Forecasting Ozone and Nitrogen Oxides for Air Quality Monitoring

César Bouças, Catarina Silva, Alberto Cardoso, Filipe Araujo, Joel Arrais, Paulo Gil, Bernardete Ribeiro

RECPAD 2020

Abstract

Ozone (O3) and nitrogen oxides (NOx) emissions can harm ecosystems, agriculture and public health through their direct and indirect effects on the air quality. Thus, the ability to predict future concentrations of such gases is of strategic importance, especially in the current climate changing scenario. This work presents three methods to predict O3 and NOx concentrations for the upcoming 24 hours, given a sequence of past window of the same gas concentrations as input: a moving average, a linear regression and a Long short-term memory (LSTM) network that exhibited the best result, being able to forecast NOx series with an average root mean squared error (RMSE) of 115ppb and mean absolute percentage error (MAPE) of 36% with respect to the ground truth series of the test set. The presented strategy was used to empower the NanoSen-AQM air quality platform.

Introduction

- Gas concentrations observed at a regular interval of time (step) consist in a time series that can be used to predict future observations in a process called forecasting [1].
- The forecast aim is to estimate how the observations will sequence into the future.
- In this work, three methods are used to forecast hourly averaged NOx concentrations and two methods were used to forecast hourly averaged O3 concentrations. The number of future steps predicted was set to 24 and only the gas measurements were used as input to forecast future concentrations.
- The proposed methods demonstrated to be simple enough to enable a smooth integration in the NanoSen-AQM online platform [4].

Proposed approach

- A moving average technique and a linear regression model were used as baseline, then a LSTM model was designed to enhance the performance.
    - A NOx series forecast model was designed as a neural network whereas the input sequence containing 72 past measurements is first transformed by a LSTM layer with 20 neurons (units) activated by a hyperbolic tangent function. Then by another LSTM layer with 8 rectified linear units and finally by an identity layer that outputs a sequence of 24 values that corresponds to the predicted future.
- Figure 1 illustrates the architecture. The network was trained through 50 epochs of backpropagation using gradient descent algorithm and mean squared error (MSE) as loss function. The hyperparameters tuned at training were the number of hidden units and its activation functions. Whereas 20 and 8 hidden units with tanh and relu activation functions demonstrated to be sufficient to reach the best average results at training phase.

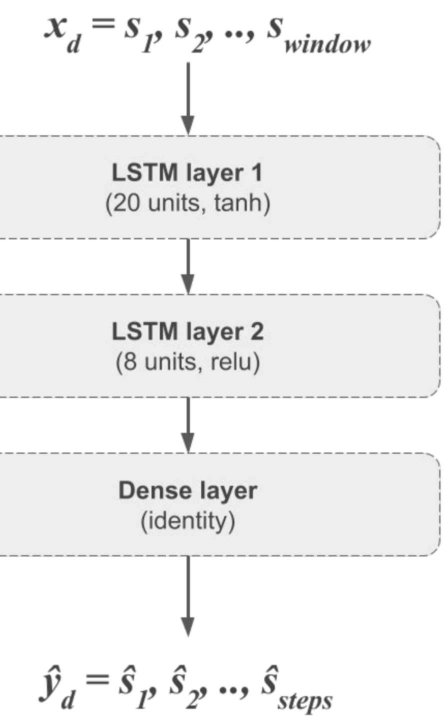

Figure 1: The neural network architecture with the number of neurons/units of each layer and its activation functions.

