor level.

II. Proposed Enhanced-A3T-GCN

Let us consider a network, deployed at a city scale, of N_s sensors measuring air pollutant concentrations. Because of climatic conditions within the area of interest, as well as due to pollution sources near and in the city, the measurements exhibit some spatial and temporal dependencies that need to be captured to better and accurately predict the pollution levels. In the remaining, prediction is done by exploiting A3T-GCN, a network able of learning spatio-temporal dependencies between sensors and that takes as input a graph embedding the aforementioned spatio-temporal information. The output of this network is the forecasted future pollutant levels. If the predicted value is too far away from the true measurement, i.e. the absolute prediction error is greater than a threshold, an anomaly is detected.

Our main contribution consists in a new graph construction given as the input of the A3T-GCN that takes advantage of the spatio-temporal available information coming from both the pollution measurements themselves, as well as from the sensor positions and environments. The main steps of our proposed anomaly detection scheme are highlighted in Fig.1.

A. Proposed enhanced spatio-temporal graph construction

The sensor network can be represented as a complete weighted graph $G = (V, E)$, where V is the set of vertices (also called nodes) and E is the set of edges connecting each pair of vertices. In this graph, each vertex v_i represents the i-th sensor in the deployed sensor network, and the edge between two vertices v_i and v_j is weighted by $\omega_{ij} \geq 0$ which quantifies the similarity or proximity between sensors i and j. In the usual graph terminology, the matrix $\mathbf{\Omega} \in \mathbb{R}^{N_s \times N_s}$ collecting all weighting parameters $\{\omega_{ij}\}_{i,j \in [1,N_s] \times [1,N_s]}$ as $\Omega(i,j) = \omega_{ij}$ is referred to as the *adjacency matrix* of the graph. $\mathbf{\Omega}$ can be defined in several ways, e.g. from distances between the position of the deployed sensors or from the temporal correlation between the time series delivered by the sensors [14]–[16].

We here propose to also exploit the similarity within the sensor environment itself to build a new adjacency matrix, that relies on:

i) *11 qualitative variables* that describe the urban environment of each sensor and that encompass information such as the type of area (e.g. downtown, residential area), the road geometry (e.g. straight road, roundabout), the urban morphology (e.g. street canyons), the way the sensor is deployed (e.g. near a tree, on a wall), etc. All this information can be gathered when deploying the sensor network. The studied sensor network was deployed by the private French company Groupe TERA and for confidentiality purposes, all the qualitative variables, as well as the values they can take, are not exhaustively presented in this paper.

ii) *5 public auxiliary variables* that describe the pollution-related environment of each sensor. We here consider the traffic density[1] in 2020, the annual mean Nitrogen Dioxide (NO_2) concentration at the sensor location in 2017 given by a model from Atmosud, the French southern region air quality observatory, approved by the Ministry of Environment, and the $PM_{2.5}$ spatial emission inventory[2] on a larger scale for 2018 also provided by AtmoSud. All these variables are used to construct a new adjacency matrix.

Sensor environment-based adjacency matrix $\mathbf{\Omega}^{\text{S. Env.}}$: For each sensor pair $(i,j), i,j \in [1,N_s] \times [1,N_s]$, we first compute their distance, denoted as $d_{ij}^{\text{S. Env.}}$, in the above mentioned 16-dimensional feature space. We then apply a Gaussian radial basis function and the entry $\omega_{ij}^{\text{S. Env.}}$ of the adjacency matrix is given as

$$\omega_{ij}^{\text{S. Env.}} = \exp(-(d_{ij}^{\text{S. Env.}})^2/(\sigma^{\text{S. Env.}})^2), \quad (1)$$

[1]https://data.ampmetropole.fr/explore/dataset/comptage-routier-base-routiere-metropole-2020/information/

