Distribution Guided Neural Disaggregation of PM10 and O3 Hourly Concentrations from Daily Statistics and Low-Cost Sensors

Evangelos Bagkis, Theodosios Kassandros, Kostas Karatzas

Distribution guided neural disaggregation of PM_{10} and O_3 hourly concentrations from daily statistics and low-cost sensors

Evangelos Bagkis[1[0000-0002-7761-6415]], Theodosios Kassandros[1[0000-0002-7931-0532]] and Kostas Karatzas[1[0000-0002-1033-5985]]

[1] Environmental Informatics Research Group, School of Mechanical Engineering, Aristotle University of Thessaloniki, 54124 Thessaloniki, Greece

Abstract. It is common for state-of-the-art research to demand higher granularity data to effectively model the atmospheric composition and personal exposure to air pollution. With the advent of Low-Cost Sensors (LCS) technology, the potential of increased spatiotemporal monitoring resolution arises, however, low cost comes with reduced measurement quality. On-site calibration via supervised machine learning (ML) is the most promising technique for the operational calibration of such devices. This study aims (a) to introduce the distribution guided neural disaggregation (DGND) method to increase the temporal resolution of air quality (AQ) low frequency data based on LCS high frequency readings and (b) simultaneously learn a calibration function with the ability to infer over the hourly resolution but with daily supervision. Towards this two-fold objective we propose an indirect training loss based on the first and second distribution moments errors to optimize a multi-layer perceptron (MLP). DGNDs generalization performance is compared against a traditionally trained MLP with the same architecture on a withheld test set in terms of errors and linearity. Furthermore, using the same metrics, the disaggregation results are evaluated on the original time series from which the reference moments originated. Results suggest that modeling the disaggregated (hourly) resolution of PM_{10} and O_3 concentrations is feasible from aggregated (daily) information indicated by modest to high linearity with coefficient of determination R^2 between 0.57 – 0.69 on the test set (except Sindos PM_{10} where $R^2<0$), and 0.49 – 0.83 on the original time series accompanied by moderate to low errors.

Keywords: multi-layer perceptron; disaggregation; nowcasting; low-cost sensors; on-site calibration; air quality.

1 Introduction

1.1 Air quality monitoring

Spatially and temporally representative monitoring of air quality (AQ) is especially relevant for the urban atmospheric environment because 1) most pollution sources such as traffic and domestic heating are present and 2) most anthropogenic activities take

place. Currently, the AQ in Europe is monitored with high precision reference analyzers on the basis of procedures and methods detailed in the EU Directive on Ambient Air (2008/50/EC) [1] however, the high costs related with this kind of instrumentation results in sparse monitoring networks. In practice, these instruments are operated by experts under strict protocols, are calibrated regularly and therefore provide the most reliable measurements of pollutant concentrations. During the last decade, the low-cost sensor (LCS) technology has developed rapidly, and several solutions have entered the market concerning ambient air quality monitoring, for a fraction of the cost of a reference analyzer. LCS is a compelling technology as it reduces the expenses of obtaining and maintaining a reference AQ monitoring instrument, and thus offers the opportunity to establish dense AQ networks in urban areas [2, 3]. Commercially available LCS can be operated by companies, citizen science projects and interested individuals. Research, however, has shown that LCS responses are subject to aging, drifts and cross sensitivity degradation effects among others, and operation requires identification of such problems and careful calibration before and after deployment [4]. To address the high costs associated with reliable data and the reduced accuracy of the LCS simultaneously, a combined AQ network can be achieved where the reference analyzers facilitate a reference network against whose measurements, a complementary LCS network is calibrated, on-site, and in real time. Several studies [5] continue to confirm that this approach can help increase the accuracy of LCS to comply with the data quality objective of the relevant EU Directive [6]. Furthermore, the resolution increases 10-fold, potentially 100-fold with allegedly minimal maintenance costs.

