Feedbacks between Air Pollution and Weather, Part 2: Effects on Chemistry

P.A. Makar[1], W. Gong[1], C. Hogrefe[2], Y. Zhang[3], G. Curci[4], R. Žabkar[5,6], J. Milbrandt[7], U. Im[8], A. Balzarini[9], R. Baró[10], R. Bianconi[11], P. Cheung[1], R. Forkel[12], S. Gravel[13], M. Hirtl[14] ,L. Honzak[6], A. Hou[1], P. Jiménez-Guerrero[10], M. Langer[14], M.D. Moran[1], B. Pabla[1], J.L. Pérez[15], G. Pirovano[9], R. San José[15], P. Tuccella[16], J. Werhahn[12], J. Zhang[1], S. Galmarini[8]

[1]*Air-Quality Research Division, Environment Canada,* Toronto, *Canada*
[2]*Atmospheric Modeling and Analysis Division, Environmental Protection Agency, Research Triangle Park, USA*
[3]*Department of Marine, Earth and Atmospheric Sciences, North Carolina State University, Raleigh, USA*
[4]*University of L'Aquila, L'Aquila, Italy*
[5]*University of Ljubljana, Ljubljana, Slovenia*
[6]*Center of Excellence SPACE-SI, Ljubljana, Slovenia*
[7]*Meteorological Research Division, Environment Canada, Montreal, Canada*
[8]*Joint Research Centre, European Commission, Ispra, Italy*
[9]*RSE, Milano, Spain*
[10]*University Murcia, MAR-UMU, Spain*
[11]*Enviroware (www.enviroware.com), Milan, Italy*
[12]*Karlsruhe Inst. of Technology, IMK-IFU, Garmisch-Partenkirchen*
[13]*Air-Quality Research Division, Environment Canada, Montreal, Canada*
[14]*ZAMG, Vienna, Austria*
[15]*Technical Univ. of Madrid, ESMG-UPM, Spain*
[16]*University L'Aquila, CETEMPS, Italy*

Corresponding author telephone/ fax and email: 1-416-739-4692 / 1-416-739-4288, paul.makar@ec.gc.ca

For submission to *Atmospheric Environment Special Issue on Phase 2 of the Air-Quality Model Evaluation International Initiative, May 28, 2014.*

Abstract

Fully-coupled air-quality models running in "feedback" and "no-feedback" configurations were compared against each other and observation network data as part of Phase 2 of the Air Quality Model Evaluation International Initiative. In the "no-feedback" mode, interactions between meteorology and chemistry through the aerosol direct and indirect effects were disabled, with the models reverting to climatologies of aerosol properties, or a no-aerosol weather simulation, while in the "feedback" mode, the model-generated aerosols were allowed to modify the models' radiative transfer and/or cloud formation processes. Annual simulations with and without feedbacks were conducted for domains in North America for the years 2006 and 2010, and for Europe for the year 2010. Comparisons against observations via annual statistics show model-to-model variation in performance is greater than the within-model variation associated with feedbacks. However, during the summer and during intense

emission events such as the Russian forest fires of 2010, feedbacks have a significant impact on the chemical predictions of the models.

The aerosol indirect effect was usually found to dominate feedbacks compared to the direct effect. The impacts of direct and indirect effects were often shown to be in competition, for predictions of ozone, particulate matter and other species. Feedbacks were shown to result in local and regional shifts of ozone-forming chemical regime, between NO_x- and VOC-limited environments. Feedbacks were shown to have a substantial influence on biogenic hydrocarbon emissions and concentrations: North American simulations incorporating both feedbacks resulted in summer average isoprene concentration decreases of up to 10%, while European direct effect simulations during the Russian forest fire period resulted in grid average isoprene changes of -5 to +12.5%. The atmospheric transport and chemistry of large emitting sources such as plumes from forest fires and large cities were shown to be strongly impacted by the presence or absence of feedback mechanisms in the model simulations. Summertime model performance for ozone and other gases was improved through the inclusion of indirect effect feedbacks, while performance for particulate matter was degraded, suggesting that current parameterizations for in- and below cloud processes, once the cloud locations become more directly influenced by aerosols, may over- or under-predict the strength of these processes. Process parameterization-level comparisons of fully coupled feedback models are therefore recommended for future work, as well as further studies using these models for the simulations of large scale urban/industrial and/or forest fire plumes.

Introduction

In the first phase of the Air-Quality Model Evaluation International Initiative (AQMEII, Galmarini *et al*, 2012a), the simulations from a large suite of air-quality models were compared against each other and observations from monitoring networks in both North America (NA) and Europe (EU). Twenty-one research groups participated in this study, which was designed to evaluate the models and ensembles of the models through the use of a common simulation period, boundary conditions and emissions data for

both NA and EU, for the year 2006. A particular focus of the intercomparison was the investigation of how to generate ensemble forecasts from the models with the minimum possible error relative to observations, for O_3 (Solazzo *et al*, 2012a), and for $PM_{2.5}$ (Solazzo *et al*, 2012b). Clustering analysis was shown to provide an improved ensemble O_3 forecast relative to the more typical averaging through investigating the predictions of 15 ensemble members (Solazzo *et al*, 2012a). All models in a ten-member ensemble had negative-biased $PM_{2.5}$ simulations, and large variations between the models' predictions of model $PM_{2.5}$, speciated $PM_{2.5}$ and its precursors were noted.

Most of the models participating in the first phase of AQMEII were "off-line" models, that is, models in which the meteorology is generated *a priori* by a weather forecast model. In contrast "on-line" models incorporate both chemical and meteorological components into a single system. While off-line models have certain advantages (e.g. the potential to use different meteorological driving models), on-line models have other advantages such as a reduction in potential interpolation errors between meteorological and chemical model grids, and the elimination of the potentially large amount of processing time required for the input of meteorological model files, c.f. Grell *et al*. (2005), Zhang (2008), Moran *et al*, 2010; and a review of models in Baklanov *et al* (2014). On-line models may be partially coupled (while both chemistry and meteorology are contained within the same model, only the meteorological variables are allowed to modify the chemistry, not vice-versa, c.f. Moran *et al*, 2010), or fully coupled (where, in addition, chemical species are also allowed to modify the meteorology). The aerosols generated by a fully coupled model's chemistry and/or emissions may thus participate in radiative transfer calculations (aerosol direct effect), and in the formation of clouds as cloud condensation nuclei, which in turn may change the radiative and other properties of the simulated clouds (aerosol indirect effect). Both of these processes have long been recognized to be of importance in the realm of global and regional climate modelling (c.f. Forster *et al*, 2007; Giorgi et al, 2003). However, the climate models typically lack the more detailed chemistry and aerosol microphysics found in regional air-quality models, due to additional computational burden associated with transporting the necessary suite of chemical species, including size-

resolved particulate matter, and the additional processing time associated with more detailed gas and aerosol chemistry as well as aerosol microphysics.

The second phase of AQMEII (AQMEII-2) compares the annual simulations of fully coupled models, which include the aerosol direct and/or indirect effects, making use of the datasets and ENSEMBLE evaluation system generated under AQMEII phase 1 (Galmarini *et al*, 2004a,b, Galmarini *et al*, 2012b)), as well as new datasets collected for the year 2010 in both NA and the EU. The performance of these fully coupled models is evaluated elsewhere in this special issue (cf. Im *et al*, 2014 (a,b), Yahya *et al*., 2014a, b; Campbell *et al*., 2014; Wang *et al*., 2014a, Brunner et al, 2014, Hogrefe et al., 2014, this issue). Here, we focus on the feedback processes themselves, and attempt to address the following questions:

(1) Does the incorporation of feedbacks in on-line models result in systematic changes to their predicted chemistry and meteorology?

(2) Do the changes vary in time and space?

(3) To what extent does the incorporation of feedbacks improve or worsen model results, compared to observations?

The final question is of importance in the context of meteorological and air-quality forecasts. The models presented here may be used in forecast mode, and the incorporation of a realistic representation of feedbacks might be expected to improve forecast accuracy in forecasts of both meteorology and air-quality. The work which follows thus provides an assessment of model accuracy from the standpoint of forecasting. In the current work (Part 2), we examine the effects of feedbacks on the model's chemical predictions. In Part 1, we examined the effects of feedbacks on the models' meteorological predictions.

Methodology

Ideally, the study of the impacts of feedbacks on coupled model simulations would make use of two versions of each air-quality model, one in which the feedback mechanisms have been disabled, and another in which the feedback mechanisms have been enabled. However, not all of the participating modelling groups in AQMEII-2 had the computational resources to carry out both non-feedback and feedback simulations. For the North American AQMEII simulations, only the group contributing the GEM-MACH model (Moran *et al*, 2010), modified for both aerosol direct and indirect effects, was able to simulate both of the years 2006 and 2010. The WRF-CMAQ model was used to generate direct-effect only feedback simulations for 2006 and 2010, but no-feedback simulations were only generated for summer periods of each year. The WRF-CHEM model with a configuration for both direct and indirect effects was used for feedback simulations of both years, but no-feedback simulations were not available for this model on this domain (simulations for the month of July, 2006, estimated the relative contributions of aerosol direct and indirect effects to chemistry and meteorology, for that model; Wang *et al*., 2014b, this issue). However, simulations of weather using the WRF model, alone, in the absence of feedbacks, were used to generate meteorological simulations which could then be used for comparison to the meteorological output of the WRF-CHEM feedback simulations (see Makar et al, 2014a, this issue). For the EU AQMEII simulations, three WRF-CHEM simulations were compared for the year 2010: a version 3.4.1 no-feedback simulation in which all aerosol interactions with meteorology were disabled, a version 3.4.1 direct effect simulation, and a version 3.4.0 simulation incorporating both direct and indirect effects.

An important difference in the "no-feedback" simulations of the models needs to be noted at the outset, in that while feedbacks are disabled, the underlying meteorological models may have parameterizations to represent aerosol effects, and these parameterizations differ between the models. The no-feedback versions of the WRF-CMAQ and WRF-CHEM models have no parameterized aerosol impacts on meteorology. The RRTMG parameterization as used here (Clough *et al*, 2005) does not include aerosol

parameterizations for radiative transfer; the aerosols are effectively set to zero concentration, unlike later versions of the WRF weather forecast portion of these models. Similarly, the aerosol indirect effect is not parameterized in the no-feedback version of these models' two-moment cloud microphysics scheme (Morrison *et al*, (2009)) as implemented here; instead, a constant cloud droplet number of 250 cm^{-3} is used (Forkel *et al*, 2012). Thus, the "no-feedback" configuration of these models has no representation of the aerosol direct effect, and a climatological or "typical conditions" cloud droplet number density in place of the aerosol indirect effect. Within GEM-MACH's radiative transfer module (Li and Barker, 2005), the no-feedback configuration makes use of specified functions, representing continental or marine air mass typical conditions, for aerosol optical depth, single-scattering albedo, and asymmetry factor (Toon and Pollack, 1976). GEM-MACH's default no-feedback indirect effect parameterization similarly makes use of a simple function linking cloud condensation nuclei numbers to supersaturation, for marine and continental air masses (Cohard *et al*., 1998) within the cloud microphysics scheme of Milbrandt and Yao (2005). Thus, the no-feedback configuration for all of the models used here does not imply no aerosol effects whatsoever, but may imply the use of parameterizations or simplifying assumptions. For the WRF-based models, the no-feedback simulations used no direct effect parameterizations and a prescribed cloud droplet number, and for the GEM-MACH model, a parameterization is used for both aerosol direct and indirect effects. Differences between the models' response to feedbacks are thus also with respect to these pre-existing parameterizations or simplifications, and differences between these approaches may influence the variation in response between the models to feedbacks.

The models, their main features with regards to feedbacks, and the details on the periods simulated are presented in Table 1. The model predictions were not free-running: GEM-MACH and WRF-CHEM followed the AQMEII-2 protocol of performing simulations for successive 48 hour periods starting from either meteorological analysis or making use of nudging, with a 12 to 24 hour meteorology-only spin up period leading to each 48-hour simulation period. In this protocol, the chemical state of the atmosphere is preserved between the 48 hour simulations, but the meteorology is constrained by observations at each re-

initialization rather than free-running. The WRF-CMAQ simulations deviated from this protocol by performing continuous simulations and applying nudging of upper layer temperature, winds, and water vapor as well as soil moisture and temperature throughout the simulation as described in Hogrefe et al. (2014). A comparison of the two approaches for July 2006 showed a small reduction in the WRF-CMAQ simulated direct feedback effect due to the use of continuous nudging but also showed improved model performance for 2m temperature (Hogrefe et al., 2014). The simulated feedback effects in all three modelling systems are therefore also constrained, and may be less than would be the case for free-running models. The models are fully coupled, but the technical details of the coupling differ: in the case of WRF-CHEM and GEM-MACH, the chemistry subroutines are incorporated into the same model code, whereas for WRF-CMAQ, the chemistry and meteorology codes share memory and pass information at every time step – these differences are not likely to impact the outcome of the simulation. Further description of the models may be found in Campbell *et al*., 2014, Im *et al*, 2014a,b, and Makar *et al*. 2014. Note that all models and/or their post-processing systems were modified to include the output of additional chemical and/or meteorological variables for AQMEII-2. Some of the models included other modifications in addition to their original code. GEM-MACH's operational configuration is 2-bin; this was converted to 12-bin for greater accuracy in the direct and indirect effect calculations, the sea-salt flux treatment was improved, as was its particle settling velocity and algorithms making use of those velocities. GEM-MACH's emissions preprocessing program was modified in order to allow hourly changes in the location and number of large "point" sources (a requirement for the forest fire emissions inputs of the AQMEII-2 emissions (Pouliot *et al*, 2014)).

The emissions used for AQMEII-2 are described in detail in Pouliot et al, 2014, and came from three sources. Inconsistencies in reporting and inventory construction between political jurisdictions meant that the emissions year could not always correspond directly to the year of the simulation. For Europe, the nearest year for which emissions data were available was 2009, with 2010 wildfire emissions provided by the Finnish Meteorological Institute. For the United States, emissions for the year 2008 were projected to

the years 2006 and 2010. In Canada, the most recent inventory available at the time of the study was for 2006; this was used to represent both years, while a 2008 Mexican inventory was used to represent the years 2006 and 2010. The mismatches between simulated year and emissions inventory year may impact the accuracy of the simulations carried out here.

The model simulations occurred on the "native" grid projection for each model, but were compared on common AQMEII latitude-longitude grids with a resolution of 0.25 degrees for the NA or EU domains, respectively. For the NA simulations, the native model grids overlapped this target grid to different degrees, so a common "mask" incorporating the union of all model projections on the common grid was employed for comparison purposes. For the EU simulations, the different versions of WRF-CHEM were operated on the same native grid, but comparisons were done using the AQMEII European grid.

Feedback and non-feedback simulations were compared to each other in three ways. First, at every hour of simulation, the feedback and non-feedback model predictions on the AQMEII grid were compared using the statistical measures described in Table 2. This comparison allowed the identification of seasonal trends in the impact of feedbacks, as well as particular time periods when these impacts were the strongest. Second, the model predictions for the years 2006 (NA) and 2010 (NA and EU) were compared to observations of air pollutants via the ENSEMBLE system (Galmarini *et al*., 2012b). These comparisons used hourly data which were subsequently time-averaged to mean daily values at each station prior to comparison for the given years, and also as hourly or daily values for shorter summer time periods described in more detail below. Third, the model predictions at each gridpoint were compared across time (for the entire simulated year and for shorter time periods), allowing the creation of spatial maps of the impact of feedbacks on the common simulation variables. These maps help identify the regions where feedbacks have the largest effect on the simulation outcome.

1. Comparison of Model Simulations by Time Series

The comparison between no-feedback and feedback simulations for Europe was limited to the direct effect simulations; insufficient computational resources were available for the direct+indirect feedback simulations within the timeframe of the AQMEII-2 project. In that respect, the EU chemical comparison can be compared in a generic sense with the WRF-CMAQ NA simulations, also made use of only the direct effect.

1.1 Ozone

Both WRF-CMAQ and GEM-MACH showed a slight decrease in mean O_3 in the summer associated with feedbacks, on the order of -0.2 to -0.4 ppb (Figure 1, (a),(b)). The change in the grid standard deviation in O_3 is negative for GEM-MACH (i.e. less variability in O_3), while WRF-CMAQ has both and negative changes in standard deviations, with most of the changes being positive (Figure 1, (e),(f)). One of the main effects of the aerosol indirect effect in GEM-MACH is an increase in cloud liquid water path – this additional cloud cover may have resulted in the reduced variability noted here. Low correlation coefficients on May 20th for both models, and for the period August 1st to August 15th, suggest that these times have disproportionately larger feedback impacts. Seasonally, the lowest correlation coefficients occur in the summer – feedbacks having the biggest impact during the summer photochemical production time (Figure 1(c,d)). Non-feedback standard deviations and change in standard deviation (Figure 1 (e,f)) show that the variability of ozone has decreased in the summer in the GEM-MACH simulation –in the WRF-CMAQ simulation standard deviations increase with occasional decreases at a lower magnitude, with direct effect feedbacks - also suggesting that aerosol indirect effects are the main cause of the changes in ozone.

Grid-averaged time series of EU mean O_3 concentration, the difference between direct effect feedback and no-feedback O_3 concentrations, and the correlation coefficient between the two simulations, are

shown in Figure 2(a,b). The simulation shows the Russian fires standing out as a major event in which the feedbacks caused the grid-average O_3 to drop by up to 2.5 ppbv on a grid-average no-feedback concentration of 70 to 85 ppbv (Fig. 2(a), compare red and blue lines). The Russian fires in the no-feedback simulation have increased ozone by about 10 ppbv relative to times before and after the fire period. The feedback-induced reduction in O_3 levels due to fires is largely limited to the period encompassing the fires. The implication is that the aerosol direct effect is capable of reducing O_3 levels, possibly through reductions in downward shortwave radiation reaching the surface due to high particulate concentrations in the atmospheric column, with consequent surface temperature reductions (see Part 1), all of which may reduce ozone formation rates.

1.2 $PM_{2.5}$

Feedbacks increased fine particulate matter for both GEM-MACH and WRF-CMAQ. For GEM-MACH, the increase to the grid-average $PM_{2.5}$ was on the order of +0.5 μg m^{-3}, while for WRF-CMAQ the increase was about an order of magnitude smaller (note change in vertical scale on Figure 3 (a) versus (b)). For GEM-MACH, most of the increase in $PM_{2.5}$ was comprised of particulate sulphate, as was approximately half of the WRF-CMAQ increase. Correlation coefficient plots for both models (Fig. 3 (d),(e)) show a significant difference between feedback and non-feedback models on May 20th and August 25th. Correlation coefficient drops for both primary and secondary organic carbon, hydrogen peroxide, and carbon monoxide occur at the same time. As will be shown below, these events correspond to an event wherein feedback effects alter the model predictions from a very large source of emissions, a forest fire.

Aerosol direct effects modify the typical EU grid-average $PM_{2.5}$ concentration of about 10 μg m^{-3} by +/- 0.5 μg m^{-3} (Figure 3 (c)). Both increases and decreases in the grid-mean concentration relative to the no-feedback simulation occur during the Russian fires period and low correlations between the simulations occur in that region (Figure 3(f)); this form of paired increases and decreases for $PM_{2.5}$ and other emitted species was also noted in NA simulations. The cause appears to be a change in wind direction,

speed,atmospheric stability and/or surface temperatures resulting from the feedbacks – these changes change the height to which the plume of emitted species may rise, the direction and speed of downwind dispersal, and the production rate of secondary particulate matter. Given this sensitivity, the accuracy of forest fire plume forecasting may in part be influenced by the aerosol direct and indirect effects incorporated in the forecasting model.

