Application of Machine Learning Approaches for High Resolution Emission Inventory and Temporal Emission Profiles

Alessandro Marongiu,
Elisabetta Angelino, Gabriele G. Distefano, Anna Gilia Collalto

ARPA Lombardia – Environmental Protection Agency of Lombardy - ITALY
Atmospheric Emission Inventory Unit
a.marongiu@arpalombardia.it

Goal Definition

Background:

- Emission Inventories at national and local scale are defined on annual basis and very often are defined on the same technical guidance (eg. EMEP EEA Guidebook).
- Bottom-up emission estimates can be very detailed considering the spatial distribution of the emissions.
- The link between emission inventories and modelling systems passes trough a time disaggregation and often by a re-spatialization on a grid
- The modelling systems can assimilate on-situ measurements end of chain from the input of the emission inventory

Goal:

- To find out a shortcut methodology able to estimate time-varying air concentrations and emission rates. The general idea is to get an hourly varying emission rate obtained by the combination of in-situ or remote sensing measures with local emission inventories.
- Find a general approach able to identify the controlling variables for adjusting time varying emission rates. They could be either meteorological parameters and pollutant concentrations.

Case studies

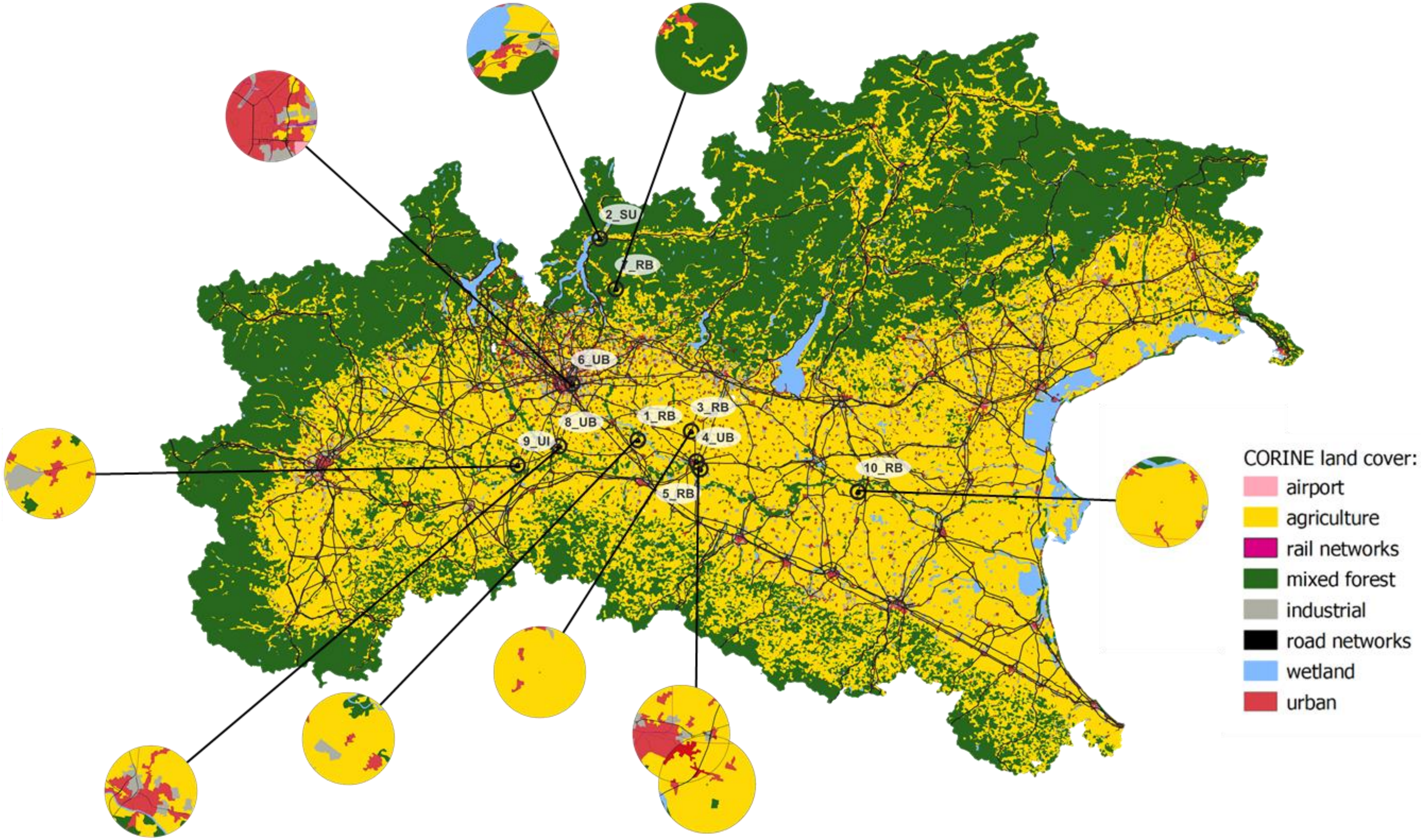
Simple One-Box (SOB): ML aims to predict concentrations and emission rates;

