22nd International Conference on
Harmonisation within Atmospheric Dispersion Modelling for Regulatory Purposes
10-14 June 2024, Pärnu, Estonia

APPLICATION OF MACHINE LEARNING APPROACHES FOR HIGH RESOLUTION EMISSION INVENTORY AND TEMPORAL EMISSION PROFILES

Alessandro Marongiu, Elisabetta Angelino, Gabriele Giuseppe Distefano and Anna Gilia Collalto

Department of Air Emission Inventories, ARPA Lombardia, Milan, 20124, Italy

Abstract: Machine Learning ML approaches are receiving a lot of interest from the scientific community due to their versatility and possibility of combining with the most traditional data processing and modelling techniques. This paper concerns the application of ML techniques in constructing temporal profiles of pollutant emissions and estimating concentrations and emission rates with high accuracy. The input data include emissions from bottom-up inventory, main meteorological parameters (such as precipitation, global solar radiation, temperature, relative humidity, wind speed and direction) and concentrations measured at fixed monitoring sites, characterized from very different local emission sources. The main assumptions in the interactive ML approach are considering a restricted area around the site (obtained by maximum distance in one hour with average wind speed) or moving downwind with respect to the emission source, training and testing of Random Forest on the measured concentrations and turbulence parameters and with a first guess value of the emission rate from the inventory. The methodology is based on the reiteration of test and training of the Random Forest model by correcting hourly emissions for the ratio of measured to estimated concentrations. Interesting results have been obtained which show how the performance of the implemented approach improves as the number of iterations increases, and how meteorological parameters influence atmospheric concentrations and emissions differently, also in relation to the type of monitoring station.

Key words: *Machine Learning, Random Forest, CALPUFF, Dispersion, Emission Rates, Inventory*

INTRODUCTION

Machine learning approaches (ML) are increasingly used in the field of air quality to distinguish the effects deriving from weather conditions from those related to emission sources on atmospheric composition, such as alterations in the PM_1 composition due to Beijing's Clean Air Action Plan (Zhang et al., 2020). They have also been implemented to assess the variation in NO_2, O_3, and $PM_{2.5}$ levels during the COVID-19 lockdown (Petetin et al., 2020; Wang et al., 2020; Shi et al., 2021). This methodology can be extended to study reactive species, such as NH_3, including the effects of gas-to-particle conversion on NH_3 concentrations (Lyu et al., 2023). Neural network-based chemical transport models capture complex relationships between emissions and atmospheric concentrations. These models have been developed to improve the accuracy of emission inventories and to enhance the performance of air quality models. The approach involves back-propagation, which adjusts the gradient of the loss function – a measure of the deviation between predicted and observed contaminant concentrations (Huang et al., 2021). As known, the two main systems for predicting ammonia concentrations include physical methods and machine learning. The physical method involves numerous variables and can be applied at the farm level or for specific manure management tasks. The CTMs are also defined on a physical and phenomenological approach and are utilized on a wider scale. They include chemical reactions, pollutant transport and emission models, incorporating data from various inventories. However, they generally simplify information on temporal variations in emissions. Annual ammonia emission estimates, as generally provided by emission inventories, are disaggregated within CTMs based on average temporary profiles to obtain emissions on an hourly base, as required by dispersion models. Only during a subsequent reanalysis phase is it possible to

recalculate the emissions by comparing the simulation results with the observed data. Machine learning techniques have been employed in both CTM combinations (Zhang et al., 2020; Petetin et al., 2020; Wang et al., 2020; Shi et al., 2021; Lyu et al., 2023) and single-source resolutions (Hempel et al., 2020). In a recent paper to estimate ammonia concentrations and emissions with high accuracy (Marongiu et al., 2024), we introduced a novel approach that iteratively applies Random Forest machine learning to an assumed homogenous aerographic region surrounding the measurement sites. The proposed model will be used to estimate the ammonia emission trend allowing the validation of annual emission estimates from the emission inventory and to obtain temporal profiles useful for CTMs. In this paper we extend the machine learning approach to different dispersion test cases:

- Case 1: Simple one box case, where ML aims to predict concentrations and emission rates;
- Case 2: Downwind multiple box model with known source, where ML aims to predict the concentration variation over time and distance with a known source and emission rate;
- Case3: Downwind multiple box model with unknown source, where ML aims to predict the variation in time and distance of the concentration and emission rate of the source.