1.2 LCS calibration as time series nowcasting

On-site LCS calibration can be viewed through the prism of nowcasting. According to the World Meteorological Organization (WMO) nowcasting is defined on the basis of meteorological systems, as the estimation of a continuous variable in the present and up to six hours in the future [7]. Supervised ML nowcasting of an AQ time series depends on the temporal resolution of official high-quality data provided mainly by local officials and the European Environmental Agency (EEA). We will refer to those measurements as ground truth concentrations (GTC) onwards. As the interest for AQ information moves from sparse and regionally representative to personal exposure assessments aided via fixed and mobile LCS, the need for higher quality LCS data arises. As a demonstrative example, there are reference methods (like the gravimetric method for PM) which by definition provides only with daily PM mean concentration estimates, and therefore equivalent automatic methods are employed capable of delivering hourly concentrations. In addition, there are cases where LCS AQ measurements are available in a finer time scale in comparison to the GTC. In all those situations, a sub-timescale calibration model is required when the standard ML supervision (each instance has one ground truth value) is insufficient.

Temporal disaggregation can be achieved either by modeling the univariate time series or through multivariate modeling assisted by higher resolution time series [8]. The former approach applies the autoregressive integrated moving average (ARIMA) model to the univariate time series and relies on modeling the autocorrelation. The latter

approach has the benefit of still including temporal information but can also be assisted with feature engineering techniques to include moving statistics from a multivariate feature space. When correlations between the low frequency target and the high frequency covariates exist then it is possible to accurately reconstruct the temporal patterns [9]. Commonly, pollutant concentrations follow 24-hour diurnal patterns depending on human activities, the weather, solar radiation, cloud coverage, and land-use. In the LCS calibration setting, temperature (T) and relative humidity (RH) are among the main factors that affect the sensors performance [10] and thus, including those into the modeling procedure can drive the model response. Additionally, other pollutant species measurements can lead to further improvement. The aforementioned observables can be considered as the high frequency covariates holding information to guide the disaggregation of the low frequency target.

1.3 Aim of the study

This investigation is a proof-of-concept study to determine the ability of a neural network to learn an hourly model indirectly from the daily GTC probability distribution instead of the hourly GTC. This formulation is interesting because 1) when combined with LCS it opens up the opportunity to temporally disaggregate low to high frequency AQ measurements and 2) learn a high-frequency model for deployment without explicitly having access to low enough granular information.

2 Materials and methods

2.1 Experimental set-up

Aiming to establish a LCS network for AQ monitoring, the KASTOM project [11] operates 33 multi-sensor devices spread throughout the Greater Thessaloniki Area (GTA). Three LoRaWAN transmitters help to collect the data in a local database. All the devices measure the temperature, relative humidity, the pressure (P) and estimate coarse (PM_{10}), fine ($PM_{2.5}$) and ultra-fine (PM_1) particulate matter levels via optical particle counters operating under the orthogonal laser scattering principle [12] (Manufacturer: Plantower PMS5003). Additionally, 20 of the devices also include electrochemical gas sensors (Alphasense) to monitor concentration levels of carbon monoxide (CO), ground level ozone (O_3) and nitrogen dioxide (NO_2) [13]. PM_{10} GTC levels are measured by an analyzer (Eberline FH 62 I-R, reference equivalence with European Standard EN 12341) that uses β-attenuation, while reference O_3 concentration levels are measured with the aid of UV photometry according to European Standard EN 14625. Three devices (including PM as well as gaseous sensors) are collocated with three reference stations. Data obtained from the LCS are collected every minute and are aggregated by their median value into hourly values. The reference data are accessed from the EEA API that exports the latest observations on an hourly basis. For the purposes of this study, the daily mean and std of the GTC are calculated and the hourly measurements are discarded from the training and are only used as test data to validate the proposed method.