Downwind Unidimensional with Know Source (DUKS): where ML aims to predict the concentration variation over time and distance with a known source and emission rate;

Downwind Unidimensional with Unknow Source (DUUS): ML aims to predict the variation in time and distance of the concentration and emission rate of the source.

Simple one-box

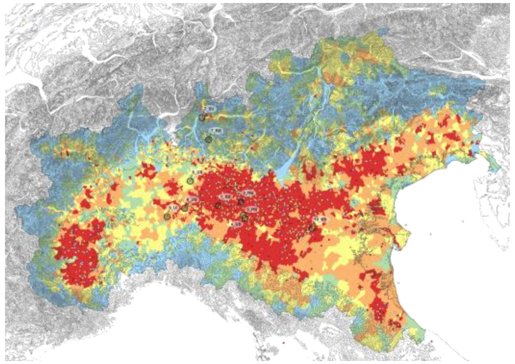
- Applied on in-situ measured data of Ammonia
- We consider an area with a radius of 3.6 km around the site (maximum distance in an hour with a wind velocity of 1 m/s).
- Training and testing of Random Forest on the measured hourly ammonia concentrations and turbulence parameters and with a first guess value of the emission rate of NH_3 from the inventory.
- Reiteration of test and training of the Random Forest model correcting the hourly emissions by the ratio between measured and estimated concentrations.

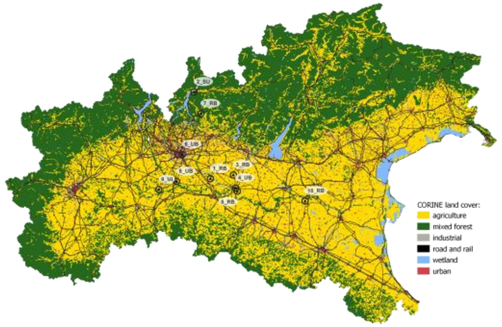


Monitoring sites: RB: rural background; UB: urban background; SU: suburban background; UI: urban industrial

NH_3 concentrations and meteorological parameters

Annual based estimates from inventory:
- Emission [t/year]





Hourly based measurements:
- wind direction (WD) [°]
- precipitation (PR) [mm]
- global solar radiation (GSR) [W/m²]
- ambient temperature (AT) [°C]
- relative humidity (RH) [%]
- wind speed (WS) [m/s]
- Pollutant concentrations [µg/m³]

Hourly emission rate
[kg/h]

ML
Random Forest
(train and test)

