Is regional air quality model diversity representative of uncertainty for ozone simulation ?

Robert Vautard, M. van Loon, M. Schaap, R. Bergstrom, Bertrand Bessagnet, J. Brandt, P. Builtjes, J.H. Christensen, C. Cuvelier, A. Graff, et al.

Is regional air quality model diversity representative of uncertainty for ozone simulation?

R. Vautard,[1] M. Van Loon,[2] M. Schaap,[3] R. Bergström,[4] B. Bessagnet,[5] J. Brandt,[6] P. J. H. Builtjes,[3] J. H. Christensen,[6] C. Cuvelier,[7] A. Graff,[8] J. E. Jonson,[1] M. Krol,[9] J. Langner,[4] P. Roberts,[10] L. Rouil,[5] R. Stern,[11] L. Tarrasón,[2] P. Thunis,[7] E. Vignati,[7] L. White,[12] and P. Wind[2]

Received 18 July 2006; revised 15 September 2006; accepted 19 October 2006; published 30 December 2006.

[1] We examine whether seven state-of-the-art European regional air quality models provide daily ensembles of predicted ozone maxima that encompass observations. Using tools borrowed from the evaluation of ensemble weather forecasting, we analyze statistics of simulated ensembles of ozone daily maxima over an entire summer season. Although the model ensemble overestimates ozone, the distribution of simulated concentrations is representative of the uncertainty. The spread of simulations is due to random fluctuations resulting from differences in model formulations and input data, but also to the spread between individual model systematic biases. The ensemble average skill increases as the spread decreases. The skill of the ensemble in giving probabilistic predictions of threshold exceedances is also demonstrated. These results allow for optimism about the ability of this ensemble to simulate the uncertainty of the impact of emission control scenarios. **Citation:** Vautard, R., et al. (2006), Is regional air quality model diversity representative of uncertainty for ozone simulation?, *Geophys. Res. Lett.*, *33*, L24818, doi:10.1029/2006GL027610.

1. Introduction

[2] Predicting air quality for the next day, or in an analysis for the future assuming anthropogenic emission reduction scenarios, is a straightforward application of regional and urban air quality modelling. However predicting the uncertainty of such model simulations or forecasts remains a challenging problem. The question of uncertainty in model predictions has been extensively addressed in weather forecasting in the last decade. Weather forecasts uncertainty strongly depends on the knowledge of the initial conditions, as initially close atmospheric states rapidly diverge. Thus uncertainty prediction has been primarily based on ensembles of forecasts differing by their initial conditions [*Molteni et al.*, 1996; *Toth and Kalnay*, 1997]. *Atger* [1999] showed that ensembles made with a limited number of different models also provide an efficient way of describing the uncertainty in weather forecasts.

[3] In air quality prediction and analysis, uncertainty in simulated concentrations results either from errors or uncertainty in model input data, physical parameters or parameterizations, or from gaps in our knowledge of the chemistry and physics of the atmosphere and its interaction with the surface. The distribution of possible concentrations has also been calculated as in meteorology with ensembles of model calculations [*Dabberdt and Miller*, 2000], or from a single model using Monte-Carlo simulations with assumed distributions of individual processes uncertainty [*Hanna et al.*, 2001]. These ensembles can also be generated by using a single model and several optimally selected parameter values [*Beekmann and Derognat*, 2003] or numerical and physical parameterizations [*Mallet and Sportisse*, 2006]. Ensembles of air quality forecasts can also be created using several models, developed independently [*Delle Monache and Stull*, 2003; *McKeen et al.*, 2005]. Calculations of air quality and its uncertainty under future European emission scenarios using model ensembles have also recently been carried out in a cooperative effort of most regional and city scale air quality modelling teams in Europe, in the projects CityDelta [*Cuvelier et al.*, 2007] and EuroDelta [*Van Loon et al.*, 2006]. It has also been shown that model ensembles can be used to improve air quality forecasts [*Delle Monache and Stull*, 2003; *Pagowski et al.*, 2005] or simulations/analysis [*Van Loon et al.*, 2006]. However the evaluation of the ability of ensembles to simulate uncertainty received interest only very recently [*Delle Monache et al.*, 2006].

