Air Quality Model Evaluation International Initiative (AQMEII): Advancing state-of-science in regional photochemical modeling and its applications

S. Trivikrama Rao[1*], Stefano Galmarini[2], and Keith Puckett[3]

1. Atmospheric Modeling and Analysis Division
   National Exposure Research Laboratory
   U.S. Environmental Protection Agency
   Research Triangle Park, NC, USA 27711

2. Institute of Environment and Sustainability
   European Commission, Joint Research Center
   Ispra, Italy

3. Air Quality Research Division
   Environment Canada
   Toronto, Ontario, M3H 5T4, Canada

[*] Corresponding author: Dr. S.T. Rao (E-mail: rao.st@epa.gov; Tel. 919-541-4541)

Abstract

During the past three decades, several workshops were held in the United States and Europe to address the evaluation of air quality models from scientific and operational perspectives. The conventions and practices for evaluation of regional-scale numerical air quality models have been evolving in recent years after a long period of routine use of decades-old model evaluation techniques that are more suited to urban-scale models and for chemically-inert pollutants. Recognizing the need for comprehensive evaluation of regional-scale numerical photochemical modeling systems, an international collaborative research project entitled "Air Quality Model Evaluation International Initiative (AQMEII)" has been initiated by scientists and managers from Europe and North America. The primary objectives of this project are to assess state-of-science in current regional-scale air quality models, rapidly advance the science in these models, and help assess model's credibility in simulating the spatial and temporal features embedded in air quality observations so that the models can be confidently used in research and policy arenas. This paper discusses the motivation for this study and approaches to help improve air quality model evaluation practices.

Introduction

Although the focus in the 1970's was primarily on urban air pollution models, it is well-known that pollution problems such as acid rain, ozone, and fine particulate matter are regional in scope, requiring regional-scale multi-pollutant models (e.g., Rao et al., 2008). In North America and in Europe, several models have been developed by different research groups. These models have undergone extensive development during the last three decades worldwide because of the increased concern regarding the impacts of atmospheric pollution on human health and sensitive ecosystems. For example, during the 1980s, regional-scale acidic deposition models were developed in Europe, Canada, and the United States. Within the framework of the National Acid Precipitation Assessment Program (NAPAP), a number of groups from government, industry, and academia were involved in addressing the so-called "acid rain" problem in the United States. Regional air quality models are now being widely used in North America and Europe for understanding the complex interactions between meteorology and atmospheric chemistry, and pollutant transport and fate. Regional models are also playing an important role in developing emission control policies to comply with the relevant standards for ozone and fine particles, forecasting air quality, and designing ambient monitoring strategies.

In the early 1980's, the American Meteorological Society (AMS) and U.S. Environmental Protection Agency (EPA) held two workshops to discuss and recommend methods for evaluating plume dispersion models (Fox, 1981 and 1984). AMS and EPA also held another workshop in 1984 to discuss evaluation issues relating to regional-scale air quality models, but the workshop participants did not recommend any specific methods for the model performance evaluation (Demerjian, 1985). Hence, the statistical metrics identified by the first AMS/EPA workshop continue to be used for evaluating Gaussian dispersion models as well as numerical regional-scale air quality models, not only in the United States but also in other countries.

The conventions and practices for evaluation of regional-scale numerical air quality models have been evolving in the past few years after a long period of routine use of decades-old model evaluation techniques that are more suited to urban-scale models and for chemically-inert pollutants. As compared to other geophysical sciences like climate, the regional character of air quality issues did not often result in the development of large-scale international collaborative research efforts with participation from groups on both continents. Regional-scale model evaluations have been conducted over the years independently in the two continents (e.g. Hogrefe et al., 2000; Hogrefe et al., 2001 a and b; Biswas et al., 2001; van Loon et al., 2004 and 2007; *Atmospheric Environment* Special Issue on Model Evaluation, 2006; Appel et al., 2007 and 2008; Irwin et al., 2008; Eder et al., 2009; Foley et al., 2009; Smyth et al., 2009) from science and policy perspectives.

While there have been some advances in model evaluation approaches, it was felt that present approaches are too often uncritically applied, without due consideration of the foundations upon which the techniques are based. Hence, AMS and EPA held a workshop in August 2007 to discuss issues relating to the evaluation of regional-scale photochemical modeling systems. Similar workshops on air quality model evaluation were also held in Europe. Arising out of the August 2007 AMS and EPA Workshop, a new framework for regional-scale air quality model evaluation has been introduced

(Dennis et al., 2010). This framework calls for *operational, diagnostic, dynamic, and probabilistic evaluations*, encompassing a full range of confidence-building exercises designed to improve the practice of model evaluation and the use of model results in a regulatory setting.

The above series of Workshops were led by a group of scientists, from Europe (EU) and North America (NA) interested in instigating a significant advance in the way regional-scale air quality modeling systems are evaluated and used. However, the North American and European modeling communities have not worked together in the past on a common model evaluation framework and activities. The initiative arising out of the USA and EU workshops is now known as the Air Quality Model Evaluation International Initiative (AQMEII). AQMEII is aimed at providing a permanent forum to constantly monitor the state of advancement of regional-scale air quality models and model evaluation methodologies (http://**aqmeii**.jrc.ec.europa.eu/). This is achieved through the organization of periodic workshops and modeling activities in which the different aspects of model performance evaluation are considered. The primary purpose of AQMEII is to coordinate international efforts in scientific research in NA and EU and help achieve the following objectives:

- exchange expert knowledge in regional-scale air quality modeling
- identify knowledge gaps in air quality science
- develop innovative methodologies to evaluate uncertainties in air quality modeling
- build a common strategy, adoptable by research communities on both sides of the Atlantic Ocean, for model development and future research priorities
- establish methodologies for model evaluation to increase knowledge on processes and to support the use of models for policy development, and
- initiate coordinated research projects and perform rigorous model inter-comparisons.