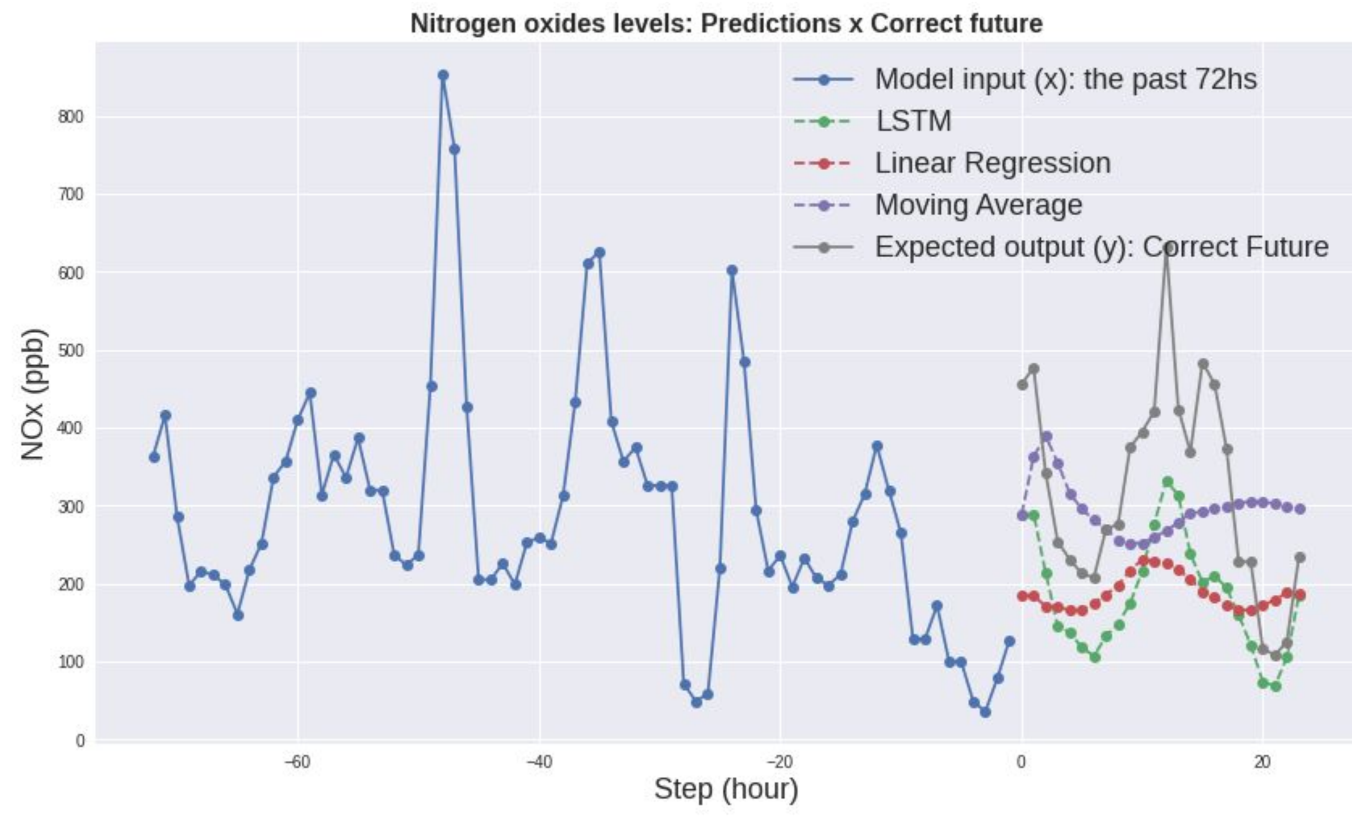

Figure 2: The predictions of the three methods for an example from the Devito's test set.

Experimental Setup

- Series from two datasets were used to develop and test the proposed methods:
    - Averaged Nitrogen Oxides (NOx) concentrations recorded from March 2004 to February 2005 in Italy, that is part of the Air Quality Data Set (Devito) [5].
    - Ozone concentrations measured with reference sensors at Extremadura University campus (Badajoz) from September 21 to September 25 of 2017.
- Two evaluation metrics were used to measure the performance of the models: root mean squared error (RMSE) and mean absolute percentage error (MAPE).
- The datasets were preprocessed to delete empty rows, reescale the values to a range between zero and one, and fill missing values with the last valid observations.

Results

- For each example in the test set, the trained models were used to make a prediction as well as the moving average was calculated.
- RMSE and MAPE were then calculated using the set of predicted values and the ground truth values of the test set. Obtaining the final average errors for each test set: Badajoz and Devito. Tables 1 and 2 summarize the obtained metrics for the test sets.
- Figure 2 illustrates an example from the Devito test set and the predicted outputs for this example. Offering a visual perception of the input, expected output and predictions of each method. The results demonstrated that all the methods should be improved, especially the moving average, which presented much worse metrics than the others methods despite the sufficient visual perception of it's predictions.

Dataset	Gas	Method	RMSE (ppb)	MAPE (%)
Devito	NOx	LSTM	115.21	36
Devito	NOx	Linear Regression	131.34	41
Devito	NOx	Moving Average	215.86	84
Badajoz	O3	Linear Regression	22.91	38
Badajoz	O3	Moving Average	44.98	57

Table1: Results over Devito's NOx test set and Badajoz's O3 test set.

Conclusion and further work

- The proposed methods demonstrated reasonable performances, and were successfully integrated into the NanoSenAQM online platform. Where the baselines showed to be attractive for its simplicity and low memory consumption.
- Methods for automatic seasonality removal should be considered instead of classic manual removal methods, since the latter would not be suitable to be implemented as part of the online platform.
- An exponential strategy may benefit the moving average, giving greater importance to recent measurements in the inputs. As well as the neural network, by developing an exponential smoothing strategy within the LSTM such as [3].
- The use of informative features about temperature, humidity, wind and other factors that have impact in such gases behaviors could benefit the Linear Regression and the LSTM methods.

Contact

César Bouçasr: cesarboucas@dei.uc.pt
Catarina Silva: catarina@dei.uc.pt
Alberto Cardoso: alberto@dei.uc.pt
Filipe Araujo: filipius@uc.pt
Joel Arrais: jpa@dei.uc.pt
Paulo Gil: pgil@dei.uc.pt
Bernardete Ribeiro: bribeiro@dei.uc.pt

References

1. Rob J Hyndman and George Athanasopoulos. Forecasting: principles and practice. OTexts, 2018.
2. Spyros Makridakis, Evangelos Spiliotis, and Vassilios Assimakopoulos. The m4 competition: 100,000 time series and 61 forecasting methods. International Journal of Forecasting, 36(1):54–74, 2020.
3. Slawek Smyl. A hybrid method of exponential smoothing and recurrent neural networks for time series forecasting. International Journal of Forecasting, 36(1):75–85, 2020.
4. Pedro Henrique Saraiva Lucas et al. Development of the server for the NanoSen-AQM Project. PhD thesis, Universidade de Coimbra, 2019.
5. Saverio De Vito, Ettore Massera, Marco Piga, Luca Martinotto, and Girolamo Di Francia. On field calibration of an electronic nose for benzene estimation in an urban pollution monitoring scenario. Sensors and Actuators B: Chemical, 129(2):750–757, 2008.