1.3 NO_2

The lowest correlations between feedback and non-feedback predictions for NO_2 occur in the summer, though these correlation decreases are larger for GEM-MACH (0.69) than for WRF-CMAQ (0.91), Figure 4(d,e). Feedbacks decreased NO_2 in the winter in GEM-MACH, while summer differences in mean NO_2 varied between positive and negative, with a maximum positive change of 0.05 ppbv. Feedbacks in WRF-CMAQ resulted in a positive shift in mean difference of 0.03 ppbv (Figure 4(b)). Feedbacks increased the variability of NO_2 for WRF-CMAQ in the summer, while GEM-MACH's variability varied between positive and negative in the summer, becoming negative (lower standard deviations; lower variability) in the winter (not shown).

For Europe, the aerosol direct effect generally resulted in increases in WRF-CHEM's NO_2 concentrations, particularly in the summer (Figure 4(c)), similar to the NA direct effect simulations with WRF-CMAQ (Fig. 4(b)). These increases in concentration probably stem from the reductions in temperature and surface-level shortwave radiation noted above, with subsequent increases in atmospheric stability. The Russian fires period has the paired +/- mean difference signature found for the NO_2 (Fig. 4(c)), indicating that the dispersion of NO_x emissions has also been affected by the feedbacks. The fires also correspond to the greatest difference in correlation coefficient (Fig. 4(f)). A second, smaller level increase in NO_2 occurs during the month of April.

1.4 Isoprene

Feedbacks resulted in a very different isoprene concentration response in the two models, with GEM-MACH showing a decrease in midsummer isoprene of up to -0.25 ppbv on concentrations ranging between 0.05 and 2.5 ppbv (i.e. >10%) decrease in midsummer grid-average isoprene, and WRF-CMAQ showing both positive and negative changes (between -0.02 and + 0.08 ppbv; about +0.4 and -1.3% of the maximum no feedback concentrations), and no overall seasonal trend (Figure5(a,b)). GEM-MACH showed summertime decreases in both temperature and downward shortwave radiation (see Part 1) associated with increased cloud liquid water paths. These in turn reduce isoprene biogenic emission rates (which are a function of temperature and photosynthetically active radiation). These effects are much less pronounced in WRF-CMAQ, due to the absence of the aerosol indirect effect in this implementation. The changes in GEM-MACH's isoprene drive similar reductions in grid average formaldehyde, the latter being a product of isoprene oxidation. Isoprene correlation coefficients in both models drop significantly between June 15th and June 26th, and from August 12th to 18th, indicating feedback-related events having a large impact during those weeks (Figure5 (d,e)).

The aerosol direct effect is shown to have a substantial impact on isoprene concentrations over the EU domain in Figure 5(c,f)), with grid-average concentration perturbations of -0.10 to +0.25 ppbv during the mid-summer upon no-feedback concentrations of up to 2.0 ppbv (-5 to +12.5%). The perturbations are the largest during the Russian fire period, and are both positive and negative. The bi-modal nature of the isoprene perturbations is of interest, given that the incoming shortwave and surface temperatures discussed earlier are both reduced by the fires, implying an overall reduction in isoprene emissions might be expected. However, the paired changes in NO_2 and $PM_{2.5}$ discussed above suggest that at least part of the changes in isoprene concentration may be ascribed to a change in the direction of the forest fire plumes due to the direct effect feedback. If the plume direction change takes the plume (and its reduction in shortwave radiation and surface temperatures) over an isoprene-emitting region, then the feedbacks will reduce isoprene concentrations. On the other hand, if the feedbacks cause the plume to move its

shadow from an isoprene-emitting region to a region with relatively low biogenic emissions, the feedbacks will increase grid-total isoprene concentrations. These results suggest that feedbacks are capable of perturbing isoprene concentrations, potentially increasing or decreasing them over continent-sized areas by up to 10%. Local changes in concentration will likely be much larger, given the spatial averaging used in these time series.

1.5 Formaldehyde

The mean differences for NA formaldehyde were negative and closely matched to the equivalent isoprene time series for GEM-MACH, while the mean HCHO levels increased during the summer for WRF-CMAQ (Figure6 (a,b)). HCHO correlation coefficient magnitudes for both models minimized in the 3rd week of May, and on April 1st (the latter corresponding to a forest fire event in the GEM-MACH simulation; Figure 6(d,e)).

Changes of EU formaldehyde associated with direct effect feedbacks are shown in Figure 6(c,f). As was the case for NA, the HCHO concentration is closely tied to the isoprene concentration (note similarity in annual time series, Fig. 5(a) versus Fig. 6(a), and Fig. 5(c) versus Fig. 6(c), blue lines). As was found for the NA direct effect feedback simulation, (Fig. 6(b)), EU formaldehyde levels increase with the aerosol direct effect. In the EU case, the negative perturbations of the isoprene concentration (Fig.5(c)) do not result in significant decreases in the predicted HCHO levels, instead, they increase (Fig. 6(e)) by approximately 10%. One possible explanation for this difference might be an increase in HCHO generated from *other* hydrocarbons, when the isoprene levels are reduced. The implications of this latter possibility are intriguing, in that feedback-induced changes in biogenic emissions may thus influence the rate of oxidation of non-biogenic hydrocarbon species, with possible similar shifts in the sources of secondary organic aerosol.

1.6 Nitric Acid and Particulate Nitrate

Feedbacks in GEM-MACH resulted in a shift of nitrate partitioning from the particle to the gas-phase in the winter, directly as a result of the feedback-derived increases in surface temperature (described in detail in the first part of this two part paper). Gaseous nitric acid increased in the winter months (Figure 7(a)), while particulate nitrate decreased (Figure 7(d)). The partitioning equilibrium of nitrate is highly temperature-sensitive, with lower temperatures favouring particulate nitrate formation, and higher temperatures favouring gaseous nitric acid. The increases in temperature in the winter in GEM-MACH have thus resulted in a shift of total nitrate from particulate towards gaseous nitrate. WRF-CMAQ's HNO_3 and particulate nitrate (Figure 16(b,e)) both increase in the summer, reflecting higher NO_x levels in this model when feedbacks are incorporated.

Given the temperature reductions associated with the Russian fires in the direct effect feedback EU simulations (see Part 1), a shift in the particulate nitrate versus HNO_3 equilibrium might be expected. Figure 7(c, f) show that this is indeed the case; the cooler temperatures result in lower HNO_3 concentrations (Fig. 7(c)), and higher particulate nitrate concentrations (Fig. 7(f)) during that period. Mid-January in the EU is another period with low correlations between EU no-feedback and feedback models for HNO_3 (not shown), though this is not echoed for particulate nitrate: presumably the particulate sulphate levels during the winter period are too high to allow particulate nitrate formation, regardless of the changes in HNO_3.

1.7 SO_2, particulate sulphate, NH_3 and particulate ammonium

Feedbacks resulted in decreases in winter mean SO_2 concentrations in GEM-MACH (Figure 8(a)) – this is associated with increased winter particulate sulphate formation (Figure 8(c)); more SO_2 is being oxidized to sulphuric acid and hence particulate sulphate with the incorporation of feedbacks. Precipitation changes showed no strong seasonality for this model (though feedbacks increased overall precipitation levels), and there was no change in wet deposition of sulphate in winter. This suggests that the winter

SO_2 oxidation increase is in the gas-phase or in non-precipitating clouds, and may in part be due to the winter temperature increase described in Part 1. This presents an alternative reason for the changes in nitrate partitioning noted above: increased sulphuric acid content in the aerosols would result in more of the available ammonia partitioning with sulphate, and nitric acid off-gassing. GEM-MACH NH_3 decreases with the feedbacks (Fig. 8 (e)) throughout the year, while $PM_{2.5}$ NH_4 increases (Fig. 8 (g)) despite the decreases in particle NO_3 noted earlier. The increases in SO_2 oxidation to sulphate are thus at least partially responsible for the shift from particulate nitrate to nitric acid noted above. Increases in summer particulate sulphate levels in GEM-MACH appear to be due to increased wet processing; the increases in cloud liquid water noted above result in more particulate sulphate formation and summer wet deposition of sulphate (Figure 8(c)). In WRF-CMAQ, summer increases in particulate sulphate (Figure 8(d)) were much lower than those from GEM-MACH (WRF-CMAQ values ranged from -0.01 to 0.03 μg m^{-3}, while GEM-MACH changes ranged from 0.0 to 1.6 μg m^{-3}). The magnitude of the differences suggests that the indirect effect processes may dominate summer formation of sulphate via feedbacks, though confirmation of this would require further model runs isolating direct and indirect effects in each model. WRF-CMAQ's NH_3 largely increased in the summer, as did its particulate ammonium (Figure 8(h)).

The perturbations caused by the aerosol direct effect on SO_2, particulate sulphate, NH_3 and particulate ammonium for the EU are shown in Figure 9. The incorporation of the direct effect has increased the SO_2 levels across the grid (which alternate between increases and decreases during the fires, Fig. 9(a), blue line versus red line). The feedbacks during the fires result in a reduction in SO_2 oxidation rates, as can be seen by the corresponding decreases in particulate sulphate concentrations at that time (Fig. 9(c)). Despite the particulate sulphate decreases, the ammonia levels decrease then increase during course of the fires (Fig. 9(e)), and particulate ammonium changes follow the ammonia changes (Fig. 9(g)). Presumably the sequence of events causing these changes starts with the feedbacks initially increasing SO_2 dispersion, reducing subsequent particulate sulphate formation, potentially freeing available ammonium for particle

nitrate formation. Towards the end of the fire period SO_2 concentrations have increased relative to the no-feedback simulation, with less particle sulphate formation, and an increase in NH_3 – despite which, particle ammonium increases. The latter may be the result of direct-effect feedback induced reductions in temperature favouring particulate nitrate formation, in addition to the reduction in particle sulphate leading to these increases in particle ammonium towards the end of the fire period. As was the case for winter in North America, feedback effects have been shown to have enough of an impact on temperatures and sulphate formation to change the particulate nitrate/nitric acid equilibria, over a large part of the continent.

2. Comparison with Observational Data from Networks

Monitoring network data were collected from a variety of sources for comparison to model simulations. North American data for 2010 were obtained from the Canadian National Atmospheric Chemistry (NAtChem) Database and Analysis Facility operated by Environment Canada (http://www.ec.gc.ca/natchem/). The NAtChem Facility obtains air quality and selected meteorological surface data from North American networks, applies quality assurance to these data, adds metadata and reformats the data from each network into a common comma-separated-variable format. The networks and data archives used for this purpose included the Canadian National Air Pollution Surveillance Network (http://maps-cartes.ec.gc.ca/rnspa-naps/data.aspx), the Canadian Air and Precipitation Monitoring Network (http://www.ec.gc.ca/natchem/), the U.S. Clean Air Status and Trends Network (http://java.epa.gov/castnet/clearsession.do), the U.S. Interagency Monitoring of Protected Visual Environments Network (http://views.cira.colostate.edu/web/DataWizard/), and the U.S. Environmental Protection Agency's Air Quality System database for U.S. air quality data (http://www.epa.gov/ttn/airs/airsaqs/detaildata/downloadaqsdata.htm). The result was a single format data set comprising Canadian and US data, making the data much more accessible for model-observation comparisons. In Europe, the monitoring network data from 2010 were obtained from European Monitoring and Evaluation Programme, http://www.emep.int/) and AirBase (European AQ database;

http://acm.eionet.europa.eu/databases/airbase/). Both 2010 NA and EU datasets were uploaded to the ENSEMBLE database and model-observations comparison system maintained by the European Commission's Joint Research Centre (JRC) in Ispra, Italy ((Galmarini *et al.*, 2004a,b), Galmarini *et al.*, 2012)). Similar comparison data for North America was obtained for the year 2006 during AQMEII Phase 1 (Galmarini *et al*, 2012b). The ENSEMBLE system greatly reduces the time required by modellers to generate comparisons to observations. Model output sent to the central collection site at JRC in the required format may be compared to the uploaded observation databases via a web browser, allowing all modelling groups participating in a study to use the same data, intercompare with each other's results, conduct independent data analyses, and conduct retrospective data-model comparisons, such as the current work. Here, ENSEMBLE was used to generate traditional scatterplots and the corresponding statistics, the latter tabulated and included in the supplemental information appendix as well as the main body of the text. The statistical quantities comparing model values to observations (as well as cross-comparing models) are given in Table 3.

The comparison to observations took place in two stages. The first stage examined model performance on an annual basis. ENSEMBLE was used to create statistical tables of the mean day averages of the measured quantities at observation stations for both model and observations, and these were compared for each simulated year (NA2006, NA2010, and EU2010), with the resulting performance table appearing in the Supplemental Information (SI) for this paper, Tables S1, S2, and S3. The second stage examined the statistics during time intervals over which the above time series analysis suggested significant impacts due to feedbacks might occur.

2.1 Annual Analysis, North America, 2006

Six models were compared to the same observation data for O_3, SO_2, NO_2, CO, $PM_{2.5}$, $PM_{2.5}$ SO_4, $PM_{2.5}$ NH_4, $PM_{2.5}$ NO_3, $PM_{2.5}$ TOM, PM SO_4, PM NO_3, PM_{10},. Two models were taken from the previous AQMEII-1 comparison (CMAQ and AURAMS), the remainder from the current set of simulations

(GEM-MACH without and with direct + indirect effect feedbacks, WRF-CMAQ (aerosol direct effect feedbacks only) and WRF-CHEM (direct and indirect effect feedbacks). The two previous intercomparison simulations were included here for reference – the intent being to determine whether the on-line, coupled models performance is better than the previous generation uncoupled models. It should be mentioned however, that both chemical boundary conditions and the emissions for the year 2006 differed from the datasets used in AQMEII-2. The differences stemming from updates to emission estimation methodologies (Pouliot *et al*., 2014) as well as boundary conditions may thus account for part of the model performance changes between AQMEII-1 and AQMEII-2. In addition, Hogrefe *et al*. (2014, this issue) present WRF-CMAQ sensitivity simulations that show that differences in monthly average ozone concentrations stemming from the different boundary conditions are 7 ppb or greater over large portions of the modeling domain in January 2006 while in July 2006 they are 3 ppb or less for most of the modeling domain though differences as large as10 ppb are simulated over the Northwestern U.S.

The statistical metrics for the NA2006 comparison are tabulated in Table S1 (Supplementary Information Appendix). The model with the highest score for each variable and each statistical metric has been identified with an italic font in the table. GEM-MACH was the only model submitting both no-feedback and feedback simulations; for these two simulations only, the model with the higher statistical score has been identified using a bold font.

From Table S1, no model is clearly superior to the other models for a given statistic, or for all statistics within one variable. There is a large amount of variation in performance between the models for the different pollutants and statistical measures, and this underlines the utility of ensembles as explored earlier (Solazzo *et al*., 2012a,b) and elsewhere in this special issue (Im *et al*, 2014(a), (b)).However, if all chemical statistical measures, for all variables, are assumed to have equal “weight”, then the Phase 1 models outperform the Phase 2 models (bearing in mind that WRF-CMAQ did not report values of NO_2 in time for writing): CMAQ 43 best values, AURAMS 24, GEM-MACH (no-feedback): 11, GEM-MACH(feedback):17, WRF-CMAQ:21, WRF-CHEM: 10. In some ways this is a sobering finding, in

that it implies that further development work is needed for the first generation fully coupled models or their emissions data. The incorporation of feedbacks did improve the overall score for the GEM-MACH model relative to its no-feedback climatological state, increasing the number of best scores by 54%. The emissions inventories between phases 1 and 2 of AQMEII were modified with more recent information, resulting in significant changes in some emissions (see Pouliot *et al*, (2014) e.g. emissions of NOx, where the phase 1 models performed better). The methodology used to generate the new emissions data may need to be reexamined, given these findings, though other model differences (such as the boundary condition updates) may also be influencing the results.

Second, the 2006 annual results of the model with both no-feedback and feedback simulations (GEM-MACH) were not always improved by the employment of feedbacks. Improvements occurred for SO_2, $PM_{2.5}$ NH_4, PM NO_3, and PM_{10}, but the no-feedback model had better overall performance (by number of higher scoring statistics) for O_3, NO_2, CO, $PM_{2.5}$, $PM_{2.5}$ SO_4, $PM_{2.5}$ NO_3, and PM SO_4. Comparing just the two GEM-MACH simulations, the total number of higher scores for the feedback model was 42, with 75 for the no-feedback GEM-MACH.

2.2 Annual Analysis, North America, 2010

The statistical metrics for the NA 2010 comparison are tabulated in Table S2 (Supplementary Information Appendix). The models compared are limited here to those participating in the current work (AQMEII-1 did not simulate the year 2010 for North America). The distribution of best scores for 2010 was similar to 2006 (aside from the absence of the phase 1 models), with GEM-MACH(no-feedback): 23, GEM-MACH(feedback): 37, WRF-CMAQ: 37, WRF-CHEM: 16. The incorporation of feedbacks improved the GEM-MACH scores by 61%, similar to the 2006 improvement. In both years, the incorporation of feedbacks in the GEM-MACH model resulted in improved SO_2 scores, while worsening the scores for NO_2 and NO. When the two GEM-MACH simulations were compared only to each other (bold-face font numbers in Table S2), the feedback model improved with 65 best scores compared to 45 with the no-

feedback model (this is in contrast to the 2006 results). It should be noted that a significant difference between the two years may be found in the boundary conditions used for the models (MACC reanalysis). Hogrefe *et al* (2014) suggests that positive winter O_3 biases in 2006 and negative O_3 biases in 2010 may in part be due to the boundary conditions used by all models in the comparison.

The NA comparisons to observations, for the variables compared here, imply that indirect + direct effect feedbacks are capable of improving a model's results relative to peer models, given that the total number of best scores for GEM-MACH improved in both years with the inclusion of feedbacks. A caveat on this finding is that the model to model variation remains high. The relative improvement between the specific model for which feedback and no-feedback simulations exist varies between the simulated year, with feedbacks improving performance in 2010, but worsening it in 2006. One possible interpretation of this latter finding is that the climatological parameterizations used in the GEM-MACH "no-feedback" simulations for the aerosol direct and indirect effects are closer to the actual averages in 2006, while the model-generated feedback values are closer to the actual averages in 2010. Differences between the boundary conditions created by global model reanalyses between the years may also cause some of the differences, particularly in winter (Hogrefe *et al*, 2014).

2.3 *Annual Analysis, EU, 2010*

The statistical metrics for the EU 2010 comparison are tabulated in Table S3 (Supplementary Information Appendix). Once again, the best scoring model of those used in this work is identified in the summary scores by italics. The SI1 and SI2 models differ only in the incorporation of direct effect feedbacks; the better scores for these two models alone are identified by bold face text. The differences between these simulations is relatively small; this is echoed in the summer-only comparisons for WRF-CMAQ; models incorporating the aerosol direct effect have smaller feedback impacts than those incorporating the aerosol indirect effect. As noted above, the WRF-CHEM direct+indirect effect feedback was for a slightly

different version of the WRF-CHEM model, so the differences shown here are not necessarily due to the indirect effect feedback alone.

The models have very different performance for gases versus particulate matter, with the model incorporating direct + indirect feedback having better performance for urban O_3, SO_2, NO (both all stations and urban stations only), as well as NO_2, while having relatively poor performance for most PM variables, with large negative biases and the lowest scores for PM_{10} (all stations and regional stations), $PM_{2.5}$, and speciated PM and $PM_{2.5}$. Overall, the no-feedback WRF-CHEM had 57 top or tied for top scores, the direct effect model had 60 (a slight improvement with the direct effect) and the direct+indirect effect model had 47 top scores.

Comparing the no-feedback and direct effect only versions of WRF-CHEM to each other, the direct effect by itself has resulted in a decrease in model performance, with the no-feedback version of the model leading with 86 higher or equal scores, and the direct effect model leading or equal with only 67 scores.

Based on the above comparison, the following conclusions may be drawn, specifically for annual performance:

(1) The incorporation of direct + indirect feedbacks in the GEM-MACH model in general improved its chemical performance relative to the suite of models compared, for both years simulated.
(2) The incorporation of direct + indirect effect feedbacks in the GEM-MACH model relative to its own no-feedback simulation, worsened its performance in 2006, but improved its performance in 2010.
(3) Comparisons between the AQMEII Phase 1 uncoupled and AQMEII Phase 2 coupled models suggests that the former had better performance, with the confounding factor that both emissions and the global model reanalysis boundary conditions changed between the two sets of simulations.

