Data-constrained uncertainty estimation in air quality simulation

Vivien Mallet (INRIA), Damien Garaud

Air quality models are 3D chemistry-transport models that rely on complex physical and chemical formulations and on large amounts of data. The physical and chemical formulations are important sources of errors. Most of the input data is provided with high uncertainties in their time evolution and spatial distribution. The numerical approximations can be another important source of errors. The numerical model's state, which contains one to ten million components, is only partially observed by a few hundreds monitoring stations.

In order to better estimate the large uncertainties in models' outputs, modern methods rely on ensembles of simulations. The simulations of the ensembles are based on different physical models (multimodel ensemble) and perturbed input data (Monte Carlo approach). Large ensembles (e.g., with 100 members) can be automatically generated, but these may not properly sample the uncertainties. The ensemble performance for uncertainty estimation can be evaluated by comparison with observations. Note that this comparison evaluates the ensemble as the whole, not the individual models.

Using the evaluation scores for ensembles, it is possible to devise algorithms for the calibration of ensembles. One strategy is to extract a sub-ensemble (from a given large ensemble) that minimizes a performance criterion over the set of available sub-ensembles. The criterion may the variance of a rank histogram, or a Brier score in case of probabilistic forecasts.

The generation and calibration of ensembles will be illustrated for 2D ozone concentration fields over Europe.