Geophysical Research Abstracts,
Vol. 11, EGU2009-9974, 2009
EGU General Assembly 2009

Ensemble forecast with machine learning algorithms

V. Mallet (1,2), G. Stoltz (3,4), É. Debry (5), B. Mauricette (3,2,1), S. Gerchinovitz (3,1,2)

(1) INRIA, Paris-Rocquencourt research center, France, (2) CEREA, joint laboratory ENPC - EDF R&D, Université Paris-Est, Marne la Vallée, France, (3) Département de mathématiques et applications, École normale supérieure, CNRS, Paris, France, (4) HEC Paris, CNRS, Jouy-en-Josas, France, (5) Institut national de l'environnement industriel et des risques (INERIS), Verneuil-en-Halatte, France

In the past few years, ensembles of air quality forecasts have been developed and studied in order to evaluate forecasts uncertainties or in order to improve the forecasts themselves. The latter goal involves aggregating the forecasts so as to produce a single forecast that is hopefully better than any individual model in the ensemble.

The most obvious option is the ensemble mean where every model is given the same weight. In many cases, this approach brings little improvements (if any, depending on the target) that may not be worth the price of running an ensemble. More successful approaches come from data assimilation, especially with the ensemble Kalman filter and its variants. Another class of methods produces a weighted average of the individual forecasts: the weights are computed based on past observations and past model forecasts. They are updated before any new forecast period, hence the approach is referred to as *sequential aggregation*. It should be noted that, in the present study, the weights do not depend on the position so that they may be applied away from the observation locations (which is actually validated in practice).

Robust and mathematically-grounded aggregation methods are developed in the *machine learning* community. These methods form efficient convex or linear combinations. The associated theory guarantees that the aggregated forecast will perform, in the long run, as well as the *best constant* (in time) convex or *linear combination*. Here, "best" refers to the best combination with respect to the quadratic discrepancy with the observations. This result holds whatever the sequence of observations may be and whatever the forecasts may be, and without any stochastic assumption. Hence the theoretical guarantees still hold in practice. This is why the learning algorithms are considered as *robust* methods, which is a nice feature for operational forecasting.

Despite their robust nature, the learning methods are also very satisfactory on the performance side since they compete, in theory and in practice, with the best constant linear combination (or convex combination, depending on the method) of models. In practice, they usually perform even better than the best constant linear combination.

Many learning methods, such as the exponentiated gradient algorithm or the discounted ridge regression, are applied to three ensembles. Two ensembles are generated using the Polyphemus system (http://cerea.enpc.fr/polyphemus/), with simulations relying on different input data, physical parameterizations and numerical schemes. The largest ensemble includes 100 members run during the whole year 2001 over Europe and mainly for ozone. The third ensemble is that of the French operational forecasting platform Prév'air (http://www.prevair.org/). It is composed of 5 to 7 simulations covering Europe in summer 2008, for ozone, PM_{10}, $\mathrm{PM}_{2.5}$ and NO_2.

The methods perform well with any of the ensembles, which shows they can be applied to ensembles of different sizes, to different time periods and to different target species. For instance, with the Prév'air ensemble and the methods applied in operational mode, the RMSE on ozone forecasts is reduced by about 32% (w.r.t. the best model for ozone RMSE) in the ensemble, and the corresponding correlation is increased by 11% (w.r.t. the best model for ozone correlation). The RMSE on PM_{10} forecasts decreases by 11% (w.r.t. the best model for PM_{10} RMSE) and the correlation increases by 29% (w.r.t. the best model for PM_{10} correlation). The improvements are of course stronger if one compares to a single model, because the best model changes with the target.

Further analysis of the methods and of their results was carried out to address the robustness of the methods, pri-

marily: spatial validity of the weights, sensitivity to the parameters of the methods, sparse aggregation, scalability in time, ability to forecast extreme events.