Concentrations prediction

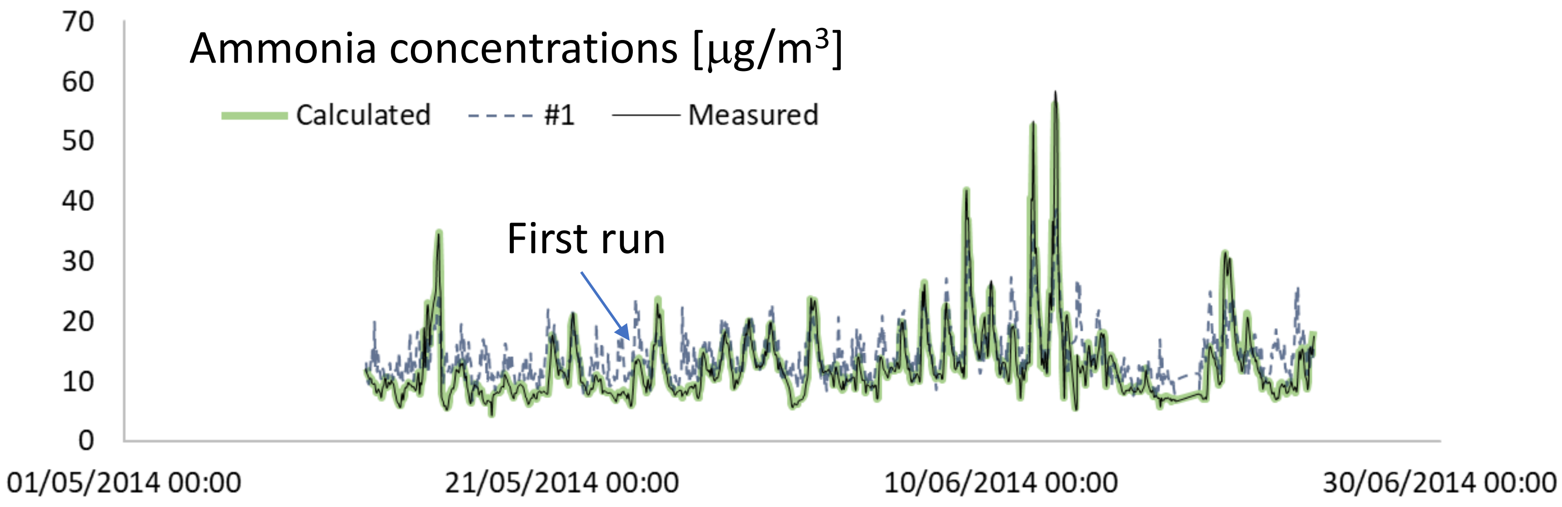
Prediction vs Measurements

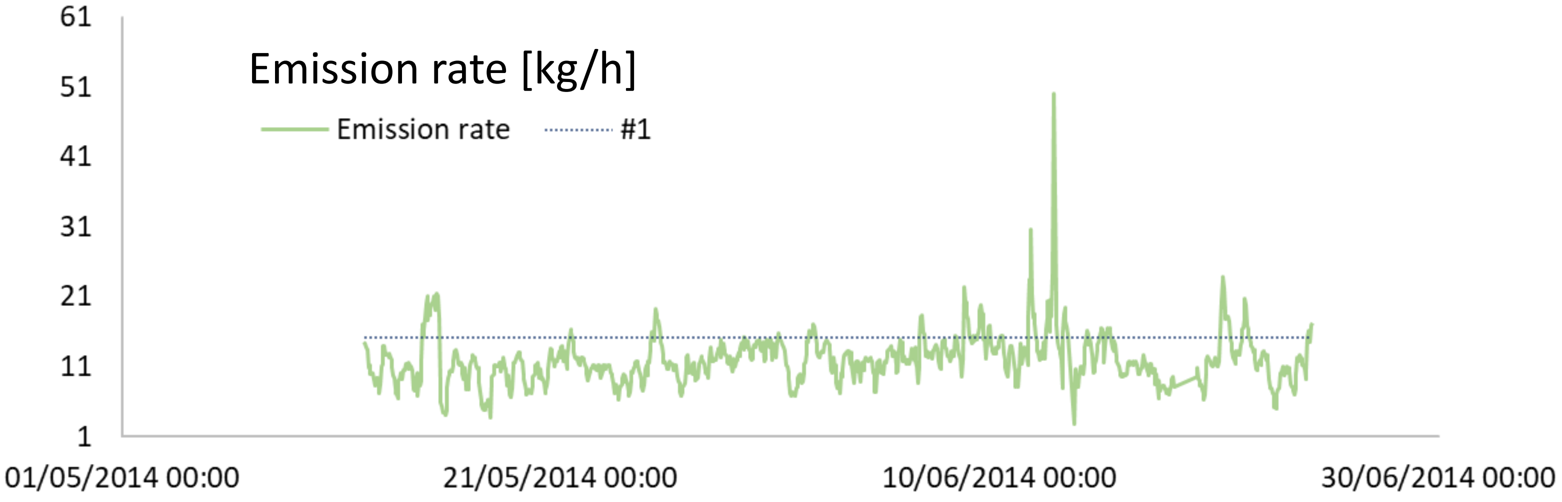
Correction of Hourly emission rate

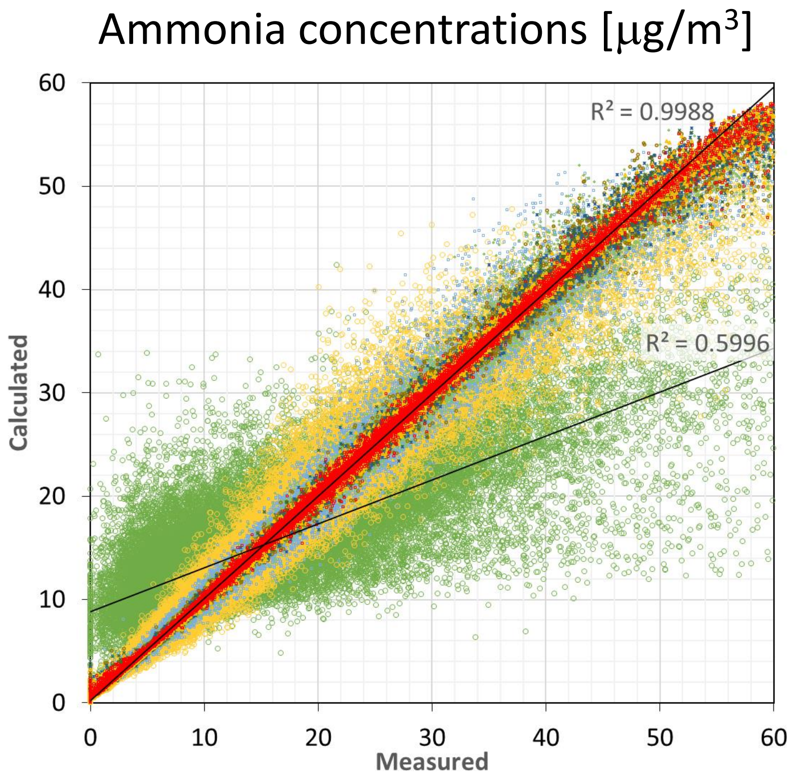
Emission rate

Simple one-box (SOB)

Iterative application of Random Forest