[2]https://opendata.atmosud.org/viewer.php?categorie=modelisation

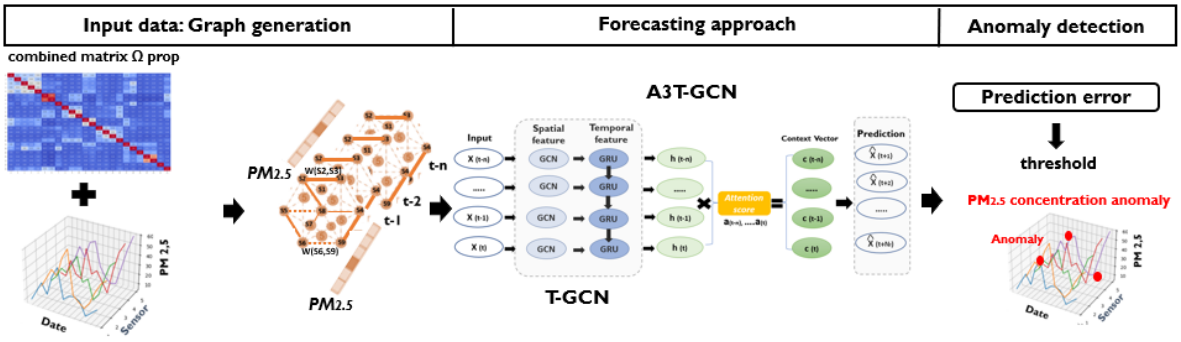

Fig. 1. Proposed Enhanced-A3T-GCN for time series forecasting and anomaly detection

where $\sigma^{\text{S. Env.}}$ is the sample standard deviation of the distance set $\{d_{ij}^{\text{S. Env.}}\}_{i,j\in[1,N_s]\times[1,N_s]}$. In this way $\mathbf{\Omega}^{\text{S. Env.}}$ embeds the sensor environment similarity.

In our proposed approach, we also exploit some previously proposed spatio-temporal adjacency matrices [12], [14]–[16].

Spatial correlation-based adjacency matrix $\mathbf{\Omega}^{Spatial}$: We follow a similar approach as above but consider the geodesic distance between all geolocated sensor pairs (i, j), denoted as d_{ij}^{Spatial}. Hence, the weighting parameter is given as

$$\omega_{ij}^{\text{Spatial}} = \exp(-(d_{ij}^{\text{Spatial}})^2/(\sigma^{\text{Spatial}})^2), \tag{2}$$

where σ^{Spatial} is the sample standard deviation of the distance set $\{d_{ij}^{\text{Spatial}}\}_{i,j\in[1,N_s]\times[1,N_s]}$.

Temporal correlation-based adjacency matrix $\mathbf{\Omega}^{Temporal}$: We exploit the temporal correlation among the concentrations measured by sensors i and j across time that are denoted as $x_i(t)$ and $x_j(t)$ respectively. We further denote by $\overline{x_i}$ and $\overline{x_j}$ their empirical temporal average. As such, the weighting parameter writes as

$$\omega_{ij}^{\text{Temporal}} = \frac{\sum_t (x_i(t) - \overline{x_i})(x_j(t) - \overline{x_j})}{\sqrt{\sum_t (x_i(t) - \overline{x_i})^2 \sum_t (x_j(t) - \overline{x_j})^2}}. \tag{3}$$

Proposed combined adjacency matrix $\mathbf{\Omega}^{Prop.}$: Finally, we propose to blend the three aforementioned adjacency matrices into a new one obtained as the product of the three, defined as $\mathbf{\Omega}^{\text{Prop.}} = \mathbf{\Omega}^{\text{S. Env.}}\mathbf{\Omega}^{\text{Spatial}}\mathbf{\Omega}^{\text{Temporal}}$. Although some values can be very close to zero, we do not apply any sparsification to avoid the tuning of the related threshold.

B. Forecasting approach

We here propose to predict air pollutant concentrations with the help of the A3T-GCN [14] model that is fed with our proposed enhanced graph accounting for the joint spatio-temporal information alongside with the sensor environment based information, as well as with a past time sequence of the air pollution measurements within the sensor network. This model exploits three key components, namely GRU, GCN and an attention mechanism, where the GRU and attention mechanism handle the temporal dimension (consideration of the temporal correlation and focus on the most relevant parts of the time series), whereas the GCN captures the spatial correlation. For a more detailed presentation of A3T-GCN, we kindly refer the interested reader to [14].