A common basis for air quality model performance evaluation is necessary for the research community to promote the use of regional-scale air quality models for decision-making and to most efficiently coordinate the modeling efforts between NA and EU, which are the primary objectives of this initiative. The spatial scales range from the city scale (~1 km) to the continental scale (1000-5000 km), and time scales range from hourly to decadal time scales. Phenomena such as acute episodes or long-term trends will also be considered. Links and feedbacks between climate change and air quality will be considered as well.

Model Performance Evaluation

Assessing Model's Credibility

The four components of model performance evaluation identified by Dennis et al (2010) are as follows. **Operational Model Evaluation** involves the direct comparison of model output with analogous observations in an overall sense. It utilizes routine observations of ambient pollutant concentrations, emissions, meteorology, and other relevant variables (e.g. Appel, 2008). **Diagnostic Model Evaluation** examines the ability of a model to predict pollutant concentrations by correctly capturing physical and chemical processes, and their relative importance as incorporated in the model (e.g., Godowitch et al., 2009). This type of model evaluation generally requires detailed atmospheric measurements that are not routinely available. **Dynamic Model Evaluation** focuses on model's ability to predict changes in air quality levels in response to changes in either source emissions or meteorological conditions (e.g., Gilliland et al., 2008; Godowitch et al., 2010; Pierce et al., 2010). This exercise requires historical case studies where known emission changes or meteorological changes occurred that could be confidently estimated. Finally, **Probabilistic Model Evaluation** attempts to capture statistical properties, including uncertainty or level of confidence in the model results for air quality management or forecasting applications (e.g, Hogrefe and Rao, 2001). This approach is necessarily based on knowledge of uncertainty imbedded in both model predictions and observations. Dennis et al (2010) provide illustrative examples of these four aspects of model performance evaluation.

Characterizing Uncertainty in Air Quality Models

A major difficulty lies in determining the uncertainty underlying a single, apparently deterministic output from grid-based models. There are several technical tools, including data assimilation, that give direct or indirect ways to use a deterministic model in a probabilistic framework. The Bayesian paradigm (Savage, 1954) provides a framework for this, since all uncertainty is represented by probabilities. In particular, all the fixed but "known unknowns" have probability distributions. In the absence of any information, these are called prior distributions but the "knowns", for example, measured values of pollutant concentrations alter these distributions in accordance with Bayes' rule to yield posterior probability distributions. This paradigm has a fundamental role in modern inductive inference since it embraces both "aleatory uncertainty" (that due to chance phenomena such as measurement error) as well as "epistemic uncertainty" (due to lack of knowledge). However, characterizing those distributions in complex dynamic systems is challenging owing to their large numbers of unknowns. One useful approach uses the Bayesian hierarchical model (e.g., Gelman et al., 2004) in which all the unknowns are arranged in a sequence of clusters, the probability distribution for each cluster being conditional on (purely hypothetical) knowledge of all the previous clusters so that Bayes' rule may be applied in steps. This allows uncertainty to be characterized in a structured fashion, one cluster at a time, and simplifies the stochastic modeler's task. Nevertheless, the resulting distribution can still be quite intractable. One example of a Bayesian hierarchical approach, called Bayesian melding, has been developed to account for the difference between the meso-scale processes simulated by the chemical transport model output and the micro-scale processes influencing the individual observations (e.g., Fuentes and Raftery, 2005 for SO_2; Zhong et al., 2007 for O_3).

The above approaches are generally based on analysis of model output as a statistical space-time process, and this analysis can be used to provide estimates of model uncertainty. A completely different approach to characterizing model uncertainty lies in sensitivity analysis methods (e.g., Saltelli et al., 2008). The greatest challenge in this type of uncertainty analysis is identifying and quantifying the different sources of uncertainty introduced into the modeling system, such as errors in model inputs or simplifications in the parameterizations used to represent complex chemical systems. These simplifications all propagate through a highly nonlinear modeling system to produce uncertainties in model output. For example, an uncertainty analysis of a chemical mechanism would include the determination of the standard deviations of that chemical mechanism rate constants and product yields. These standard deviations could then be used in a sensitivity analysis, for example, a constrained Monte-Carlo analysis, to determine the probability of a model's estimated chemical concentrations. Likewise, different cloud schemes or PBL formulations could be used in air quality models to assess the sensitivity of the estimated pollutant concentrations to various physics options. Here, the identification of the sources contributing to significant uncertainty is important because it will help guide where model improvements are needed. While brute-force† methods do exist for quantifying the nature and magnitude of these uncertainties, a comprehensive, theoretically-based and computationally-efficient framework remains to be defined.