2.2 Preprocessing and temporal feature engineering

Only days with a full 24 cycle of observations were identified and included in the modeling thus, the missing values were dropped along with incomplete days. For each pollutant and each station, different time periods are evaluated (Table 1). As the focus is concentrated on the algorithm presentation and evaluation, a subset of all the measurements was selected to demonstrate the method. This analysis is not concerned with the evaluation of the LCS network and the identification of the fittest calibration methods as these results are already published in [11]. As nowcasting is an inherently time dependent task, several feature identification techniques are employed to engineer an input vector that includes temporal patterns and moving distribution properties as well as interactions between the covariates. Initially, the interactions are represented by fractions between the covariates (the LCS readings) and supplement the input vector. Specifically, the following fractions are included: PM_{10}/RH, NO_2/O_3, NO_2/T, O_3/T, RH/PM_{10}, T/P, T/RH, P/RH. To inject the temporal pattern first, 12 lags are extracted from each covariate (original and fraction) to complement the input vector. This technique lets the model infer the short-term evolution. Furthermore, for each of the original variables, the statistical properties (mean, median, std, min, max, and the difference between the min and max) in the last 24-hour and 48-hour range represent the long-term behavior of the system. To elaborate further, for example the fraction PM_{10}/RH can potentially provide information about the relationship between particulate matter and relative humidity, which is a known factor affecting the performance of optical particle counters [14]. The following features are extracted: $(PM_{10}/RH)_{t-1}$ … $(PM_{10}/RH)_{t-12}$, mean(PM_{10}/RH, 24h), mean(PM_{10}/RH, 48h), median(PM_{10}/RH, 24h), median(PM_{10}/RH, 48h), … min_max_difference(PM_{10}/RH, 24h), min_max_difference(PM_{10}/RH, 48h). In this way a total of 320 features were constructed.

Table 1. The measurement periods included in the study.

Pollutant/Station	Start (dd/mm/yyyy)	End (dd/mm/yyyy)	Training days
PM_{10} Agia Sofia	06/11/2019	30/4/2020	70
PM_{10} Kordelio	06/11/2019	17/11/2021	500
PM_{10} Sindos	06/11/2019	30/4/2020	250
O_3 Agia Sofia	06/11/2019	30/05/2020	100
O_3 Kordelio	06/11/2019	30/09/2020	300
O_3 Sindos	06/11/2019	30/10/2020	150

2.3 Distribution guided neural disaggregation

Apart from the state-of-the-art performance in a multitude of tasks [15, 16], artificial neural networks (ANN) are particularly interesting for their flexibility in processing high dimensional or/and unstructured data (e.g., text, medical imaging data). On the one hand, architectures like MLP do not understand temporal relationships by design in contrast to other architectures such as the long-short term memory (LSTM) ANNs [17]. On the other hand, feature identification/engineering techniques have been

established for time series that can summarize and present the temporal dimension to simpler architectures. Although the LSTM and its successor, the transformer architecture have proven very useful in modeling language from colossal amounts of text, in time series problems it struggles to surpass simpler approaches with expert feature engineering input [18]. This is partly because most time series of interest are non-stationary in contrast to language which is complex and high dimensional but static (the words and syntax do not change with time). Taking advantage of the flexibility we design a simple MLP and modify the training procedure as follows:

Let $X \in \mathbb{R}^N$, where N is the dimension (number of columns) of the dataset, be a sample of the input space and $y \in \mathbb{R}$ the GTC time series. Suppose that an underlying generating mechanism produces (X, y) pairs in discrete time steps. The pairs obey the identically and independently distributed (iid) assumption [19] and the joint distribution probability P(X, y) represents the unknown generating function. The goal is to approximate the true generating function with the MLP function F_w: $X \rightarrow y$, where w are the weights (parameters) of the network. MLP is a directed graph of perceptrons [20] ordered in layers, with each layer interacting with the previous and the next layer via the weights w and the bias b. Forward propagation from a n-layer MLP translates into the following equation.

$$F_w(X) = f_n\big(w^{(n)} f_{n-1}\big(\dots f_2\big(w^{(2)} f_1\big(w^{(1)} X^T + b^{(1)}\big) + b^{(2)}\big) \dots + b^{(n-1)}\big) + b^{(n)}\big) \quad (1)$$

Suppose that M1, M2, ..., Mn are the successive hidden layer sizes with the output layer comprised of one neuron (with linear activation for regression settings). The weight matrix $w^{(1)}$ with dimensions M1xN between the input dataset X with N features, called input layer, and the first hidden layer with M1 nodes, is multiplied with X^T and the bias vector of length M1 is added. A non-linear activation function f_1 is applied resulting in the first layer output. The first layer is connected with the second layer via the $w^{(2)}$ weight matrix with dimensions M2xM1; following the same steps the result passes through the activation function f_2 and so on. Weights are initialized either pseudo-randomly from a known distribution or with more sophisticated approaches such as the Glorot [21] initialization. Training refers to the optimization of a loss function calculated between the estimations the MLP produces and the GTC and is usually achieved via the gradient descent algorithm.

