Dimensionality Reduction for Pollutant Forecasting on a Real Urban Area Test Case

Nixarlidis, Christos; Cotteleer, Léo; Gambale, Alessandro; De Troyer, Tim; Parente, Alessandro

Published in:
Proceedings of the First International Conference Math 2 Product (M2P 2023)

Publication date:
2023

Document Version:
Submitted manuscript

Dimensionality Reduction for Pollutant Forecasting on a Real Urban Area Test Case

Christos Nixarlidis*, 1, 2, 3, Léo Cotteleer[1, 2], Alessandro Gambale[4], Tim de Troyer[2, 3], Alessandro Parente[1, 2]

[1]Université Libre de Bruxelles (ULB), Laboratory of Aero-Thermo-Mechanics, Belgium
[2]Brussels Institute for Thermal-Fluid Systems and Clean Energy (BRITE), Belgium
[3]Vrije Universiteit Brussel (VUB), FLOW Research Group, Brussels, Belgium
[4]BuildWind SPRL, Brussels, Belgium
*Corresponding Author: christos.nixarlidis@ulb.be

ABSTRACT

In recent years, the developments in transportation and industrialization have constituted air pollution as one of the main widespread environmental and health issues. Air quality forecasting has become an important aspect of smart city planning and management [1]. Digital twin technologies, such as sensor networks coupled with machine learning algorithms, can be used to predict and monitor air pollution levels in real-time. This way, government agencies and organizations can take proactive measures to reduce air pollution, to identify its sources and target interventions accordingly to mitigate its negative effects on human health and the environment [2]. An urban air quality prediction framework is developed as a digital twin of the district of Ixelles in Brussels, Belgium, using state-of-the-art Machine/Deep Learning (ML/DL) methodologies to predict the concentrations of key air pollutants.

Throughout the district of Ixelles in Brussels, Belgium, a network of 11 QSENSE-Air sensors by Macq are installed at key locations that cover a total test case area of ~1 km^2. The local meteorological conditions (Temperature, Pressure, Humidity, Precipitation, Wind Direction and Speed) and pollutant concentrations ($PM_{2.5}$, PM_{10}, NO_2 and O_3) are measured with a 10-minute sampling frequency. The date and time of the samples are then converted to numerical values (such as day of the year, day of the week, minute of the day etc.) before they are used as input for the proposed methodologies, so that the different seasonal/temporal patterns can be captured. Dimensionality reduction techniques such as Principal Component Analysis (PCA), Auto Encoders (AE) and Convolutional Auto Encoders (CAE) are employed to reduce the vast volume of input data and therefore drastically decrease the training time of the proposed models, in order to predict the local air quality in real-time.

The aim of the present work is to enable the real-time prediction of the pollutant concentrations ($\boldsymbol{y}$) as a function of the local meteorological conditions ($\boldsymbol{x}$) and time ($\boldsymbol{t}$). Two different approaches have been used in this research. The first one is to teach the direct mapping $\tilde{\boldsymbol{f}}: \boldsymbol{x} \longmapsto \boldsymbol{y}$ with $\tilde{\boldsymbol{f}} \approx \boldsymbol{f}$, considering the full meteorological and time input data. For this purpose, both Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) neural network architectures are employed. The second methodology relies on the assumption that a reduced and optimal basis exists to represent the input data. Therefore, the mapping is trained from a data subset x^{red} to reduce the computational cost: $\boldsymbol{f}: \boldsymbol{x}^{red} \longmapsto \boldsymbol{y}$.

Hence, two steps are required: the reduction of the input data and the regression from the latent space to the output $\boldsymbol{y}$. As far as the dimensionality reduction is concerned, Principal Component Analysis (PCA) [3], Auto-Encoders (AE) [4] and Convolutional Auto-Encoders (CAE) [5] have been used. An extensive sensitivity analysis study was conducted to fine-tune the parameters of the above methodologies, such as the depth of the network, number of trainable parameters etc., which suggested a network depth of 3 layers as the optimal for all NN architectures as a compromise between training time and retained information. Regarding the CNN and LSTM architectures, the number of trainable parameters was kept

to 1024 and 512 respectively.