C. Anomaly detection

To detect an anomaly, we compute the Mahalanobis distance between the pollution concentration forecasted by the aforementioned Enhanced-A3T-GCN model to the true one measured by the sensor network, which is referred to as Prediction error in Fig. 1. The latter is then compared to a predetermined threshold and if above, an anomaly is detected. Note that we here exploit the Mahalanobis distance given its good performance in anomaly detection [17].

III. Numerical results

A. Dataset construction and pre-processing

We consider real data provided by a sensor network recently deployed in an urban area near Aix-en-Provence and composed of $N_s = 100$ NextPM[3] sensors designed by the company Groupe TERA for PM pollution monitoring. Each sensor delivers $PM_{2.5}$ concentration measurements every 15 minutes.

In order to work with a reasonably-sized dataset, we select a limited number of sensors from all the sensors in the network while considering various sensor environments. To this end, we first identify the main categories of environments by performing sensor clustering via the k-means algorithm using the above computed distances within the 16-dimensional feature space. The optimal number of clusters, obtained thanks to the elbow method, was found to be 7. Then to ensure data representativeness, we select $N = 24$ sensors out of the $N_s = 100$ sensors within the deployed network such that each cluster, i.e. each major category of environment, is represented by at least 2 (and up to 5) sensors. Finally, the used dataset contains 9464 $PM_{2.5}$ concentrations measured between June, 10th and Sept., 26th 2023 for each of the $N = 24$ sensors.

Due to the presence of imperfections and unknown anomalies in these real measurements, pre-processing steps are performed for a proper model evaluation:

i) We exclude all observations that deviate significantly from the empirical average by considering the well-known 3σ criterion. Indeed, these outlier values are not relevant to assess the forecasting performance.

ii) We fill in all missing values, that correspond to values that were either removed in the previous step or not measured or

[3]https://tera-sensor.com/fr/categorie-produit/capteurs-oem/

not transmitted by the sensor network. To do so, we use a linear interpolation method.

B. Forecasting performance

In simulations, we use 80% of the $PM_{2.5}$ measurements, i.e. $N_t = 7571$ samples, for training and the remaining 20%, i.e. $N_f = 1893$ samples, for testing. To preserve the temporal correlation of the time series, the order of samples is maintained and the test set is placed just after the training set. Moreover prior to this choice of splitting, we checked the stationarity of the time series over the whole considered period via the Dickey-Fuller test. In the remaining, we focus on the forecasting performance in terms of *root mean square error* (RMSE) and *mean absolute error* (MAE), given as

$$\text{RMSE} = \frac{1}{N}\sum_{i=1}^{N}\sqrt{\frac{1}{N_f}\sum_{t=1}^{N_f}(x_i(t)-\hat{x}_i(t))^2}, \quad (4)$$

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{N_f}\sum_{t=1}^{N_f}|x_i(t)-\hat{x}_i(t)|. \quad (5)$$

First, we compare the performance obtained with our proposed Enhanced-A3T-GCN to the one obtained with LSTM, GRU and GCN single models. All these algorithms were implemented on a computer equipped with 2 RTX A6000 cards by Nvidia, both having 10752 CUDA cores and 48 Go (DDR6) of RAM. Hyperparameters were tuned to obtain the best performance on our dataset through extensive numerical simulations. The selected values for the four considered deep learning approaches are summarized in Table I.

Hyperparameters	Enhanced-A3T-GCN	LSTM	GRU	GCN
Number of Epochs	50	50	50	50
Learning Rate	0.01	0.01	0.01	0.01
Hidden dimensions	64	256	256	64
Optimizer	Adam	Adam	Adam	Adam
Loss Function	MSE	MSE	MSE	MSE

TABLE I
CONFIGURATION OF THE FOUR MODELS

Table II presents the obtained performance including GPU time computed over the learning process. The results show that our Enhanced-A3T-GCN clearly outperforms all other deep learning techniques in terms of RMSE and MAE, at the cost of a higher computational complexity. Thus they emphasize the benefits of combining GRU and GCN models to capture both temporal and spatial dependencies between sensors and of incorporating an attention mechanism.