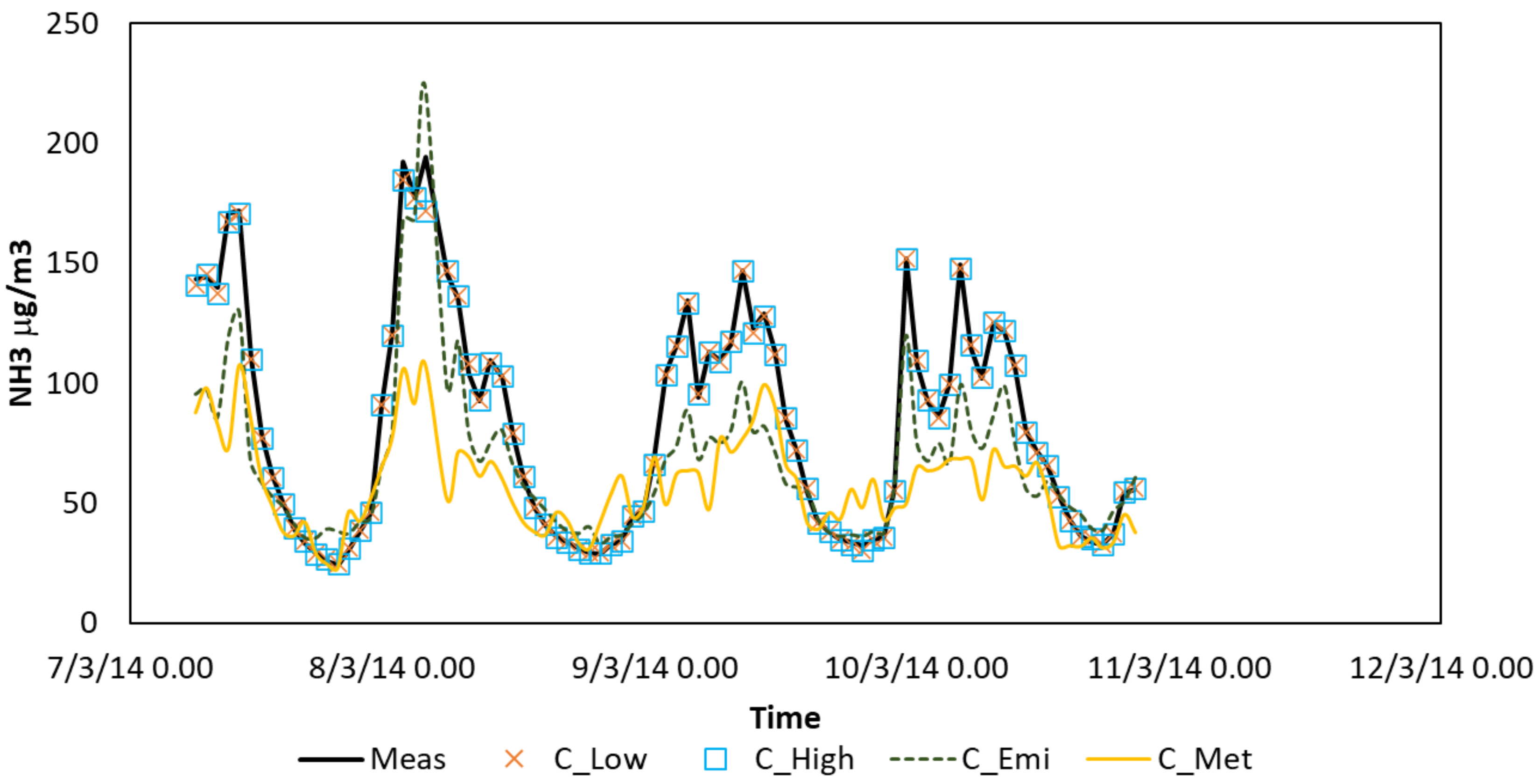




Marongiu, A.; Collalto, A.G.; Distefano, G.G.; Angelino, E. Application of Machine Learning to Estimate Ammonia Atmospheric Emissions and Concentrations. Air 2024, 2, 38–60. https://doi.org/10.3390/air2010003