[4] In this article we evaluate whether an ensemble of long-term simulations, performed independently with seven European state-of-the-art regional air quality models, simulates spreads of daily ozone maxima that are representative of the uncertainty of simulated concentrations, i.e., of their closeness to observed concentrations. Models use the

[1]Laboratoire des Sciences du Climat et de l'Environnement, Institut Pierre Simon Laplace Laboratoire, CEA, CNRS, UVSQ, Gif-sur-Yvette, France.
[2]European Monitoring and Evaluation Programme, Meteorological Synthesizing Centre-West, Oslo, Norway.
[3]Built Environment and Geosciences, Netherlands Organization for Applied Scientific Research, Apeldoorn, Netherlands.
[4]Swedish Meteorological and Hydrological Institute, Norrköping, Sweden.
[5]Institute National de l'Environnement Industriel et des Risques, Verneuil en Halatte, France.
[6]National Environmental Research Institute, Roskilde, Denmark.
[7]European Commission-Directorate General Joint Research Centre, Institute for Environment and Sustainability, Ispra, Italy.
[8]Umweltbundesamt, Berlin, Germany.
[9]Netherlands Institute for Space Research, Utrecht, Netherlands.
[10]Conservation of Clean Air and Water in Europe, Health Safety and Environment Department, Shell Global Solutions, Chester, UK.
[11]Institut für Meteorologie, Freie Universität, Berlin, Germany.
[12]Conservation of Clean Air and Water in Europe, Haywards Heath, UK.

0094-8276/06/2006GL027610

equations of physics, but also a number of parameterizations, with parameters determined from limited sets of observations or empirically. The uncertainty on all these values translates into a global uncertainty on the simulated concentrations. In the best case with respect to estimation of uncertainty, modellers have, independently from one another, selected model options or parameter values with a range of choices that is representative of the uncertainty on these parameters. In the worst case, all modellers have selected the same options or parameter values, or missed the same key processes. In the former case one expects observations to lie within the range of simulated concentrations, while in the latter observations should be "outliers" of the simulations distribution. Therefore the consistency between observations and the distribution of ensemble simulated concentrations measures our ability to represent uncertainty of simulations. In order to explore these questions we use the tools developed in the evaluation of uncertainty estimates using ensemble weather forecasting [*Talagrand et al.*, 1998; *Jollife and Stephenson*, 2003].

[5] For the sake of conciseness we focus here on ozone daily maxima simulated at 97 specific air quality monitoring sites over Europe throughout an entire summer season (April to September 2001). These simulations are the control simulations of the EuroDelta experiment [*Van Loon et al.*, 2006].

[6] In section 2 models and simulations are described. In section 3 we examine the global properties of the ensemble distributions and their relation to observations. In section 4 the time variability of the uncertainty is discussed. Section 5 contains conclusions.

2. Models, Observations, and the EuroDelta Experiment

[7] The EuroDelta experiment [*Van Loon et al.*, 2006] is designed to evaluate the impact of regional-scale emission changes for 2020 on air quality. Seven state-of-the-art chemistry-transport models are used to calculate the differences between predicted concentrations under several emission change scenarios for 2020 and concentrations issued from control simulations using emissions for a reference year, 2000. In this article we only use the results of the control simulations, and we focus on ozone daily maxima over the summer period (April to September), using the meteorology of Summer 2001, instead of 2000, because its more anticyclonic weather led to more photochemical episodes. Daily maxima ensembles therefore consist of seven ozone concentrations per day and station, the total number of simulated days being 183. Ozone daily maxima typically range between 30 and 120 ppb. We use observations gathered at 97 European Monitoring and Evaluation Program (EMEP) sites which lie in the intersection of all model domains, and whose altitude is less than 1000 m. The total number of available observations and ensembles of seven concentrations is 17069.

[8] Participating regional-scale models are EMEP (available at http://www.emep.int), MATCH [*Andersson et al.*, 2006, and references therein], LOTOS-EUROS [*Schaap et al.*, 2006], CHIMERE [*Schmidt et al.*, 2001], RCG [*Stern et al.*, 2003], DEHM [*Christensen*, 1997; *Frohn et al.*, 2002], with horizontal resolutions of about 30–50 km, and the global TM5 model [*Krol et al.*, 2005], zoomed to a 1×1 degree over Europe. Vertical resolution varies from four to 25 layers. Although the meteorological year used for the simulations was 2001, all models use the EMEP annual emissions totals for Year 2000 [*Vestreng*, 2003], as the modelling project was intended to study emission changes between 2000 and 2010. All other driving parameters differ: meteorology, boundary conditions, land use, etc. For more details on the model configurations and the EuroDelta experiment the reader is referred to *Van Loon et al.* [2006].

