CrossMark

Impacts of using an ensemble Kalman filter on air quality simulations along the California–Mexico border region during Cal–Mex 2010 Field Campaign

Naifang Bei [a,b,*], Guohui Li [b,c], Zhiyong Meng [d], Yonghui Weng [e], Miguel Zavala [b], L.T. Molina [b,f,**]

[a] School of Human Settlements and Civil Engineering, Xi'an Jiaotong University, Xi'an, China
[b] Molina Center for Energy and the Environment, La Jolla, CA, USA
[c] Key Laboratory of Aerosol Science & Technology, SKLLQG, Institute of Earth Environment, Chinese Academy of Sciences, Xi'an, China
[d] School of Physics, Peking University, Beijing, China
[e] Department of Meteorology, The Pennsylvanian State University, University Park, PA 16802, USA
[f] Department of Earth, Atmospheric and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, MA, USA

HIGHLIGHTS

- The forecast initialized with EnKF improved the meteorological simulations.
- EnKF produced more reasonable simulations for nitrate and ammonium aerosols.
- Discrepancies between EnKF and the measurements indicate rooms for improvement in the data assimilation and/or modeling systems.

ARTICLE INFO

ABSTRACT

The purpose of this study is to investigate the impact of using an ensemble Kalman filter (EnKF) on air quality simulations in the California–Mexico border region on two days (May 30 and June 04, 2010) during Cal–Mex 2010. The uncertainties in ozone (O_3) and aerosol simulations in the border area due to the meteorological initial uncertainties were examined through ensemble simulations. The ensemble spread of surface O_3 averaged over the coastal region was less than 10 ppb. The spreads in the nitrate and ammonium aerosols are substantial on both days, mostly caused by the large uncertainties in the surface temperature and humidity simulations. In general, the forecast initialized with the EnKF analysis (EnKF) improved the simulation of meteorological fields to some degree in the border region compared to the reference forecast initialized with NCEP analysis data (FCST) and the simulation with observation nudging (FDDA), which in turn leading to reasonable air quality simulations. The simulated surface O_3 distributions by EnKF were consistently better than FCST and FDDA on both days. EnKF usually produced more reasonable simulations of nitrate and ammonium aerosols compared to the observations, but still have difficulties in improving the simulations of organic and sulfate aerosols. However, discrepancies between the EnKF simulations and the measurements were still considerably large, particularly for sulfate and organic aerosols, indicating that there are still ample rooms for improvement in the present data assimilation and/or the modeling systems.

© 2014 Elsevier B.V. All rights reserved.

1. Introduction

The meteorological condition is one of the key components in photochemical air quality models and its simulation is critical for understanding the formation, transformation, diffusion, transport, and removal of the atmospheric pollutants.

Previous studies on the photochemical sensitivity to meteorological uncertainties generally include Monte Carlo simulations and adjoint sensitivity studies (e. g., Bergin et al., 1999; Dabberdt and Miller, 2000; Hanna et al., 2001; Beekmann and Derognat, 2003; Menut, 2003). Recently, using both meteorological and photochemical ensemble forecast, Zhang et al. (2007a) demonstrated that the predicted ozone (O_3) in Houston and the surrounding areas exhibit large uncertainties due to meteorological initial condition uncertainties. Bei et al. (2010) explored the O_3 predictability due to meteorological

* Correspondence to: N. Bei, School of Human Settlements and Civil Engineering, Xi'an Jiaotong University, Xi'an, China.
** Correspondence to: L. T. Molina, Molina Center for Energy and the Environment, La Jolla, CA, USA.
E-mail addresses: bei.naifang@mail.xjtu.edu.cn (N. Bei), ltmolina@mit.edu (L.T. Molina).

uncertainties in the Mexico City Basin using ensemble forecasts. They have shown that the least predictability in O₃ simulations is attributable to the increasing uncertainties in meteorological fields during the peak O₃ period, and the effects of wind speeds and planetary boundary layer (PBL) height on the O₃ simulations are more straightforward. The uncertainties in O₃ simulations also vary with different PBL schemes and meteorological episodes. Bei et al. (2012a) further investigated the uncertainties in modeling the secondary organic aerosol (SOA) in Mexico City due to meteorological initial uncertainties using the Weather Research and Forecasting model coupled with Chemistry (WRF-CHEM) through ensemble simulations, and showed that uncertainties in meteorological initial conditions substantially influence SOA simulations. These works have demonstrated the importance of accurate representation of meteorological conditions in the air pollution modeling studies.

