17th International Conference on
Harmonisation within Atmospheric Dispersion Modelling for Regulatory Purposes
9-12 May 2016, Budapest, Hungary

AN APPLICATION OF THE SCHAAKE SHUFFLE TECHNIQUE TO GENERATE SPACE-TIME CONSISTENT AQ PREDICTIONS

S. Alessandrini[1], L. Delle Monache[1], Irina Djalalova[2], Jim Wilczak[2]

[1]National Center for Atmospheric Research, Boulder, Colorado, USA
[2]National Oceanic and Atmospheric Administration, Boulder, Colorado, USA

Abstract: Several statistical methods used to generate ensemble forecast of air pollutants or different meteorological variables do not represent accurately the covariance between neighbouring stations or temporal correlation between subsequent lead times. Probabilistic predictions of O_3 and $PM_{2.5}$ surface concentrations over the U.S. are generated with the analog ensemble (AnEn) technique. The ensemble members provided by AnEn are statistically indistinguishable and they are generated without space-time correlation. We apply the Schaake Shuffle technique to reorder the ensemble members and recover space-time variability of $PM_{2.5}$ and O_3 forecast time-series.

Key words: *Analog Ensemble, Schaake Shuffle, air quality ensemble forecasting*

INTRODUCTION

Several statistical methods used to generate ensemble forecast of air pollutants or different meteorological variables do not represent accurately the covariance between neighbouring stations or temporal correlation between subsequent lead times. This is an important issue for air quality ensemble forecast when generating, for example, two-dimensional maps of the concentration of a given pollutant, or when trying to predict the risk of consecutive days/hours with hazardous concentrations. Probabilistic predictions of O_3 and $PM_{2.5}$ surface concentrations over the U.S. are generated with the analog ensemble (AnEn) technique. The ensemble members provided by AnEn are statistically indistinguishable and they are generated without space-time correlation. We apply the Schaake Shuffle technique, widely used for hydrological application, to reorder the ensemble members and recover space-time variability of $PM_{2.5}$ and O_3 forecast time-series. With this technique, the ensemble members for a given forecast lead-time are ranked and matched with the rank of $PM_{2.5}$ or O_3 past observations at the same hours appropriately selected across the historical record. The ensembles are then reordered to match the original order of the selected historical data. Using this technique, the observed inter-station correlation and the observed temporal auto-correlation are almost completely recovered.

METHODOLOGY AND RESULTS

The AnEn technique (Delle Monache et al. 2011, 2013) has been extensively tested for the probabilistic prediction of both meteorological variables and renewable energy (Alessandrini et al., 2015).

The AnEn is built from a historical set of deterministic predictions and observations of the quantity to be predicted. For each forecast lead time and location, the ensemble prediction of a given variable is constituted by a set of measurements of the past (i.e., 1-hour averages of $PM_{2.5}$ and O_3 concentrations). These measurements are those concurrent to past deterministic predictions for the same lead time and location, chosen based on their similarity to the current forecast. The forecasted variables used to identify the past forecast similar to the current one are called analog predictors. In this application we use as predictors, among others meteorological variables, the O_3 and $PM_{2.5}$ concentrations forecasts over the continental US generated by the U.S. EPA CMAQ CTM (Byun and Schere, 2006) model. The forecasts are issued at 12 UTC with a 24 hours frequency for lead times between 0-48 hours ahead over the period 01 July 2014-31 July 2015. The AnEn forecasts are issued for all the 564 stations of the AIRNow EPA network whose locations is depicted in Figure 1. The first 6 months of this period are used for training purposes while the remaining part for the verification. The AnEn provides reliable, sharp, and statistical consistent probabilistic AQ predictions, at a fraction of the real-time computational cost of traditional ensemble methods.

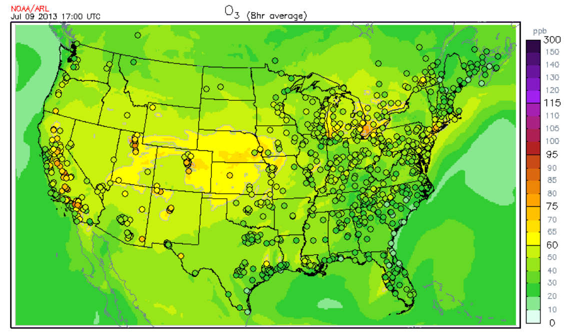

Figure 1. Maximum daily 8-hour average surface O_3 on July 9 2013. Shown are the model prediction (shading) from CMAQ, and measurement (filled solid circle) from the AIRNow EPA network.