3. Ensemble Distribution of Ozone Daily Maxima and Uncertainty

[9] Using several models the hope is that, for each station and day, the seven-member ensemble of ozone daily maxima represents of the uncertainty in the prediction. Roughly speaking, the observation can be any of the ensemble members. In this case, the distribution of the observation rank within the seven-member ensemble of values, cumulated over many cases must be equiprobable. One useful tool to check this property is the rank histogram, often called in meteorology the "Talagrand diagram" [*Talagrand et al.*, 1998]. Using seven models, the rank of the observed daily maxima among the simulated ones takes an integer value between zero (for the interval below the lowest value) and seven (for the interval above the highest of the seven simulated values). The rank histogram counts the occurrences of the rank for each integer between zero and seven. If the distribution of the observation within the ensemble is equiprobable, on average, the rank histogram must bear constant values. Note that the ensembles can give very poor predictions of the actual values (large spreads) but still satisfy the rank histogram condition. The aim here is not to evaluate the skill of the ensemble itself, but the coincidence between ensemble spread and uncertainty.

[10] Figure 1a shows the rank histogram of the summertime (April to September) daily ozone maxima, all stations and days being put together. The first two bins (0 and 1) have a number of counts much larger than the other bins, reflecting a difficulty of models to simulate low daily maxima, which is a bias of the ensemble. This bias could result from the larger modelling effort put for skilful prediction of high concentrations rather than of lower ones. This "ensemble bias" can be removed, at each station, by subtracting the average difference between simulated (all models together) and observed ozone daily maxima. After this operation, daily maxima ensembles are only shifted but their distribution, spread and model rank are unchanged. In this case, the rank histogram becomes flatter (Figure 1a). Therefore the bias-free ensemble gives a fair average account of the uncertainty. However the bumps in the histogram reveal heterogeneities in the distribution of ensemble ozone values relative to the distribution of observations.

[11] When calculated station by station, rank histograms exhibit variable shapes. Most of them display rather equiprobable distribution. However in many cases they exhibit either a "U" shape or a bump shape. The former case results from an underestimation of the ensemble spread relative to the observations. In the latter case the uncertainty is overrepresented by the ensemble.