Approaches	RMSE	MAE	GPU time (s)
Enhanced-A3T-GCN	**0,397**	**0.286**	17,198
LSTM	0.573	0.426	9,436
GRU	0.482	0.352	1,317
GCN	0,631	0.448	**503**

TABLE II
COMPARISON BETWEEN OUR PROPOSED APPROACH AND SEVERAL DEEP LEARNING FORECASTING METHODS

In the following, we investigate the impact of the adjacency matrix used by the A3T-GCN model and we compare in Table III the RMSE and MAE achieved when considering only the spatial correlation-based adjacency matrix $\mathbf{\Omega}^{\text{Spatial}}$, the temporal correlation-based one $\mathbf{\Omega}^{\text{Temporal}}$, the sensor environment-based one $\mathbf{\Omega}^{\text{S. Env.}}$ and finally our proposed adjacency matrix $\mathbf{\Omega}^{\text{Prop.}}$ combining all the three above. We can note that our proposed new matrix $\mathbf{\Omega}^{\text{S. Env.}}$, that takes into account the sensor environment similarity, gives better results than the conventional matrices embedding the spatial or temporal correlations. Finally the A3T-GCN with our proposed combined adjacency matrix $\mathbf{\Omega}^{\text{Prop.}}$ clearly outperforms all other versions of A3T-GCN using adjacency matrices only focusing on one of the above relationships.

Adjacency matrices used in A3T-GCN	MAE	RMSE
$\mathbf{\Omega}^{\text{Spatial}}$	0.413	0.521
$\mathbf{\Omega}^{\text{Temporal}}$	0.402	0.515
$\mathbf{\Omega}^{\text{S. Env.}}$	0.343	0.461
$\mathbf{\Omega}^{\text{Prop.}}$	**0.286**	**0.397**

TABLE III
IMPACT OF THE ADJACENCY MATRIX FED TO THE A3T-GCN

In the following, we investigate the impact of the training set size, i.e. the number of samples, used to predict the $N_f = 1893$ values of the test set (which represents 20% of the whole dataset). We here consider time series cross-validation based on a rolling window: fixed-size training and test windows are moved across the dataset. Table IV presents both the averaged RMSE and the standard deviation (SD) obtained over 5 splits, except when 80% of the dataset is used for training because the splitting is unique. We can see a slight improvement in prediction with increasing training set size. Prediction remains very good with a training set size smaller than 80%.

Train	Test	RMSE (mean $\pm$ S.D)
20%	20%	0.405 ± 0.044
40%	20%	0.391 ± 0.051
60%	20%	0.396 ± 0.040
80%	20%	0.397

TABLE IV
STABILITY OF ENHANCED-A3T-GCN UNDER DIFFERENT TRAINING AND TESTING RATIOS

C. Anomaly detection performance

Now we investigate the ability of our proposed Enhanced-A3T-GCN to detect anomalies. To do so, simulated anomalies are injected in the test set. To simulate a temporary change in level and variance, they are obtained by adding 100 consecutive random values drawn according to a Poisson distribution of parameter $\lambda = 8$ to the pre-processed real $PM_{2.5}$ concentrations. The resulting time series is considered as the truth-valued data. We consider two types of anomalies:
i) *local* anomalies, injected on data from one sensor;
ii) *global* anomalies, injected on data from all sensors belonging to a given cluster.

Fig.2 depicts the general detection performance in terms of ROC curves and AUC values in both cases. We can observe that our Enhanced-A3T-GCN is able to detect anomalies very well at both sensor and cluster levels, with AUC values respectively equal to 0.98 and 0.97. Further, the figure indicates for each ROC curve the optimal threshold (that maximizes the true positive rate while minimizing the false positive rate) above which a measurement is considered as an anomaly.

