Response to the Referees + Marked-up Manuscript

Response to the Comments by Referees on Manuscript gmd-2014-212 titled "Forecast error covariance structure in coupled atmosphere-chemistry data assimilation"

1. Response to the Comments by Referees #1:

General comments:

The authors examined the structure of an ensemble-based coupled atmosphere-chemistry forecast error covariance using WRF model coupled with Chemistry (WRF-Chem), a coupled atmospherechemistry model, was used to create an ensemble error covariance. The control variable vector includes both the dynamical and chemistry model variables. A synthetic single observation experiment was designed in order to evaluate the cross-variable components of a coupled error covariance.

It appears to this reviewer that a very short Appendix describing WRF-Chem latest version might contribute to the paper being self contained.

Otherwise the results are impressive and indicate that the coupled error covariance has important cross-variable components allowing a physically meaningful adjustment of all control variables. An additional benefit of the coupled error covariance is that a cross-component impact is allowed, e.g., atmospheric observations can exert impact on chemistry analysis, and vice-versa.

It remains to see if the results carry to realistic cases.

I recommend publication with very minor revision.

⇒ We do appreciate the positive comments by Referee #1. The referee fully understands the contents of our research. We agree with the referee's suggestion to include a description of the WRF-Chem. Following the referee's suggestion, we have included a short Appendix to describe the WRF-Chem in our revised manuscript. We recently applied this technique to a real data case, and will submit our results in a separate manuscript soon.

2. Response to the Comments by Referees #2:

General comments:

The authors examine an ensemble based coupled atmposphere-chemistry forecast error covariance matrix in the context of WRF-Chem model. A synthetic single observation test is carried out and the results are discussed in terms of cross-variable components of a coupled error covariance. The procedure is clear and well-defined as well as the structure of the paper. There is no major complains regarding the grammar. I would recommend to accept this paper but I will address the next questions to the authors:

1. *How well your idea scales regarding the dimension of the model? Observations?*
2. *What happen in the context of realistic scenarios? For your reference, here I cite two:*
   * *SPEEDY Model: http://www.ictp.it/research/esp/models/speedy.aspx*
   * *QGCM Model: http://www.q-gcm.org/*

⇒ We appreciate the positive and constructive comments by Referee #2 (Dr. E. Nino). The referee seems to fully understand the contents of our research. An item-by-item reply to the referee's questions is provided below:

1. *How well your idea scales regarding the dimension of the model? Observations?*

   ⇒ In principle, there is nothing in the described methodology that would prevent its use with larger dimensions of model or observations. Please note that the actual experiment specifications in our paper include $132 \times 147 \times 28 \approx 0.5 \times 10^6$ grid points and 11 control variables (10 of which are 3-dimensional), making the actual state vector dimension about 0.5×10^7, and thus the error covariance being a 0.5×10^7 by 0.5×10^7 matrix. We address high dimensionality of state by using the ensemble-based square-root covariance matrix, which has dimensions $0.5 \times 10^7 \times 32$ (for 32 ensemble columns), which can be processed column-by-column if the computer memory is restrictive. Although the high dimensional observations were not used in this manuscript due to our aim to describe the structure of the error covariance, this issue is generally addressed within the ensemble data assimilation algorithm. In our case, this is the Maximum Likelihood Ensemble Filter (MLEF – Zupanski 2005; Zupanski et al. 2008). As in similar ensemble filters without the perturbed observations (e.g., square-root ensemble filters), the high dimensional observations are processed in the low-dimensional local ensemble subspace. The practical advantage is that the matrix inversion is done in ensemble space, with dense and well-conditioned matrices. Therefore, the anticipated scaling of computing due to dimension change is likely sub-linear. For the error covariance the scaling is probably linear since an increase of the column dimension implies a proportionally longer I/O, and the required matrix-vector product has also a proportionally more term to calculate. For observation processing, although there is a linear scaling of the observation vectors and the involved calculations, the cost is ultimately governed by the ensemble dimension, implying a negligible cost increase due to increased observation dimensions.