(4) In the EU domain, the incorporation of feedbacks had a less discernable benefit, with a slight increase in the number of best scores going from no feedback to direct effect feedback, and a substantial decrease in the number of best scores going to direct+indirect feedback. The incorporation of direct + indirect effects resulted in a substantial improvement in gas-phase statistics, while significantly degrading the aerosol performance of the model. The latter performance degradation may be due to other model differences aside from feedbacks.

2.4 Summer 2010 Analysis, North America

The time series comparison of feedback and no-feedback simulations for North America consistently showed the summer period as having the largest impacts for both direct and indirect feedback models, hence suitable for a focused comparison to observations. The ENSEMBLE database was used to generate summary statistics during the period July 15th through August 15th, 2010 (Table 4). Here, hourly observations were paired with model values where possible; $PM_{2.5}$ and speciated $PM_{2.5}$ values are daily averages. The "validity cutoff" mentioned in Tables 4, 5, SI1, SI2, SI3 refers to the percentage of observations available at a given monitoring site relative to the highest number of observations possible. A 75% validity criterion for hourly data thus means that only those stations with 6570 or more hourly observations during the year were used for the comparison. Some of the PM monitoring networks report daily average values at only a 1 day in 6 frequency, hence a 16.6% validity cutoff was used for daily PM observations.

Examining the GEM-MACH performance in Table 4, the performance was improved with the implementation of feedbacks for most of the gases and PM_{10}; regional and urban/suburban O_3 (8 and 7 out of 9 statistics improved), SO_2 (7 out of 9 statistics), NO (5 out of 9), NO_2 (7 out of 9), all PM_{10} stations (7 out of 9), and regional PM_{10} stations (7 out of 9). Carbon monoxide performance is degraded (5 out of 9 stations had better performance with the no-feedback model). For $PM_{2.5}$, the addition of feedbacks had a negative effect on model performance, with total $PM_{2.5}$ performance scores being better

with the no-feedback model for all measures (9 out of 9), as was the case for $PM_{2.5}$ SO_4 and NH_4. $PM_{2.5}$ NO_3 and total organic carbon had a smaller decrease in performance with the incorporation of feedbacks (only 3 and 2 out of 9 measures improved with feedbacks, respectively) For the GEM-MACH model, the inclusion of feedbacks has improved the gas-phase chemistry and PM_{10} performance, but reduced the performance for $PM_{2.5}$.

Another important finding from Table 4 is that the magnitude of the change in model performance associated with interactive feedbacks relative to climatological aerosol properties without feedbacks is often smaller than the changes in performance going from one model to another. That is, the change the magnitude of the performance statistics between the two GEM-MACH runs is often less than the differences between GEM-MACH, WRF-CMAQ and WRF-CHEM "(for example, the mean biases for urban/suburban O_3 for the GEM-MACH no-feedback, GEM-MACH feedback, WRF-CMAQ and WRF-CHEM simulations are 2.86, 2.47, 2.61 and -4.32 ppbv, respectively). The difference between a climatological approach to aerosol direct and indirect effects and that of "fully coupled" direct + indirect effect feedbacks, has less of an impact on model performance than the model architecture employed.

The findings suggest that targeted studies examining specific species where the performance between different models is examined in detail would be of great benefit to the community. For example, the advection, dispersion, gas and aqueous phase oxidation of SO_2 likely differs between the three modelling frameworks examined here, and a process study of the production and losses of SO_2 would help explain the observed performance differences. Similarly the differences in PM performance between the models should be examined using process analysis. Future ensemble studies such as AQMEII phases 1 and 2 should include process analysis as a focus, in order to improve understanding of these differences, and improve overall model performance.

2.5 Russian Fires Analysis: EU domain, July 25th to August 19th, 2010

Statistics for the EU domain were regenerated for the period corresponding to the large deviation in grid average values between feedback and no-feedback simulations noted in the above analysis on the EU results, from July 25th through August 19th. The results of this analysis are shown in Table 5. It must be remembered at the outset that the direct + indirect effect simulation here was carried out with a slightly less recent version of WRF-CHEM, hence some differences noted may be due to other model parameterizations aside from the institution of indirect effect feedbacks.

During this period, the best overall performance for the gas-phase species was usually with the direct+indirect effect simulation. Regional O_3 was the exception, with 8 best or tied scores being attributable to the no-feedback model, compared to 3 for the direct effect model and one for the combined direct + indirect model. However, for urban O_3, the number of best scores (no-feedback, direct effect, direct + indirect effect) was in favour of direct+indirect effect model (2,1,7), as was the case for SO_2 (1, 1, 9), for all NO stations (2,1,7), urban NO stations (1,2,7), regional NO stations (0,0,9), urban NO_2 stations (0,1,8). For CO, the no-feedback model had the highest number of best scores (6,5,1).

For particulate matter variables, the direct + indirect effect model often, but not always, had the least number of best scores across the metrics considered, while the relative impact of the direct effect varied according to the particulate species or size range considered. For all PM_{10} stations, the direct effect simulation had the highest number of best scores, (no-feedback, direct effect, direct + indirect effect) was (6,4,1), while for regional PM_{10} stations scored (5,4,1), $PM_{2.5}$ (2,4,3), PM SO_4 (4,6,3), PM NH_4 (0,6,3), and PM NO_3 (2,5,2).

While the model architecture used in the EU simulations differs from the GEM-MACH model, it is worth noting here that the pattern of changes associated with going from a no-feedback model to the direct + indirect feedback model was similar for both EU and NA summer comparisons: improvements took place in most gas-phase species, the performance was equivocal for CO, and the performance decreased

for PM. The gas-phase improvements also tended to manifest themselves more for statistics other than correlation coefficient in both cases, with slight decreases for PCC while the other statistical metrics improved (NO_2 being one exception).

Combined indirect + direct effect feedbacks tend to improve gas-phase simulation accuracy while decreasing PM simulation accuracy, at this stage in the fully coupled models' development. It is worth noting here that both of the indirect effect models showing this effect (GEM-MACH and the EU/IT2 WRF-CHEM simulation) make use of the cloud condensation nucleation parameterization of Abdul-Razzak and Ghan (2002). Moreover, detailed analysis by Gong *et al* (2014) using ICARTT 2004 in-cloud observations suggests that this parameterization is highly sensitive to the choices made in describing the standard deviation of cloud updraft velocity. It seems likely, then, that the degraded performance in the mean PM performance statistics with both models is linked to the models' rate of uptake of aerosols into clouds, aqueous processing, and rainout/washout of the aerosols . In GEM-MACH, the chemical processing may be dominating, hence creating positive biases in aerosol sulphate. In WRF-CHEM/IT2, the particle removal processes may be dominating, leading to excessive particle removal and negative biases. A comparison of the process parameterizations for the "in-and-below-cloud" processes between these two models may thus be a fruitful avenue for future research. In both cases, changes in intensity and location of precipitation events are also linked to improvements in ozone formation statistics, implying an aerosol indirect effect feedback impact on ozone formation, in both cases leading to a reduction in positive O_3 biases seen in the no-feedback model (See Makar *et al* (2014), Part 1, this special issue).

3. Spatial Analysis of Feedbacks, Annual and Events

In this section, we return to the no-feedback versus feedback comparison, this time analyzing the model results averaged over time at each model gridpoint, rather than averaged over space. The resulting model-to-model comparison statistics are described in Table 2, where N is now the number of hours, rather than

the number gridpoints. Due to space limitations, not all statistical comparisons created will be shown here, with mean differences and correlation coefficients being the primary means of displaying the regions with the greatest impact of feedbacks. This portion of the analysis pairs NA and EU contour maps of feedback influences. The maps were generated for the period July 15th through August 15th , 2010 for the NA domain, and July 25th through August 19^{th,} 2010 for the EU domain, in order to allow all three models to be compared for NA, and to focus on the Russian fires period for EU.

3.1 O_3

The GEM-MACH and WRF-CMAQ NA domain mean concentration differences are shown in Figure 10 (a,b), the correlation coefficients in Figure 10 (c,d) and the change in standard deviation (feedback – basecase) in Figure 10 (WRF –CHEM comparisons are only available for meteorological variables over North America, see Part 1). The equivalent EU fields for the WRF-CHEM direct effect simulation during the Russian fires period (July 25th through August 19th) is shown in Figure 11. Both the GEM-MACH simulation with both direct and indirect feedbacks and the WRF-CMAQ direct effect simulation have resulted in the largest regional changes in mean O_3 in eastern NA. In the direct+indirect effect feedback GEM-MACH simulation (Fig. 10(a), ozone has decreased over much of this region, with the largest decreases over the Great Lakes, upstate New York, and many of the urban regions along the Mississippi valley and the SE USA, with the largest decreases in mean O_3 during the period of -2.93 ppbv. The direct effect feedback WRF-CMAQ (Fig 10(b)) has a smaller range of O_3 changes (WRF-CMAQ: +0.6 to -0.5 ppbv, GEM-MACh+1.9 to -2.93 ppbv - note the scales change between panels in the figure). The direct effect feedback changes are less organized into a regional pattern; both positive and negative regions are side-by-side in the direct effect (Fig. 10(b)) as opposed to the indirect+direct effect (Fig. 10(a)). The changes noted with the direct effect simulation represents shifts in local wind direction or cloud amounts. In contrast, the indirect+direct effect feedback simulation results in an overall decrease in O_3 over most of the eastern half of the continent. The GEM-MACH simulation mean difference (Fig. 10(a)) also shows increases in O_3 in the cities of San Francisco, Los Angeles, while the

WRF-CMAQ simulation shows decreases or no change there. Both simulations show O_3 decreases in cities in the SE USA (e.g. Atlanta, New Orleans). Increases in O_3 in northern Canada may reflect decreases in isoprene concentrations noted in the above time series analysis: northern Canada includes large boreal forest regions, with few large regional sources of NO_x emissions – a reduction in biogenic emissions due to decreased temperatures and increased cloud cover would result in less O_3 destruction by alkene + O_3 reactions in that area.

The lower value correlation coefficients (Fig. 10(c,d)) highlight the regions where the feedbacks are having the greatest impact in O_3 concentrations. Both models show the Los Angeles area as being significantly affected by feedbacks (and in the GEM-MACH simulation comparison, this region extends up the entire California coast). Other areas signficiantly impacted by feedbacks in the GEM-MACH simulation include central Washington state (possibly due to a forest fire during the period), Phoenix, Denver, Chicago, central Lake Superior, Georgia just north of Atlanta, and Jacksonville and Orlando in Florida. The correlation coefficients from the WRF-CMAQ direct effect simulation have less of a tendency to relate to the position of large cities aside from Los Angeles and Jacksonville; minima occur in the state of northern Montana, and the south of the provinces of Alberta and Saskatchewan, possibly related to oil and gas extraction activities in those regions, and northern Lake Michigan. The magnitude of the changes in correlation coefficient differ – the GEM-MACH values dropping to 0.565, while the direct-effect-only WRF-CMAQ values reach 0.90.

The standard deviations (Fig. 10 (e,f)) show regional increases in standard deviation of hourly O_3 (orange areas) over much of North America for the direct + indirect feedback GEM-MACH simulation (Fig 10(e)), with smaller regions in which the variability has either increased or decreased relative to the no-feedback simulation. The Washington State fire event shows a paired increase/decrease in variability, indicating a change in direction of a large plume resulting from the feedbacks. Lake Michigan's O_3 variability decreases, while the region to the north-east of Atlanta noted above has increased variability. The direct effect WRF-CMAQ simulation has smaller magnitude variability changes – with decreases in

Los Angeles, again, a change opposite to that of the direct+indirect effect simulation with GEM-MACH, a paired set of increases and decreases near Minneapolis, south-eastern Indiana, Columbus Ohio, and to the north-west of Montreal. These paired changes in variability seem to reflect changes in the locations of plumes in the direct effect-only simulation.

The differences in magnitude of the impacts between the available simulations suggests that the indirect effect may have a larger impact on O_3 concentrations than the direct effect. Confirmation of this finding will require further 'direct-only' and 'indirect-only' simulations within individual models incorporating both direct and indirect effects (see Conclusions and Recommendations). The changes in O_3 mean value and variability are also often in different directions between the two model runs for large cities and plumes. This suggests that the direct and indirect effects may sometimes act in *competition*, with the direct effect increasing O_3, the indirect effect decreasing O_3, and vice-versa. The direct effect will increase the amount of scattering of light, potentially increasing photo-oxidation rates hence increasing surface O_3 concentrations, while the indirect effect may increase the amount of clouds, hence leading to decreases in photo-oxidation rates, in turn decreasing O_3 concentrations.

The substantial direct effect impact of the Russian fire event on O_3 mean values, correlation coefficents and changes in standard deviation is shown in Figure 11 (a-c), with the largest feature in the model grid corresponding to the fires and their downwind plumes. Mean differences are both positive and negative, with decreases in O_3 dominating (Fig. 11 (a), note that most of the colour scale encompasses negative numbers, with the greatest decrease in the time-averaged O_3 in excess of 7 ppbv). Correlation coefficients (Fig. 11(b)) show local decreases far larger than elsewhere on the grid (most of the grid having values higher than 0.975 while the Russian fires have values as low as 0.85). The direct effect feedback decreases O_3 variability (Fig. 11(c)) in the region of the fires – the emissions were of sufficiently long term and the chemical effects relatively uniform over time to decrease the variability by 10 ppbv. The magnitude of these changes can be compared to the direct effect simulations in the previous figure with the WRF-CMAQ model – the changes in the European grid are far larger than either of the NA

simulations (direct or direct + indirect effect feedbacks), indicating the very substantial impact of the Russian fires via the direct effect feedback, and the likely dominating influence of large fires of this nature on chemistry downwind. Similar findings were noted by Wong *et al* (2012), for fires in California.

The distribution of the changes also explains the reason why the impact of the fires relative to observations in the analysis above is not larger – all of the observation stations used in the comparison were in the EU, none within Russia and downwind, hence the more dramatic effects did not appear in the measurement record for most of Europe. From Figure 11, Northern Finland would have experienced some of the fire impact –observations from Finland or Russia are needed to evaluate the feedback effects against measurements.

Feedback-induced changes in chemical regime are examined for the NA GEM-MACH and EU WRF-CHEM simulations in Figure 12. The branching ratio describes the relative importance of the NO versus HO_2 and RO_2 pathways for organic radical reactions, numbers closer to unity being representative of more VOC-limited regimes:

$$Branching\ Ratio = \frac{k_{RO2+NO}(NO)}{(k_{RO2+NO}(NO)+k_{RO2+HO2}(HO_2)+k_{RO2+RO2}(RO_2))} \quad (1)$$

Negative changes in the mean branching ratio thus represent shifts towards a more NOx-limited regime, and positive changes represent a shift towards a more VOC-limited regime. A second measure of atmospheric chemistry changes with regards to ozone formation pathways is the net VOC reactivity, defined here as the sum of non-methane VOC concentrations multiplied by their OH rate constants. Positive changes in the VOC reactivity indicate higher concentrations of VOCs, negative changes indicate lower concentrations.

Figure 12 (a) and (b) contrast the changes in the branching ratio for NA with changes in NO_2 concentrations for the summer time period of interest. The mean difference in the branching ratio (Fig. 12(a)) has become substantially more negative for the cities of San Francisco and Los Angeles (shifted

towards more NOx-limited conditions), and more positive (shifting towards VOC-limited conditions) for the cities and industrial regions of the province of Alberta, the cities of New Mexico, Arizona and Colorado, as well as Vancouver/Seattle, Detroit, Toronto, Montreal, and Birmingham. NO_2 changes in the same locations follow the reverse pattern. This allows interpretation of the O_3 changes noted above: in San Francisco and Los Angeles, already VOC-limited areas, the feedbacks lead to reductions in NO_x, shifting these cities towards a more NOx-limited environment. However, the very VOC-limited starting point of these changes means that the net result is a decrease in NO_x titration of O_3, hence increasing local O_3 levels. The other cities show a shift towards *more* VOC-limited regimes, suggested reduced O_3 concentrations there may be the result of increased NOx titration. This is borne out in Figure 12(b), showing the changes in NO_2. In Europe, the central region of the Russian fires has become more NO_x-limited immediately under the plume and more VOC-limited on the periphery (Fig. 12(c)) –the direct effect feedbacks have resulted in higher levels of VOCs (Fig. 12(d)) due to less surface reactions, possibly due in part to shadowing effects of the smoke plumes, and lower concentrations of NO_x (Fig. 12(e)) near the surface close to the fires, possibly as a result of increased strength of plume convection and vertical transport under the direct effect scenario. This in turn results in more NO_x dispersion downwind, shifting the outlying regions in the direction of VOC limitation.

3.2 $PM_{2.5}$

Figures 13 and 14 compare the feedback-induced changes in $PM_{2.5}$ mean differences, the correlation coefficients and the changes in standard deviation for the two summer periods on the NA and EU domains, respectively.

For GEM-MACH (Fig 13(a)), the increases in $PM_{2.5}$ are the largest along the California coast, at the fire location in Washington State, and over the Great Lakes, though an overall increase in “background” $PM_{2.5}$ can be seen across the domain. For the WRF-CMAQ direct effect feedback simulation (Fig. 13(b), increases in $PM_{2.5}$ can be seen at an intense hot-spot change at Portland, Oregon and to a lesser degree

over a broad region in the north-eastern part of the study areas (the same region as the ozone changes described above). Both models again show the California coast and coastal cities as being strongly affected by the feedbacks (Fig. 13(c,d)); GEM-MACH increasing $PM_{2.5}$ there, and WRF-CMAQ decreasing it. GEM-MACH shows much broader regions of low correlation values than WRF-CMAQ; the addition of indirect effect feedbacks has resulted in changes in $PM_{2.5}$ over a much larger portion of NA.. The changes in the standard deviation between the simulations (Fig. 13 (e,f)) are dominated on a linear scale such as used here by the "hot-spots" in Washington State (GEM-MACH) and Oregon (WRF-CMAQ).

The EU WRF-CHEM direct effect simulations again show the dominating influence of the Russian fires. With the addition of the direct effect feedbacks, the $PM_{2.5}$ concentrations generally increase in the vicinity of the fire centres (Fig. 14(a)) – a similar pattern seen for NO_2 (Fig. 12(e)), and consistent with a greater vertical rather than horizontal dispersion at the surface, with subsequent downmixing further downwind. The largest impact on correlation coefficients (Fig 14(b) again corresponds to the fire locations. The variability in $PM_{2.5}$ shows paired increases and decreases at the fire hot-spots, indicating a local change in plume location and strength; a shift in the location of a highly time-varying source, as opposed to an increase in inherent variability.

These findings highlight a common theme amongst the models – the simulation of the height and dispersion pattern of very large emissions sources is clearly highly sensitive to the local meteorological conditions. An examination of other time periods with the models shows that these changes in plume height and direction, particularly from forest fires, following the incorporation of feedbacks, commonly occur in the models. Given this high degree of sensitivity, the accurate simulation of large plume dispersion may require fully coupled models such as those studied here. At the same time, the work shows that the plume rise algorithms used in the models are also very sensitive to changes in meteorological conditions, a sensitivity that is increased when the emissions are allowed to modify the

meteorology via feedbacks. We therefore recommend the use of feedback models for the testing and improvement of forest fire and large urban plume rise and dispersion simulations.

3.3 Isoprene

Mean differences in isoprene concentrations are shown in Figure 15, for NA/GEM-MACH, NA/WRF-CMAQ, and EU/WRF-CHEM. The GEM-MACH decreases in isoprene (Fig. 15(a)) align well with the location of the main emitting regions, the Canadian boreal forest, and south-eastern USA. This suggests that the changes in isoprene concentrations noted earlier correspond to continental-scale changes in the emitting conditions (photosynthetically active radiation and temperature). In contrast, the WRF-CMAQ direct effect isoprene changes and those in the EU WRF-CHEM simulation (Fig. 15 (b,c)) are much more localized. For WRF-CMAQ, the changes are both positive and negative, likely indicating a shift of local clouds. For WRF-CHEM, the feedbacks have resulted in areas of isoprene decreases and increases in the vicinity of the Russian fires, again suggesting that changes in the location of the plumes are having a large impact on local chemistry, in this case via changes to the emissions of isoprene, hence to the relative importance of biogenic versus anthropogenic hydrocarbons in the atmosphere.