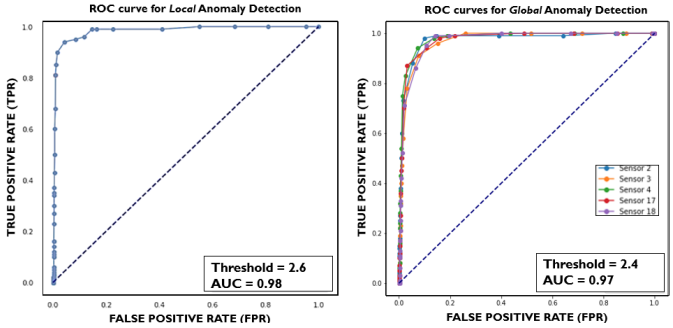

Fig. 2. ROC curves for local and global anomaly detection

An example of local anomaly detection affecting one sensor is depicted in Fig. 3. The yellow curve represents the actual data in which the 100 consecutive anomalies are injected from observation 1450. The blue curve represents the accurate prediction provided by our Enhanced-A3T-GCN and all red points are detected anomalies due to their distance to the actual data above the optimal threshold obtained from the ROC curve. We can note that our proposed approach detects a very high proportion of anomalies, as well as a few points that were not artificially injected but deviate from the true data and could correspond to unknown anomalies already in the original data.

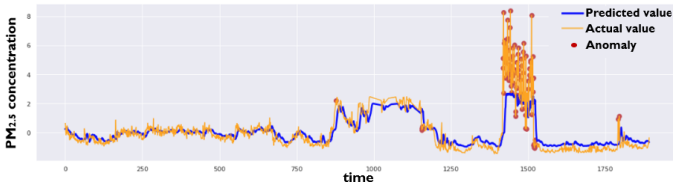

Fig. 3. Local anomaly detection visualization

IV. Conclusion and Future Work

This paper proposes an Enhanced-A3T-GCN pollutant forecasting approach, that embeds in a graph both spatial and temporal correlations in $PM_{2.5}$ measurements alongside with additional information regarding the sensor environment. The latter is obtained by exploiting various qualitative information regarding the location of the sensors, as well as public auxiliary variables. The accuracy of our proposed approach is validated via numerical simulations, highlighting the benefits of exploiting the sensor environment knowledge compared to using more classical spatio-temporal correlations. Further, our Enhanced-A3T-GCN is shown to outperform other deep learning forecasting methods from the literature. Finally, our proposed approach is also shown to efficiently help detecting simulated anomalies, corresponding to a temporary change in level and variance of measurements, at both single and multiple sensor levels. In future work, we will investigate its ability to detect other types of anomalies, such as punctual ones, e.g. corresponding to engine start or pollution peak. We will also study the effect of seasonality on the forecasting performance by considering data acquired over different times of the year. Additionally, we intend to further enhance the accuracy of our forecasting and anomaly detection method by including several meteorological parameters, as well as by considering other pollutant concentrations.

Acknowledgment

This work has been supported by the French ANR agency under contract No. ANR-21-LCV1-0007-01 IAM-Lab (Innovative Air Monitoring Laboratory). The authors would also like to thank Mohammed Achbouk for the implementation of the A3T-GCN algorithm during his research internship at IMT Nord Europe within the Centre for Digital Systems.

References

[1] M. A. Zaidan, Y. Xie, N. H. Motlagh, B. Wang, W. Nie, P. Nurmi, S. Tarkoma, T. Petäjä, A. Ding, and M. Kulmala, "Dense air quality sensor networks: Validation, analysis, and benefits," *IEEE Sensors Journal*, vol. 22, no. 23, pp. 23 507–23 520, 2022.

[2] T.-B. Ottosen and P. Kumar, "Outlier detection and gap filling methodologies for low-cost air quality measurements," *Environmental Science: Processes & Impacts*, vol. 21, no. 4, pp. 701–713, 2019.

