Geophysical Research Abstracts
Vol. 21, EGU2019-517, 2019
EGU General Assembly 2019

Correcting air quality forecasts with machine learning algorithms

Hervé Petetin, Albert Soret, Alicia Sanchez-Lorente, Oriol Jorba Cassellas, Kim Serradell, and Carlos Pérez Garcia-Pando

Barcelona Supercomputing Center (BSC), Barcelona, Spain (hervepetetin@gmail.com)

Air pollution is a major environmental problem affecting human health and ecosystems. Mitigating the effect of pollution episodes requires reliable air quality forecasting (AQF) systems for both warning vulnerable populations and taking short-term measures of emission control (e.g. traffic reduction, interruption of some industrial facilities). The rise of chemistry-transport models (CTMs) over the last decades have allowed major improvements in AQF, supplanting the use of purely statistical forecasting systems. However, AQF systems remain affected by numerous sources of uncertainty (e.g. emissions, meteorology, initialization). In order to improve the forecasts, so-called model output statistics (MOS) methods are commonly applied to correct CTM outputs. Various methods have been proposed in the literature, including moving averages, Kalman filter and, more recently, analogs. Although these methods allow removing a large part of the bias, they still suffer from some limitations, in particular when it comes to the detection of fast changes in atmospheric conditions.

In this work, we explore the use of machine learning (ML) techniques for correcting the raw AQF produced by CTMs based on the past AQF errors along with ancillary information (e.g. meteorology). The objective is to build a ML model able to estimate the error of the AQF in advance based on multiple ancillary variables (hereafter called features). In ML terms, we plan to solve a supervised regression problem in which the past forecast errors along with their associated features are used to train a model that estimates the future forecast errors on the basis of new (unseen) features. Several popular ML algorithms will be tested, including gradient boosting machine, support vector machine and artificial neural networks. Results will be compared with the MOS methods previously mentioned. The analysis will be applied on the MONARCH model currently developed at the Barcelona Supercomputing Center, as well as on the CAMS regional forecasts.