METHODOLOGY

The methodology that recursively applies Random Forest machine learning assumes that aerographic domain around measurement sites is homogeneous to predict concentrations and emissions of ammonia, an atmospheric pollutant that causes acidification and eutrophication of soil and water and contributes to secondary $PM_{2.5}$. The recursive implementation of machine learning is described in table 1. According to Voulgarakis (2021), the NH_3 dispersion model was simplified to a box system, where the main flows that determined the potential accumulation of gaseous ammonia were the rate of emission, the chemical reactions considered completed shifted to the conversion and the deposition.

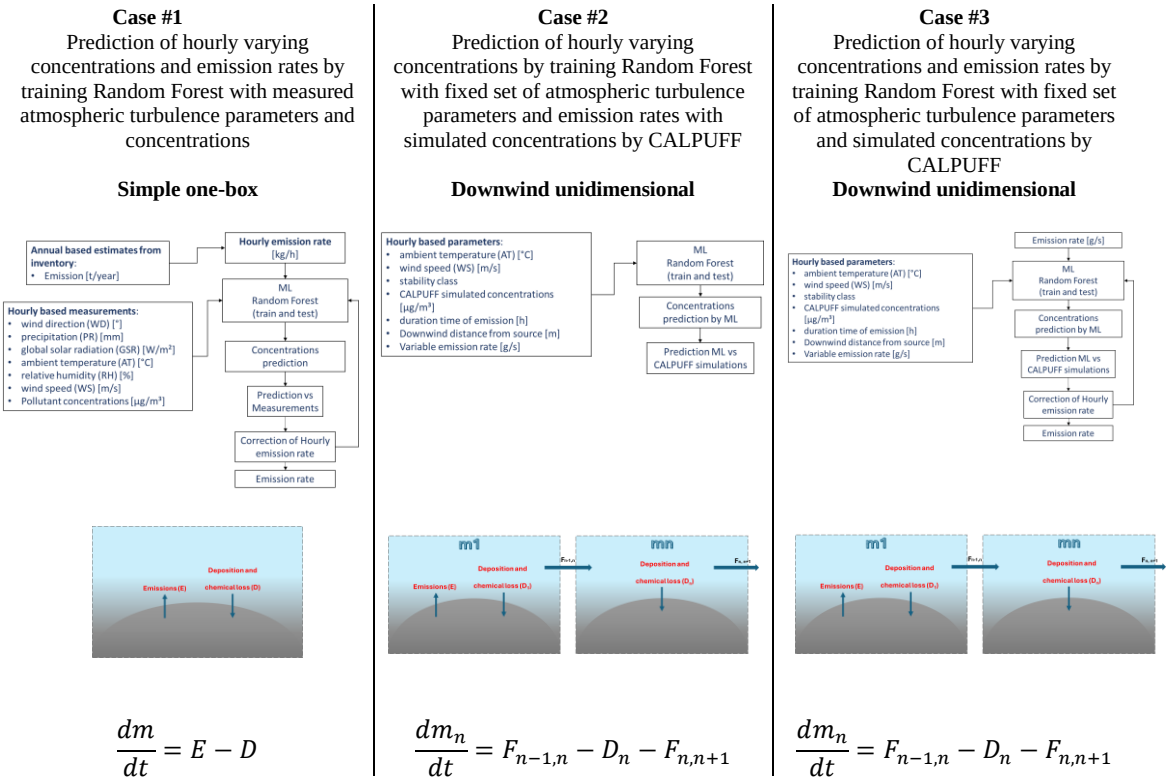

Figure 1. Definition of the test cases for the application of machine learning.

The methodology was implemented to understand the effects of changes in weather and emission on atmospheric ammonia concentrations. The model was trained and tested using hourly measurements of

ammonia concentrations and atmospheric turbulence parameters. This process started with a constant emission scenario. Initially, emissions were calculated based on a detailed bottom-up emission inventory at the municipal level. The calculation considered a circular area with a radius of approximately 4 kilometers centered around the measurement sites. In each iteration, predicted and measured concentrations were compared, leading to changes in emissions. The training and testing of the model were then repeated, achieving a convergence towards remarkably high performance in predicting ammonia concentrations and defining time-varying emission profiles. The validity of this approach has been demonstrated under certain conditions: low dispersion of gaseous ammonia far away from sources and high complementary rate in the transformation of NH_3 in the atmosphere. These types of conditions are well understood in the scientific literature (Asman et al., 1998; Renard et al., 2004) and verified in the Po-basin, where wind velocity is generally very low, in the range of 1 m/s (Marongiu et al., 2024). In this work we want to investigate the application of ML to test more complex dispersion cases and to evaluate if the approach can be able to:

- Simulate downwind concentrations with known emission rates and meteorological parameters;
- Simulate downwind concentrations and emission rate of a single source.