Producing accurate air quality prediction suitable for decision makers and the public through improving meteorological forecasts is also a major challenge. Dabberdt et al. (2004) have listed the meteorological research needs for improved air quality forecasting, one of which is to yield an optimal representation of the state of the atmosphere using data assimilation tools. Various data assimilation approaches have been developed, such as nudging method, three-dimensional variational method (3DVAR), four-dimensional variational method (4DVAR), and ensemble Kalman filter (EnKF). These methods are promising and have been extensively used in meteorology and oceanography (Navon, 1998; Kalnay, 2003). The EnKF (Evensen, 1994), which estimates the background error covariance with a short-term ensemble forecast, has demonstrated promising performance in the meteorological community (e.g., Snyder and Zhang, 2003; Zhang et al., 2006; Meng and Zhang, 2007, 2011; Zhang et al., 2011a,b). Studies have also shown that EnKF compares favorably with variational methods (Meng and Zhang, 2008a,b; Weng et al., 2011; Zhang et al., 2011a; Meng and Zhang, 2011; Weng and Zhang, 2012).

The above-mentioned data assimilation methods have also been applied to chemical transport models (CTMs) to improve the chemical initial conditions and emission inventories and eventually to produce better air quality simulations or forecasts (e.g., Stuart et al., 2007; Tang et al., 2011; Curier et al., 2012). Most of the previous works regarding improvement of the representation of meteorological fields in air quality studies have primarily focused on observational nudging, which is one of the four-dimensional data assimilation (FDDA) schemes (e.g., Stauffer et al., 1990, 1991; Stauffer and Seaman, 1994; Fast, 1995; Seaman, 2000). Bei et al. (2008) used the 3DVAR data assimilation system to improve the O₃ simulations in the Mexico City basin during the Mexico City Metropolitan Area 2003 field campaign (Molina et al., 2007). They found that the simulated wind circulation, temperature, and humidity fields in the basin with the 3DVAR data assimilation are more consistent with the observations than those without any data assimilation. In addition, Bei et al. (2012a) highlighted that the meteorological ensemble is an efficient method to reduce the meteorological uncertainties in simulations of CTMs, but computationally expensive. The EnKF has been widely used in meteorological simulations and forecasts, but few studies have been performed using the EnKF in air quality modeling studies.

The Cal–Mex 2010 field campaign is a US–Mexico collaborative research project to investigate cross-border transport of emissions in the California–Mexico border region and its impact on regional air quality and climate. A comprehensive data set has been obtained during the campaign from May 15 to June 30, 2010, including highly time-resolved meteorological variables, ambient gas phase species and aerosols. Bei et al. (2013) presented an overview of the meteorological conditions and plume transport patterns during the campaign through the measurements analyses and modeling study. They identified four representative plume transport patterns, namely "plume-southeast", "plume-southwest", "plume-east" and "plume-north". They also pointed out that the WRF model often underestimates the surface

temperature and PBL height during nighttime compared with observations at central supersite Parque Morelos (PQM, Tijuana Municipal System Theme Park in the center of Tijuana). In addition, certain discrepancies also existed between the observed and simulated surface wind directions. These discrepancies may be induced by uncertainties in the model initial conditions as well as the forecast model itself, particularly the PBL parameterization schemes as are commonly found in past air quality modeling studies (e.g., Zhang et al., 2007b; Nielsen-Gammon et al., 2010; Hu et al., 2010a,b).