Convergence of the solutions

- The approach gives the same result independently from the first guess value of the emission rate.
- The implemented methodology normalizes the input variables: atmospheric turbulence variables and emission rates.
- A calibration curve based on annual average measured concentrations and annual emissions could provide a plausible scaling and forecast of the initial values for emissions.



Marongiu, A.; Collalto, A.G.; Distefano, G.G.; Angelino, E. Application of Machine Learning to Estimate Ammonia Atmospheric Emissions and Concentrations. Air 2024, 2, 38–60. https://doi.org/10.3390/air2010003

Downwind unidimensional with know source

Hourly based parameters:
- ambient temperature (AT) [°C]
- wind speed (WS) [m/s]
- stability class
- **CALPUFF simulated concentrations [μg/m³]**
- duration time of emission [h]
- Downwind distance from source [m]
- Variable emission rate [g/s]

The database of CALPUFF simulations are taken from:
A. Marongiu, E. Angelino, S. A. Bellinzona, G. Fossati, G. Lanzani, E. Peroni, Simulation Database for Atmospheric Dispersion of Pollutant and Toxic Compounds during Accidental Fires, Harmonization within Atmospheric Dispersion Modelling for Regulatory Purposes (HARMO 15), Madrid, 6-9 May 2013

ML
Random Forest
(train and test)