Conclusions and Recommendations

In our Introduction, we posed three questions for investigating the impacts of feedbacks between weather and chemistry. The work we have conducted here suggests that the direct and indirect effects may have significant impacts on air-quality predictions, and allows us to provide initial answers to these questions, as follows:

(1) The incorporation of feedbacks resulted in systematic changes in the predictions of chemistry. The largest impact on the model results, as inferred by hourly-calculated spatial correlation coefficients between feedback and no-feedback models, occurred during the summer season, when the most active photochemistry takes place, and when forest fire emissions are the highest.

(2) The feedback-induced changes vary spatially – the largest changes associated with feedbacks corresponded to the regions with the highest emissions, significantly changing the local to regional concentrations of O_3, $PM_{2.5}$ and other pollutants. For example, feedback effects associated with large forest fires in Russia in the summer of 2010 resulted in larger impacts on chemical predictions than the feedback effects associated with anthropogenic emissions in Europe during the same time period. Similarly, the impact of feedbacks in North America was usually greatest in the industrialized east of the continent, the region of highest overall emissions and downwind chemical processing. Feedback effects were also shown to have the largest impacts near cities, with defined shifts in ozone production regimes towards more/less NOx or VOC-sensitive regimes for individual cities.

(3) At the current state of fully coupled model development, the incorporation of feedback effects did not always result in improvements in model performance, depending on the year and time period of comparison to observations. The differences in annual performance between the different models' predictions with respect to observations were usually larger than the changes resulting from implementing feedbacks within a given model. This suggests that the implementation details of other processes, such as chemical mechanisms, particle microphysics, etc., have a larger effect on model performance than feedbacks, when annual simulations are considered. During the summer season, the incorporation of feedbacks was shown to significantly improve predictions of atmospheric gas concentrations in both North America and Europe. Predictions of summer particulate matter became slightly worse in North America with the incorporation of both direct and indirect feedbacks. In Europe, summer simulations including the both direct and indirect effects improved the gas prediction performance relative to observations, while the best $PM_{2.5}$ performance was for the direct-effect only model. The aerosol indirect effect feedback was shown to be the dominant process in modifying atmospheric chemistry compared to the direct effect feedback, consistent with Wang *et al*. (2014b, this issue). The direct and indirect effect feedbacks were also shown to often be in competition with regards to the resulting

chemistry of the atmosphere, with opposing changes in O_3 and $PM_{2.5}$ occurring in direct-effect-only versus direct+indirect effect simulations.

The above work also suggests several directions for further research to improve our understanding of feedback processes, given the potential improvements seen here from this first intercomparison of fully coupled feedback models. Some of these recommendations are also made in order to address uncertainties resulting from the limitations of the above work, as described below:

(1) *Shorter timer period "event" modelling studies.* Future studies making use of a broader array of models, but simulating a shorter specified time period (such as the Russian Forest Fire period during the summer of 2010), with a focus on mass tracking and comparison of indirect and direct effect parameterizations, would be of great value to the community. The shorter time period would allow for the participation of more modelling groups, and simulations of no-feedback, direct-effect, indirect effect and direct+indirect effect conditions for each of the participating models. We note here that a considerable source of uncertainty in our results stems from the limited number of simulations available for each model, this in turn stemming from the computational resources needed for annual simulations, required under the AQMEII-2 protocol.

(2) *Indirect effect algorithm and process studies.* Further work is clearly needed to improve the representation of aerosol indirect effect in feedback models. For example, while all of the indirect-effect models employed here made use of the Abdul-Razzak and Ghan (2002) formulation as the basis for parameterizations for the formation of cloud condensation nuclei from aerosols, the response of the models relative to observations when the indirect effect is incorporated varied widely. The GEM-MACH model in North America had an increased positive bias with the incorporation of feedbacks, while the WRF-CHEM $PM_{2.5}$ simulations in Europe had a large negative bias with the indirect effect implementation. Incorporating the indirect effect has the potential to improve the distribution and radiative effects of clouds (improving the radiative budget and hence ozone formation accuracy as noted above). However, the models' in-cloud

aerosol formation and removal processes may create or remove too much aerosol mass. The inter-comparison of the different in-and-below-cloud aerosol formation and removal parameterizations used in the current generation of fully coupled feedback models should therefore be a focus for continued research.

(3) *Directed studies of feedback effects for large emission sources.* The work carried out here showed that feedback effects are strongest for sources such as large forest fires and industrial/urban plumes. This suggests that short-time-period studies for these sources will provide the best conditions for the improvement and testing of feedback models.

(4) *Detection of feedback effects in existing observation data.* The O_3 formation regime was shown to be sensitive to feedbacks, as was winter inorganic particle formation. It therefore may be possible to detect feedback effects in observation data through careful analysis of NOx and VOC sensitivity of O_3 (e.g. through comparing observed O_3 and particle nitrate formation regimes on days with high aerosol column loadings to otherwise similar days with low aerosol column loadings). Similarly, the work undertaken here suggests that indications of feedback effects may be present in observations of inorganic aerosol partitioning and biogenic hydrocarbon emissions, and may be identified through re-analysis of such data, particularly when coupled with observations of aerosol column optical properties. Such analysis would help identify useful periods for further model evaluation.

(5) *Further studies on the interaction between aerosol direct and indirect effects.* The work undertaken here suggested the direct and indirect effects may have competing influences on both ozone and $PM_{2.5}$ formation, though the manner in which this takes place has not been investigated. Short term case studies such as the ones described above should examine this competition at a process level, using separate direct, indirect and combined simulations, and existing or new observations.

We discuss the meteorological impacts of feedbacks (and their relationship to the above chemical impacts), in Part 1 of this work.

Acknowledgements

The authors gratefully acknowledge funding from Environment Canada's Clean Air Regulatory Agenda. The Centre of Excellence for Space Sciences and Technologies SPACE-SI is an operation partly financed by the European Union, European Regional Development Fund and Republic of Slovenia, Ministry of Higher Education, Science, Sport and Culture.

References

Abdul-Razzak, H. and Ghan,S.J., 2002. A parameterization of aerosol activation 3. Sectional representation. *Journal of Geophysical Research, Atmospheres*, vol 107, D3, pp AAC-1-1 to AAC1-6, DOI 10.1029/2001JD000483.

Baklanov, A., Schlunzen, K., Suppan., P., Baldasano, J., Brunner, D., Aksoyoglu, S., Carmichael, G., Douros, J., Flemming, J., Forkel, R., Galmarini, S., Gauss, M., Grell, G., Hirtl, M., Joffre, S., Jorba, O., Kaas, E., Kaasik, M., Kallos, G., Kong., X., Korsholm., U., Kurganskiy, A., Kushta, J., Lohmann, U., Mahura, A., Manders-Groot, A., Murizi, A., Moussiopoulos, N., Rao, S.T., Savage, N., Seigneur, C., Sokhi, R.S., Solazzo, E., Solomos, S., Sorenson, B., Tsegas, G., Vignati, E., Vogel, B., and Zhang, Y., 2014. Online coupled regional meteorology chemistry models in Europe: current status and prospects. Atm. Chem. Phys. 14, 317-398.

Brunner, D., Savage, N., Jorba, O., Eder, B., Giordano, L., Badia, A., Balzarini, A., Baro, Rocio, Bianconi, R., Chemel, C., Forkel, R., Jimenez-Guerrero, P., Hirtl, M., Hodzic, A., Honzak, L., Im., U., Knote, C., Kuensen, J.J.P., Makar, P.A., Manders-Groot, A., Neal, L., Perez, J.L., Pirovano, G., San Jose, R., Schroder, W., Sokhi, R.S., Syrakov, D., Torian, A., Werhahn, J., Wolke, R.., van Meijgaard, E., Yahya, K., Zabkar, R., Zhang, Y., Hogrefe, C., and Galmarini, S., Evaluation of the meteoroloogical performance of coupled chemistry-meteorology models in phase 2 of the Air Quality Model Evaluation International Initiative, Atmos. Env., *submitted*.

Campbell, P., K. Yahya, K. Wang, Y. Zhang, C. Hogrefe, G. Pouliot, C. Knote, R. San Jose and J. L. Perez, P. J. Guerrero, R. Baro, and P. Makar, 2014, Indicators of the Sensitivity of O_3 and $PM_{2.5}$ Formation to Precursor Gases over the Continental United States: A Multi-Model Assessment for the 2006 and 2010 Simulations under the Air Quality Model Evaluation International Initiative (AQMEII) Phase 2, *Atmospheric Environment*, this issue.

Clough, S. A., Shephard, M. W., Mlawer, E. J., Delamere, J. S., Iacono, M. J.,Cady-Pereira, K., Boukabara, S., Brown, P. D. ,2005. Atmospheric radiative transfer modeling: a summary of the AER codes. J. Quant. Spectrosc. Radiat. Transfer 91 (2): 233–244, doi:10.1016/j.jqsrt.2004.05.058.

Cohard, J.-M. Pinty, J.-P., and Bedos, C, 1998. Extending Twomey's analytical estimate of nucleated cloud droplet concentrations from CCN spectra. J. Atmos. Sci., 55, 3348–3357.

Forkel, R., Werhahn, J., Hansen, A.B., McKeen, S., Peckham, S., Grell, G., Suppan, P., 2012. Effect of aerosol-radiation feedback on regional air quality – A case study with WRF/Chem. Atmos. Environ., 53, 202-211.

Forster, P.M., Ramaswamy, V., *et al*., 2007. Changes in atmospheric constituents and in radiative forcing. In: Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K.B., Tignor, M., Miller, H.L. (Eds.), Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press.

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

Galmarini, S., Bianconi, R., Klug, W., Mikkelsen, T., Addis, R., Andronopoulos, S., Astrup, P., Baklanov, A., Bartniki, J., Bartzis, J.C., Bellasio, R., Bompay, F., Buckley, R., Bouzom, M., Champion, H., D'Amours, R., Davakis, E., Eleveld, H., Geertsema, G.T., Glaab, H., Kollax, M., Ilvonen, M., Manning, A., Pechinger, U., Persson, C., Polreich, E., Potemski, S., Prodanova, M., Saltbones, J., Slaper, H., Sofiev, M.A., Syrakov, D., Sørensen, J.H., Van der Auwera, L., Valkama, I., Zelazny, R., 2004b. Ensemble dispersion forecasting, Part I: concept, approach and Indicators. Atmos. Environ. 38 (28), 4607-4617.

Galmarini, S., Rao, S.T., and Steyn, D.G., 2012a. Preface, Atmospheric Environment Special Issue on the Air Quality Model Evaluation International Initiative, Atmos. Environ., 53, 1-3.

Galmarini, S., Bianconi, R., Appel, W., Solazzo, E., Mosca, S., Grossi, P., Moran, M., Schere, K., Rao, S.T., 2012b. ENSEMBLE and AMET: Two systems and approaches to a harmonized, simplified and efficient facility for air quality models development and evaluation, Atmospheric Environment 53, 51-59.

Giorgi, F., Bi, X., and Qian, Y., 2003. Indirect vs. Direct Effects of Anthropogenic Sulfate on the Climate of East Asia as Simulated with a Regional Coupled Climate-Chemistry/Aerosol Model, Climate Change, 345-376, 2003.

Gong, W., Makar, P.A., Zhang, J., Milbrandt, J., and Gravel, S., 2014. Modelling aerosol-cloud-meteorology interaction: a case study with a fully coupled air-quality model (GEM-MACH), *submitted*, Atmos. Env., 2014.

Grell, G.A., Peckham, S.E., Schmitz, R., McKeen, S.A., Frost, G., Skamarock, W.C., Eder, B., 2005. Fully coupled online chemistry within the WRF model. Atmos. Environ. 39, 6957-6975.

Hogrefe. C., G. Pouliot, D. Wong, A. Torian, S. Roselle, J. Pleim, and R. Mathur, 2014, Annual Application and Evaluation of the Online Coupled WRF-CMAQ System over North America under AQMEII Phase 2, Atmospheric Environment, under review

Im U., Bianconi, R., Solasso, E., Kioutsioukis, I., Badia, A., Balzasrini, A., Brunner, D., Chemel, C., Curci, G., Davis, L., van der Gon, H.D., Esteban., R.B., Flemming, J. Forkel, R., Giordano, L., Geurro, P.J., Hirtl, M., Hodsic, A., Honzka, L., Jorba, O., Knote, C., Kuenen, J.J.P, Makar, P.A., Manders-Groot, A., Pravano, G., Pouliot, G., San Jose, R., Savage, N., Schorder, W., Syrakov, D., Torian, A., Werhan, J., Wolke, R., Yahya, K., Žabkar, R., Zhang, J., Zhang, Y., Hogrefe, C., Galmarini, S., 2014a. Evaluation of operational online-coupled regional air quality models over Europe and North America in the context of AQMEII phase 2. Part I: Ozone, Atmos. Environ. *submitted.*

Im U., Bianconi, R., Solasso, E., Kioutsioukis, I., Badia, A., Balzasrini, A., Brunner, D., Chemel, C., Curci, G., Davis, L., van der Gon, H.D., Esteban., R.B., Flemming, J. Forkel, R., Giordano, L., Geurro, P.J., Hirtl, M., Hodsic, A., Honzka, L., Jorba, O., Knote, C., Kuenen, J.J.P, Makar, P.A., Manders-Groot, A., Pravano, G., Pouliot, G., San Jose, R., Savage, N., Schorder, W., Syrakov, D., Torian, A., Werhan, J., Wolke, R., Yahya, K., Žabkar, R., Zhang, J., Zhang, Y., Hogrefe, C., Galmarini, S., 2014a. Evaluation of operational online-coupled regional air quality models over Europe and North America in the context of AQMEII phase 2. Part 2: Particulate Matter, Atmos. Environ. *submitted*

Li, J. and Barker, H.W., 2005. A radiation algorithm with correlated-k distribution. Part I: Local Thermal Equilibrium. *J. of the Atmospheric Sciences*, 62, 286-309.

Makar, P.A., Gong, W., Milbrandt, J., Hogrefe, C., Zhang, Y., Curci, G., Zakbar, R., Im, U., Galmarini, S., Gravel, S., Zhang, J., Hou, A., Pabla, B., Cheung, P., Bianconi, R., 2014. Feedbacks between air pollution and weather, part 1: effects on weather. *Submitted, Atmos. Env.*

Milbrandt, J. A.,M. K. Yau, 2005. A multimoment bulk microphysics parameterization. Part II: A proposed three-moment closure and scheme description, J.Atmos. Sci., 62, 3065-3081.

Moran M.D., S. Ménard, D. Talbot, P. Huang, P.A. Makar, W. Gong, H. Landry, S. Gravel, S. Gong, L-P. Crevier, A. Kallaur, M. Sassi, 2010. Particulate-matter forecasting with GEM-MACH15, a new Canadian air-quality forecast model. In: Steyn DG, Rao ST (eds) *Air Pollution Modelling and Its Application XX*, Springer, Dordrecht, pp. 289-292.

Morrison, H., and Pinto, J.O., 2005. Mesoscale modeling of springtime Arctic mixed-phase stratiform clouds using a new two-moment bulk microphysics scheme. J. Atm. Sci., 62, 3683-3704.

Morrison, H., Thompson, G., and Tatarskii, V., 2009. Impact of Cloud Microphysics on the Development of Trailing Stratiform Precipitation in a Simulated Squall Line: Comparison of One- and Two-Moment Schemes. *Mon. Wea. Rev.*, 137, 991–1007.doi: http://dx.doi.org/10.1175/2008MWR2556.1

Solazzo, E., Bianconi, R., Vautaurd, R., Wyat Appel, K., Moran, M.D., Hogrefe, C., Bessagnet, B., Brandt, J., Christensen, J.H., Chemel, C., Coll, I., van der Gon, H.D., Ferreira, J., Forkel, R.,

Francis, X.V., Grell, G., Grossi, P., Hansen, A.B., Jericevic, A., Kraljevic, L, Miranda, A.I, Nopmongcol, U., Pirovano, G., Prank, M., Riccio, A., Sartelet, K.N, Schaap, M., Silver, J.D., Sokhi, R.S., Vira, J., Werhahn, J., Wolke, R., Yarwood, G., Zhang, J., Rao, S.T., Galmarini, S., 2012(a). Model evaluation and ensemble modelling of surface-level ozone in Europe and North America in the context of AQMEII. Atmos. Environ., 53, 60-74.

Solazzo, E., Bianconi, R., Pirovano, G., Matthias, V., Vautard, R., Moran, M.D., Appel, K.W., Bessagnet, B., Brandt, J., Christensen, J.H., Chemel, C., Coll, I., Ferreira, J., Forkel, R., Francis, X.V., Grell, G., Grossi, P., Hansen, A.B., Miranda, A.I., Nopmongcol, U., Prank, M., Sartelet, K.N., Schaap, M., Silver, J.D., Sokhi, R.S., Vira, J., Werhahn, J., Wolke, R., Yarwood, G., Zhang, J., Rao, S.T., Galmarini, S., 2012b. Operational model evaluation for particulate matter in Europe and North America in the context of AQMEII. Atmos. Environ. 53, 75-92.

Toon, O.B., and Pollack, J.B., 1976. A global average model of atmospheric aerosols for radiative transfer calculations. J. App. Met., 15, 225-246.

Wang, K., K. Yahya, Y. Zhang, C. Hogrefe, G. Pouliot, C. Knote, R. San Jose and J. L. Perez, P. J. Guerrero, R. Baro, and P. Makar, 2014a, Evaluation of Column Variable Predictions Using Satellite Data over the Continental United States: A Multi-Model Assessment for the 2006 and 2010 Simulations under the Air Quality Model Evaluation International Initiative (AQMEII) Phase 2, *Atmospheric Environment*, this issue.

Wang, K., K. Yahya, Y. Zhang, S.-Y. Wu, and G. Grell, 2014b, Implementation and Initial Application of A New Chemistry-Aerosol Option in WRF/Chem for Simulation of Secondary Organic Aerosols and Aerosol Indirect Effects, *Atmospheric Environment*, , this issue.

Yahya, K., K. Wang, M. Gudoshava, T. Glotfelty, and Y. Zhang, 2014a, Application of WRF/Chem over the continental U.S. under the AQMEII Phase II: Comprehensive Evaluation of 2006 Simulation, *Atmospheric Environment*, in review.

Yahya, K., K. Wang, Y. Zhang, C. Hogrefe and G. Pouliot, and T. E. Kleindienst, 2014b, Application of WRF/Chem over the continental U.S. under the AQMEII Phase II: Comprehensive Evaluation of 2010 Simulation and Responses of Air Quality and Meteorology-Chemistry Interactions to Changes in Emissions and Meteorology from 2006 to 2010, *Atmospheric Environment*, , this issue.

Zhang, Y., 2008, Online Coupled Meteorology and Chemistry models: History, Current Status, and Outlook, *Atmospheric Chemistry and Physics*, 8, 2895-2932.

Table 1. Methodologies used in simulating aerosol direct and indirect effects and feedbacks in the suite of models.