There are three main approaches to implementing the gradient descent algorithm. In stochastic gradient descent, backpropagation of the loss is applied for every instance and has the ability to find global minima for a specific dataset but with increased time requirements during training. In batch gradient descent the full training dataset is processed at once by the neural network but this leads to slow convergence, possibly to local minima [22] and high memory requirements. To compensate, commonly the third approach is followed namely, the mini-batch gradient descent, where the dataset is partitioned randomly and processed into batches leading to faster training times and introduces enough stochasticity to effectively avoid local minima. In this setting, we assign temporal meaning to each batch by including a fixed number (24) of hourly LCS measurements (high frequency time series) of complete days. By representing the time

dimension of the problem with the batch size we can effortlessly train an hourly model with daily (low frequency time series) supervision. Therefore, forward propagation of each batch produces a vector $\hat{y}$ of 24 estimations for every hour of the day. We incorporate the MAE or L1 loss function to construct the full loss (Loss) as follows.

$$Loss1 = \lambda \left(\frac{1}{m} \sum_{i=1}^{m} |mean(\hat{y}_{00:00:00}, \dots, \hat{y}_{23:00:00}) - \mu_i| \right) \quad (2)$$

$$Loss2 = \lambda \left(\frac{1}{m} \sum_{i=1}^{m} |std(\hat{y}_{00:00:00}, \dots, \hat{y}_{23:00:00}) - \sigma_i| \right) \quad (3)$$

$$Loss = Loss1 + Loss2 \quad (4)$$

Where i represents the day, μ_i is the true mean, σ_i the true std of the day, m is the total number of training days and λ is the learning rate. It is therefore evident that the full loss we apply, takes into account deviations both in terms of mean and std values. After manual hyperparameter tuning the MLP is chosen with 4 layers, 20 nodes at each layer, λ = 0.0001 (a typically low learning rate), and the leaky rectified linear unit (Leaky ReLu), an activation function not suffering from the "vanishing gradient" or the "dying neuron" problem, [23] for all the studied cases. Finally, the Adam [24] optimizer is employed to implement the backpropagation of the loss.

Algorithm 1: Distribution Guided Neural Disaggregation

Input: Training tensors X_{MxN} , $\mu_{M/24}$, $\sigma_{M/24}$, N features, M hourly steps

Step 1: Construct MLP graph with n layers

Step 2: Initialize weights from the uniform distribution $U(-\sqrt{1/N}, \sqrt{1/N})$

Step 3: For epoch until max_epochs:

For each complete day i:

$x_i \leftarrow X^i_{(00:00:00,\ 23:00:00)xN}$

$\hat{y}_i \leftarrow F_w(x_i)$ calculated with eq. (1)

$\hat{\mu}_i \leftarrow mean(\hat{y}_i)$

$\hat{\sigma}_i \leftarrow std(\hat{y}_i)$

$loss \leftarrow Loss(\hat{\mu}_i, \hat{\sigma}_i, \mu_i, \sigma_i)$ calculated with eqs (2), (3), (4)

$w_{i+1} \leftarrow Adam(w_i, loss)$

Output: $\hat{y}_M \leftarrow F_w(X_{MxN})$

2.4 Evaluation

ML models are usually validated with train-validation-test splits or with various forms of cross-validation [25]. In the former approach, the model is trained on the training subset and continuously evaluated on the validation subset after every epoch. The validation subset serves as a test proxy for the metrics to be compared against the training metrics aiming to find an equilibrium to the bias-variance tradeoff. The generalization error is then calculated on the test subset to provide a more robust evaluation on data that are completely new for the model. In cross-validation (CV), the dataset splits into k randomly separated subsets; the first k-1 subsets provide the fitting set and the last subset the validation set. Once the metrics are obtained from the first run, the procedure repeats another k-1 times until every instance has been validated as a test instance once. This technique is used either for validation or for hyperparameter tuning. Several extensions exist such as time-block CV, spatial CV and forward CV. Forward CV is indicated as the more appropriate method for time series problems [26]. In this setting however, all of the aforementioned methods are not directly applicable because disaggregation is applied to historical data. Therefore, the evaluation metrics are calculated on the hourly time series from which the daily statistics were calculated. Addressing the second research objective, a direct comparison responding to generalization performance between the traditional MLP and the DGND approach must be made on a withheld test set thus we split the original dataset into training and test; the disaggregation performance is evaluated on the training subset and the generalization performance on the test set. Moreover, the daily and the hourly loss are calculated during the DGND training to investigate and compare their behavior. The evaluation metrics are presented in Table 2.