Estimating Uncertainties from Ensemble Modeling

The combination of several model results in what is normally defined as ensemble modeling has proven to produce an improvement in the model results when compared with measurements and with the individual model ensemble members. While there has been some progress in ensemble modeling regardless (Dabbert, Miller, 2000; Delle Monache, Stull, 2003; Delle Monache et al, 2003; Galmarini et al, 2001; Galmarini et al, 2004a; Galmarini et al, 2004b; Killip et al, 2003; Krishnamurti et al, 2000; Mallet, Sportise, 2006; McKeen at al, 2005; Mutemi et al, 2007; Potempski et al 2008; Potempski, Galmarini, 2009; Riccio et al 2007; Straume, 2001; Van Loon et al, 2007; Vautard et al, 2008; Vijaya Kumar et al, 2003; Wang et al, 2008; Yun et al, 2004; Ziehmann, 2000) of the way in which the ensemble has been constructed and what the generic expression "ensemble" means, it seems the whole ensemble modeling technique has more the character of a practice than a theoretical framework as pointed out by Potempski and Galmarini (2009). The use of ensemble and in particular multi-model ensembles is felt to be an excellent opportunity for a model inter-comparison exercise, with the added value of improving individual model performance (e.g., Pinder et al., 2009). It is acknowledged that multi-model ensembles can provide some consensus on the model results based on multiple models and, therefore, it should have a prominent role as one of the techniques to be used in future. Given the level of attention gathered around multi-model ensembles (not only in regional-scale air quality models, but also in global-scale and climate models), it is becoming increasingly clear that research efforts need to focus on a more rigorous theoretical framework for ensemble modeling. There is a series of fundamental

† By "brute-force" we mean the estimation of uncertainties by performing multiple runs of the entire modeling system with small, but realistic changes of the independent variables in question.

questions that must addressed; for example, in what way should an ensemble of model results be assembled? what is the minimum number of model results necessary to define the group of an ensemble? how can we get around the model dependence issue? can we diagnose a-priori the ensemble properties based on model characteristics? how can we guarantee a priori maximum coverage of the measurement probability density functions by the ensemble and compensate for missing portions? Given the fact that ensembles are in operational use in weather forecasting, it is suggested that the air quality modeling community work with the meteorological community in advancing air quality ensemble theory.

Representing and communicating model uncertainty

Much of the motivation for enhanced model evaluation approaches stems from the need to use model output to provide guidance in the policy-making realm. Communication within the scientific realm is a well-established practice, and makes use of particular media and language. By contrast, communication between scientific and policy realms is difficult because these two realms use different media and language. If model results, as evaluated by scientists are to be transferred to the policy realm, it is clear that particular attention must be paid to the language of communication. For this reason, scientists need to work with communications specialists, journalists and psychologists to develop communications strategies that will be effective in the policy realm. This work must include the development of methods for display and presentation of model output, including animations, spaghetti plots and other devices. Of particular difficulty will be the communication of the linked space-time nature of air pollution fields, and the use of probabilities in making environmental decisions.

Operational, diagnostic, and dynamic evaluation approaches complement one another by not only characterizing how well the model simulated the air quality levels at that time, but how well the model captures the role and contributions of individual inputs and processes and the ability of the air quality model to respond correctly to changes in these factors (Dennis et al., 2010). While it is true that all evaluation approaches use some form of statistical techniques, under this framework probabilistic evaluation is viewed as a comprehensive approach to evaluation that goes beyond the mere application of statistical tools.

Current Focus of AQMEII

As a result of the May 2009 AQMEII Workshop in Stresa, Italy (see Galmarini et al., 2010), an initial exercise has been launched in which participating modeling groups in the United States and Canada and across Europe, will be using their regional-scale air quality modeling systems to simulate full-years (years 2002 or 2003 and 2006) retrospective continental applications on both sides of the Atlantic Ocean with a common reference model input data set , namely the emissions inventory and lateral boundary conditions, and applying the four elements of the above model evaluation framework to inter-compare results across the models as well as compare results with routine observations and special field studies. To this end, a reference database consisting of satellite and aircraft observations has been created. Evaluation of emissions, meteorological and air quality models within the photochemical modelling systems will also be considered. The modeling domains that will be used for NA and EU simulations are displayed in Figures 1 and 2.

Invitations to participate in AQMEII were sent during mid-2009 and commitments have been secured thus far for the participation of regional models such as CMAQ and CAMx models (United States), AURAMS (Canada), CHIMERE (France), RAMS-CAMx (Greece), EMEP (Norway), EURAD (Germany), HIRLAM (Denmark, Finland), and LOTOS (Netherlands). Other modeling groups are currently securing funding support and are expected to join AQMEII. So far, 23 modeling groups from 15 countries are engaged in the AQMEII activity. All models will be utilizing the same base emissions inventory and lateral air quality boundary conditions, which have been developed specifically for the AQMEII project. Also, common analysis grids have been defined for each continent to facilitate model inter-comparisons. Data from standard meteorological and air quality monitoring networks across NA and EU for 2006 will be used for model evaluation. Emphasis for air quality is placed on ozone, nitrogen oxides, $PM_{2.5}$ and PM_{10} including chemical species components, as well as surface deposition of key pollutants. Other measurements from regional field studies, profilers, satellites, and commercial aircraft (including the MOZAIC program of measurements on-board commercial European aircraft) will be used in the evaluation. The Joint Research Center (JRC) of the European Commission will house the data archive and make available to participants the web-based statistical and graphical facilities of its ENSEMBLE analysis system (Galmarini et al., 2004). Several technical documents and protocol for participation in the AQMEII activity can be found on the AQMEII web site. It is envisioned that the participating scientists will run probe multiple models, rather than running their own, to elucidate the similarities and differences among different models from the diagnostic model evaluation perspective. Thus, this major collaborative effort is a first of its kind in bringing together modelers and data analysts from Europe and North America in an attempt to assess the current state-of-science in air quality models and work together to rapidly advance the science in regional-scale numerical photochemical models.