Domain	**Model (AQMEII-2 ID)**	**Direct Effect Methodology**	**Indirect Effect Methodology**	**Time Period, Data available for comparisons**
NA	GEM-MACH (CA2, CA2f)	Mie scattering (Bohren and Huffman, 1983), homogeneous aerosol assumed, complex refractive indexes from bilinear interpolation in aerosol water content; detailed code used to generate high resolution lookup tables tested to be within 1% accuracy of the original Mie code.	Milbrandt-Yao 2 moment microphysics scheme (Milbrandt and Yao, 2005). No-feedback uses Cohard *et al* (1998) 'typical continental aerosol' cloud condensation nucleii tables. Feedback uses the aerosol size and speciation-dependent formulation of Abdul-Razzak and Ghan (2002), operating across bins. Aerosol activation determined by comparing the upper and lower bounds of critical supersaturation for each size bin to the maximum supersaturation in an updraft, through a number-weighted critical supersaturation (See Gong *et al*, 2014, this issue of Atm. Env)).	2006, 2010, feedback and non-feedback. Both chemical and meteorological variables available for comparisons
	WRF-CHEM 3.4.1 (US8)	Fast-Chapman *Fast et al.* [2006] *Chapman et al.* [2009]	Indirect effects simulated following Chapman *et al*. (2009), using the Morrison 2-moment microphysics scheme (Morrison *et al*., 2009), with aerosol activation based on the parameterization of Abdul-Razzak (2002), operating across the each mode of the WRF-CHEM aerosol distribution.	2006, 2010 feedback simulations, weather-only simulations. Meteorological variables available for comparisons
	WRF-CMAQ (US6)	CMAQ Feedback *Bohren and Huffman* [1998]; *Wong et al.* [2012]	None; the cloud droplet concentration is assumed to be 250 cm^{-3}.	June 1 to September 1, 2006; May 1 to October 1, 2010. Both chemical and meteorological variables available for comparison.
EU	WRF-CHEM 3.4.1 (Feedback: SI1,basecase: SI2)	Fast-Chapman *Fast et al.* [2006] *Chapman et al.* [2009]	None; the cloud droplet concentration is assumed to be 250 cm^{-3}.	2010, feedback and non-feedback. Both chemistry and meteorological models available for comparison.
	WRF-CHEM 3.4 + (New experimental version based on v 3.4; IT2)	Direct effects simulated following Fast *et al*. (2006). The lognormal modes are divided in bins. Each aerosol constituent is associated with a complex index of refraction. The refractive index of each bin is calculated with viavolume averaging. Mie theory is used to calculate the extinction and scattering efficiency.	Indirect effects simulated following Chapman *et al*. (2009), using the Morrison 2-moment microphysics scheme (Morrison *et al*., 2009), with aerosol activation based on the parameterization of Abdul-Razzak (2002), operating across the each mode of the WRF-CHEM aerosol distribution. When indirect effects are de-activated (no-feedback simulation), it is assumed that the cloud droplet concentration is 250 cm^{-3}.	2010, feedback and weather-only simulation. Meteorological variables available for comparison.

Table 2 Statistical measures used to compare Feedback (F) and No-Feedback (NF) simulations

Statistical Measure	**Description**	**Formula**
PCC	Pearson Correlation Coefficient	$PCC = \frac{N\sum_{i=1}^{N}(NF_i \cdot F_i) - \sum_{i=1}^{N}(F_i)\sum_{i=1}^{N}(NF_i)}{\sqrt{N\sum_{i=1}^{N}(F_i \cdot F_i) - \sum_{i=1}^{N}(F_i)\cdot\sum_{i=1}^{N}(F_i)}\sqrt{N\sum_{i=1}^{N}(NF_i \cdot NF_i) - \sum_{i=1}^{N}(NF_i)\cdot\sum_{i=1}^{N}(NF_i)}}$
MD	Mean Difference	$MD = \frac{1}{N}\sum_{i=1}^{N}(F_i - NF_i)$
MAD	Mean Absolute Difference	$MAD = \frac{1}{N}\sum_{i=1}^{N}\lvert F_i - NF_i \rvert$
MSD	Mean Square Difference	$MSD = \frac{1}{N}\sum_{i=1}^{N}(F_i - NF_i)^2$
Intercept	Intercept of observations vs. model best-fit line	$a = \overline{F} - b \cdot \overline{NF}$
NMD	Normalized Mean Difference	$NMD = \frac{\sum_{i=1}^{N}(F_i - NF_i)}{\sum_{i=1}^{N} NF_i} x100$
NMAD	Normalized Mean Absolute Difference	$NMAD = \frac{\sum_{i=1}^{N}\lvert F_i - NF_i \rvert}{\sum_{i=1}^{N} NF_i} x100$
RMSD	Root Mean Square Difference	$RMSD = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(F_i - NF_i)^2}$

Slope	Slope of observations vs. model best-fit line	$b = \frac{\sum_{i=1}^{N}\left[\left(NF_i - \overline{NF}\right)\left(F_i - \overline{F}\right)\right]}{\sum_{i=1}^{N}\left[\left(NF_i - \overline{NF}\right)^2\right]}$
STD	Standard Deviation (Feedback and No-Feedback)	$STD = \frac{\sum_{i=1}^{N}\left(F_i - \overline{F}_i\right)^2}{N}, \frac{\sum_{i=1}^{N}\left(NF_i - \overline{NF}_i\right)^2}{N}$
DSTD	Change in standard deviation (used to compare two model's variability, where F and NF are the Feedback and No-Feedback models, respectively)	$DSTD = \frac{\sum_{i=1}^{N}\left(F_i - \overline{F}_i\right)^2}{N} - \frac{\sum_{i=1}^{N}\left(NF_i - \overline{NF}_i\right)^2}{N}$

Table 3 Statistical measures used for model – observation performance estimates. N is the number of paired observed-model values. For comparisons between observations and model values, $\overline{\boldsymbol{O}}$ is the mean observed value, $\overline{\boldsymbol{M}}$ is the mean model value.

Statistical Measure	**Description**	**Formula**
FA2	Fraction (percentage) of model values within a factor of two of observations.	-
FA5	Fraction (percentage) of model values within a factor of five of observations.	
MB	Mean Bias	$MB = \frac{1}{N}\sum_{i=1}^{N}\left(M_i - O_i\right)$
FB	Fractional Bias	$FB = 2\left(\frac{\overline{M} - \overline{O}}{\overline{M} + \overline{O}}\right)$
NMB	Normalized Mean Bias	$NMB = \frac{\sum_{i=1}^{N}\left(M_i - O_i\right)}{\sum_{i=1}^{N} O_i} x100$
PCC	Pearson Correlation Coefficient	$PCC = \frac{N\sum_{i=1}^{N}(O_i \cdot M_i) - \sum_{i=1}^{N}(M_i)\sum_{i=1}^{N}(O_i)}{\sqrt{N\sum_{i=1}^{N}(M_i \cdot M_i) - \sum_{i=1}^{N}(M_i)\cdot\sum_{i=1}^{N}(M_i)}\sqrt{N\sum_{i=1}^{N}(O_i \cdot O_i) - \sum_{i=1}^{N}(O_i)\cdot\sum_{i=1}^{N}(O_i)}}$
ME	Mean Error	$ME = \frac{1}{N}\sum_{i=1}^{N}\left\|M_i - O_i\right\|$
NMSE	Normalized mean square error	$NMSE = \frac{\frac{1}{N}\sum_{i=1}^{N}\left(M_i - O_i\right)^2}{\overline{M}\,\overline{O}}$
NME	Normalized Mean Absolute Error	$NME = \frac{\sum_{i=1}^{N}\left\|M_i - O_i\right\|}{\sum_{i=1}^{N} O_i} x100$

Table 4 Model Evaluation, 2010, July 15 0:00 to August 15th 0:00. Values hourly unless otherwise noted. Bold face indicates the best performing model of the GEM-MACH no-feedback and feedback pair, italics the best performing model of all four examined here. A bold face font is used to identify the best performing model of the two GEM-MACH simulations, and italics to identify the highest performing model of the suite of four models examined here.

Variable	Statistic	GEM-MACH non-feedback (CA2)	GEM-MACH feedback (CA2f)	WRF-CMAQ (US6)	WRF-CHEM (US8)
O_3 (Regional) 75% validity cutoff, 257 stations.	NP	187330	187287	188017	188269
	FA2 (%)	83.49	**83.72**	85.03	*86.11*
	FA5 (%)	96.75	***98.77***	97.21	97.64
	MB	4.21E+00	**3.81E+00**	*3.36E+00*	-4.05E+00
	FB	1.20E-01	**1.09E-01**	*9.72E-02*	-1.31E-01
	NMB (%)	12.78	**11.57**	*10.21*	-12.30
	PCC	0.60	0.60	*0.70*	0.67
	ME	1.15E+01	**1.13E+01**	1.01E+01	*9.98E+00*
	NMSE	1.81E-01	**1.77E-01**	*1.42E-01*	1.82E-01
	NME (%)	34.96	**34.42**	30.79	*30.30*
	N. Scores	*0* **(0)**	*1* **(8)**	*5*	*3*
O_3 (Urban + suburban) 75% validity cutoff, 494 stations	NP	333840	334317	345222	345649
	FA2 (%)	79.26	**79.31**	78.45	*80.65*
	FA5 (%)	**95.06**	95.04	94.17	*95.68*
	MB	2.86E+00	***2.47E+00***	2.61E+00	-4.32E+00
	FB	8.72E-02	***7.59E-02***	8.09E-02	-1.50E-01
	NMB (%)	9.11	***7.88***	8.43	-13.96
	PCC	0.63	0.63	*0.69*	*0.69*
	ME	1.16E+01	**1.15E+01**	1.11E+01	*1.01E+01*
	NMSE	2.12E-01	**2.10E-01**	*1.94E-01*	2.16E-01
	NME (%)	37.03	**36.65**	35.87	*32.70*
	N. Scores	*0* **(1)**	*3* **(7)**	*2*	*5*
SO_2, all stations, 75% validity cutoff, 181 stations	NP	86816	86789	81894	81896
	FA2 (%)	**39.09**	38.97	40.97	*42.84*
	FA5 (%)	**75.26**	74.96	74.99	*76.64*
	MB	1.31E+00	***1.18E+00***	-1.29E+00	-1.58E+00
	FB	3.34E-01	***3.05E-01***	-4.88E-01	-6.37E-01
	NMB (%)	40.10	***36.01***	-39.23	-48.33
	PCC	0.12	**0.13**	0.19	*0.21*
	ME	4.03E+00	**3.95E+00**	2.63E+00	*2.45E+00*
	NMSE	4.89E+00	***4.88E+00***	7.00E+00	8.18E+00
	NME (%)	122.95	**120.48**	79.97	*75.13*
	N. Scores	*0* **(2)**	*4* **(7)**	*0*	*5*
NO, all stations, 16%	NP	45481	45311	53593	51701
	FA2 (%)	***31.50***	30.18	27.14	21.02
	FA5 (%)	***62.95***	60.88	56.07	47.79

validity cutoff, 135 stations	MB	3.76E-01	***1.62E-01***	1.56E+00	-2.33E+00
	FB	9.14E-02	***4.03E-02***	3.48E-01	-8.98E-01
	NMB (%)	9.58	***4.12***	42.15	-61.96
	PCC	***0.26***	0.24	0.15	0.19
	ME	4.40E+00	**4.35E+00**	5.60E+00	*3.30E+00*
	NMSE	***4.72E+00***	4.77E+00	8.50E+00	7.64E+00
	NME (%)	112.22	**110.79**	151.42	*87.46*
	N. Scores	*4* **(4)**	*3* **(5)**	*0*	*2*
NO_2, 75% validity cutoff, 198 stations, all stations	NP	131961	131961	130776	130776
	FA2 (%)	**49.97**	49.96	*50.31*	49.03
	FA5 (%)	87.49	**87.60**	*87.76*	86.18
	MB	1.14E+00	**1.01E+00**	2.38E+00	*6.38E-01*
	FB	1.35E-01	**1.20E-01**	2.62E-01	*7.77E-02*
	NMB (%)	14.45	**12.79**	30.20	*8.09*
	PCC	**0.47**	0.46	*0.50*	0.46
	ME	6.03E+00	**5.99E+00**	6.40E+00	*5.77E+00*
	NMSE	1.33E+00	**1.32E+00**	1.25E+00	*1.16E+00*
	NME (%)	76.59	**76.11**	81.14	*73.22*
	N. Scores	*0* **(2)**	*0* **(7)**	*3*	*6*
CO, 75% validity cutoff, 108 stations, urban, suburban and regional	NP	48037	48037	48029	48029
	FA2 (%)	**68.95**	68.55	69.24	*72.72*
	FA5 (%)	95.83	**95.84**	95.82	*96.28*
	MB	***-1.88E+01***	-2.67E+01	-2.97E+01	-5.38E+01
	FB	***-7.27E-02***	-1.05E-01	-1.18E-01	2.23E-01
	NMB (%)	***-7.02***	-9.95	-11.10	-20.09
	PCC	**0.15**	0.14	*0.21*	0.21
	ME	1.62E+02	**1.59E+02**	1.50E+02	*1.36E+02*
	NMSE	1.15E+00	**1.12E+00**	8.94E-01	*8.27E-01*
	NME (%)	60.53	**59.45**	56.00	*50.60*
	N. Scores	*3* **(5)**	*0* **(4)**	*1*	*5*
PM_{10}, 16% validity cutoff, 350 stations, all station types	NP	3896	3896	3896	3896
	FA2 (%)	***54.26***	53.77	39.66	11.32
	FA5 (%)	94.12	***94.79***	77.10	61.78
	MB	-1.64E+00	***-1.43E+00***	-9.09E+00	-1.93E+01
	FB	-6.55E-02	***-5.69E-02***	-4.28E-01	-1.19E+00
	NMB (%)	-6.34	***-5.53***	-35.22	-74.64
	PCC	**0.09**	0.08	0.13	*0.26*
	ME	1.74E+01	***1.72E+01***	1.74E+01	1.93E+01
	NMSE	1.08E+00	***1.05E+00***	1.40E+00	3.47E+00
	NME (%)	67.46	***66.66***	67.59	74.77
	N. Scores	*1* **(2)**	*7* **(7)**	*0*	*1*
PM_{10}, 16% validity cutoff, stations, regional stations only, 72 stations	NP	1088	1088	1088	1088
	FA2 (%)	***49.92***	41.64	37.22	8.00
	FA5 (%)	89.89	***91.54***	77.30	61.49
	MB	6.02E+00	**5.66E+00**	*-2.49E+00*	-2.02E+01
	FB	2.08E-01	**1.97E-01**	*-1.01E-01*	-1.27E+00
	NMB (%)	23.20	**21.82**	*-9.59*	-77.74
	PCC	**0.02**	0.01	0.06	*0.25*
	ME	2.49E+01	**2.44E+01**	2.17E+01	*2.02E+01*

	NMSE	1.64E+00	***1.60E+00***	1.64E+00	4.77E+00
	NME (%)	96.18	**93.90**	83.60	*77.74*
	N. Scores	*1* **(2)**	*2* **(7)**	*3*	*3*
$PM_{2.5}$, daily average, 16% validity cutoff (to capture 1 day in 6 stations): 900 stations, all stations combined	NP	11754	11754	11798	11798
	FA2 (%)	**78.06**	75.63	79.27	*82.48*
	FA5 (%)	**99.06**	98.91	99.08	*99.28*
	MB	**3.37E+00**	4.02E+00	*-1.71E+00*	-2.03E+00
	FB	**2.66E-01**	3.09E-01	*-1.69E-01*	-2.03E-01
	NMB (%)	**30.66**	36.53	*-15.59*	-18.44
	PCC	**0.52**	0.51	0.63	*0.72*
	ME	**5.80E+00**	6.25E+00	4.01E+00	*3.59E+00*
	NMSE	**5.13E-01**	5.51E-01	3.06E-01	*2.27E-01*
	NME (%)	**52.78**	56.86	36.49	*32.66*
	N. Scores	*0* **(9)**	*0* **(0)**	*3*	*6*
$PM_{2.5}$ SO_4, 16% validity cutoff (to capture 1 day in 6 stations): 297 stations, all station types	NP	2468	2468	2492	2492
	FA2 (%)	**46.03**	42.30	*86.88*	81.54
	FA5 (%)	**91.69**	89.47	99.12	*99.20*
	MB	**2.25E+00**	2.61E+00	-2.09E-01	*-1.34E-01*
	FB	**7.34E-01**	8.03E-01	-1.14E-01	*-7.18E-02*
	NMB (%)	**116.03**	134.21	-10.77	*-6.93*
	PCC	**0.72**	0.71	*0.84*	0.80
	ME	**2.41E+00**	2.74E+00	*6.07E-01*	7.27E-01
	NMSE	**1.89E+00**	2.17E+00	*2.74E-01*	3.97E-01
	NME (%)	**124.27**	141.08	*31.33*	37.55
	N. Scores	*0* **(9)**	*0* **(0)**	*5*	*4*
$PM_{2.5}$ NH_4, 16% validity cutoff (to capture 1 day in 6 stations): 142 stations, all station types	NP	1359	1359	1380	1380
	FA2 (%)	**62.69**	60.41	53.99	*64.57*
	FA5 (%)	***95.58***	95.00	94.06	92.75
	MB	***1.73E-01***	2.16E-01	-3.65E-01	-2.93E-01
	FB	***2.13E-01***	2.59E-01	-6.73E-01	-5.07E-01
	NMB (%)	***23.80***	29.73	-50.38	-40.44
	PCC	**0.66**	0.65	*0.82*	0.80
	ME	**4.46E-01**	4.72E-01	3.83E-01	*3.38E-01*
	NMSE	***7.10E-01***	7.43E-01	1.35E+00	8.73E-01
	NME (%)	**61.37**	64.94	52.93	*46.70*
	N. Scores	*4* **(9)**	*0* **(0)**	*1*	*3*
$PM_{2.5}$ NO_3, 16% validity cutoff (to capture 1 day in 6 stations): 139 stations, all types	NP	1281	1284	1342	885
	FA2 (%)	***18.58***	17.99	12.67	15.93
	FA5 (%)	***37.39***	35.12	32.19	35.03
	MB	***-1.22E-01***	-1.32E-01	-2.45E-01	-2.23E-01
	FB	***-4.59E-01***	-5.09E-01	-1.27E+00	-7.89E-01
	NMB (%)	***-37.32***	-40.55	-77.59	56.60
	PCC	**0.16**	0.15	*0.26*	0.22
	ME	3.62E-01	**3.57E-01**	*2.91E-01*	3.68E-01
	NMSE	7.42E+00	**7.25E+00**	1.46E+01	*6.59E+00*
	NME (%)	110.61	**109.19**	*92.31*	93.69
	N. Scores	*5* **(6)**	*0* **(3)**	*3*	*1*
$PM_{2.5}$ TOC	NP	1525	1525	1549	1549
	FA2 (%)	***60.46***	57.11	47.45	51.00

16% validity cutoff (to capture 1 day in 6 stations): 160 stations, all types	FA5 (%)	*94.89*	*94.89*	91.28	87.15
	MB	**6.92E-01**	8.01E-01	-5.12E-01	*-1.29E-03*
	FB	**4.21E-01**	4.72E-01	-4.98E-01	*1.00E-03*
	NMB (%)	**53.38**	61.76	-39.87	*-0.10*
	PCC	0.38	**0.40**	*0.55*	0.26
	ME	**1.02E+00**	1.08E+00	*6.97E-01*	9.10E-01
	NMSE	1.24E+00	**1.16E+00**	*1.07E+00*	1.18E+00
	NME (%)	**78.38**	83.04	*54.29*	70.89
	N. Scores	*2* **(6)**	*1* **(2)**	*4*	*3*

Table 5 Model Evaluation, EU, 2010, July 25th 00:00 to August 19th 00:00. The relative performance of the no-feedback and direct-effect only feedback simulations with WRF-CHEM v3.4.0 are highlighted using a bold font, while the best scores over all three models are highlighted with an italic font.