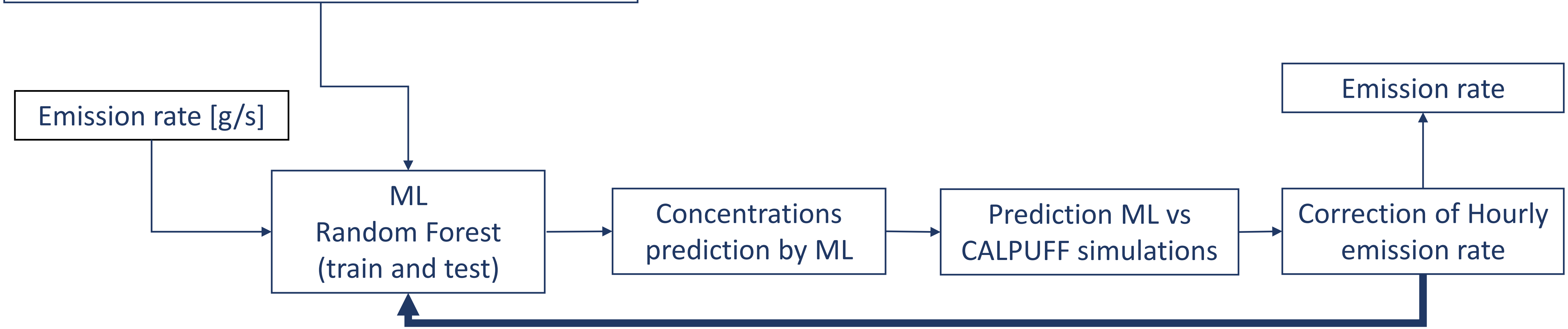
→ Concentrations prediction by ML

→ Prediction ML vs CALPUFF simulations

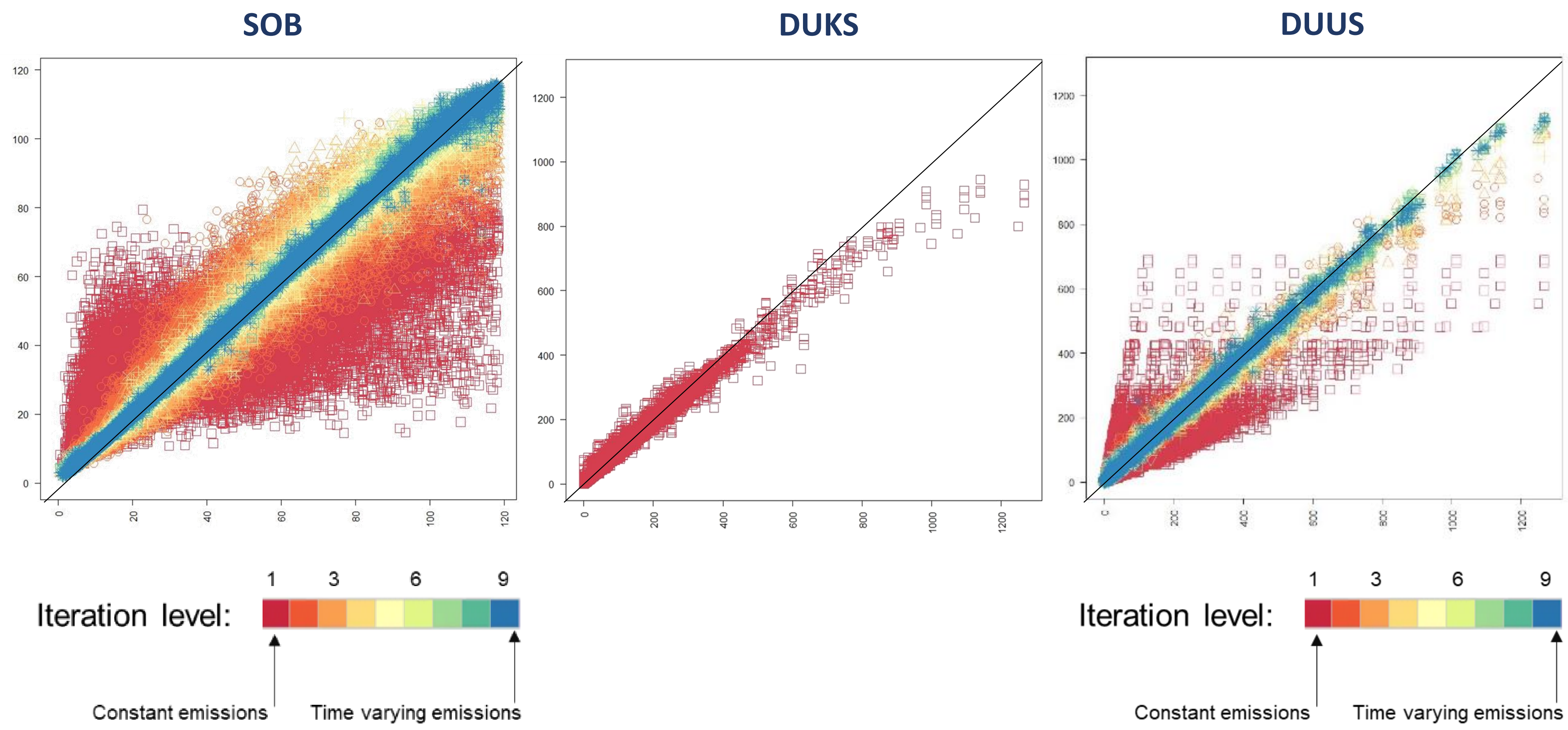
Downwind unidimensional with unknow source

Hourly based parameters:
- ambient temperature (AT) [°C]
- wind speed (WS) [m/s]
- stability class
- **CALPUFF simulated concentrations [μg/m³]**
- duration time of emission [h]
- Downwind distance from source [m]
- Variable emission rate [g/s]

The database of CALPUFF simulations are taken from:
A. Marongiu, E. Angelino, S. A. Bellinzona, G. Fossati, G. Lanzani, E. Peroni, Simulation Database for Atmospheric Dispersion of Pollutant and Toxic Compounds during Accidental Fires, Harmonization within Atmospheric Dispersion Modelling for Regulatory Purposes (HARMO 15), Madrid, 6-9 May 2013