2. *What happen in the context of realistic scenarios?*

   ⇒ Please note that the WRF-CHEM model is itself a very complex coupled model, with complex chemical and atmospheric interactions. We believe that the use of ensemble-based error covariance is a huge advantage for describing complex and typically unknown correlations in a coupled system. The only required input is a set of (nonlinear) ensemble forecasts used to construct the square-root forecast error covariance, which automatically produces the most complex correlations. Although this is a flow-dependent error covariance, its relevance is quite important even if the processes are changing slowly with time, because the structure of cross-component covariance brings a wealth of dynamically-based correlations as represented by the employed model. The only important restriction is that a covariance localization is implemented due to a low-dimensional ensemble space, as is in our system. Therefore, using an atmosphere-ocean model (such as the mentioned QGCM), or the SPEEDY model, would automatically reveal all the relevant structure of correlations in such a modeling system.

Manuscript prepared for Geosci. Model Dev.
with version 2014/07/29 7.12 Copernicus papers of the LaTeX class copernicus.cls.
Date: 22 March 2015

~~Forecast~~ Structure of forecast error covariance ~~structure~~ in coupled atmosphere-chemistry data assimilation

Seon Ki Park[1,2,3,4], Sujeong Lim[2,3], and Milija Zupanski[5]

[1]Department of Environmental Science and Engineering, Ewha Womans University, Seoul, Republic of Korea
[2]Department of Atmospheric Science and Engineering, Ewha Womans University, Seoul, Republic of Korea
[3]Center for Climate/Environment Change Prediction Research, Ewha Womans University, Seoul, Republic of Korea
[4]Severe Storm Research Center, Ewha Womans University, Seoul, Republic of Korea
[5]Cooperative Institute for Research in the Atmosphere, Colorado State University, Fort Collins, Colorado, USA

Correspondence to: Seon Ki Park (spark@ewha.ac.kr)

Abstract. In this study, we examined the structure of an ensemble-based coupled atmosphere-chemistry forecast error covariance. The Weather Research and Forecasting (WRF) model coupled with Chemistry (WRF-Chem), a coupled atmosphere-chemistry model, was used to create an ensemble error covariance. The control variable includes both the dynamical and chemistry model variables. A synthetic single observation experiment was designed in order to evaluate the cross-variable components of a coupled error covariance. The results indicate that the coupled error covariance has important cross-variable components that allow a physically meaningful adjustment of all control variables. The additional benefit of the coupled error covariance is that a cross-component impact is allowed, e.g., atmospheric observations can exert impact on chemistry analysis, and vice versa. Given the realistic structure of ensemble forecast error covariance produced by the WRF-Chem, we anticipate the ensemble-based coupled atmosphere-chemistry data assimilation will respond similarly to assimilation of real observations.

1 Introduction

The regional air quality is affected by synoptic weather situations or air masses with special chemical properties (Grell et al., 2000). In prediction of air quality, the coupled physical and chemical processes are essential, which include transport, deposition, emission, chemical transformation, aerosol interactions, photolysis, and radiation (Grell et al., 2005). Optimized initial conditions for a numerical model, including such coupled processes, can be obtained by data assimilation (DA;

e.g., Houtekamer and Mitchell, 1998; Eibern and Schmidt, 1999; Wang et al., 2001; Evensen, 2003; Park and Zupanski, 2003; Navon, 2009; Zupanski, 2009). Therefore, DA for an air quality prediction system could be approached as a coupled atmosphere-chemistry DA, with interaction between atmospheric and chemistry components. In typical data assimilation methodologies, such as variational and ensemble, the interaction between different variables is achieved by forecast error covariance, in particular its cross-variable components. Therefore, it is of fundamental interest for the development of atmosphere-chemistry DA to investigate the coupled forecast error covariance. Here, we investigate the structure of the atmosphere-chemistry forecast error covariance using ensemble forecasting, which corresponds to the prediction step of an ensemble data assimilation algorithm (e.g., Zupanski, 2005, 2009).

2 Methodology and Synoptic Case

In this research, we use the Weather Research and Forecasting (WRF) model coupled with Chemistry (WRF-Chem) as a prediction model (Grell et al., 2005). The chosen chemistry option is the Carbon-Bond Mechanism version Z (CBMZ), which simulates the emission, transport, mixing, and chemical transformation of trace gases and aerosols simultaneously with meteorology and investigates the regional-scale air quality. More details on the WRF-Chem and corresponding options used in this study are described in Appendix A.