Unlike case number one (table 1), no measured data were available for this type of application and a database of previously simulated cases was adopted. Simulations on the atmospheric dispersion of pollutants were carried out with CALPUFF (Scire et al., 2000). The downwind maximum concentration ratio at different distances from the source is simulated by CALPUFF for more than 5880 different cases. Each record of the simulation is obtained by varying the wind intensity in the range between 1 and 5 m/s, stability class between A and F, time in the range of 1-8 h for a point source of 1 m diameter, 1 m height, temperature 200°C, exit velocity of 5 m/s and emission rate of 1 g/s. For the simulation of case 3 the concentrations and emission rates are multiplied from 1 to 10 times to simulate variable and unknown emission rates.

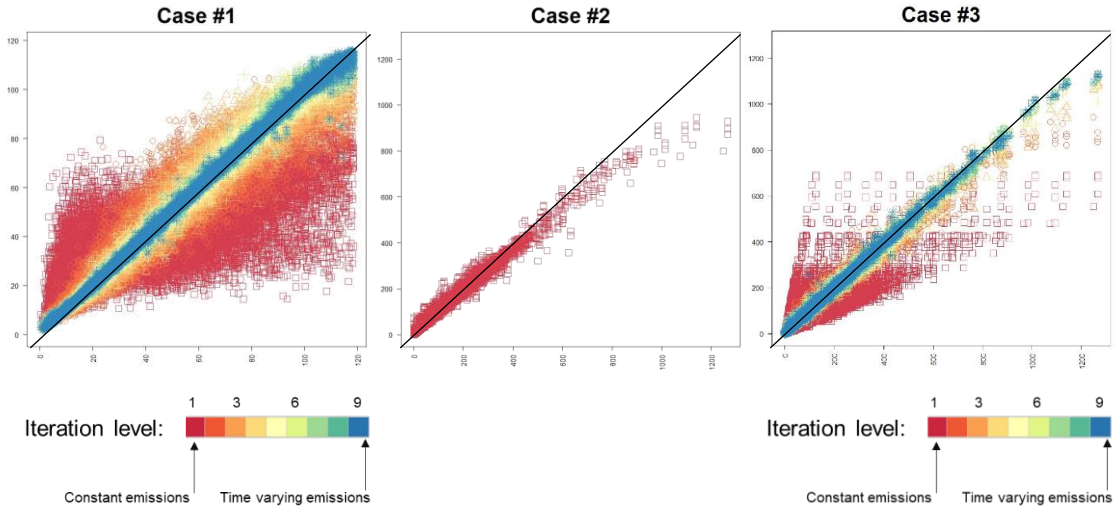

Figure 2. Comparison between concentrations from measurements or CALPUFF simulations and the predicted ones from Machine Learning [$\mu g/m^3$].

RESULTS

For case 1 in Figure 2, it is evident that the correlation between measured concentrations (on the x-axis) and calculated concentrations (on the y-axis) tends to improve as emissions are iteratively corrected by the model. A similar behavior is shown for case 3 while in simulations with known emission rate only one iteration step is needed to train and test the random forest (case 2). The prediction performances for concentrations are similar for all the investigated cases. The effect on the iteration steps is shown in figure 3. The progressive correction of emission rates leads to a better fit of the concentration prediction in case study number 1. Similar results are shown in case number 3, where downwind concentrations are predicted simulated by CALPUFF using machine learning through recursive correction of emission rates. In case study number 2, emission rates are known parameters and considered as input for machine learning predictions. The model's concentration predictions showed accuracy and reliability compared to measured

or simulated values. Furthermore, the relationship between concentrations and calculated emission rates aligns with the physical parameters of atmospheric turbulence.