Next Steps

A workshop to present preliminary analyses of the first phase of the AQMEII project is scheduled to take place during September 26-27, 2010 in Torino, Italy in conjunction with the 31st NATO/SPS International Technical Meeting Air pollution Modeling and Its Application. It is expected that this Workshop will determine future analyses for the continental model simulations as well as suggest additional model simulations and evaluation datasets for the next phase of AQMEII. A subsequent Workshop will be held in Research Triangle Park, NC, USA during 2011 for discussion of final analyses for the 2006 model evaluations as part of the first phase of AQMEII, to plan publications and communications of the results, and to plan and coordinate the next phase of AQMEII to involve simulations of other years and field-study periods, and to include new sources of data for the model evaluations. A special issue of *Atmospheric Environment*, devoted to AQMEII-related papers will be published in 2012. The first phase of AQMEII will be concluded with a workshop in Europe in 2012.

Future Focus of AQMEII

Recently, regional models have been linked to global models for examining the impacts of climate change on future air and water resources. Weaver et al. (2009) provide a summary of studies relating to downscaling of global climate models to regional climate models for assessing the effects of climate change on ozone air quality. Also, coupled meteorology-chemistry models are being developed by various research groups in EU and NA to better simulate the effects of atmospheric loading of aerosols on the radiative forcing. Hence, it is envisioned that the second phase of AQMEII will consider global-to-regional models to better understand the interactions of climate change and air quality and the linkages to human health and ecosystems. The intended outcome of AQMEII is a set of international collaborations involving model evaluation exercises carried out on shared data sets using different state-of-science multiscale photochemical models.

Acknowledgements

We gratefully acknowledge the contribution of various groups to the first air Quality Model Evaluation international Initiative (AQMEII) activity. The following agencies have prepared the data sets used in the preparation phase of this study: U.S. EPA (North American emissions processing and gridded meteorology); U.S. EPA, Environment Canada, Mexican Secretariat of the Environment and Natural Resources (Secretaría de Medio Ambiente y Recursos Naturales-SEMARNAT) and National Institute of Ecology (Instituto Nacional de Ecología-INE) (North American national emissions inventories); TNO (European emissions processing); Laboratoire des Sciences du Climat et de l'Environnement, IPSL, CEA/CNRS/UVSQ (gridded meteorology for Europe); ECMWF/GEMS project & Météo-France/CNRM-GAME (Chemical boundary conditions). Ambient North American concentration measurements were extracted from Environment Canada's National Atmospheric Chemistry Database (NAtChem) PM database and provided by several U.S. and Canadian agencies (AQS, CAPMoN, CASTNet, IMPROVE, NAPS, SEARCH and STN networks); North American precipitation-chemistry measurements were extracted from NAtChem's precipitation-chemistry data base and were provided by several U.S. and Canadian agencies (CAPMoN, NADP, NBPMN, NSPSN, and REPQ networks); the WMO World Ozone and Ultraviolet Data Centre (WOUDC) and its data-contributing agencies provided North American and European ozonesonde profiles; NASA's AErosol RObotic NETwork (AeroNet) and its data-contributing agencies provided North American and European AOD measurements; the MOZAIC Data Centre and its contributing airlines provided North American and European aircraft takeoff and landing vertical profiles; for European air quality data the following data centers were used: EMEP European Environment Agency/European Topic Center on Air and Climate Change/AirBase provided European air- and precipitation-chemistry data. Data from meteorological station monitoring networks were provided by NOAA and Environment Canada (for the US and Canadian meteorological network data) and the National Center for Atmospheric Research (NCAR) data support section. Joint Research

Center Ispra/Institute for Environment and Sustainability provided its ENSEMBLE system for model output harmonization and analyses and evaluation.

Disclaimer

The views expressed here are those of the authors and do not necessarily reflect the views and policies of the United States Environmental Protection Agency, European Commission, and Environment Canada.

References

Appel, K.W., P. Bhave, A. Gilliland, G. Sarwar, S. Roselle, 2008: Evaluation of the Community Multiscale Air Quality (CMAQ) model version 4.5: Sensitivities impacting model predictions: Part II – Particulate matter, *Atmos. Environ*., 42, 6057-6066

Appel, W., A. Gilliland, G. Sarwar, R. Gilliam, 2007: Evaluation of the Community Multiscale Air Quality (CMAQ) model version 4.5: Sensitivities impacting model predictions: Part I - Ozone, *Atmos. Environ*., 41, 9603-9615.

Biswas, J., C. Hogrefe, S.T. Rao, W. Hao, and G. Sistla, 2001: Evaluating the performance of regional-scale photochemical modeling systems: Part III – Precursor predictions, *Atmos. Environ*., 35, 6129-6149.