Convergences in the models – predicted concentrations

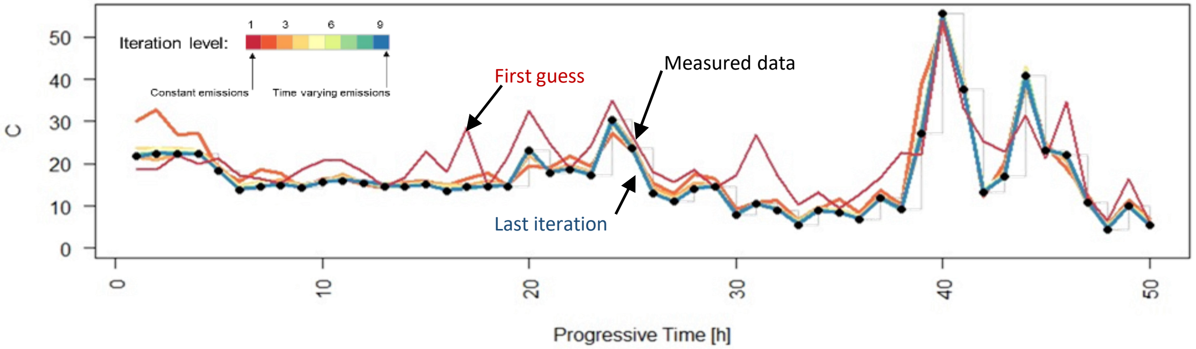


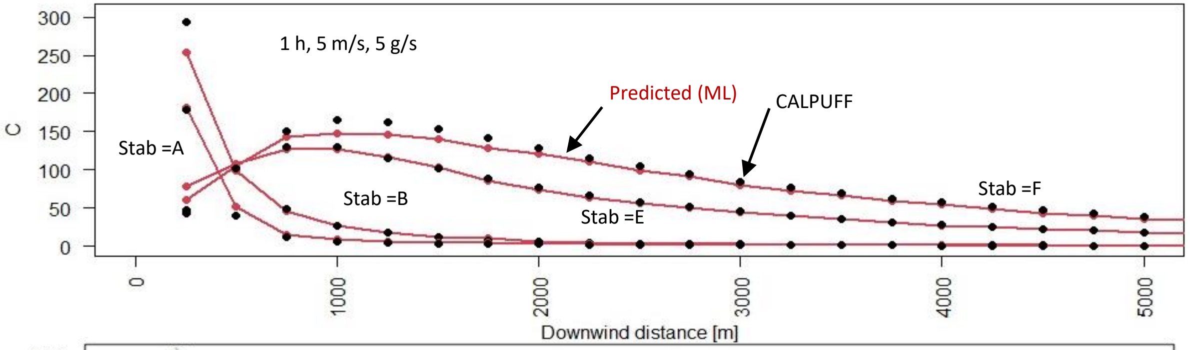
Atmospheric concentrations modelling

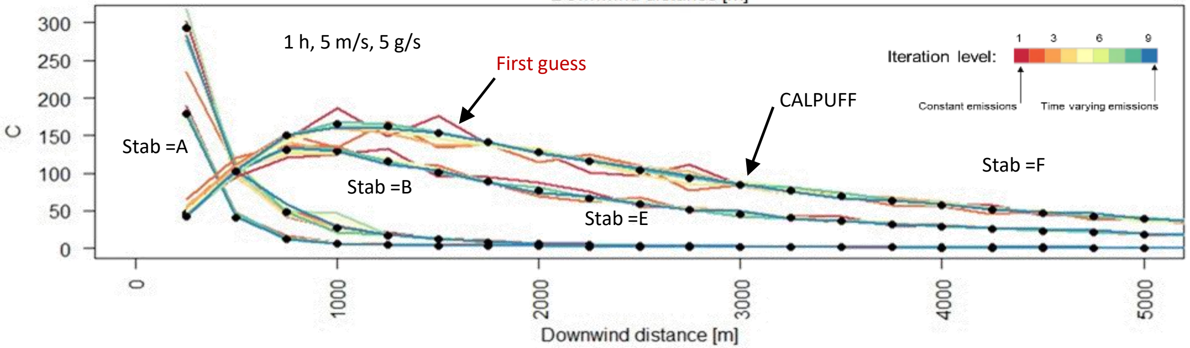
SOB

DUKS

DUUS





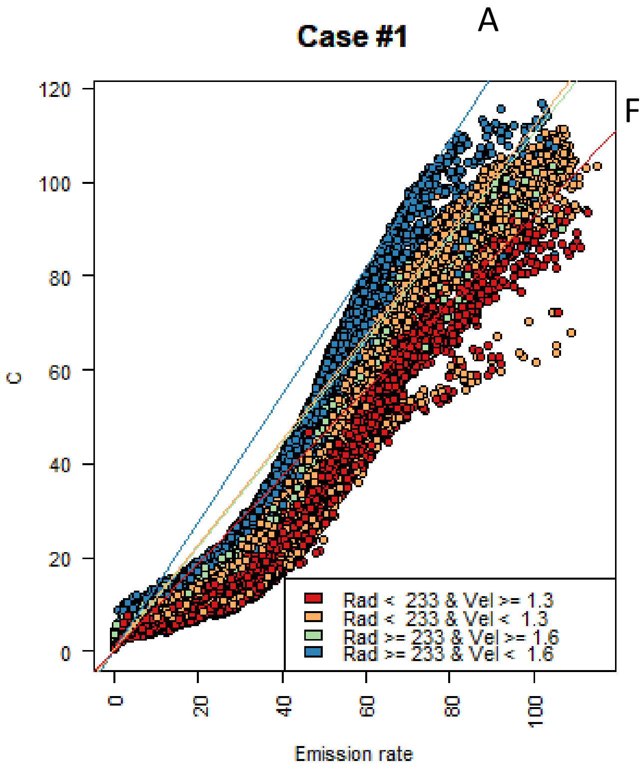


Unkonwn emission rates

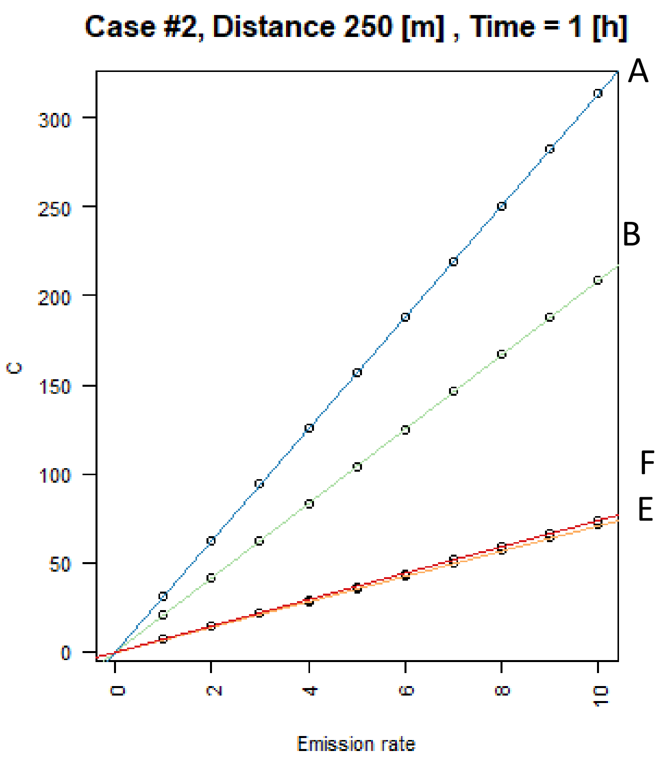
(they are calculated iteratively as the best input values to fit ML in predicting concentrations)

Emission rates estimates

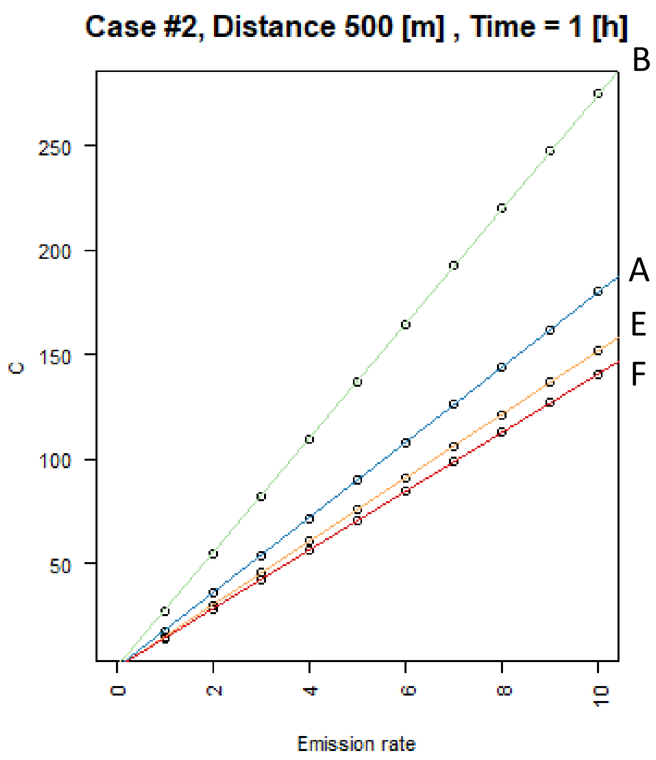
SOB



DUKS



DUUS



Variability observed in concentration levels is a result of the different meteorological conditions that can occur at the same emission level. Atmospheric stability classes on the intensity of solar radiation and wind speed. Divergences in emission rate calculation can be relevant increasing distance from the source.

Conclusions

- 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.
- ML 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.

Acknowledgments

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).

More details are available in:

Conference proceedings

Marongiu, A.; Collalto, A.G.; Distefano, G.G.; Angelino, E. Application of Machine Learning to Estimate Ammonia Atmospheric Emissions and Concentrations. Air 2024, 2, 38–60. https://doi.org/10.3390/air2010003

A. Marongiu, E. Angelino, S. A. Bellinzona, G. Fossati, G. Lanzani, E. Peroni, Simulation Database for Atmospheric Dispersion of Pollutant and Toxic Compounds during Accidental Fires, Harmonization within Atmospheric Dispersion Modelling for Regulatory Purposes (HARMO 15), Madrid, 6-9 May 2013