Table 2. Evaluation metrics.

Metric	Symbol	Formula
Pearson correlation	R	$\frac{\sum_{i=1}^{n} y_i \hat{y}_i - n\mu_i\hat{\mu}_i}{\sqrt{(\sum_{i=1}^{n} y_i^2 - n\mu_i^2)(\sum_{i=1}^{n} \hat{y}_i{}^2 - n\hat{\mu}_i{}^2)}}$
Coefficient of determination	R^2	$1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$
Root mean squared error	RMSE	$\sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}}$
Mean absolute error	MAE	$\frac{1}{n}\sum_{i=1}^{n} \lvert y_i - \hat{y}_i \rvert$

3 Results

The evaluation results of the DGND approach are presented in Table 3 and for the standard MLP in Table 4. Regarding the former approach, we observe that in all stations and for both pollutants, the disaggregated time series show high linearity as the R^2 and R indicate. When compared against the training metrics the errors increase substantially which is expected considering that one approach "sees" the GTC directly while the other indirectly. Therefore, to provide a more representative comparison on performance degradation, we employ a test set to calculate the metrics. The errors slightly increase for the DGND method which indicates that even though some information is lost in aggregation, a larger part can be reconstructed through this method. Overall, the Ozone GTC is modeled with higher precision than PM_{10} in all cases. This is explained partly because Ozone follows a clear diurnal (daily) pattern due to the photochemistry production mechanism that is catalyzed by ultraviolet radiation. On the contrary, PM_{10} diurnal pattern at the city center is strongly correlated with human activities and in Thessaloniki, shows two peaks usually at 10:00 and at 24:00 local time [27], directly related with traffic and commercial activities in the area [28]. This is a more complex pattern to model as it evolves according to human irregular schedules (reduced traffic on weekends, holidays, daily peaks shifted according to season, etc.). Moreover, focusing on Sindos PM_{10} metrics, we observe that in the test set of both approaches the linearity is lost in terms of R^2. Interestingly, the disaggregation still works as well as for the other two stations. This indicates that if a time series can be modeled, even when the extrapolation errors are high on the test set, the disaggregation can work in case the metrics are adequate in the training set. One caveat that we didn't address here is that when the actual hourly measurements are unknown, then the model can't be validated in this resolution.

Table 3. The evaluation metrics of the DGND approach. Abbreviations: Kord stands for Kordelio, Sind for Sindos and Agia stands for Agia Sofia.

Disaggregation metrics (training set)					Generalization metrics (test set)			
	RMSE	MAE	R^2	R	RMSE	MAE	R^2	R
				O_3				
Kord	14.93	11.00	0.76	0.88	15.65	12.19	0.70	0.86
Agia	11.10	7.96	0.83	0.91	15.50	11.94	0.69	0.84
Sind	15.44	11.23	0.71	0.85	15.95	12.26	0.67	0.83
				PM_{10}				
Kord	15.25	10.54	0.49	0.73	13.43	9.76	0.58	0.77
Agia	14.44	9.94	0.75	0.87	13.94	8.85	0.57	0.87
Sind	13.78	8.27	0.59	0.79	18.72	15.52	-3.52	0.45

Table 4. The evaluation metrics of the traditional MLP approach. Abbreviations: Kord stands for Kordelio, Sind for Sindos and Agia stands for Agia Sofia.