[3] M. Van Poppel, P. Schneider, J. Peters, S. Yatkin, M. Gerboles, C. Matheeussen, A. Bartonova, S. Davila, M. Signorini, M. Vogt *et al.*, "SensEURCity: A multi-city air quality dataset collected for 2020/2021 using open low-cost sensor systems," *Scientific Data*, vol. 10, no. 1, p. 322, 2023.

[4] Y. Zhang, N. Meratnia, and P. Havinga, "A taxonomy framework for unsupervised outlier detection techniques for multi-type data sets," *Computer*, vol. 49, no. 3, pp. 355–363, 2007.

[5] A. Blázquez-García, A. Conde, U. Mori, and J. A. Lozano, "A review on outlier/anomaly detection in time series data," *ACM Computing Surveys (CSUR)*, vol. 54, no. 3, pp. 1–33, 2021.

[6] V. Van Zoest, A. Stein, and G. Hoek, "Outlier detection in urban air quality sensor networks," *Water, Air, & Soil Pollution*, vol. 229, pp. 1–13, 2018.

[7] M. E. Aslan and S. Onut, "Detection of Outliers and Extreme Events of Ground Level Particulate Matter Using DBSCAN Algorithm with Local Parameters," *Water, Air, & Soil Pollution*, vol. 233, no. 6, p. 203, 2022.

[8] D. J. Hill and B. S. Minsker, "Anomaly detection in streaming environmental sensor data: A data-driven modeling approach," *Environmental Modelling & Software*, vol. 25, no. 9, pp. 1014–1022, 2010.

[9] D. Topping, D. Watts, H. Coe, J. Evans, T. J. Bannan, D. Lowe, C. Jay, and J. W. Taylor, "Evaluating the use of Facebook's Prophet model v0. 6 in forecasting concentrations of NO_2 at single sites across the UK and in response to the COVID-19 lockdown in Manchester, England," *Geoscientific Model Development Discussions*, pp. 1–22, 2020.

[10] J. Park, Y. Seo, and J. Cho, "Unsupervised outlier detection for time-series data of indoor air quality using LSTM autoencoder with ensemble method," *Journal of Big Data*, vol. 10, no. 1, pp. 1–24, 2023.

[11] Z. Z. Darban, G. I. Webb, S. Pan, C. C. Aggarwal, and M. Salehi, "Deep learning for time series anomaly detection: A survey," *arXiv preprint arXiv:2211.05244*, 2022.

[12] X. Lin, H. Wang, J. Guo, and G. Mei, "A Deep Learning Approach Using Graph Neural Networks for Anomaly Detection in Air Quality Data Considering Spatiotemporal Correlations," *IEEE Access*, vol. 10, pp. 94 074–94 088, 2022.

[13] K. Choi, J. Yi, C. Park, and S. Yoon, "Deep learning for anomaly detection in time-series data: review, analysis, and guidelines," *IEEE Access*, vol. 9, pp. 120 043–120 065, 2021.

[14] J. Bai, J. Zhu, Y. Song, L. Zhao, Z. Hou, R. Du, and H. Li, "A3T-GCN: Attention temporal graph convolutional network for traffic forecasting," *ISPRS International Journal of Geo-Information*, vol. 10, no. 7, p. 485, 2021.

[15] S. Rana, N. H. Barna, and J. A. Miller, "Exploring the Predictive Power of Correlation and Mutual Information in Attention Temporal Graph Convolutional Network for COVID-19 Forecasting," in *International Conference on Big Data*. Springer, 2023, pp. 18–33.

[16] D. Iskandaryan, F. Ramos, and S. Trilles, "Graph Neural Network for Air Quality Prediction: A Case Study in Madrid," *IEEE Access*, vol. 11, pp. 2729–2742, 2023.

[17] J.-H. Liu, N. T. Corbita Jr, R.-M. Lee, and C.-C. Wang, "Wind turbine anomaly detection using Mahalanobis distance and SCADA alarm data," *Applied Sciences*, vol. 12, no. 17, p. 8661, 2022.