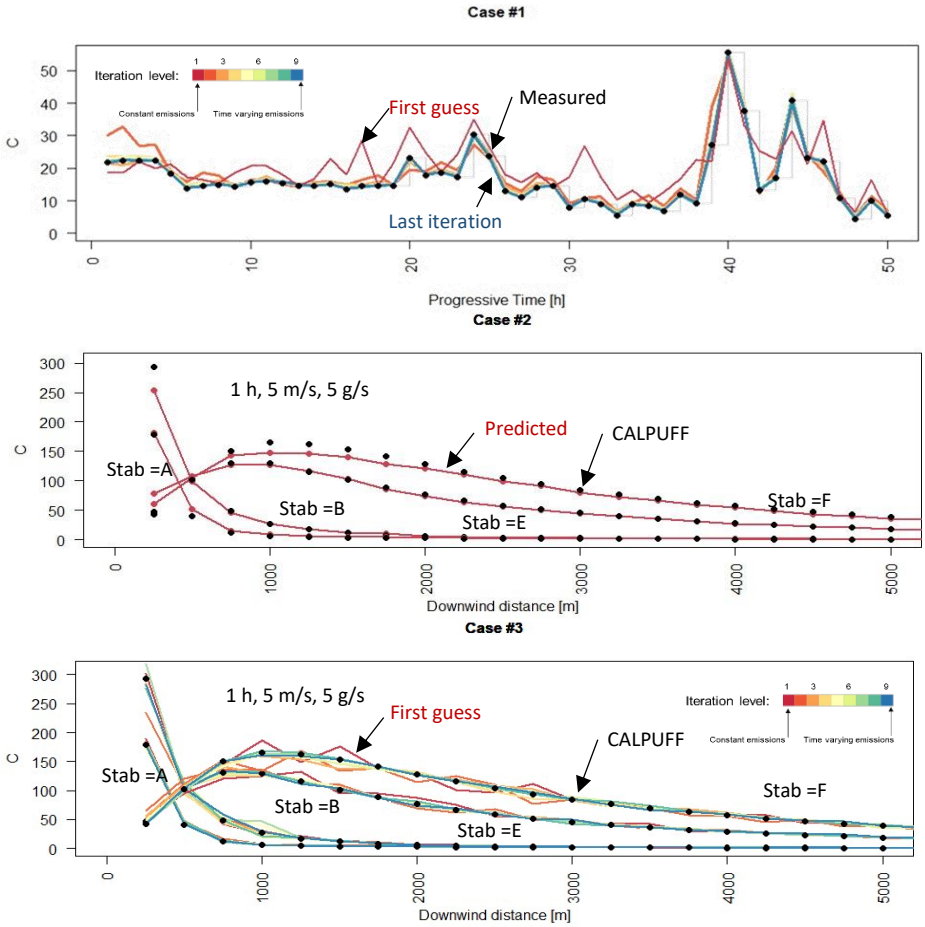

Figure 3. Simulation with time variation and distance of concentrations [µg/m^3].

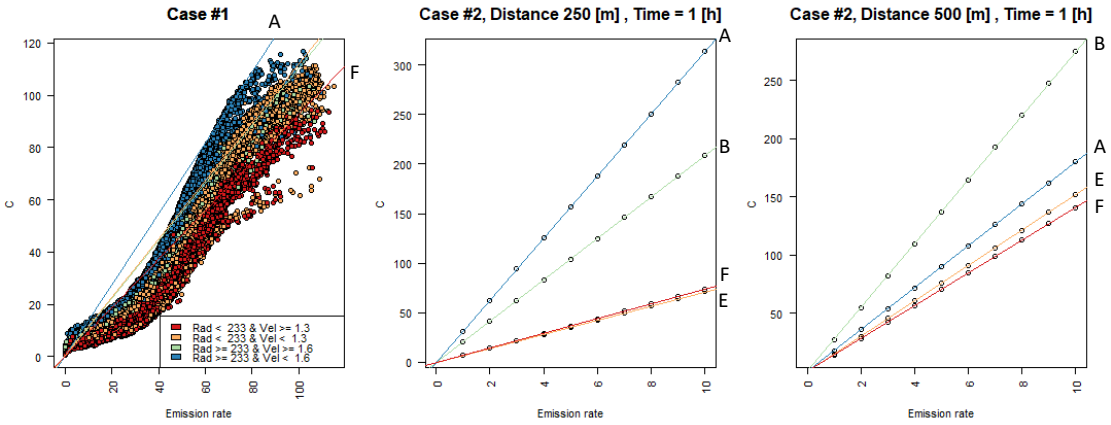

Figure 4. Simulation of concentrations [µg/m^3] at varying emission rates [g/s] and stability classes.

Figure 4 illustrates, for the three cases investigated, the relationship between the emission rates (on the x-axis) and the concentrations (on the y-axis). It highlights various atmospheric stability classes, which depend on the intensity of solar radiation and wind speed, using different color shades. Specifically, in case 1, it is evident that a certain emission value corresponds to a range of concentration values rather than a

single concentration value. This variability observed in concentration levels is a result of the different meteorological conditions that can occur at the same emission level.
It is interesting to note that the classification of possible cases of concentration levels defined on emission rates in case 1 is identified based on solar radiation and wind speed. Higher accumulation in the atmosphere is simulated for higher values of thermal radiation and lower wind velocity, which are typically associated with stability class A. Data obtained from real-world measurements seem describing similar relationships between concentrations and emission rates at the variation of stability classes with a time step of one hour and relative short distance from the emission sources.