In the present study, we investigate the impact of assimilating both the routine and additional meteorological observations on the air quality simulations in the California–Mexico border region during Cal–Mex 2010 by using EnKF in the meteorological modeling. In Section 2, we describe the observations and methodologies, and the synoptic overview for the selected two days is given in Section 3. The results are presented in Section 4. Conclusions and discussions are included in Section 5.

## 2. Observations and methodology

2.1. Observations and models

Both routine and additional meteorological observations during the Cal–Mex 2010 campaign were used in this study. The routine observations included sounding observations and wind profilers. The additional meteorological observations during the campaign included the additional sounding at PQM and additional wind profiler from CalNex 2010 conducted in California (Ryerson et al., 2013) (see Fig. 1a).

The Advanced Research WRF (ARW) v3.2 (Skamarock et al., 2008) adopted two one-way nested grids with horizontal resolutions of 12, 2-km and 35 sigma levels in the vertical direction (Fig. 1b). The grid cells used for the two domains were 181 × 151 and 151 × 103, respectively. Two experiments were initialized at 0000 UTC on May 30 and June 04, 2010, respectively, and integrated for 30 h. The National Centers for Environmental Prediction (NCEP) final operational global gridded analysis (FNL) was used to produce the initial and boundary conditions for the WRF model. The physical process parameterization schemes used in simulations included the Grell–Devenyi ensemble scheme for cumulus scheme (Grell and Devenyi, 2002), the WRF Single Moment (WSM) three-class microphysics (Hong et al., 2004), and Mellor–Yamada–Janjic (MYJ) TKE scheme (Janjic, 2002) for the PBL processes.

A specific version of the WRF-CHEM model (Grell et al., 2005) was used in the present study, which was developed by Li et al. (2010, 2011b,c, 2012) at the Molina Center for Energy and the Environment, with a new flexible gas phase chemical module and the CMAQ (version 4.6) aerosol module developed by US EPA (Binkowski and Roselle, 2003). The inorganic aerosols were predicted in the WRF-CHEM model using ISORROPIA ("equilibrium" in Greek, here referred to as an improved thermodynamic equilibrium aerosol model) Version 1.7 (http://nenes.eas.gatech.edu/ISORROPIA/). The SOA formation was simulated using a non-traditional SOA model including the volatility basis-set modeling method in which primary organic components are assumed to be semi-volatile and photochemically reactive and are distributed in logarithmically spaced volatility bins (Li et al., 2011a). Detailed description of the WRF-CHEM model can be found in Li et al. (2010, 2011b,c, 2012). The emission inventory used in this study was developed at the Molina Center (Li et al., 2014). The chemical initial and boundary conditions for the WRF-CHEM model simulations were interpolated from the 6-h output of a global chemical transport model for O₃ and related chemical tracers (MOZART). Considering that we mainly concentrated on the effects caused by changes in the meteorological fields, the initial and boundary conditions for chemical fields and the emission inventory were the same for all model experiments.