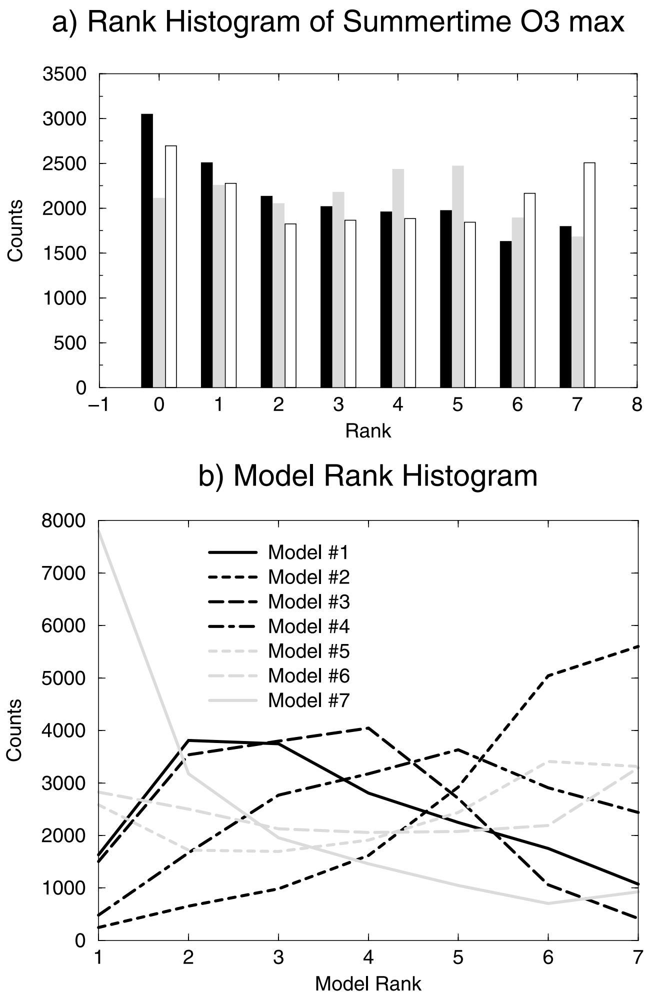

Figure 1. (a) Histograms of the rank of observed summertime ozone daily maxima among the seven simulated values, all stations and days, from April to September 2001 being put together. The black bars show the rank histogram from the raw ensemble. Gray bars stand for the histogram of unbiased estimates, the ensemble bias being removed. Empty bars show histograms for simulated values where bias has been removed separately for each model and each station. (b) Histograms of the rank of each model within the ensemble, for the simulation of ozone daily maxima, for the raw ensemble.

[12] When removing the bias of each model at each monitoring site, daily maxima ensemble spread are changed, as the spread of model biases is removed. In this case the rank histogram displays a clear U-shaped curve (Figure 1a), indicating an underestimation of ensemble spread. Relative to the previous case where only the ensemble bias is removed, the spread of ensembles is reduced to its "random" component, as it does not contain the contribution of the spread due to individual model biases.

[13] Therefore from Figure 1a it is clear that part of the fair representation of uncertainty with ensemble bias removed is due to the spread of model biases. These individual biases lead some models to be often found at one extreme rank of the ensemble, as shown by the histogram of models rank in the biased ensembles in Figure 1b. For instance model 2 is often found at a high rank, while model 7 is very often found at the first rank.

4. Space and Time Variability of the Uncertainty

[14] The uncertainty in simulated ozone concentration can be quite small on a windy and cloudy day or much larger in stagnant anticyclonic conditions as it becomes sensitive to many parameters, such as wind direction and emissions. The question we address here is whether the model ensemble is able to reproduce the space and time variability of uncertainty. Figure 2 shows a "spread-skill diagram," where the skill, characterized by the root mean square RMS error of the ensemble average concentration, is plotted against the ensemble spread, defined as the standard deviation of the ensemble. In order to avoid statistical noise the results are averaged over equally populated spread bins of 40 cases. If the variability of uncertainty was perfectly simulated by the ensembles, the points should lie along the diagonal in Figure 2. By contrast if the variability of the ensemble spread is not correlated with that of uncertainty, curves should be horizontal lines. Figure 2 shows that the situation is "in between." For all types of ensembles (biased or unbiased), skill decreases with increasing spread. However for largest spread ensembles, RMS error is smaller than ensemble spread meaning that when models strongly disagree the uncertainty is overestimated. The reverse is true: when spread is small, RMS error is larger than spread. Thus

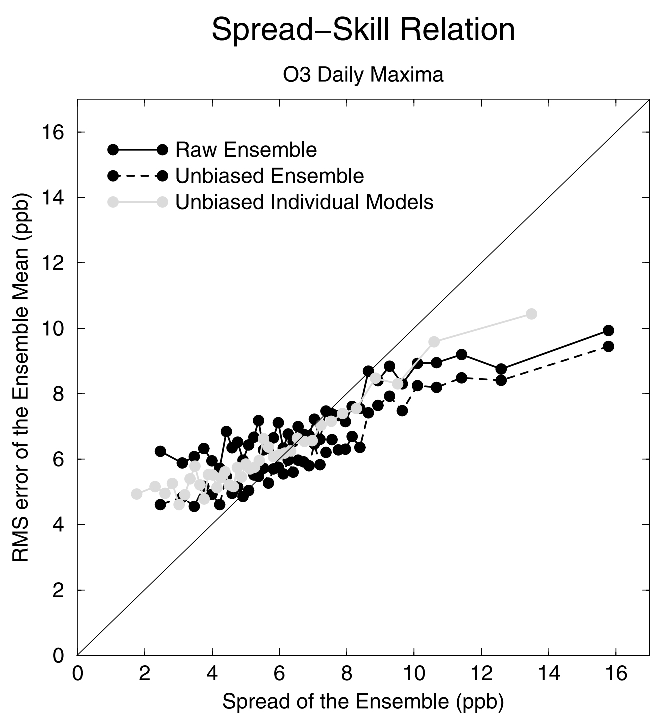

Figure 2. Root mean square error of the ensemble average daily maxima versus the "spread" of the ensemble taken as the standard deviation of the seven simulated values. Values have been averaged in bins of 40 consecutive values of the spread. The three curves stand respectively for the raw ensembles, the unbiased ensemble, and the ensemble with unbiased individual models.