We chose a synoptic case on 03 September 2005 related to Typhoon Nabi (2005), characterized by an increased impact on the Korean Peninsula. The experiment begins at 0000 UTC and ends at 0600 UTC on 03 September 2005. The WRF-Chem is set up with a horizontal resolution of 30 km and 28 vertical levels. Model domain is centered over the Korean Peninsula, covering an area of approximately 3900 km x 4400 km with 132 x 147 horizontal grid points.

The ensemble forecast includes 32 ensemble members with a 6-hour assimilation window. The lateral boundary conditions are provided by the National Center for Environmental Prediction (NCEP) Global Forecasting System (GFS). The control variables defined in DA (i.e., variables adjusted during DA) are the WRF-Chem prognostic variables that include dynamical variables such as winds, perturbation potential temperature, perturbation geopotential, water vapor mixing ratio and perturbation dry air mass in column, and the chemical variables such as ozone (O_3), nitrates (NO, NO_2, NO_3) and sulfur dioxide (SO_2) as well.

3 Experimental Design

A common approach to investigating forecast error covariance in data assimilation is to conduct a single observation experiment (Thepaut et al., 1996; Whitaker et al., 2009; Buehner et al., 2010), in which only one observation is assimilated using the full DA system. The analysis increments (i.e., analysis minus guess) from such an experiment show how the observation information is distributed spatially and among different analysis variables (e.g., Buehner, 2005). However, in order to investigate the structure of a coupled forecast error covariance before real observations are available and even before the full DA algorithm is developed, one can consider the assimilation of a single synthetic observation located at a chosen model grid point. In particular, we define the synthetic observation as

$$y_{synth} = x^f + \sigma_o \tag{1}$$

where x^f is the forecast and σ_o is the observation error standard deviation. Following *Thepaut et al.* [1996, Eq. (3)], with some modifications and using (1), the analysis increment in a single synthetic observation experiment is

$$x^a - x^f = \mathbf{P}_f \left(\frac{\sigma_o}{\sigma_f^2 + \sigma_o^2} \right)_{ijk} \tag{2}$$

where x^a is the analysis, σ_f is the forecast error standard deviation, and the subscript ijk defines the grid location of the pseudo-observation point. Equation (2) indicates that analysis increment represents the ijk-th column of the forecast error covariance scaled by standard deviations of observation error and forecast error. In our experiments the forecast error covariance is ensemble-based, as defined in Zupanski (2005) as:

$$\mathbf{P}_f = \mathbf{P}_f^{1/2} (\mathbf{P}_f^{1/2})^T, \qquad \mathbf{P}_f^{1/2} = \left(p_1^f \cdots p_N^f \right), \qquad p_n^f = m(x_0^n) - m(x_0) \tag{3}$$

where the superscript T denotes the transpose, the index n refers to ensemble member, N is the total number of ensemble forecasts, m represents the nonlinear WRF-Chem model, and the subscript 0 denotes the initial time of the forecast with corresponding initial conditions x_0 and ensemble initial conditions x_0^n. In this experiment, the control initial conditions are obtained by interpolation from the NCEP GFS model, while the initial ensemble perturbations are created using the lagged forecast outputs.

Since we are interested in the coupled atmosphere-chemistry forecast error covariance, we design two experiments with: (i) synthetic temperature observation at 250 hPa located at a grid point near (132E, 23N), on the northwest side of the typhoon, and (ii) synthetic ozone observation at 250 hPa located at a grid point near the eye of the typhoon (134E, 21N).

4 Results

We show the impact of single synthetic temperature (T) and ozone (O_3) observations in terms of the analysis increments $x^a - x^f$ impacting all control variables. As mentioned earlier, our main interest is to examine the cross-variable covariance structure between atmospheric and chemistry variables, since the cross-variable analysis impact is possible only because of the multivariate structure of the coupled ensemble forecast error covariance.