CONCLUSIONS

In this paper we have discussed the methodology of applying ML to describe typical physical processes of atmospheric dispersion from known and unknown sources. As reported in previous literature, machine learning seems to be confirmed as a useful tool for estimating emission rates at high temporal resolution, allowing to discriminate high emission episodes from meteorological turbulence effects. All these preliminary investigations suggest that ML must be applied considering physical and conceptual constrains to preserve the relationship as source-receptors effects. These approaches applied in brutal force can fit measurements or other model outputs with very high performances, but the preservation of the key physical constrains suggests defining specific performance indicators for evaluating ML approaches.

AKNWOLEDGMENTS

This research was funded by LIFE-IP PREPAIR (LIFE15 IPE/IT/000013) and REGIONE LOMBARDIA "Progetto di monitoraggio ammoniaca dal comparto agricolo 2022/2023" (2022.0043891—18/03/2022).

REFERENCES

Asman, W. A. H., M. A. Sutton andJ. K. Schjørring, 1998: Ammonia: Emission, atmospheric transport and deposition. *New Phytol.*, **139**, 27–48.

Hempel, S., J. Adolphs, N. Landwehr, D. Janke and T. Amon, 2020: How the Selection of Training Data and Modeling Approach Affects the Estimation of Ammonia Emissions from a Naturally Ventilated Dairy Barn—Classical Statistics Versus Machine Learning. *Sustainability*, **12**, 1030.

Huang, L., S. Liu, Z. Yang, J. Xing, J. Zhang, J. Bian, S. Li, S. K. Sahu, S. Wang and T.-Y. Liu, 2021: Exploring deep learning for air pollutant emission estimation. *Geosci. Model Dev.*, **14**, 4641–4654.

Lyu, Y., Q. Zhang, Q. Sun, M. Gu, Y. He, W. W. Walters, Y. Sun and Y. Pan, 2023: Revisiting the dynamics of gaseous ammonia and ammonium aerosols during the COVID-19 lockdown in urban Beijing using machine learning models. *Sci. Total Environ.*, **905**, 166946.

Marongiu, A., A. G. Collalto, G. G. Distefano and E. Angelino, 2024: Application of machine learning to estimate ammonia atmospheric emissions and concentrations. *Air*, **2**, 38-60.

Petetin, H., D. Bowdalo, A. Soret, M. Guevara, O. Jorba, K. Serradell and C. P. García-Pando, 2020: Meteorology-normalized impact of the COVID-19 lockdown upon NO2 pollution in Spain. *Atmos. Chem. Phys.*, **20**, 11119–11141.

Renard, J. J., S. E. Calidonna, and M. V. Henley, 2004: Fate of ammonia in the atmosphere – a review for applicability to hazardous release. *J. of Hazard. Mat.*, **B108**, 29-60.

Scire, J. S., D. G. Strimaitis and R. J. Yamartino, 2000: A User's Guide for the CALPUFF Dispersion Model, v.5.

Shi, Z., C. Song, B. Liu, G. Lu, J. Xu, T. Van Vu, R. J. R. Elliott, W. J. Bloss and R. M. Harrison, 2021: Abrupt but smaller than expected changes in surface air quality attributable to COVID-19 lockdowns. *Sci. Adv.*, **7**, eabd6696.

Voulgarakis, A., Atmospheric modelling, 2021, *3rd ECMWF, EUMETSAT and ESA Training in Atmospheric Composition*, 6-17 December 2021.

Wang, Y., Y. Wen, Y. Wang, S. Zhang, K. M. Zhang, H. Zheng, J. Xing, Y. Wu and J. Hao, 2020: Four-Month Changes in Air Quality during and after the COVID-19 Lockdown in Six Megacities in China. *Environ. Sci. Technol. Lett.*, 7, 802–808.

Zhang, Y., T. V. Vu, J. Sun, J. He, X. Shen, W. Lin, X. Zhang, J. Zhong, W. Gao, Y. Wang et al., 2020: Significant Changes in Chemistry of Fine Particles in Wintertime Beijing from 2007 to 2017: Impact of Clean Air Actions. *Environ. Sci. Technol.*, 54, 1344–1352.