N/A: Data not available in the ENSEMBLE archive

Variable	Statistic	WRF-CHEM (no direct effect feedback-SI1)	WRF-CHEM (direct effect feedback – SI2)	WRF-CHEM (direct and indirect effect feedback – IT2)
O_3 (Regional) 75% validity cutoff, 498 stations	NP	284959	284959	284914
	FA2 (%)	***89.50***	89.45	88.94
	FA5 (%)	97.87	97.87	*98.19*
	MB	***7.53E-01***	8.90E-01	-8.65E+00
	FB (%)	***1.12E-02***	1.33E-02	-1.39E-01
	NMB (%)	***1.13***	1.35	-12.98
	PCC	*0.55*	*0.55*	0.53
	ME	*1.88E+01*	*1.88E+01*	2.06E+01
	NMSE (%)	*1.28E-01*	*1.28E-01*	1.75E-01
	NME (%)	***28.16***	28.20	30.85
	N. Scores	*8* **(5)**	*3* **(0)**	*1*
O_3 (Urban + suburban) 75% validity cutoff, 1005 stations	NP	1294806	1294806	1294713
	FA2 (%)	**81.02**	80.97	*83.15*
	FA5 (%)	**94.58**	94.57	*95.64*
	MB	**1.02E+01**	1.03E+01	*-1.18E-01*
	FB (%)	**1.70E-01**	1.72E-01	*-2.16E-03*
	NMB (%)	**18.55**	18.78	*-0.21*
	PCC	*0.59*	*0.59*	0.54
	ME	1.98E+01	1.98E+01	*1.89E+01*
	NMSE (%)	*1.78E-01*	*1.78E-01*	1.91E-01
	NME (%)	**35.90**	35.98	*34.36*
	N. Scores	*2* ***(7)***	*1***(0)**	*7*
	NP	445468	445697	474455

SO_2, all stations, 75% validity cutoff, 1000 stations	FA2 (%)	22.82	**22.55**	*32.72*
	FA5 (%)	51.78	**51.39**	*68.27*
	MB	**-2.44E+00**	-2.46E+00	*-1.97E+00*
	FB (%)	**-9.28E-01**	-9.42E-01	*-7.02E-01*
	NMB (%)	**-63.39**	-64.02	*-51.94*
	PCC	0.17	0.17	0.17
	ME	3.28E+00	**3.27E+00**	*3.01E+00*
	NMSE (%)	1.09E+01	1.09E+01	*8.08E+00*
	NME (%)	85.08	**84.98**	*81.47*
	N. Scores	*1* **(3)**	1 **(4)**	*9*
NO, all stations, 75% validity cutoff, 904 stations	NP	416717	416563	609303
	FA2 (%)	**17.46**	17.23	*18.70*
	FA5 (%)	**44.97**	44.32	*42.01*
	MB	**-3.14E+00**	-3.16E+00	*-2.84E+00*
	FB (%)	**-1.41E+00**	-1.42E+00	*-1.23E+00*
	NMB (%)	**-82.57**	-83.13	*-76.01*
	PCC	*0.08*	*0.08*	0.06
	ME	3.37E+00	3.37E+00	*3.16E+00*
	NMSE (%)	**2.30E+01**	2.37E+01	*2.11E+01*
	NME (%)	***88.66***	88.74	91.03
	N. Scores	*2* **(9)**	*1* **(0)**	*7*
NO, urban stations only, 75% validity cutoff, 383 stations	NP	177863	177802	263704
	FA2 (%)	14.17	***13.80***	16.31
	FA5 (%)	39.76	**39.07**	*36.27*
	MB	**-4.44E+00**	-4.46E+00	*-3.73E+00*
	FB (%)	**-1.52E+00**	-1.53E+00	*-1.37E+00*
	NMB (%)	**-86.28**	-86.67	*-81.46*
	PCC	*0.11*	*0.11*	0.09
	ME	**4.58E+00**	4.59E+00	*4.08E+00*
	NMSE (%)	**2.46E+01**	2.53E+01	*2.10E+01*
	NME (%)	**88.97**	89.16	*88.29*
	N. Scores	*1* **(7)**	2 **(3)**	7
NO_2, 75% validity cutoff, regional stations, 366 stations	NP	281752	281752	281752
	FA2 (%)	**48.33**	48.01	*52.63*
	FA5 (%)	**85.67**	85.50	*88.61*
	MB	**-3.04E+00**	-3.11E+00	*-2.61E+00*
	FB (%)	**-5.01E-01**	-5.16E-01	*-4.15E-01*
	NMB (%)	**-40.04**	-41.00	*-34.37*
	PCC	0.23	**0.25**	*0.30*
	ME	4.83E+00	**4.82E+00**	*4.66E+00*
	NMSE (%)	2.62E+00	**2.33E+00**	*1.83E+00*
	NME (%)	63.64	**63.49**	*61.45*
	N. Scores	0 **(5)**	0 **(4)**	9
NO_2, 75% validity cutoff, urban stations, 721 stations	NP	403113	403113	403113
	FA2 (%)	**30.73**	30.37	*32.77*
	FA5 (%)	**71.29**	70.86	*73.98*
	MB	**-9.94E+00**	-1.00E+01	*-9.87E+00*
	FB (%)	**-8.69E-01**	-8.80E-01	*-8.60E-01*
	NMB (%)	**-60.58**	-61.11	*-60.14*

	PCC	0.34	***0.36***	0.35
	ME	1.13E+01	1.13E+01	*1.11E+01*
	NMSE (%)	**2.40E+00**	2.42E+00	*2.32E+00*
	NME (%)	**68.66**	68.82	*67.61*
	N. Scores	0 **(8)**	1 (2)	8
CO, 75% validity cutoff, 431 stations, all station types	NP	233292	233292	233292
	FA2 (%)	57.14	***57.24***	46.90
	FA5 (%)	*94.47*	*94.47*	92.11
	MB	***-9.07E+01***	-9.27E+01	-1.11E+02
	FB (%)	***-4.17E-01***	-4.28E-01	-5.36E-01
	NMB (%)	***-34.53***	-35.27	-42.25
	PCC	***0.06***	0.05	*0.06*
	ME	*1.67E+02*	*1.67E+02*	1.80E+02
	NMSE (%)	2.62E+00	***2.40E+00***	3.10E+00
	NME (%)	63.74	***63.44***	68.52
	N. Scores	*6* **(5)**	*5* **(4)**	*1*
PM_{10}, daily average, 95% validity cutoff, 887 stations, all station types	NP	22521	22521	22521
	FA2 (%)	***68.13***	67.19	26.85
	FA5 (%)	*97.20*	*97.20*	90.82
	MB	***-4.01E+00***	-4.19E+00	-1.16E+01
	FB (%)	***-2.36E-01***	-2.47E-01	-8.73E-01
	NMB (%)	***-21.08***	-22.01	-60.79
	PCC	0.35	0.35	*0.39*
	ME	*9.45E+00*	*9.45E+00*	1.23E+01
	NMSE (%)	1.04E+00	***1.01E+00***	2.28E+00
	NME (%)	49.67	***49.66***	64.65
	N. Scores	*6* **(6)**	*4* **(4)**	*1*
PM_{10}, 75% validity cutoff, stations, regional stations only, 307 stations	NP	7534	7534	7534
	FA2 (%)	***70.95***	70.87	36.06
	FA5 (%)	*96.60*	*96.60*	92.49
	MB	***-1.51E+00***	-1.70E+00	-8.95E+00
	FB (%)	***-9.84E-02***	-1.11E-01	-7.69E-01
	NMB (%)	***-9.38***	-10.56	-55.53
	PCC	0.32	0.33	*0.35*
	ME	8.18E+00	*8.13E+00*	9.95E+00
	NMSE (%)	1.20E+00	***1.08E+00***	2.16E+00
	NME (%)	50.76	***50.48***	61.78
	N. Scores	*5* **(5)**	*4* **(3)**	*1*
$PM_{2.5}$, daily average, 75% validity cutoff: 499 stations, all stations combined	NP	12041	12041	12041
	FA2 (%)	***65.92***	66.37	51.76
	FA5 (%)	***97.67***	97.61	97.18
	MB	1.12E+00	***9.58E-01***	-5.04E+00
	FB (%)	9.57E-02	***8.23E-02***	-5.83E-01
	NMB (%)	10.05	***8.58***	-45.11
	PCC	0.26	0.25	*0.32*
	ME	6.50E+00	6.43E+00	*6.03E+00*
	NMSE (%)	7.72E-01	***7.49E-01***	1.22E+00
	NME (%)	58.22	57.84	*54.05*
	N. Scores	*2* **(2)**	*4* **(4)**	*3*

PM SO_4, 16% validity cutoff (to capture 1 day in 6 stations): 38 stations, all types	NP	909	909	909
	FA2 (%)	***47.52***	47.08	32.23
	FA5 (%)	*85.04*	*85.04*	73.38
	MB	4.33E-02	***4.24E-02***	-1.47E+00
	FB (%)	1.69E-02	***1.65E-02***	-8.11E-01
	NMB (%)	1.70	***1.66***	-57.71
	PCC	*0.23*	*0.23*	0.17
	ME	1.94E+00	1.94E+00	*1.90E+00*
	NMSE (%)	*1.22E+00*	*1.22E+00*	3.23E+00
	NME (%)	**76.09**	76.24	*74.73*
	N. Scores	*4* **(5)**	*6* **(5)**	*3*
PM NH_4, 16% validity cutoff (to capture 1 day in 6 stations): 25 stations, all types	NP	567	567	567
	FA2 (%)	**37.39**	38.10	*29.63*
	FA5 (%)	**73.19**	73.54	*69.49*
	MB	5.61E-01	***5.27E-01***	-9.13E-01
	FB (%)	3.12E-01	***2.96E-01***	-8.63E-01
	NMB (%)	37.00	***34.79***	-60.28
	PCC	0.35	***0.36***	0.20
	ME	1.47E+00	1.45E+00	*1.22E+00*
	NMSE (%)	1.78E+00	***1.72E+00***	5.20E+00
	NME (%)	97.17	***95.63***	80.59
	N. Scores	*0* **(2)**	*6* **(6)**	*3*
PM NO_3, 16% validity cutoff (to capture 1 day in 6 stations): 19 stations, all types	NP	349	351	406
	FA2 (%)	***34.96***	34.47	10.34
	FA5 (%)	***61.60***	60.40	27.09
	MB	5.65E-01	***3.85E-01***	-2.80E+00
	FB (%)	1.36E-01	***9.47E-02***	-1.33E+00
	NMB (%)	14.57	***9.94***	-79.81
	PCC	0.28	***0.31***	0.16
	ME	3.72E+00	**3.61E+00**	*3.25E+00*
	NMSE (%)	2.74E+00	***2.55E+00***	1.26E+01
	NME (%)	95.91	**93.22**	*92.59*
	N. Scores	2 **(2)**	5 **(7)**	2

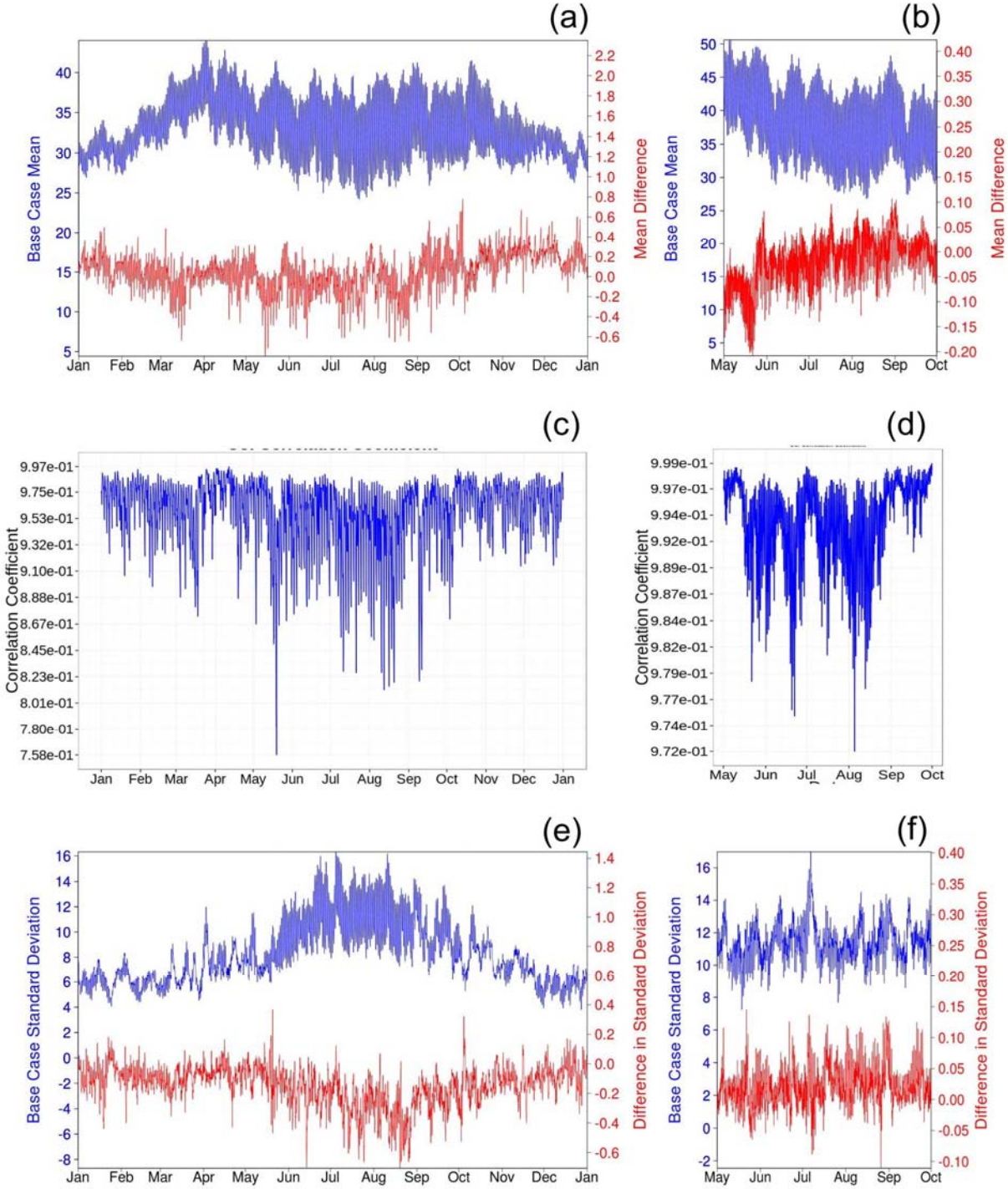

Figure 1. Grid-average O_3 time series for GEM-MACH (left column) and WRF-CMAQ (right column). Top row: mean non-feedback (blue) and mean differences (red), middle: correlation coefficients. Bottom row: non-feedback standard deviation (blue) and difference in standard deviation (feedback – non-feedback, red).

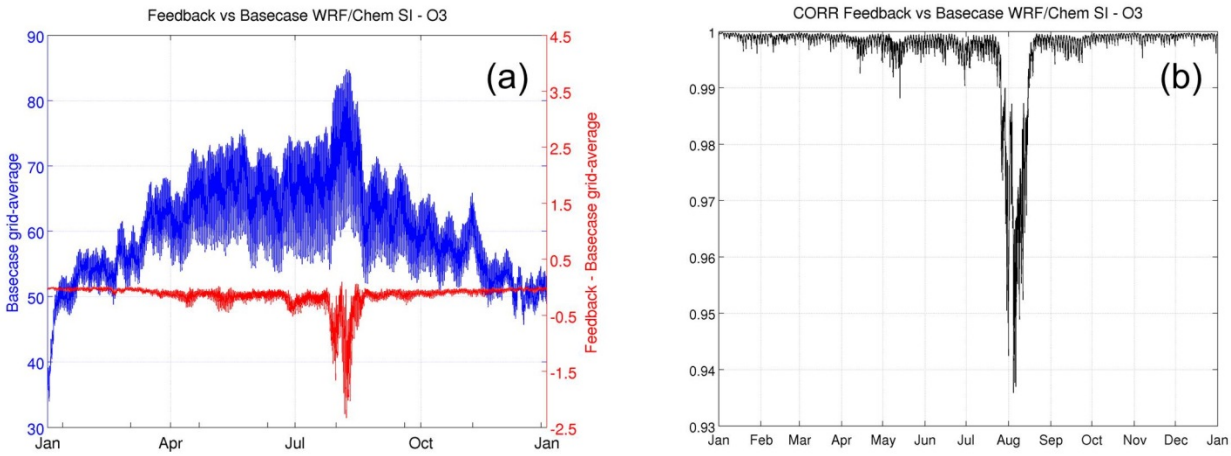

Figure 2. (a,b): Hourly grid-average O_3 no-feedback mean concentrations, mean differences (feedback – no-feedback), and simulation correlation coefficients, EU domain, 2010 (ppbv).

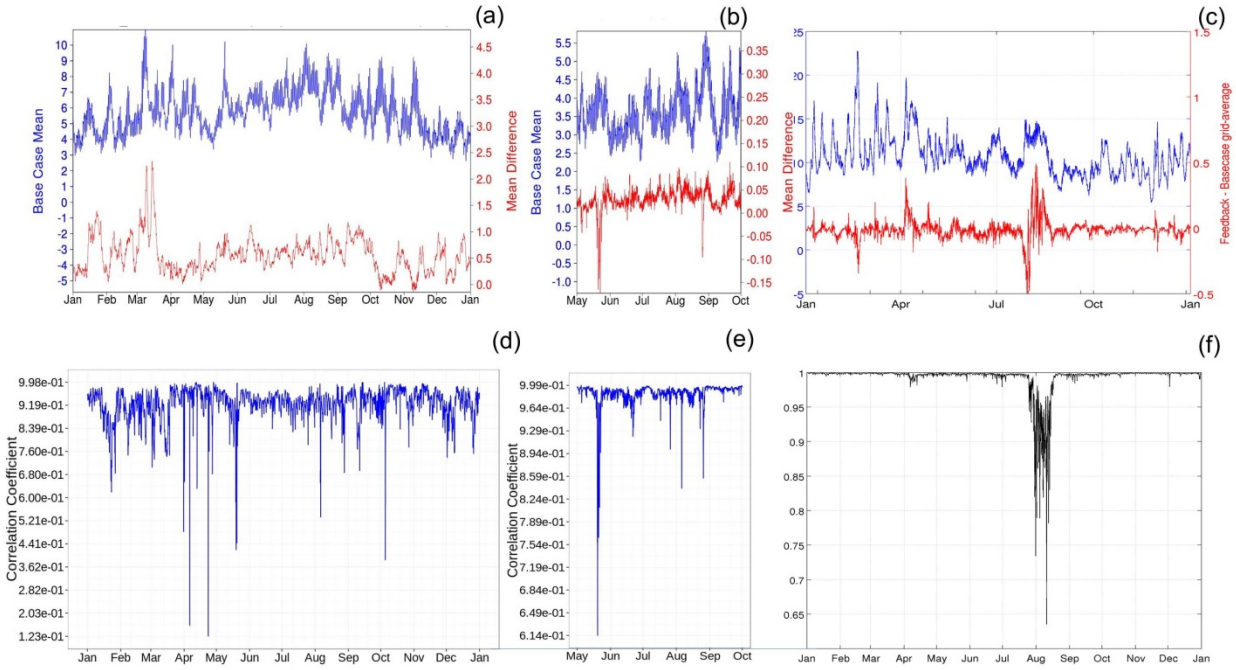

Figure 3. Grid mean $PM_{2.5}$, non-feedback (blue) and mean difference (red), for (a) NA/GEM-MACH, (b) NA/WRF-CMAQ, (c) EU/WRF-CHEM. (d,e,f): Correlation coefficients for these models.