In Fig. 1 we show the impact of synthetic T observation at 250 hPa on the analysis increments of T, O_3, nitrogen-dioxide (NO_2), and sulfur dioxide (SO_2). The analysis increment of T at 250 hPa (e.g., at the same level of synthetic T observation) shows a typical response with nearly circular isolines with the maximum of 0.4 K at the observation location (Fig. 1a). The analysis increments of O_3, NO_2, and SO_2 are also shown in vertical cross-sections. One can see that O_3 (Fig. 1b) and NO_2 (Fig. 1c) analyses have the largest change at the level of single T observation, while the SO_2 analysis (Fig. 1d) is mostly impacted near 700 hPa (approximately σ-level 13). This is likely a consequence of the vertical structure of O_3 and NO_2 with the largest values in the upper troposphere and the stratosphere, while SO_2 has typically the largest values in the lower troposphere (e.g., Meena et al., 2006). The strongest impact of T observation is on O_3, with the magnitude up to 0.001 ppmv, while the magnitude is somewhat smaller for NO_2 and SO_2. One can also infer that an increase of T will imply a decrease of O_3, NO_2, and SO_2. Probably the most important implication of these results is that observations of an atmospheric variable (e.g., temperature) can change the analysis of chemical variables in a physically meaningful way. This means that even with no chemistry observations in the local area, the analysis of chemical variables can still be adjusted in agreement with standard dynamical variables of the model. On the other hand, if there are chemistry observations in the area, the chemistry analysis change introduced by atmospheric observations will act as an additional dynamical constraint to the final analysis.

In Fig. 2 the impact of O_3 single observation at 250 hPa on itself and the other variables is shown. As before, we focus on the vertical cross-section of the analysis response. The impact of O_3 observation on its own analysis shows the anticipated response with the largest magnitude at observation location, approximately 0.02 ppmv (Fig. 2a). Although smaller in magnitude, the analysis increments of NO_2 (Fig. 2b) and SO_2 (Fig. 2c) show the vertical structure with maxima in the upper and lower troposphere, respectively. It is also notable that an increase of O_3 brings about an increase of NO_2 and SO_2, confirming the direct relationship between these variables as noticed in Fig. 1. The T analysis increment indicates that there is a cooling at the level of O_3 observation, while there is a warming above and below (Fig. 2d).

The results shown in Figs. 1 and 2 indirectly confirm that the improved stratospheric ozone distribution by DA can make a better representation of stratospheric winds, temperature and other constituents (e.g., Lahoz et al., 2007).

5 Conclusions

The structure of an ensemble-based coupled atmosphere-chemistry forecast error covariance was examined in the context of the WRF-Chem model. A synthetic single observation experiment was designed in order to evaluate the cross-variable components of the coupled error covariance. Our results indicate that the coupled error covariance has important cross-variable components that allow a physically meaningful adjustment of all control variables, and a much wider impact of observations (e.g., atmospheric observation on chemistry analysis, and vice versa). The analysis increments created in response to synthetic temperature and ozone observations illustrate the complexity of atmosphere-chemistry cross-correlations and the forecast error covariance structure. Given the realistic structure of ensemble forecast error covariance produced by the WRF-Chem, we anticipate the ensemble-based coupled atmosphere-chemistry data assimilation will respond similarly to assimilation of real observations. Therefore, our next step is to apply the WRF-Chem with an ensemble-based data assimilation algorithm (e.g., the maximum likelihood ensemble filter (MLEF); Zupanski, 2005) to assimilation of real chemical and atmospheric observations.

Appendix A: Description on the WRF-Chem

In this research, we use the Weather Research and Forecasting (WRF) model coupled with Chemistry (WRF-Chem) version 3.4.1 as a prediction model in a regional-scale. As a coupled model, it simulates the emission, transport, mixing and chemical transformation of trace gases and aerosols simultaneously with meteorology using the governing equations with mass and scalar conserving flux form and the terrain-following mass vertical coordinate system (Grell et al., 2005; Fast et al., 2006) . Therefore, it uses the same transport scheme, horizontal and vertical coordinates, and physics schemes with the same time step (Grell et al., 2005; Fast et al., 2006) . Figure A1 represents the flow chart of the WRF-Chem model. It is made up of the WRF Pre-processing System (WPS), the WRF-Chem model, and the visualization processes. The WPS creates the meteorology data with the terrestrial data and the meteorology initial conditions (ICs) and boundary conditions (BCs) which are provided by the National Center for Environmental Prediction (NCEP) Global Forecasting System (GFS) producing the global latitude/longitude 1 degree resolution and terrestrial data. In the WRF-Chem model, the chemical ICs and BCs are automatically obtained by the climatology. Furthermore, it simulates the evolution of chemical species with the prognostic variables that include both dynamical and chemical variables. For the ensemble experiments, the initial ensemble perturbations are created using the 12-hour lagged forecast outputs. Finally, the Advanced Research WRF post-processing (ARWpost) along with Grid Analysis and Display System (GrADS) is used for the visualization process.