	Training metrics (training set)				Generalization metrics (test set)			
	RMSE	MAE	R^2	R	RMSE	MAE	R^2	R
				O_3				
Kord	9.28	6.45	0.91	0.95	14.55	10.94	0.74	0.88
Agia	7.76	5.07	0.92	0.96	14.64	11.62	0.72	0.86
Sind	10.12	7.46	0.88	0.94	11.77	8.88	0.82	0.91
				PM_{10}				
Kord	10.79	6.77	0.74	0.87	11.65	8.01	0.69	0.83
Agia	11.00	6.87	0.86	0.93	12.13	7.81	0.68	0.83
Sind	11.99	6.39	0.69	0.84	12.76	10.77	-1.10	0.59

In Fig. 1 we present the evaluation of the DGND method for O_3 in Agia Sofia. The slope of the regression analysis represents the accuracy in the LCS calibration setting; therefore, the disaggregated time series accuracy is 92%. The shape of the predicted distribution resembles the observed distribution, and the peaks are accurately estimated; however, in the lower concentration range (0 - 10 $\mu g/m^3$) predictions are noisier. More importantly, it can be observed that the hourly loss (calculated only for comparison) is driven by the daily loss. This is evident as the daily loss decreases slowly and remains high (>10 $\mu g/m^3$) but the hourly loss continues its decreasing trajectory. Given that the DGND method outputs a trained MLP, the generalization performance is depicted in Fig. 2; the accuracy is 88% and the distributions match as well. Interestingly, the training and the testing GTC distributions differ visibly but nonetheless the diurnal pattern is present in both datasets and the model discovers it from the LCS measurements. It should be mentioned that the DGND approach needs approximately twice the number of epochs to converge compared with the traditional MLP training. This is expected considering that in the former approach the MLP learns with 12 times less supervision because for every 24 hours only the mean and std are available.

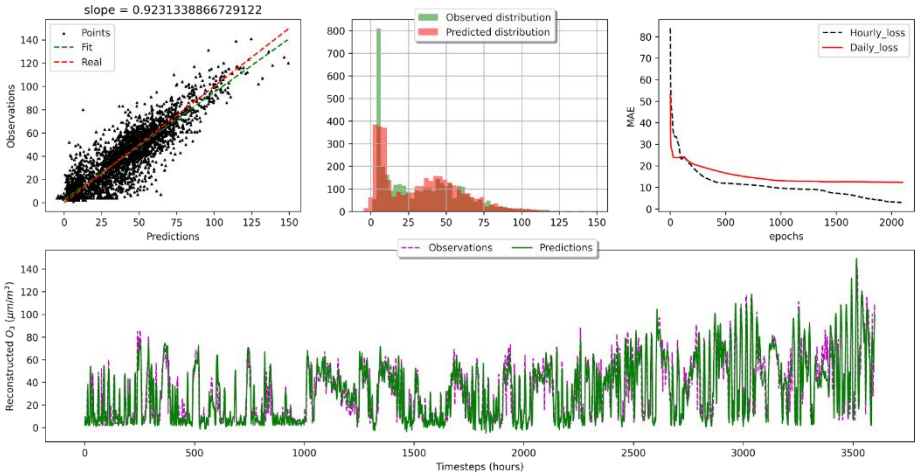

Fig. 1. Disaggregation results for O_3 in Agia Sofia station. Upper left panel: scatter plot, upper center panel: comparison between the real and predicted distributions, upper right panel: loss progression against epochs calculated for both resolutions, bottom panel: the reconstructed O_3 time series against the reference time series (hourly).