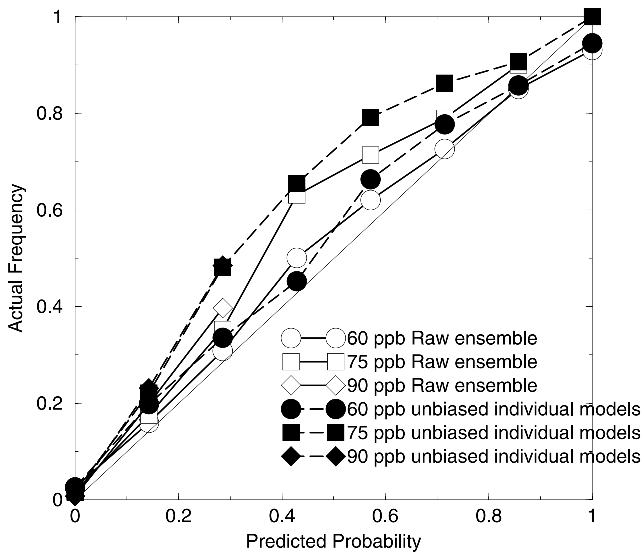

Figure 3. Reliability diagram, representing the actual frequency versus the predicted probability of thresholds exceedances, with three thresholds: 60 ppb (circles), 75 ppb (squares), and 90 ppb (diamonds). Solid curves stand for the raw ensemble and dashed curves for the ensemble with unbiased individual models.

a part of the variability of the spread is not due to actual uncertainty, but to model differences uncorrelated with their skill. The best fit to the diagonal is found with the model-bias free ensembles.

[15] Another way to explore the variability of uncertainty is through probabilistic prediction of concentration exceedances. Given the seven daily maxima values a probabilistic prediction of the exceedance of a given threshold can be made by counting the number of ensemble members that exceed the threshold and dividing this number by seven to obtain a probability p. In perfect ensembles, the frequency of the actual occurrence of the exceedance, given the predicted probability p, should be equal to p. This property can be verified using reliability diagrams [*Talagrand et al.*, 1998] which displays the frequency of occurrence as a function of predicted probability. These diagrams are shown in Figure 3 for the biased and model-unbiased ensembles, and for 3 thresholds: 60 ppb, 75 ppb, and 90 ppb. For a 90 ppb threshold, there were no cases where more than 2 models simultaneously predicted the event. For a 60 ppb threshold the reliability of the probabilistic prediction is high, thus the representation of the uncertainty in threshold exceedance is accurate. When threshold increases, the occurrence frequency is larger than the predicted probability, indicating an underestimation of high ozone values by all models. This effect is more pronounced when model biases are removed (not shown). The removal of general positive biases increases the underestimation of high ozone concentrations.

5. Conclusions

[16] We have examined the spread of long-term simulations of daily ozone maxima performed by an ensemble of seven state-of-the-art regional air quality models. The main issue of this article was to assess whether this spread is representative of the uncertainty of ozone prediction. We used throughout this study statistical tools developed for the evaluation of ensemble weather forecasts. The analysis of rank histograms showed that (1) there is a global positive bias of the ensemble, (2) when ensemble bias is removed at each monitoring station the spread of simulated values is fairly representative of the uncertainty, that is, of the spread of the simulation errors, and (3) this spread is partly due to the spread of individual model systematic biases.

[17] The variability of the uncertainty from day to day or from station to station is reproduced by the simulated ensembles, as the simulation skill decreases as ensemble spread increases. The ability of the ensemble in predicting uncertainty and its variability is also shown by an evaluation of the reliability of probabilistic prediction of threshold exceedances.