We also discuss various physical and chemical processes employed in the WRF-Chem model in more detail. Table A1 summarizes the WRF-Chem configuration options what are used in this study. To evaluate the cross-variable component of forecast error covariance, we select the simplified dynamics rather than sophisticated physical processes. Regarding the atmospheric processes, we use the recommended physics options for the regional climate case at 30 km grid size in our experiments. As the chemical options, the Carbon Bond Mechanism version Z (CBM-Z) without Dimethylsulfide scheme is used for the gas-phase chemistry. The CBM-Z photochemical mechanism contains 55 prognostic species and 134 reactions having the lumped structure approach for condensing organic chemical species and reactions (Fast et al., 2006) . It also uses a regime dependent approach based on the partitioned kinetics, such as background, anthropogenic, and biogenic submechanisms for saving the computational time (Fast et al., 2006) . Furthermore, we consider the chemical tendency diagnostic for equation budget analysis. However, we did not consider the convective parameterization which can simulate the subgrid convective transport, wet scavenging, and aqueous chemistry due to simple experiment setting, even with a typhoon case.

Acknowledgements. This work is supported by the Korea Environmental Industry & Technology Institute through the Eco Innovation Program (ARQ201204015), and partly by the National Research Foundation of Korea grant (No. 2009-0083527) funded by the Korean government (MSIP). The third acknowledges a partial support from the National Science Foundation Collaboration in Mathematical Geosciences Grant 0930265 and from the NASA Modeling, Analysis and Prediction (MAP) Program Grant NNX10AG92G.

References

Buehner, M.: Ensemble-derived stationary and flow-dependent background-error covariances, Q. J. R. Meteorol. Soc., 131, 1013–1043, 2005.

Buehner, M., Houtekamer, P. L., Charette, C., Mitchell, H. L., and He, B.: Intercomparison of variational data assimilation and the ensemble kalman filter for global deterministic NWP. Part I: Description and single-observation experiments, Mon. Wea. Rev., 138, 1567–1586, 2010.

Eibern, H., and Schmidt, H.: A four-dimensional variational chemistry data assimilation scheme for Eulerian chemistry transport modeling, J. Geophys. Res., 104(D15), 18583–18598, 1999.

Evensen, G.: The ensemble Kalman filter: theoretical formulation and practical implementation, Ocean Dyn., 53, 343–367, 2003.

Fast, J. D., Gustafson Jr., W. I., Easter R. C., Zaveri R. A., Barnard J. C., Chapman E. G., Grell G. A., and Peckham S. E.: Evolution of ozone, particulates, and aerosol direct radiative forcing in the vicinity of Houston using a fully coupled meteorology-chemistry-aerosol model, J. Geophys. Res., 111, D21305, doi:10.1029/2005JD006721, 2006.

Grell, G. A., Emeis, S., Stockwell, W. R., Schoenemeyer, T., Forkel, R., Michalakes, J., Knoche, R., and Seidl, W.: Application of a multiscale, coupled MM5/chemistry model to the complex terrain of the VTALP valley campaign, Atmos. Environ., 34, 1435–1453, 2000.

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

Houtekamer, P. L., and Mitchell, H. L.: Data assimilation using an ensemble Kalman filter technique, Mon. Wea. Rev., 126, 796–811, 1998.

Lahoz, W. A., Errera, Q., Swinbank, R., and Fonteyn, D.: Data assimilation of stratospheric constituents: a review, Atmos. Chem. Phys., 7, 5745–5773, 2007.

Meena, G. S., Bhosale, C. S., and Jadhav, D. B.: Retrieval of stratospheric O_3 and NO_2 vertical profiles using zenith scattered light observations, J. Earth Syst. Sci., 115(3), 333–347, 2006.