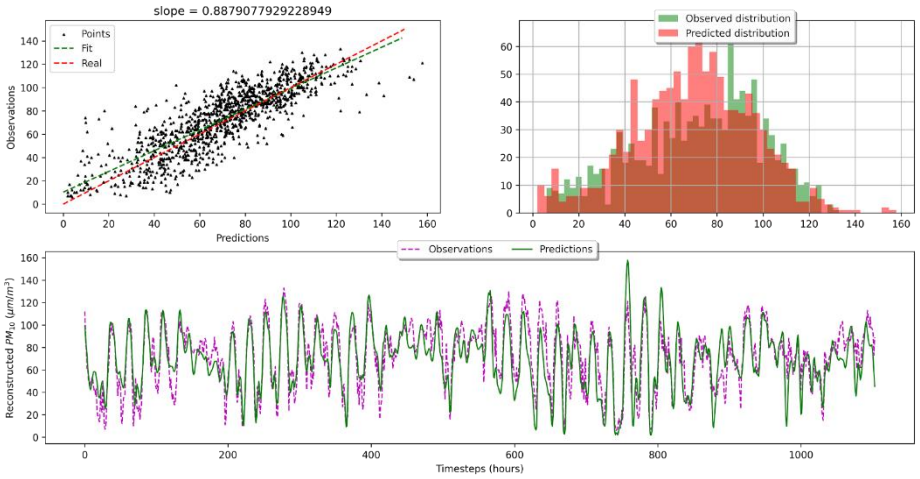

Fig. 2. Generalization results for O_3 in Agia Sofia station. Upper left panel: scatter plot, upper right panel: comparison between the real and predicted distributions, bottom panel: the reconstructed O_3 time series against the reference time series (hourly).

4 Conclusions

The AQ modeling community focuses on increasing the spatiotemporal monitoring resolution to accurately describe the atmospheric composition of urban environments. Towards this goal, we demonstrated the DGND method, capable of increasing the

temporal resolution of highly sought reference GTC. Overall, incorporating LCS to guide the disaggregation procedure based on the flexibility of neural networks shows a promising path in increasing the granularity of expensive measurements. Furthermore, the constructed MLP model has the potential to be deployed without significant loss of accuracy and serve as an on-site calibration function which is suggested from the metrics in all but one cases. Further insights on a consistent way to terminate the training will be the focus of future studies.

Acknowledgments

This research has been co-financed by the European Union and Greek national funds through the Operational Program Competitiveness, Entrepreneurship and Innovation, under the call RESEARCH—CREATE—INNOVATE. Project code T1EDK-01697; project name Innovative system for air quality monitoring and forecasting (KASTOM, www.air4me.eu, accessed on 28 January 2022).

References

1. EUD (European Union Directive). Directive 2008/50/EC of the European Parliament and of the Council of 21 May 2008 on ambient air quality and cleaner air for Europe. Off. J. Eur. Union 2008, L152 (2008).
2. Castell, N.: Low-cost sensors and networks-Overview of current status by the Norwegian Reference Laboratory for Air Quality. NILU report 15/2020, (2020).
3. RIVM data portal, https://sensors.rivm.nl/dataportaal/, last accessed 2022/03/05
4. Peltier, R., Castell, N., Clements, A., Dye, T., Hüglin, Ch, Kroll, J., Candice Lung, SC., Ning, Z., Parsons, M., Penza, M., Reisen, F. von Schneidemesser, E.: An Update on Low-cost Sensors for the Measurement of Atmospheric Composition. WMO, (2021).
5. Bigi, A., Mueller, M., Grange, S., Ghermandi, G., & Hueglin, C.: Performance of NO, NO2 low cost sensors and three calibration approaches within a real world application. Atmospheric Measurement Techniques, 11(6) (2018).
6. Bagkis, E., Kassandros, T., Karteris, M., Karteris, A., & Karatzas, K.: Analyzing and Improving the Performance of a Particulate Matter Low Cost Air Quality Monitoring Device. Atmosphere, 12(2), 251 (2021).
7. Wapler, K., de Coning, E., & Buzzi, M.: Nowcasting. Reference Module In Earth Systems And Environmental Sciences (2019).
8. Moauro, F., & Savio, G.: Temporal disaggregation using multivariate structural time series models. The Econometrics Journal, 8(2), 214-234 (2005).
9. Banbura, M., Giannone, D., Modugno, M. and Reichlin, L.: "Now-casting and the real-time data flow," in Handbook of Economic Forecasting, Volume 2 Part A, ed. by G. Elliott, and A. Timmermann, Elsevier, 195–237 (2013).
10. Samad, A., Obando Nuñez, D., Solis Castillo, G., Laquai, B., & Vogt, U.: Effect of Relative Humidity and Air Temperature on the Results Obtained from Low-Cost Gas Sensors for Ambient Air Quality Measurements. Sensors, 20(18), 5175 (2020).
11. Bagkis, E., Kassandros, T., & Karatzas, K.: Learning Calibration Functions on the Fly: Hybrid Batch Online Stacking Ensembles for the Calibration of Low-Cost Air Quality Sensor Networks in the Presence of Concept Drift. Atmosphere, 13(3), 416 (2022).