Dabbert W.F., Miller E., Uncertainty, ensembles and air quality dispersion modelling: applications and challenges, *Atmos. Environ.* 34, 4667-4673, 2000

Delle Monache L, Deng X, Zhou Y., Stull R, Ozone ensemble forecasts: 1. A new ensemble design. Journal of Geophysical Research 111(d5):D05307, 2006

Delle Monache L., Stull R.B., An ensemble air-quality forecast over western Europe during an ozone episode, *Atmos. Environ.* vol. 37, Issue 25, pp. 3469-3474, 2003

Demerjian, K.L., 1985: Quantifying uncertainty in long range transport models: A summary of the AMS workshop on sources and evaluation of uncertainty in long range transport models. *Bull. of the Amer. Meteor. Soc.*, 66, 1533-1540.

Dennis, R., T. Fox, M. Fuentes, A. Gilliland, S. Hanna, C. Hogrefe, J.Irwin, S.T. Rao, R. Scheffe, K.Schere, D. Steyn, and A.Venkatram, 2010: A framework for evaluating of regional-scale numerical photochemical modeling systems. *Environ. Fluid Mech.,* In Press, DOI: 10.1007/s10652-009-9163-2

Eder, B., D. Kang, S.T. Rao, R. Mathur, S. Yu, T. Otte, K. Schere, R. Wayland, S. Jackson, P. Davidson, J. McQueen, and G. Bridgers, 2009: A demonstration of the use of national air quality forecast guidance for developing local air quality index forecasts. *Bull. of the Amer. Meteor. Soc.*, DOI: 10.1175/2009BAMS2734.1

Foley, K. M., Roselle, S. J., Appel, K. W., Bhave, P. V., Pleim, J. E., Otte, T. L., Mathur, R., Sarwar, G., Young, J. O., Gilliam, R. C., Nolte, C. G., Kelly, J. T., Gilliland, A. , and Bash, J. O, 2009: Incremental testing of the community multiscale air quality (CMAQ) modeling system version 4.7, *Geosci. Model Dev. Discuss*., 2, 1245-1297.

Fox, D.G., 1981: Judging air quality model performance: A summary of the AMS Workshop on Dispersion Model Performance, *Bull. of the Amer. Meteor. Soc.,* 62, 599-609.

Fox, D.G., 1984: Uncertainty in air quality modeling, *Bull. of the Amer. Meteor. Soc*., 65, 27-36.

Fuentes M, and Raftery A., 2005: Model evaluation and spatial interpolation by Bayesian combination of observations with outputs from numerical models, *Biometrics*, 61, 36-45.

Galmarini S., Bianconi R., Bellasio R., Graziani G., Forecasting consequences of accidental releases from ensemble dispersion modelling, *Journal of Environmental Radioactivity* 57, 203-219, 2001

Galmarini S., D.G. Steyn, K. Schere, and M. Moran, 2010: Advancing the evaluation of regional-scale air quality models, EUR24245, ISBN 978-92-79-15007-4.

Galmarini S., R. Bianconi, R.Addis, S. Andronopoulos, P. Astrup, J. C. Bartzis, R. Bellasio, R. Buckley, H. Champion, M. Chino, R. D'Amours, E. Davakis, H. Eleveld, H. Glaab, A. Manning, T. Mikkelsen, U. Pechinger, E. Polreich, M. Prodanova, H. Slaper, D. Syrakov, H. Terada, L. Van der Auwera, Ensemble dispersion forecasting, Part II: Application and evaluation, *Atmos. Environ.*, 38, 28, 4619-4632, 2004b

Galmarini S., R. Bianconi, W. Klug, T. Mikkelsen, R. Addis, S. Andronopoulos, P. Astrup, A. Baklanov, J. Bartniki, J. C. Bartzis, R. Bellasio, F. Bompay, R. Buckley, M. Bouzom, H. Champion, R. D'Amours, E. Davakis, H. Eleveld, G. T. Geertsema, H. Glaab, M. Kollax, M. Ilvonen, A. Manning, U. Pechinger, C. Persson, E. Polreich, S. Potemski, M. Prodanova, J. Saltbones, H. Slaper, M. A. Sofiev, D. Syrakov, J. H. Sørensen, L.

Van der Auwera, I. Valkama, R. Zelazny, Ensemble dispersion forecasting, Part 1: Concept, Approach and indicators, *Atmos. Environ.*, 38, 28, 4607-4617, 2004a

Galmarini, S., R. Bianconi, R. Addis, S. Andronopoulos, P. Astrup, J. C. Bartzis, R. Bellasio, R. Buckley, H. Champion, M. Chino, R. D'Amours, E. Davakis, H. Eleveld, H. Glaab, A. Manning, T. Mikkelsen, U. Pechinger, E. Polreich, M. Prodanova, H. Slaper, D. Syrakov, H. Terada and L. Van der Auwera, 2004: Ensemble dispersion forecasting – Part I: Concept, approach, and indicators, 2004, *Atmos. Environ*., 38, 4607-4617.

Gelman, A, J. Carlin, H. Stern, and D. Rubin, 2004: *Bayesian data analysis*, Second Edition. London: Chapman and Hall/CRC Press.