Navon, I. M.: Data assimilation for numerical weather prediction: A review, in: Data Assimilation for Atmospheric, Oceanic and Hydrologic Applications, Park, S. K. and Xu, L., Eds., Springer, Berlin-Heidelberg, 21–65, 2009.

Park, S. K., and Zupanski, D.: Four-dimensional variational data assimilation for mesoscale and storm-scale applications, Meteorol. Atmos. Phys., 82, 173–208, 2003.

Thepaut, J.-N., Courtier, P., Belaud, G., and Lemaitre, G.: Dynamical structure functions in a four-dimensional variational assimilation: A case study, Q. J. R. Meteorol. Soc., 122, 535–561, 1996.

Wang, K.-Y., Lary, D. J., Shallcross, D. E., Hall, S. M., and Pyle, J. A.: A review on the use of the adjoint method in four-dimensional atmospheric-chemistry data assimilation, Q. J. R. Meteorol. Soc., 127, 2181–2204, 2001.

Whitaker, J. S., Compo, G. P., and Thepaut, J.-N.: A comparison of variational and ensemble-based data assimilation systems for reanalysis of sparse observations, Mon. Wea. Rev., 137, 1991–1999, 2009.

Zupanski, M.: Maximum likelihood ensemble filter: Theoretical aspects, Mon. Wea. Rev., 133, 1710–1726, 2005.

Zupanski, M.: Theoretical and practical issues of ensemble data assimilation in weather and climate, in: Data Assimilation for Atmospheric, Oceanic and Hydrologic Applications, Park, S. K. and Xu, L., Eds., Springer, Berlin-Heidelberg, 67–84, 2009.

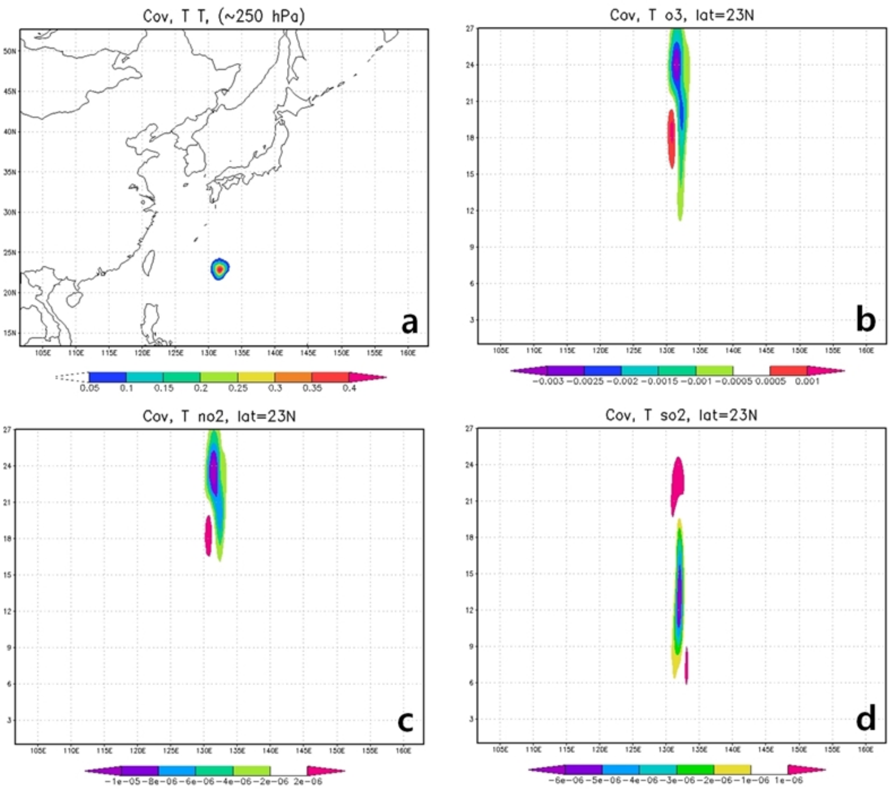

Figure 1. The analysis increments $(x^a - x^f)$ in response to a single T observation at 250 hPa (near σ-level 24): (a) horizontal response of T at 250 hPa, and vertical responses of (b) O_3, (c) NO_2 and (d) SO_2. In (b)-(d), the vertical axis represents the vertical σ-levels. Units are ppmv for chemical variables and K for temperature.