[18] There are several limitations to this study. First, all models use the same emissions yearly total, the EMEP emissions. The ensembles are therefore missing part of the spread, the amplitude and relative importance of which remaining undetermined. If it is significant, the spread of the ensembles should increase, while uncertainty may not, depending on the quality of EMEP emissions. If emissions closer to actual ones were used in the ensemble uncertainty should decrease. Finally the distribution and relatively small number of sites used for estimating ensemble representativeness may not allow us to extrapolate (or interpolate) our conclusions to the whole of Europe. In particular results do not apply to areas with complex terrain or emission patterns (coasts, cities, industrial areas).

[19] The evaluation of European emission reduction strategies is now carried out using ensembles of models, as in the CityDelta [*Cuvelier et al.*, 2007] and EuroDelta [*Van Loon et al.*, 2006] projects. What we learn from our findings is that apart from general biases problems the diversity of models used in these evaluations gives a fair account of the uncertainty in the simulated ozone daily maxima. It is hoped that this representativeness extends to other air quality parameters and pollutants, a question that will be addressed in future work. It is also hoped that it extends to results of emission reduction scenarios, in which case the ensemble provides an efficient way to evaluate uncertainty in our simulations of future air quality.

[20] The good correspondence between ensemble spread and uncertainty also indicates that, for ozone daily maxima, regional air quality models developed in Europe are complementary and their (unintentional) diversity reflect the uncertainty in our knowledge of air quality processes.

References

Andersson, C., J. Langner, and R. Bergström (2006), Interannual variation and trends in air pollution over Europe due to climate variability during 1958–2001 simulated with a regional CTM coupled to the ERA40 reanalysis, *Tellus, Ser. B*, doi:10.1111/j.1600-0889.2006.00196.x, in press.

Atger, F. (1999), The skill of ensemble prediction systems, *Mon. Weather Rev.*, *127*, 1941–1953.

Beekmann, M., and C. Derognat (2003), Monte Carlo uncertainty analysis of a regional scale transport chemistry model constrained by measurements from the ESQUIF campaign, *J. Geophys. Res.*, *108*(D17), 8559, doi 10.1029/2003JD003391.

Christensen, J. (1997), The Danish eulerian hemispheric model—A three-dimensional air pollution model used for the Arctic, *Atmos. Environ.*, *31*, 4169–4191.

Cuvelier, C., et al. (2007), CityDelta: A model intercomparison study to explore the impact of emission reductions in European cities in 2010, *Atmos. Environ*, *41*(1), 189–207.

Dabberdt, W. F., and E. Miller (2000), Uncertainty, ensembles and air quality dispersion modelling: Applications and challenges, *Atmos. Environ.*, *34*, 4667–4673.

Delle Monache, L., and R. B. Stull (2003), An ensemble air quality forecast over western Europe during an ozone episode, *Atmos. Environ.*, *37*, 3469–3474.

Delle Monache, L., J. P. Hacker, Y. Zhou, X. Deng, and R. B. Stull (2006), Probabilistic aspects of meteorological and ozone regional ensemble forecasts, *J. Geophys. Res.*, doi:10.1029/2005JD006917, in press.

Frohn, L. M., J. H. Christensen, and J. Brandt (2002), Development of a high-resolution nested air pollution model—The numerical approach, *J. Comput. Phys.*, *179*, 68–94.

Hanna, S. R., Z. Lu, H. C. Frey, N. Wheeler, J. Vukovich, S. Arumachalam, and M. Fernau (2001), Uncertainties in predicted ozone concentration due to input uncertainties for the UAM-V photochemical grid model applied to the July 1995 OTAG domain, *Atmos. Environ.*, *35*, 891–903.

Jollife, I. T., and D. B. Stephenson (Eds.) (2003), Forecast Verification: A Practitioner's Guide in Atmospheric Science, 240 pp., John Wiley, Hoboken, N. J.

Krol, M., S. Houweling, B. Bregman, M. van den Broek, A. Segers, P. van Velthoven, W. Peters, F. Dentener, and P. Bergamaschi (2005), The two-way nested global chemistry-transport zoom model TM5: Algorithm and applications, *Atmos. Chem. Phys.*, *5*, 417–432.

Mallet, V., and B. Sportisse (2006), Uncertainty in a chemistry-transport model due to physical parameterizations and numerical approximations: An ensemble approach applied to ozone modeling, *J. Geophys. Res.*, *111*, D01302, doi:10.1029/2005JD006149.