Both meteorological and photochemical simulations were conducted on May 30 and June 04, 2010 according to data availability. Additionally, these two cases represented different plume transport pattern (Bei et al.,
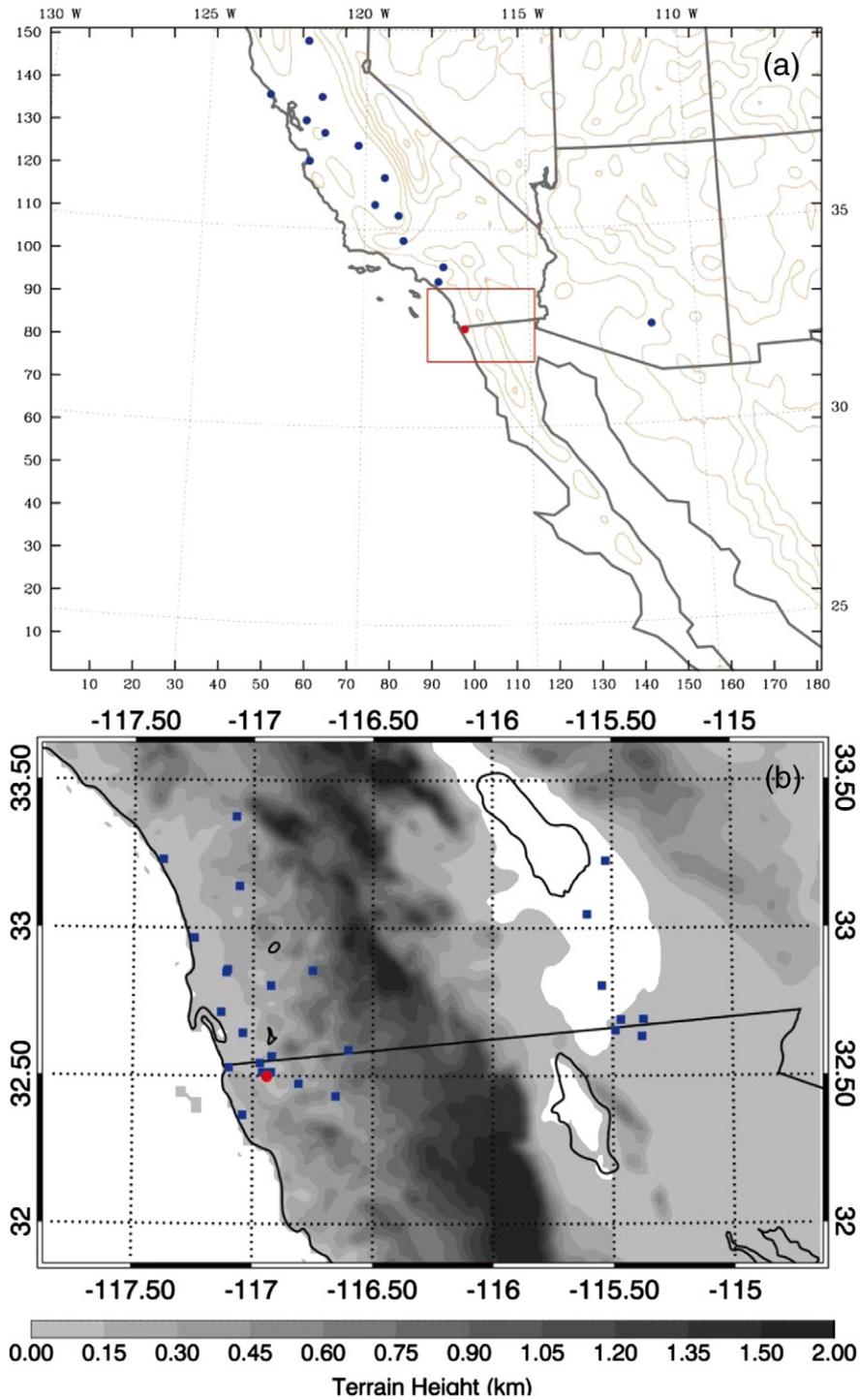
Fig. 1. (a) WRF and (b) WRF-CHEM simulation domains. In (a), blue dots represent additional wind profiler sites and the red dot is the additional sounding at Parque Morelos (PQM, Tijuana Municipal System Theme Park in the center of Tijuana, Mexico); In (b), green dots are the O_3 monitoring sites and the red dot is the supersite PQM with highly time-resolved meteorological variables, ambient gas phase species, and aerosols.

2013). The physical process parameterization schemes used in the WRF-CHEM simulations were the same as those used in the WRF model.

2.2. Ensemble initial and boundary conditions

The ensemble initialization method was similar to the one employed in our previous study on O_3 predictability due to meteorological uncertainty (Bei et al., 2010). Detailed information about the method can be found in Bei et al. (2010). The average initial ensemble spread was 0.4–1.2 m s^{-1} for horizontal winds (u, v), 0.5–1.8 K for temperature (T), 0–0.5 hPa for pressure (p), and 0.5–1.8 g kg^{-1} for the water vapor mixing ratio (q). We used 30 members, which was both affordable and reasonable based on previous studies (e.g., Meng and Zhang, 2007; Zhang et al., 2009a).