Gilliland, A, C. Hogrefe, R. Pinder, J. Godowitch, K. Foley, and S.T. Rao, 2008: Dynamic evaluation of regional air quality models: Assessing changes in O3 stemming from changes in emissions and meteorology, *Atm. Environ*., 42, 5110-5123.

Godowitch, J., G. Pouliot, and S.T. Rao, 2010: Assessing multi-year changes in modelled and observed NOx concentrations from a dynamic evaluation perspective, *Atm. Environ.*, 44, 2894-2901

Hogrefe, C. and S.T. Rao, 2001: Demonstrating of the air quality standards: Integration of observations and model predictions into the probabilistic framework, J. Air & Waste Manage. Assoc., 51, 1060-1072.

Hogrefe, C., S.T. Rao, I. Zurbenko, and P. Porter, 2000: Investigating the information in ozone observations and model predictions relevant to regulatory policies in the eastern United States, *Bull. Amer. Meteor*., Soc., 81, 2083-2106.

Hogrefe, C., S.T. Rao, P. Kasibhatla, G. Kallos, C. Trembach, W. Hao, D. Olerud, A. Xiu, J. McHenry, and K. Alapaty, 2001a: Evaluating the performance of regional-scale photochemical modeling systems: Part I – Meteorological predictions, *Atmos. Environ*., 35, 4159-4174.

Hogrefe, C., S.T. Rao, P. Kasibhatla, W. Hao, G. Sistla, R. Mathur, and J. McHenry, 2001b: Evaluating the performance of regional-scale photochemical modeling systems: Part II – Ozone predictions, Atmos. Environ., 35, 2175-4188.

Irwin, J., K. Civerolo, C. Hogrefe, W. Appel, K. Foley, J. Swall, 2008: A procedure for inter-comparing the skill of regional-scale air quality model simulations of daily maximum 8-h ozone concentrations, *Atmos. Environ*., 42, 5403-5412.

Krishnamurti, T.N., C.M. Kishtawal, Z. Zhang, T. LaRow, D. Bachiochi, E. Williford, S. Gadgil, S. Surendran, Multimodel ensemble forecasts for weather and seasonal climate, *J. Clim*. 13 issue 23, pp. 4196-4216, 2000

Loon M. van, et al., 2007, Evaluation of long term ozone simulations from seven regional scale air quality models and their ensemble average, *Atmos. Environ*. 41, 2083-2097

Loon M. van, M.G.M. Roemer and P.J.H. Builtjes, 2004, Model intercomparison in the framework of the review of the unified EMEP-model, TNO-rep R2004/282

Mallet, V., and B. Sportisse, Ensemble-based air quality forecasts: A multimodel approach applied to ozone, *J. Geophys. Res.*, 111, D18302, doi:10.1029/2005JD006675, 2006

McKeen S., Wilczak J., Grell G., Djalalova I., Peckham S., Hsie E.-Y., Gong W., BouchetV., Menard S., Moffet R., McHenry J., McQueen J., Tang Y., Carmichael G.R., Pagowski M., Chan A., Dye T., Frost G., Lee P., Mathur R., Assessment of an ensemble of seven real-time ozone forecasts over eastern North America during the summer of 2004, *Journal of Geophysical Research D: Atmospheres*, vol. 110, is. 21, pp. 1-16, 2005

Mutemi, J.N., Ogallo, L.A., Krishnamurti, T.N., Mishra, A.K., Vijaya Kumar, T.S.V., 2007, Multimodel based superensemble forecasts for short and medium range NWP over various regions of Africa, *Meteorol. Atmos. Phys.*, vol. 95 issue 1-2, pp 87-113

Pierce, T., C. Hogrefe, S.T. Rao, P.S. Porter, and J.Y. Ku 2010: Dtnamic evaluation of a regional air quality model: Assessing the emissions-induced weekly cycle, *Atm. Environ*., DOI:10.1016/j.atmosenv.2010.05.046.

Pinder, R., R. Gilliam, K.W. Appel, S. Napelenok, K. Foley, and A. Gilliland, 2009,: Efficient probabilistic estimates of surface ozone concentration using ensemble of model configurations and direct sensitivity calculations, *Environ. Sci. Technol.*, 43, 2388-2393.

Potempski S., S. Galmarini, R. Addis, P. Astrup, S. Bader, R. Bellasio, R. Bianconi, F. Bonnardot, R. Buckley, R. D'Amours, A. van Dijk, G. Geertsema, A. Jones, P. Kaufmann, U. Pechinger, C.Persson, E. Polreich, M. Prodanova, L. Robertson, J. Sørensen, D. Syrakov, 2008, Multi-model ensemble analysis of the ETEX-2 experiment, *Atmospheric Environment* 42, 7250-7265

Potempski, S. and S. Galmarini, 2009: *Est modus in rebus*: analytical properties of multi-model ensembles, Atmos. Chem. Phys., 9, 9471-9489.

Rao, S.T., C. Hogrefe, T. Holloway, and G. kallos, 2008: Long-range transport of atmospheric pollutants and transboundary pollution, *World Atlas of Atmospheric Pollution*, Edited by R.S. Sokhi, Anthem Press, ISBN 978-1- 84331-289-5, 120p

Riccio A., G. Giunta, S. Galmarini, 2007: Seeking for the rational of the Median Model: the optimal combination of multi-model ensemble, *Atmos. Chem. Phys,* 7:1-14.

Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D. Saisana, M., Tarantola, S., 2008: *Global Sensitivity Analysis. The Primer*, John Wiley & Sons publishers.

Savage, L.J., 1954: *The Foundations of Statistics*. New York, Wiley.

Smyth, S., W. Jiang, H. Roth, M. Moran, P. Makar, F.Yang, V. Bouchet, H.Landry, 2009: A comparative performance evaluation of the AURAMS and CMAQ air-quality modelling systems, *Atmos. Environ.*, 43, 1059-1070.

Straume A.G., 2001: A more extensive investigation of the use of ensemble forecasts for dispersion model evaluation. *Journal of Applied Meteorology* 40 (2001), pp. 425–445.

Van Loon M, R. Vautard, M. Schaap, R. Bergström, B. Bessagnet, J. Brandt, P.J.H. Builtjes, J.H. Christensen, C. Cuvelier, A. Graff, J.E. Jonson, M. Krol, J. Langner, P. Roberts, L. Rouil, R. Stern, L. Tarrasón, P. Thunis, E. Vignati, L. White, P. Wind, 2007: Evaluation of long-term ozone simulations from seven regional air quality models and their ensemble, *Atmos. Envir.* 41: 2083-2097.

Vautard, R., M. Schaap, R. Bergström, B. Bessagnet, J. Brandt, P.J.H. Builtjes, J. Christensen, C. Cuvelier, V. Foltescu, A. Graff, A. Kerschbaumer, M. Krol,, P. Roberts, L. Rouïl, R. Stern, L. Tarrason, P. Thunis, E. Vignati, P. Wind, 2008: Skill and uncertainty of a regional air quality model ensemble, *Article in press*, DOI: 10.1016/j.atmosenv.2008.09.083.

Vijaya Kumar, T.S.V., T.N. Krishnamurti, M. Fiorino, M. Nagata, 2003: Multimodel superensemble forecasting of tropical cyclones in the Pacific, *Mon. Weather Rev.* 131 issue 3, pp. 574-583.

Wang, B., J-Y. Lee, I.S. Kang, J. Shukla, C-K. Park, A. Kumar, J. Schemm, S. Cocke, J-S Kug, J-J.Luo, T. Zhou, B. Wang, X. Fu, W-T. Yun, O. Alves, E.K. Jin, J. Kinter, B. Kirtman, T.N. Krishnamurti, N.C. Lau, W. Lau, P. Liu, P. Pegion, T. Rosati, S. Schubert, W. Stern, M. Suarez, T. Yamagata, 2008; Advance and prospectus of seasonal prediction: assessment of the APCC/CliPAS 14-model ensemble retrospective seasonal prediction (1980-2004), *Climate Dynamics*, pp. 1-25.

Weaver, C., X.-Z. Liang, J. Zhu, P. J. Adams, P. Amar, J. Avise, M. Caughey, J. Chen, R. C. Cohen, E. Cooter, J. P. Dawson, R. Gilliam, A. Gilliland, A. H. Goldstein, A. Grambsch, D. Grano, A. Guenther, W. I. Gustafson, R. A. Harley, S. He, B. Hemming, C. Hogrefe, H.-C. Huang, S. W. Hunt, D. J. Jacob, P. L. Kinney, K. Kunkel, J.-F. Lamarque, B. Lamb, N. K. Larkin, L. R. Leung, K.-J. Liao, J.-T. Lin, B. H. Lynn, K. Manomaiphiboon, C. Mass, D. McKenzie, L. J. Mickley, S. M. O'neill, C. Nolte, S. N. Pandis, P. N. Racherla, C. Rosenzweig, A. G. Russell, E. Salathé, A. L. Steiner, E. Tagaris, Z. Tao, S. Tonse, C. Wiedinmyer, A. Williams, D. A. Winner, J.-H. Woo, S. Wu, and D. J. Wuebbles, 2009: A Preliminary Synthesis of Modeled Climate Change Impacts on U.S. Regional Ozone Concentrations, *Bull. Amer. Meteor. Soc*., 90, 1843-1863.

Yun, W.T., L. Stefanova, A.K. Mitra, T.N. Krishnamurti, 2004: Multi-model synthetic superensemble prediction system, *Bull. of Amer. Met. Soc.* 4119-4122.

Zhong L, N. Le, and J. Zidek, 2007: Calibrating Deterministic Modeling Output with Application to Ozone Fields. TR 232, Department of Statistics, University of British Columbia, Canada.

Ziehmann, C., Comparison of a single model EPS with a multi-model ensemble consisting of a few operational models, *Tellus* 52A, 280-299, 2000