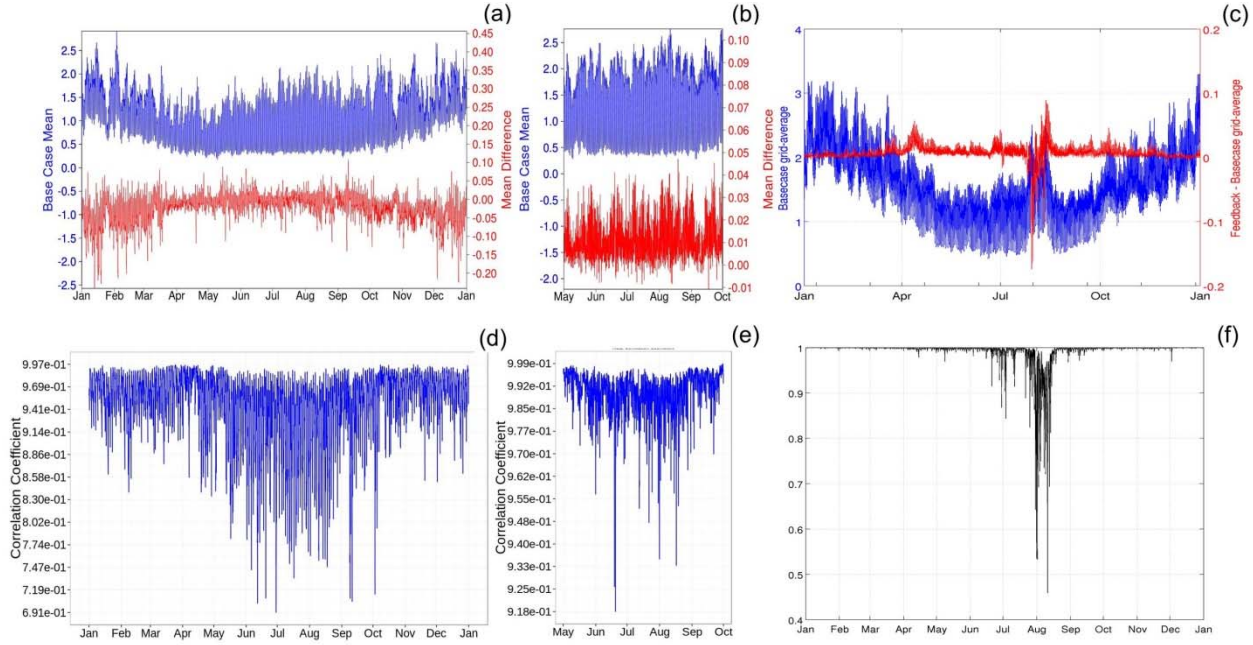

Figure 4. As for Figure 3, NO_2.

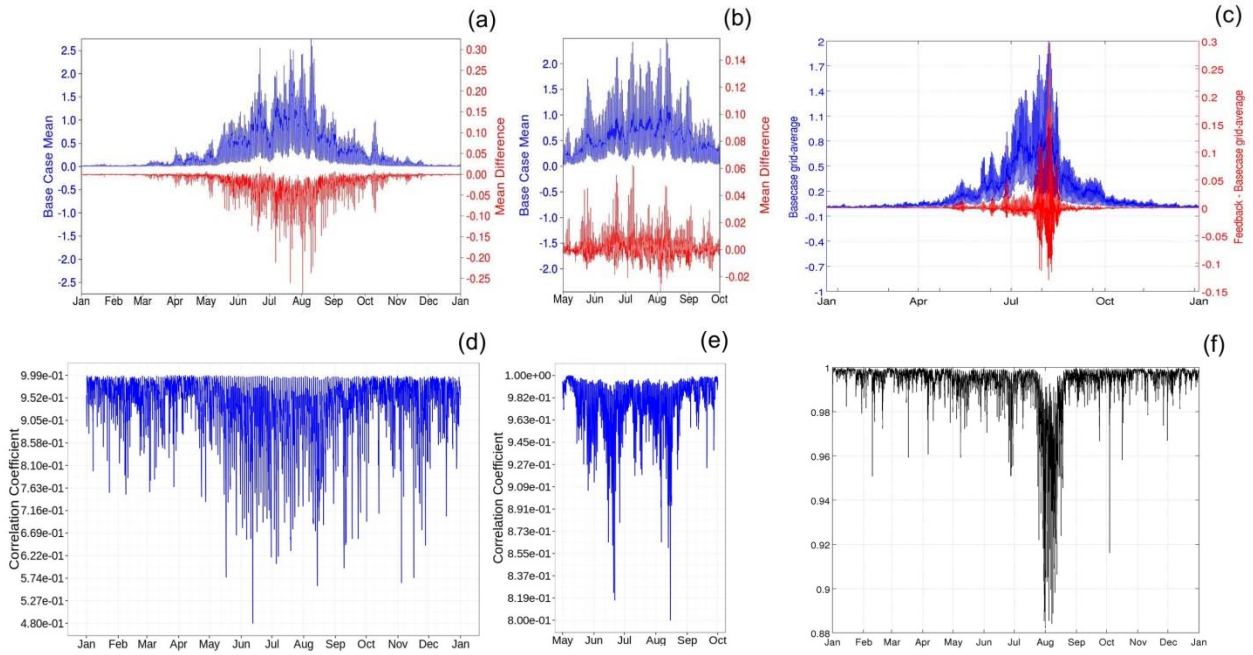

Figure 5. As for Figure 3, Isoprene.

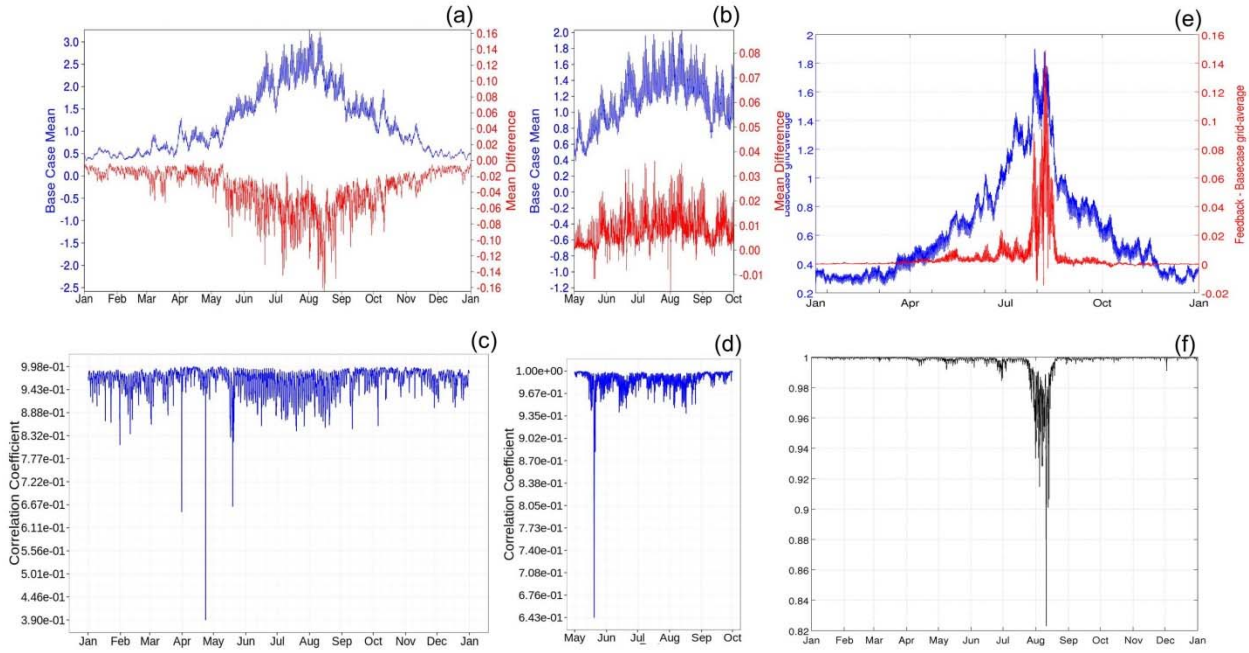

Figure 6. As for Figure 3, HCHO.

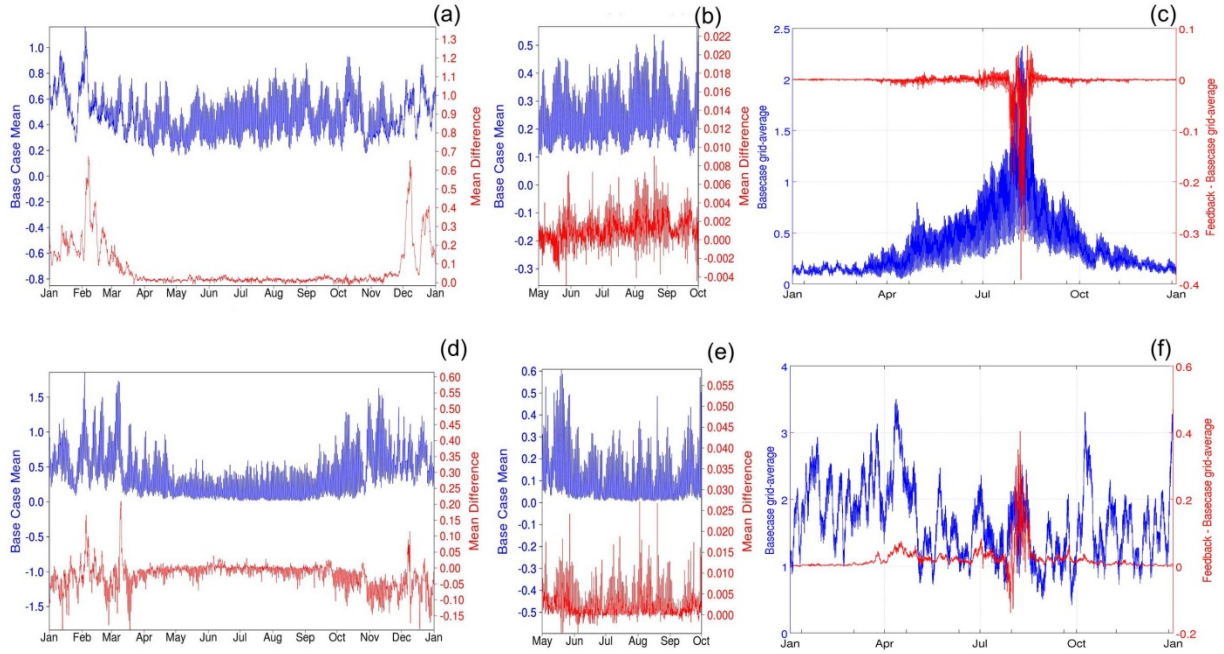

Figure 7. Non-feedback mean HNO_3 (blue) and mean differences (red) for NA/GEM-MACH(a), NA/WRF-CMAQ(b), EU/WRF-CHEM(c), followed by non-feedback mean $PM_{2.5}$ NO_3 (blue) and mean differences (red) for the same three models.

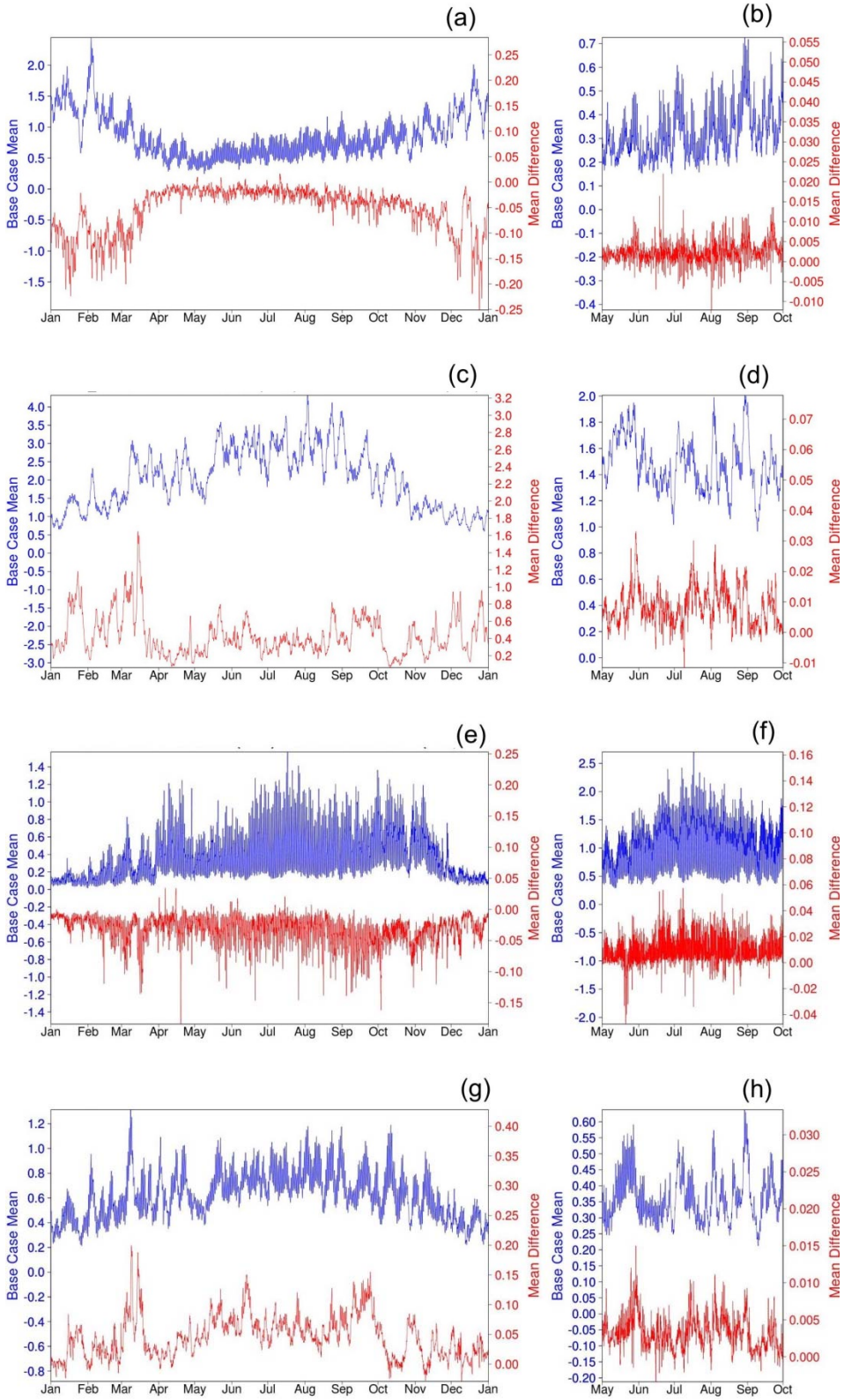

Figure 8. GEM-MACH (left column) and WRF-CMAQ (right column) non-feedback grid mean values (blue) and mean differences (red) for SO_2 (a,b), $PM_{2.5}$ SO_4 (c,d), NH_3 (e,f) and $PM_{2.5}$ NH_4 (g,h).

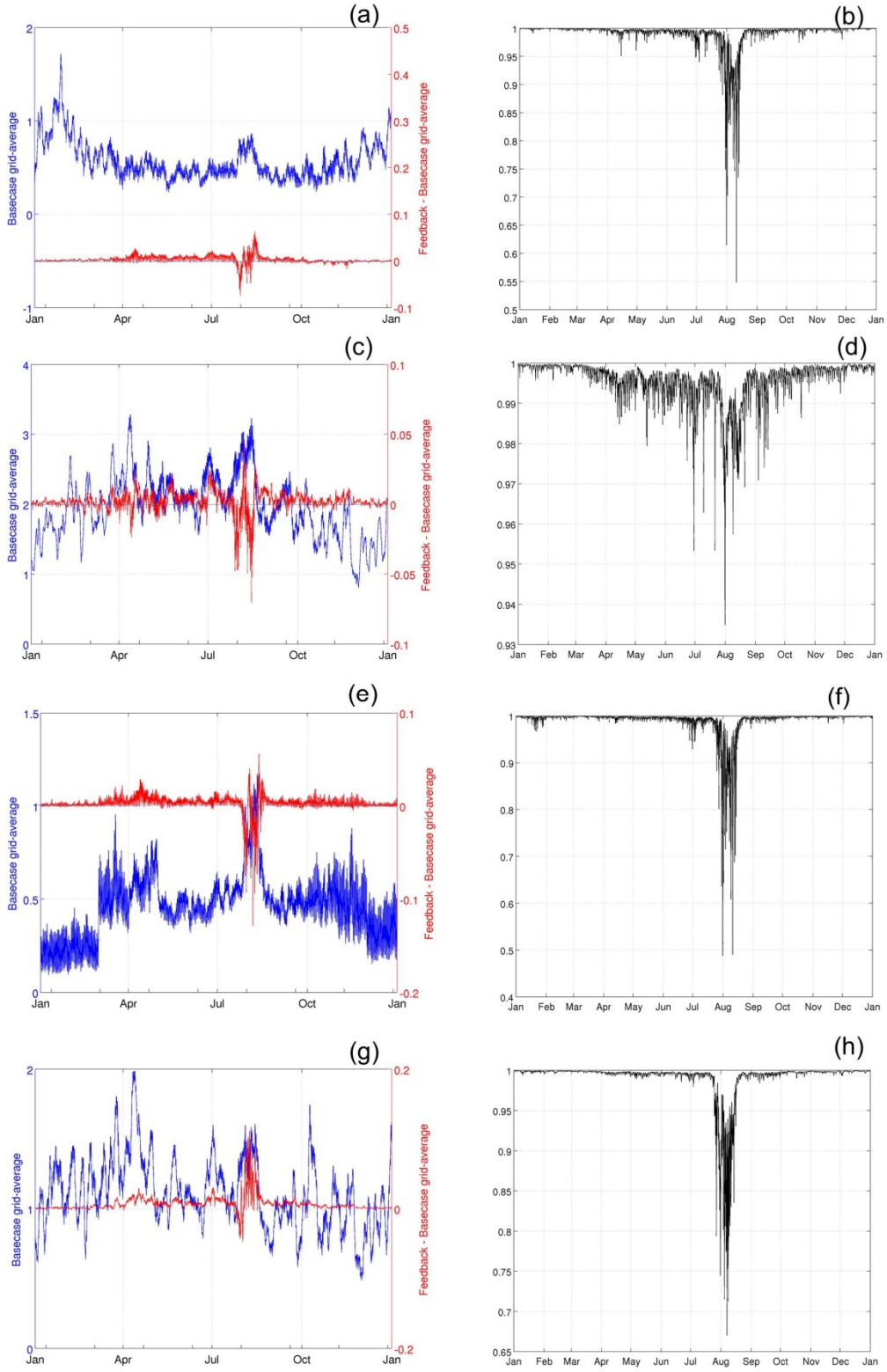

Figure 9. (a,b): SO_2 domain average concentrations, domain average concentration differences with direct effect feedback, and correlation coefficients, AQMEII-2 EU domain, 2010. (c,d): $PM_{2.5}$ SO_4. (e,f): NH_3. (g,h): $PM_{2.5}$ NH_4.