2.3. The EnKF

The EnKF is a sequential data assimilation method, first proposed for geophysical applications by Evensen (1994). The regional scale EnKF used in the present study was developed in Zhang et al. (2006) and Meng and Zhang (2008a,b). It evolved from the first limited-area EnKF study of Snyder and Zhang (2003) but employs the covariance relaxation of Zhang et al. (2004) to inflate the background error covariance.

Further information on the EnKF configurations used in this study can be found in Meng and Zhang (2008a,b), an overview of the EnKF technique can be found in Snyder and Zhang (2003), and an extensive review of limited-area ensemble based data assimilation can be found in Meng and Zhang (2011).

2.4. Experimental design

Four experiments were conducted for the selected two cases in this study, including the reference deterministic forecast experiment initialized with NCEP-FNL data (hereafter referred to as FCST), a simulation using observational nudging (hereafter referred to as FDDA), in which the observational nudging was used during the 12 h before the model initial time, an ensemble simulation (for the convenience, hereafter we refer to the ensemble mean as EnsM), and an ensemble simulation with EnKF data assimilation (hereafter referred to as EnKF), in which EnKF was applied during the 12 h before the model initial time with a 6-h interval. The routine sounding observations and wind profilers, the additional sounding at PQM, and the additional wind profiler during CalNex 2010 were assimilated in domain 1 (Fig. 1a). The rest of the input set-ups for the three simulations were the same. The results from domain1 were interpolated to the 2-km domain to produce the initial and boundary condition for the WRF-CHEM model. It should be noted that the ensemble mean of EnKF in domain1 was used to drive the 2-km WRF-CHEM run.

2.5. Statistical methods for comparisons

In order to evaluate the performance of the different methods in simulating meteorological parameters and aerosols against measurements, the mean bias (MB), R squared (R^2), and the root mean square error (RMSE) were utilized in the study.

$$MB = \frac{1}{N} \sum_{i=1}^{N} (P_i - O_i)$$

$$R^2 = \frac{\left[\sum_{i=1}^{N} (P_i - \overline{P})(O_i - \overline{O}) \right]^2}{\sum_{i=1}^{N} (P_i - \overline{P})^2 \sum_{i=1}^{N} (O_i - \overline{O})^2}$$

$$RMSE = \left[\frac{1}{N} \sum_{i=1}^{N} (P_i - O_i)^2 \right]^{\frac{1}{2}}$$

where P_i and O_i are the prediction and observation, respectively. N is the total number of the predictions used for comparisons, and $\overline{P}$ and $\overline{O}$ denote the average of the prediction and observation, respectively.