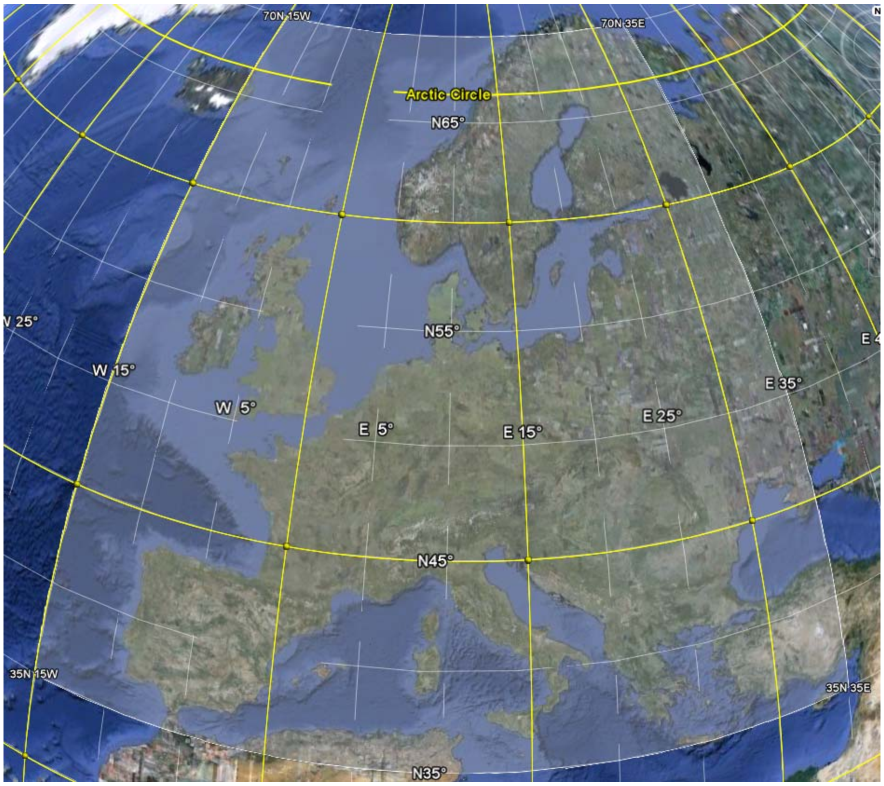

Figure 1. Model domain covering Europe

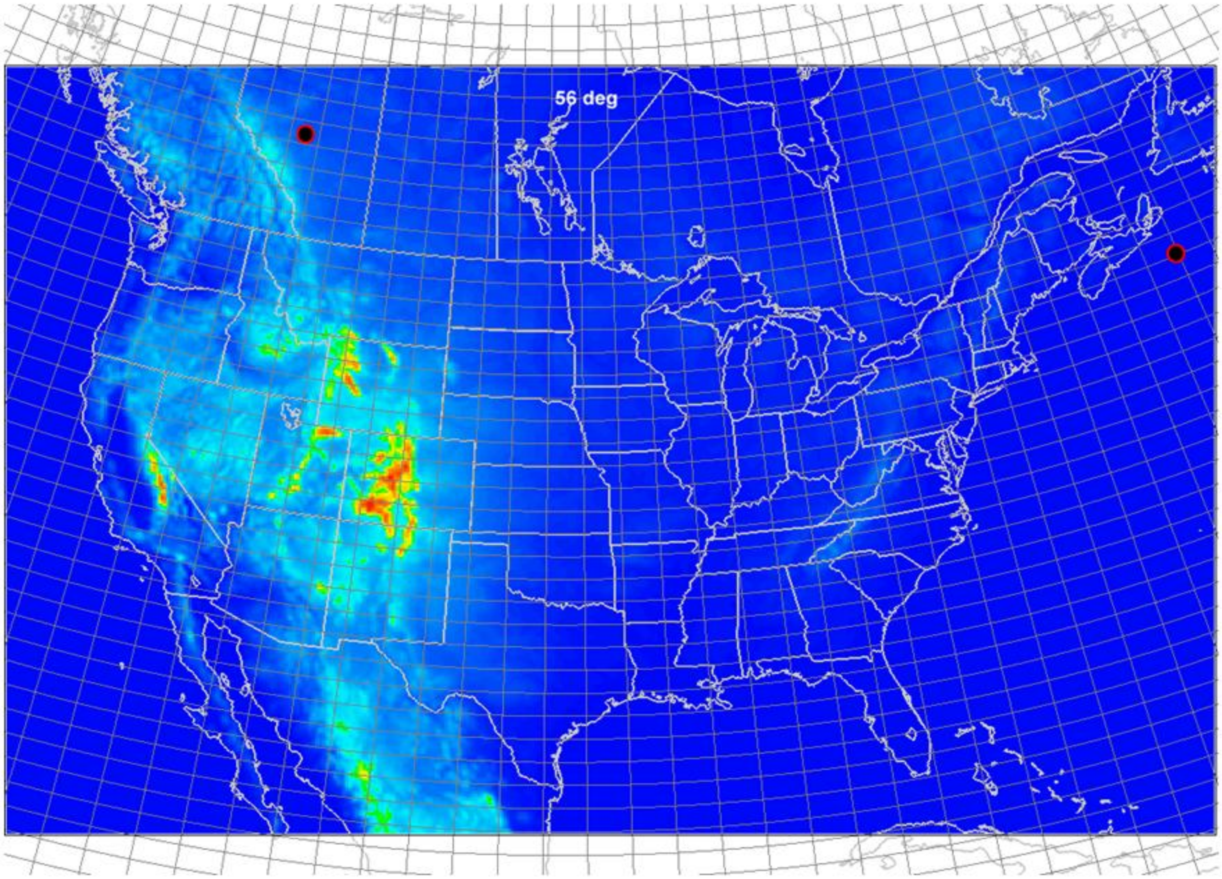

Figure 2. Model domain covering continental United States and Southern Canada