12. Yu, X., Shi, Y., Wang, T., & Sun, X.: Dust-concentration measurement based on Mie scattering of a laser beam. PLOS ONE, 12(8), e0181575 (2017).
13. Park, C., Fergus, J., Miura, N., Park, J., & Choi, A.: Solid-state electrochemical gas sensors. Ionics, 15(3), 261-284 (2009).
14. Di Antonio, A., Popoola, O., Ouyang, B., Saffell, J., & Jones, R.: Developing a Relative Humidity Correction for Low-Cost Sensors Measuring Ambient Particulate Matter. Sensors, 18(9), 2790 (2018).
15. Perone, C., Calabrese, E., & Cohen-Adad, J.: Spinal cord gray matter segmentation using deep dilated convolutions. Scientific Reports, 8(1) (2018).
16. Han, J., Liu, H., Xiong, H., & Yang, J.: Semi-Supervised Air Quality Forecasting via Self-Supervised Hierarchical Graph Neural Network. IEEE Transactions On Knowledge And Data Engineering, 5966 (2022).
17. Jiang, X., Luo, Y., & Zhang, B.: Prediction of PM2.5 Concentration Based on the LSTM-TSLightGBM Variable Weight Combination Model. Atmosphere, 12(9), 1211 (2021).
18. Elsayed, S., Thyssens, D., Rashed, A., Jomaa, H.S., Schmidt-Thieme, L.: Do We Really Need Deep Learning Models for Time Series Forecasting?. https://arxiv.org/abs/2101.02118, last accessed on 2022/03/05
19. Stats exchange, https://stats.stackexchange.com/questions/213464/on-the-importance-of-the-i-i-d-assumption-in-statistical-learning, accessed on 2022/03/05
20. Hounmenou, C., Gneyou, K., & GLELE KAKAÏ, R: A Formalism of the General Mathematical Expression of Multilayer Perceptron Neural Networks. Preprint, (2021).
21. Glorot, X., Bengio, Y.: Understanding the difficulty of training deep feedforward neural networks, Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics, PMLR 9:249-256, (2010).
22. Ge, R., Huang, F., Jin, C., Yuan, Y., Escaping From Saddle Points --- Online Stochastic Gradient for Tensor Decomposition. https://arxiv.org/abs/1503.02101, accessed on 2022/03/05
23. Maas, A. L.: Rectifier Nonlinearities Improve Neural Network Acoustic Models. (2013).
24. Kingma, D.P., Ba, J.: Adam: A Method for Stochastic Optimization. https://arxiv.org/abs/1412.6980, accessed on 2022/03/05
25. Maleki, F., Muthukrishnan, N., Ovens, K., Reinhold, C., & Forghani, R.: Machine Learning Algorithm Validation. Neuroimaging Clinics Of North America. 30(4), 433-445 (2020).
26. Fan, K., Dhammapala, R., Harrington, K., Lamastro, R., Lamb, B., & Lee, Y.: Development of a Machine Learning Approach for Local-Scale Ozone Forecasting: Application to Kennewick, WA. Frontiers In Big Data, 5 (2022).
27. Giannakidou P.A.: Study on aerosols PM10 concentrations in the region of Thessaloniki, MSc Thesis, in Greek, (2021). (DOI: 10.26262/heal.auth.ir.335294)
28. Athanasakis, E., Kassandros, Th., Karatzas, K.: Investigation of traffic and air pollution in Thessaloniki, Greece, under ordinary and COVID-19 pandemic conditions. Environmental Informatics. New perspectives in Environmental Information Systems: Transport, Sensors, Recycling (Kamilaris A., Wohlgemuth V., Karatzas K., Athanasiadis I., eds.), pp. 84-92 (2020). Adjunct Proceedings of the 34th edition of the EnviroInfo, Shaker Verlang, Kassel, Germany, ISBN: 978-3-8440-7628-8.