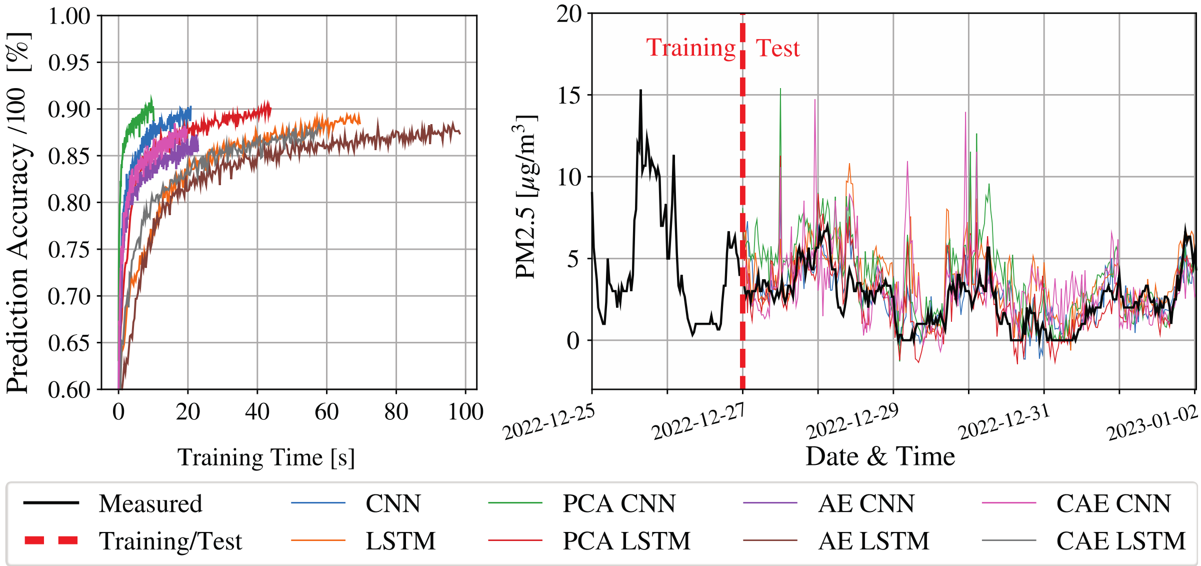

Fig 1. Prediction accuracy vs training time of the proposed methodologies (left) and PM2.5 forecasting results using CNN/LSTM with ROM compared to measured values (right).

One can observe that PCA reduces the training time of both neural network architectures almost by half, while Auto Encoders may actually increase it, as in the case of AE-LSTM. As seen in Fig. 1 (left). The predicted values for $PM_{2.5}$ in Fig. 1 (right) are indicative to showcase the capability of the forecasting framework. Similar predictions are performed for PM_{10}, NO2 and O_3. Even though all NN and ROM model combinations demonstrated a prediction accuracy of over 85%, PCA with CNN was shown to outperform both of the proposed Auto-Encoder methodologies in both time and prediction accuracy.

REFERENCES

[1] Dembski, F., Wössner, U., Letzgus, M., Ruddat, M., and Yamu, C., 2020. Urban digital twins for smart cities and citizens: The case study of Herrenberg, Germany. Sustainability 12, 2307.

[2] Fenger, J., 1999. Urban air quality. Atmospheric environment 33, 4877–4900.

[3] Jolliffe, I. T. and Cadima, J., 2016. Principal component analysis: a review and recent developments. Philosophical transactions of the royal society A: Mathematical, Physical and Engineering Sciences 374.

[4] Wang, Y., Yao, H., and Zhao, S., 2016. Auto-encoder based dimensionality reduction. Neurocomputing 184, 232–242.

[5] Masci, J., Meier, U., Cires ̧an, D., and Schmidhuber, J., 2011. Stacked convolutional auto-encoders for hierarchical feature extraction. Proceedings of Artificial Neural Networks and Machine Learning–ICANN 2011: 21st International Conference on Artificial Neural Networks, Espoo, Finland, June 14-17, 2011, Proceedings, Part I 21. Springer, 52–59.