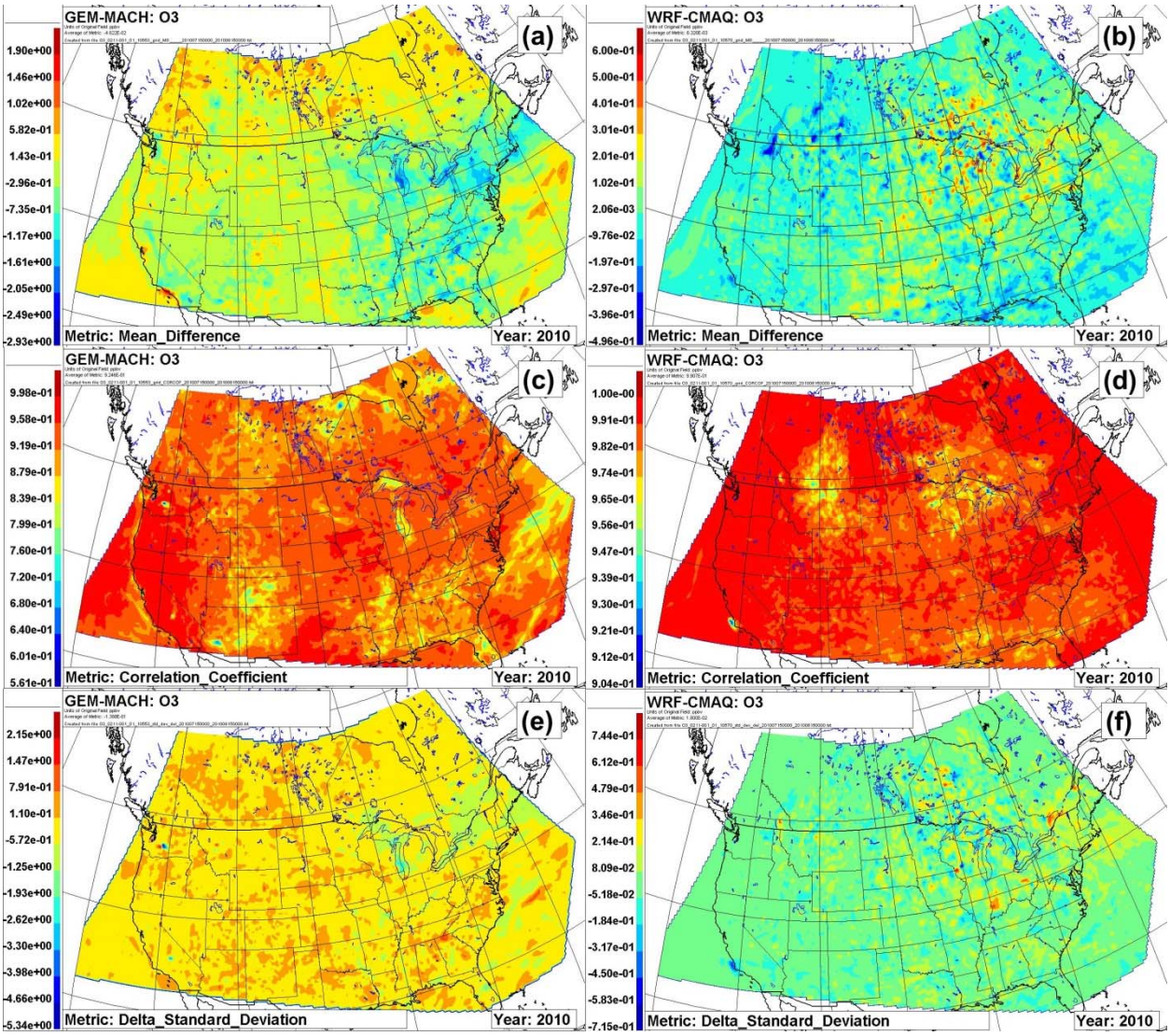

Figure 10. Comparison between O_3 feedback and no-feedback simulations for GEM-MACH (a,c,e) and WRF-CMAQ (b,d,f) , AQMEII-2 NA domain, July 15th to August 15th, 2010. (a,b): Mean differences from no-feedback simulations. (c,d): Correlation coefficients between feedback and no-feedback simulations. (e,f): Changes in standard deviation (feedback s – no-feedback). Note that the scales differ between the panels depicting the two model simulations.

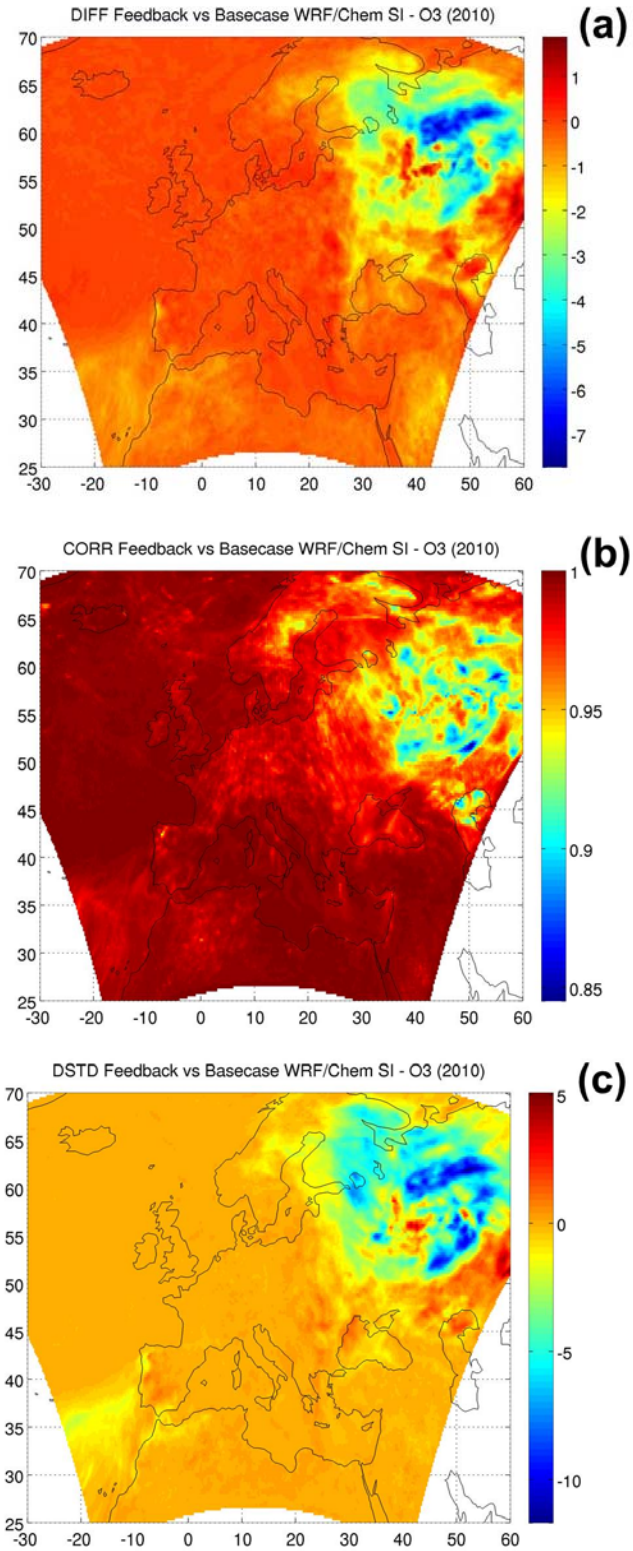

Figure 11. Comparison between WRF-CHEM direct effect feedback and no feedback O_3 simulations for the AQMEII-2 EU domain, July 25th to August 19th. (a) Mean differences, (b) Correlation coefficients, (c) changes in standard deviation (feedback – no-feedback). Compare scales to those in Figure 10.

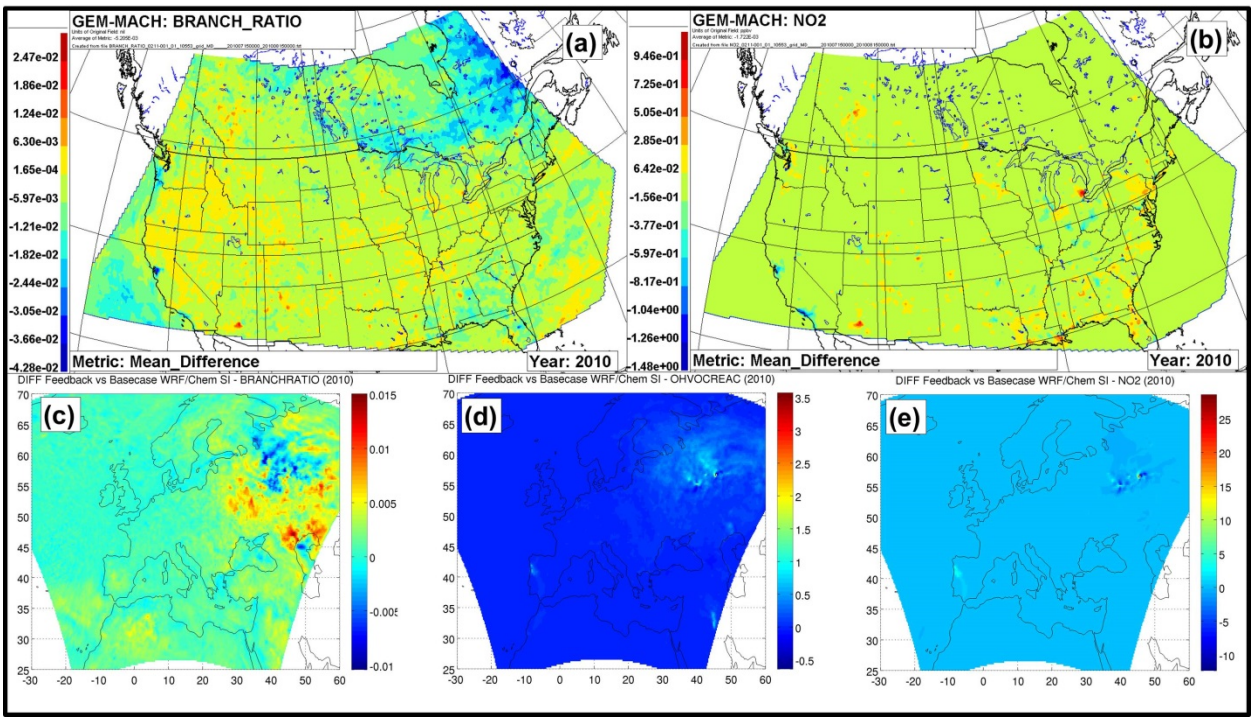

Figure 12. Analysis of O_3 changes, NA and EU. (a,b): NA mean differences in branching ratio and NO_2 concentrations. (c,d,e) EU changes in branching ratio, VOC reactivity and NO_2 concentration.

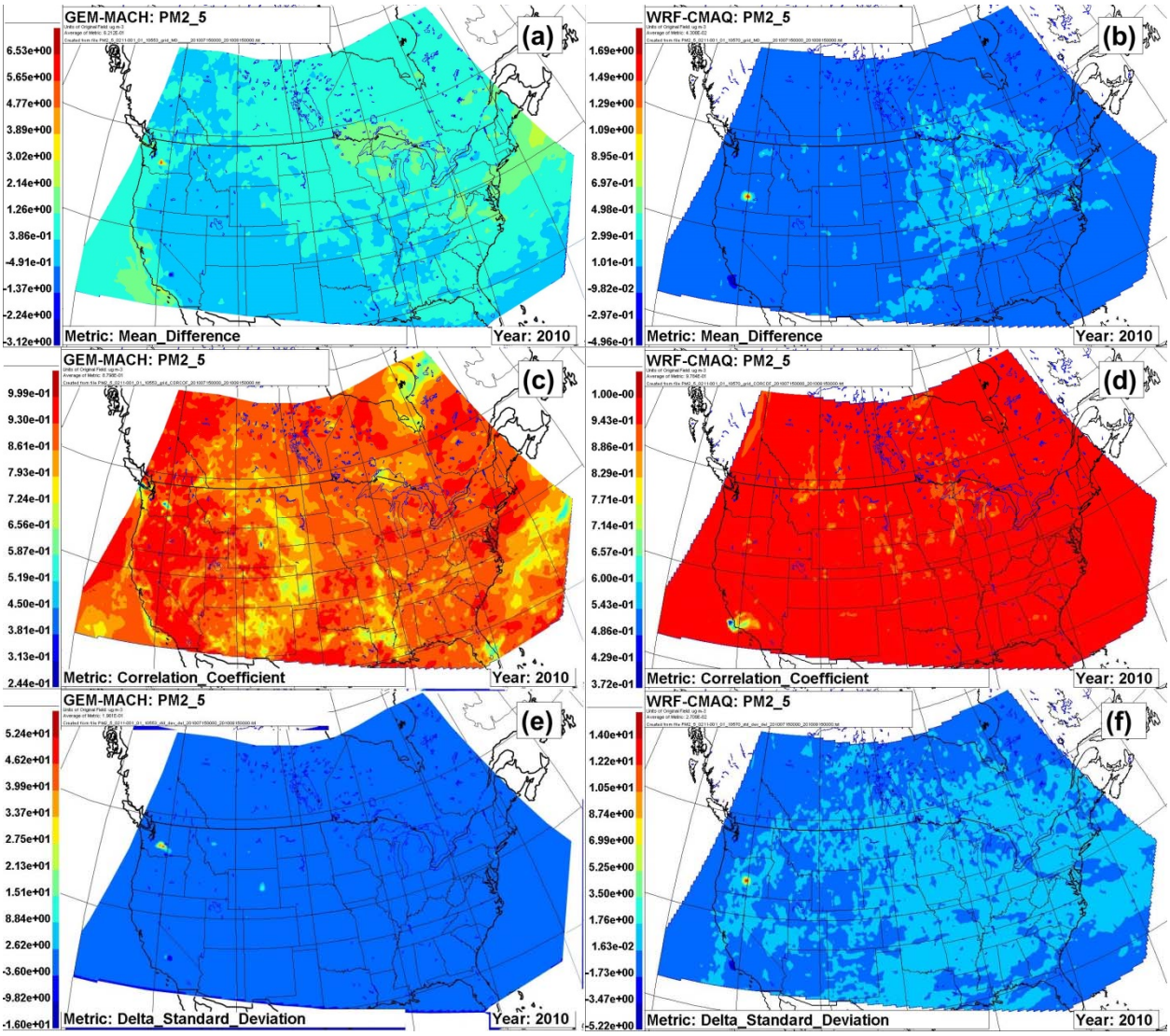

Figure 13. As for Figure 10, $PM_{2.5}$.

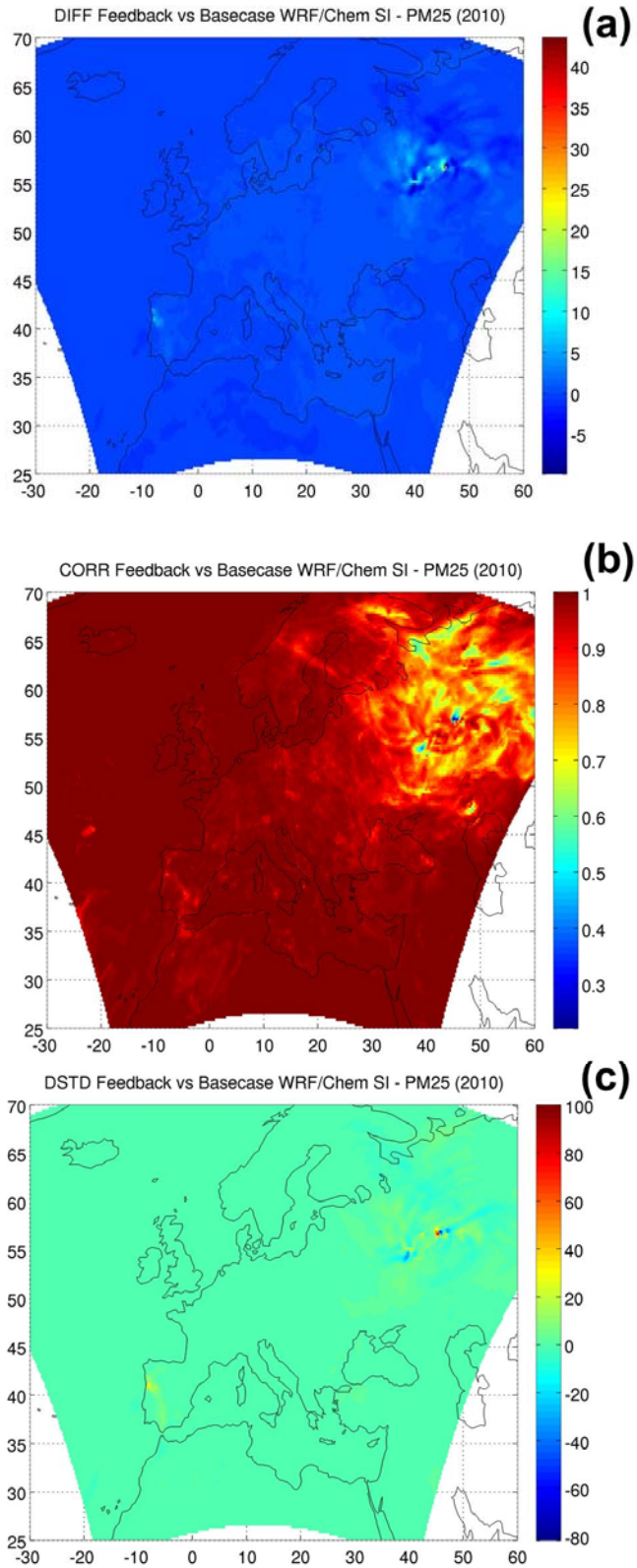

Figure 14. As for Figure 11, $PM_{2.5}$.

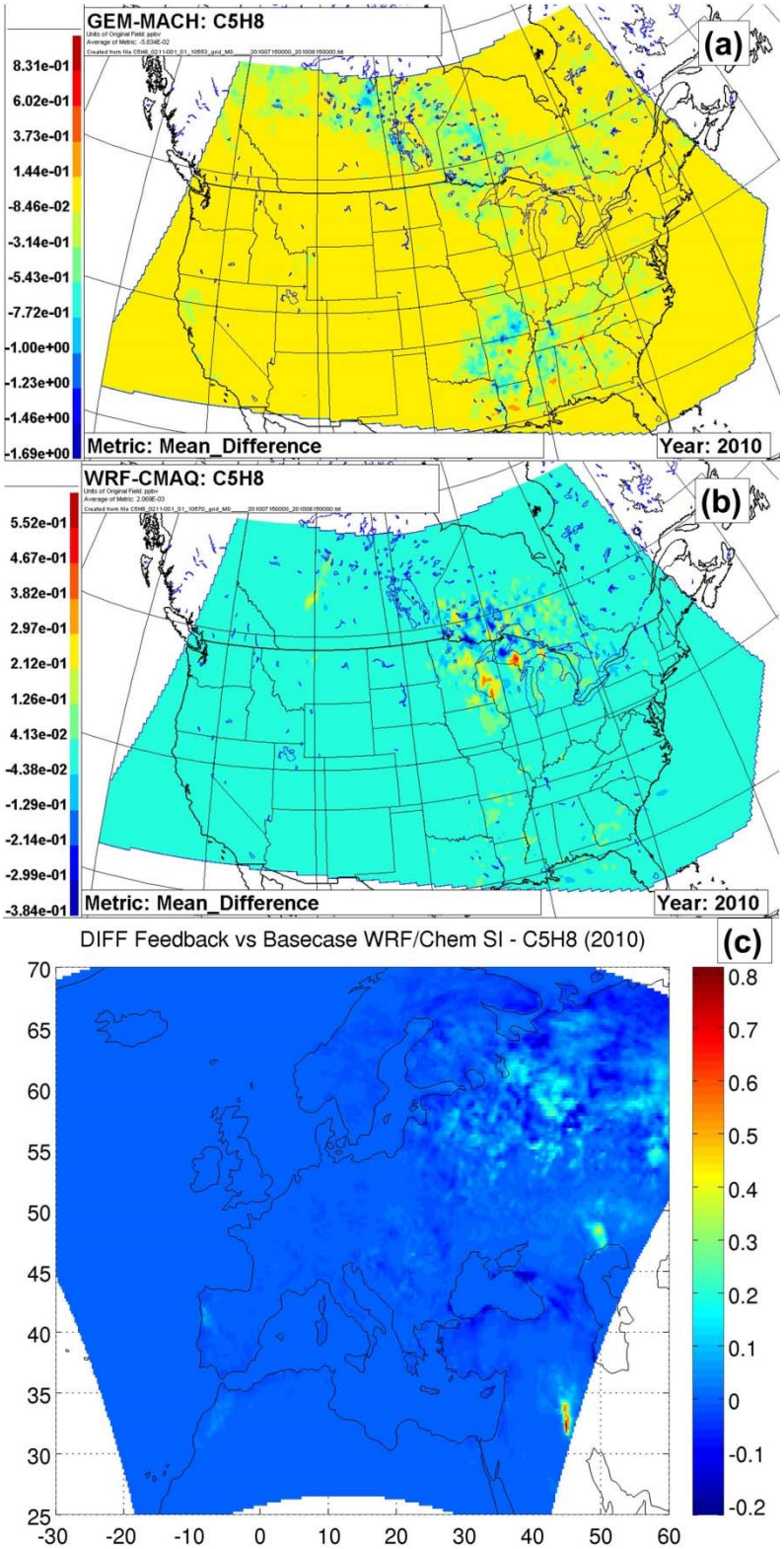

Figure 15. Mean isoprene differences, summer analysis periods. (a) GEM-MACH (direct+indirect effect), (b) WRF-CMAQ, (direct effect) , (c) WRF-CHEM/EU (direct effect).