The ensemble members provided by AnEn are statistically indistinguishable and they are generated without space-time correlation. We apply the Schaake Shuffle (ScS) technique (Clark et al., 2004), widely used for hydrological application, to reorder the ensemble members and recover space-time variability of $PM_{2.5}$ and O_3 forecast time-series. With this technique, the ensemble members for a given forecast lead-time are ranked and matched with the rank of $PM_{2.5}$ or O_3 past observations at the same hours appropriately selected across the historical record. The ensembles are then reordered to match the original order of the selected historical data. Using this technique, the observed inter-station correlation and the observed temporal auto-correlation are almost completely recovered.

In Figure 2 the autocorrelation function of $PM_{2.5}$ concentrations is plotted for all the 20 members (red line) generated by the AnEn and the measurements (black line) for one station. On the left, the auto-correlations are computed for the AnEn members not yet reordered by the ScS technique which has been instead applied to plot the chart on the right. The observed temporal auto-correlation is significantly underestimated by all the 20 members when computed without the ScS reordering. When ScS is applied, the observed auto-correlation is better reproduced by all AnEn members.

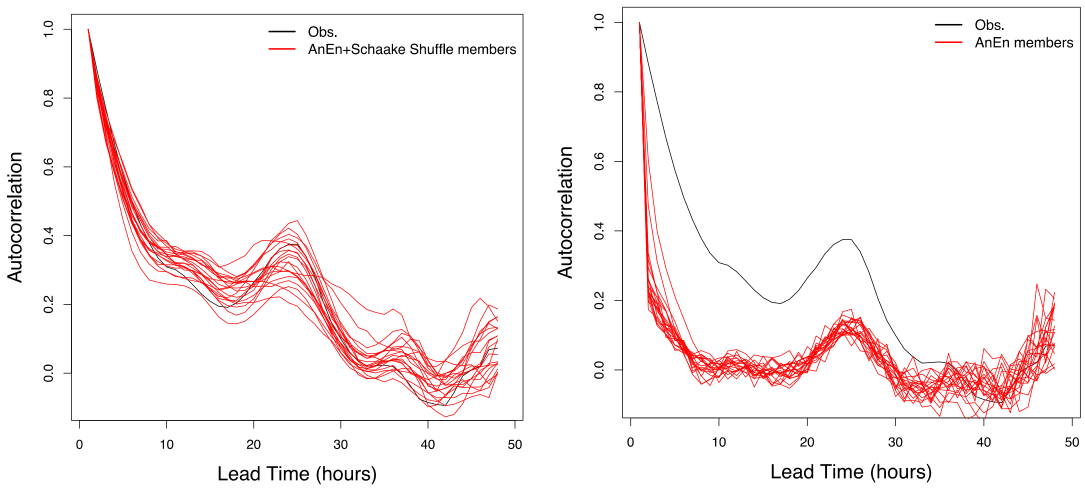

Figure 2. The autocorrelation function of $PM_{2.5}$ concentrations is plotted for all the 20 members (red line) generated by the AnEn and the measurements (black line) for one station. On the left the autocorrelations are computed for the AnEn members not yet reordered by the Schaake Shuffle (ScS) technique. The ScS technique has been instead applied to plot the chart on the right.

CONCLUSION

We have applied the Schaake Shuffle (ScS) to recover the observed temporal auto-correlation of the $PM_{2.5}$ ensemble forecast generated by the analog ensemble (AnEn) technique. Preliminary results show that the ScS can recover the observed temporal auto-correlation of $PM_{2.5}$ hourly concentrations. Future work will extend the ScS application to O_3 concentration forecasts and to verify the ScS ability to recover the observed inter-station correlation.

REFERENCES

Alessandrini, S., Delle Monache, L., Sperati, S., & Nissen, J. N. (2015). A novel application of an analog ensemble for short-term wind power forecasting. *Renewable Energy*, **76**, 768-781.

Byun D. W., and K. L. Schere (2006), Description of the Models-3 Community Multiscale Air Quality (CMAQ) Model: System overview, governing equations, and science algorithms, *Appl. Mech. Rev*., **59**, 51-77.

Clark, M., Gangopadhyay, S., Hay, L., Rajagopalan, B., & Wilby, R. (2004). The Schaake shuffle: A method for reconstructing space-time variability in forecasted precipitation and temperature fields. *Journal of Hydrometeorology*, **5**(1), 243-262.

Delle Monache, L., T. Nipen, Y. Liu, G. Roux, and R. Stull (2011), Kalman filter and analog schemes to post-process numerical weather predictions, *Mon. Wea. Rev*., **139,** 3554–3570.

Delle Monache, L., T. Eckel, D. Rife, B. Nagarajan, K. Searight (2013), Probabilistic weather predictions with an analog ensemble, *Mon. Wea. Rev*., **141**, 3498-3516.