McKeen, S., et al. (2005), Assessment of an ensemble of seven real-time ozone forecasts over eastern North America during the summer of 2004, *J. Geophys. Res.*, *110*, D21307, doi:10.1029/2005JD005858.

Molteni, F., R. Buizza, T. N. Palmer, and T. Petroliagis (1996), The ECMWF ensemble system: Methodology and validation, *Q.J.R. Meteorol. Soc.*, *122*, 73–119.

Pagowski, M., et al. (2005), A simple method to improve ensemble-based ozone forecasts, *Geophys. Res. Lett.*, *32*, L07814, doi:10.1029/2004GL022305.

Schaap, M., R. M. A. Timmermans, F. J. Sauter, M. Roemer, G. J. M. Velders, G. A. C. Boersen, J. P. Beck, and P. J. H. Builtjes (2006), The LOTOS-EUROS model: Description, validation, and latest developments, *Int. J. Environ. Pollut*, in press.

Schmidt, H., C. Derognat, R. Vautard, and M. Beekmann (2001), A comparison of simulated and observed ozone mixing ratios for the summer of 1998 in western Europe, *Atmos. Environ.*, *36*, 6277–6297.

Stern, R., R. Yamartino, and A. Graff (2003), Dispersion modelling within the European community's air quality directives: Long term modelling of O_3, PM10, and NO_2, paper presented at 26th International Technical Meeting on Air Pollution Modelling and its Application, Natl. Atl. Treaty Organ., Istanbul, Turkey, 26–30 May.

Talagrand, O., R. Vautard, and B. Strauss (1998), Evaluation of probabilistic prediction systems, paper presented at Seminar on Predictability, Eur. Cent. for Medium Weather Forecasting, Reading, UK.

Toth, Z., and E. Kalnay (1997), Ensemble forecasting at NCEP and the breeding method, *Mon. Weather Rev.*, *125*, 3297–3319.

Van Loon, M., et al. (2006), Evaluation of long-term ozone simulations from seven regional air quality models and their ensemble average, *Atmos. Environ*, in press.

Vestreng, V. (2003), Review and revision of emission data reported to CLRTAP, *MSC-W Tech. Rep. 1/03*, Norw. Meteorol. Inst., Oslo.

R. Bergström and J. Langner, SMHI, Folkborgsvägen 1, SE-601 76 Norrköping, Sweden.

B. Bessagnet and L. Rouil, Institute National de l'Environnement Industriel et des Risques, Parc Technologique Alata, F-60550 Verneuil-en-Halatte, France.

J. Brandt and J. H. Christensen, National Environmental Research Institute, Frederiksborgvej 399, P.O. Box 358, DK-4000 Roskilde, Denmark.

P. J. H. Builtjes and M. Schaap, Built Environment and Geosciences, Netherlands Organization for Applied Scientific Research, Postbus 342, NL-7300 AH Apeldoorn, Netherlands.

C. Cuvelier, P. Thunis, and E. Vignati, European Commission-Directorate General Joint Research Centre, Institute for Environment and Sustainability, Via E. Fermi 1, I-21020 Ispra, Italy.

A. Graff, Umweltbundesamt, Bisckarckplatz 1, D-14193 Berlin, Germany.

J. E. Jonson and R. Vautard, LSCE/IPSL Laboratoire, CEA, CNRS, UVSQ, Orme les Mérisiers, Batiment 701, F-91198 Gif-sur-Yvette, France. (robert.vautard@cea.fr)

M. Krol, Netherland Institute for Space Research, Sorbonnelaan 2, NL-3584 CA Utrecht, Netherlands.

P. Roberts, Health Safety and Environment Department, Shell Global Solutions, P.O. Box 1, Chester CH1 3SH, UK.

R. Stern, Institut für Meteorologie, Freie Universität, Carl-Heinrich-Becker-Weg 6-10, D-12165 Berlin, Germany.

L. Tarrasón, M. Van Loon, and P. Wind, EMEP, MSC-W, P.O. Box 43, Blindern 0313, Oslo, Norway.

L. White, CONCAWE, 42 Blunts Wood Road, Haywards Heath RH16 1NB, UK.