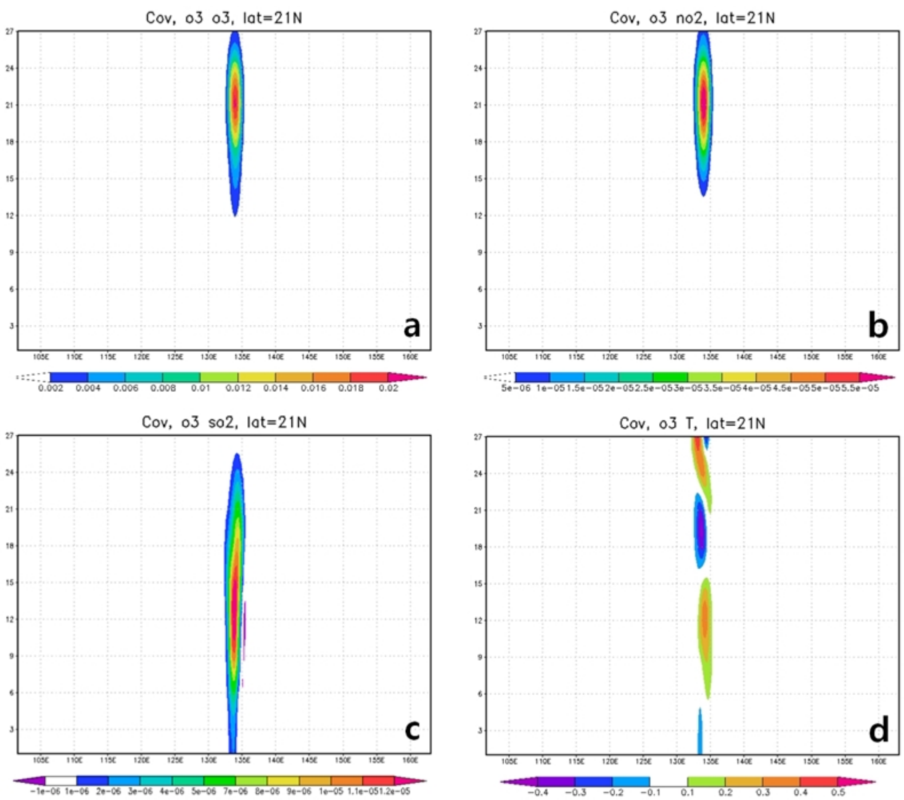

Figure 2. Same as in Fig. 1 but for vertical cross-section of the analysis increments $(x^a - x^f)$ in response to a single O_3 observation at 250 hPa for (a) O_3, (b) NO_2, (c) SO_2 and (d) T.

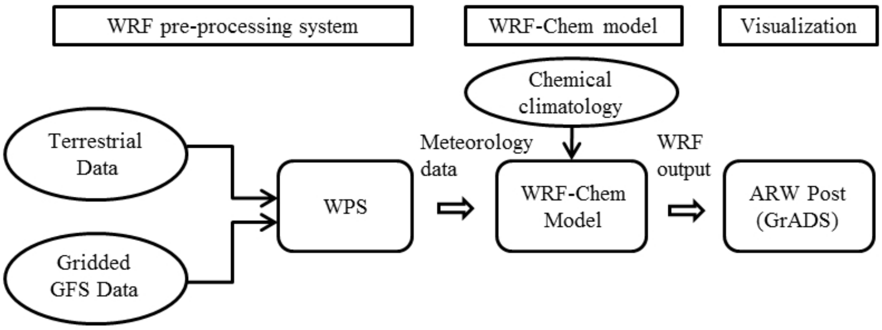

Figure A1. Flowchart of the WRF-Chem model.

Table A1. Selected WRF-Chem configuration options.

Atmospheric Process	WRF-Chem Option
Microphysics	WSM 6-class graupel
Longwave radiation	CAM
Shortwave radiation	CAM
Surface layer	Revised MM5 Monin-Obukhov
Land surface	Unified Noah LSM
Planetary boundary layer	YSU
Cumulus parameterization	Kain-Fritsch (new Eta)
Gas-phase chemistry	CBM-Z