3. Synoptic overview for the two selected days

The two days (May 30 and June 04, 2010) selected in this study represent two meteorological episode types in the California–Mexico border region, corresponding to two plume transport patterns "plume-southwest" and "plume-east", respectively (Bei et al., 2013). Fig. 2 shows the synoptic conditions on 0500 PDT May 30 and June 04 on 850 hPa. In general, the plume transport directions were consistent with the prevailing surface wind directions as analyzed in Bei et al. (2013). On May 30, the border region was located at the north of a
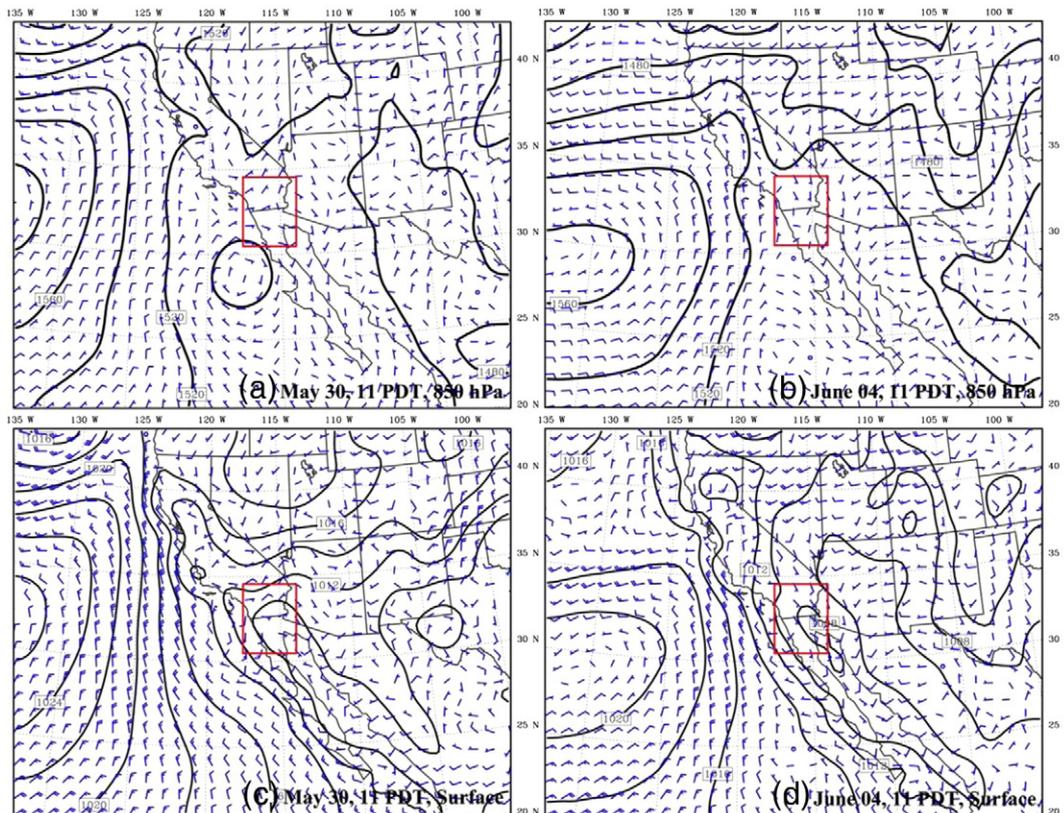

Fig. 2. 850 hPa geopotential heights and winds (a–b), sea level pressure and surface winds (c–d) from NCEP-FNL reanalysis data at 05 PDT on May 30 and June 04, 2010. The California–Mexico border region is indicated by the red box.

low on 850 hPa and the northwest of the low pressure area at the surface, causing the prevailing northeasterly wind at the low levels, same as those on 850 hPa. On June 04, the border region was located at the southeast of a trough on 850 hPa and the southwest of a surface low, leading to different prevailing winds on 850 hPa and at the surface. The plume transport pattern on June 04 is "plume-east", mainly decided by the surface wind.

4. Model results

4.1. Uncertainties in air quality simulations due to the meteorological initial errors

Our previous studies have examined the O_3 and SOA predictabilities due to meteorological uncertainties in Houston and Mexico City using ensemble forecasts (Zhang et al., 2007a; Bei et al., 2010, 2013). The meteorological initial uncertainties influence substantially O_3 and SOA predictions, particularly with regard to the simulation at one or several

supersites. However, the relationship between the meteorological uncertainties and air quality simulations is flow-dependent and complicated. Using similar method, we have investigated the uncertainties in O_3 and aerosol simulations in the border region due to the meteorological initial uncertainties.

Fig. 3 shows the temporal evolutions of the ensemble mean and the spread of the surface meteorological fields (relative humidity, temperature, wind direction and speed, PBL height) and the observations (black dots) at PQM on May 30 and June 04, 2010. On both days, the PBL height (Fig. 3a and f) and temperature (Fig. 3d and i) exhibited large ensemble spreads in the afternoon compared to those in the morning, while the wind speed (Fig. 3b and g) and direction (Fig. 3c and h), and relative humidity (Fig. 3d and h) have large uncertainties in the morning. The ensemble mean generally reproduced the diurnal variation of observations, but there were still discrepancies between the ensemble forecasts and observations. The ensemble predictions failed to yield the abrupt changes of the PBL height in the morning and late afternoon, which was likely caused by the transition of sea breeze, and also considerably
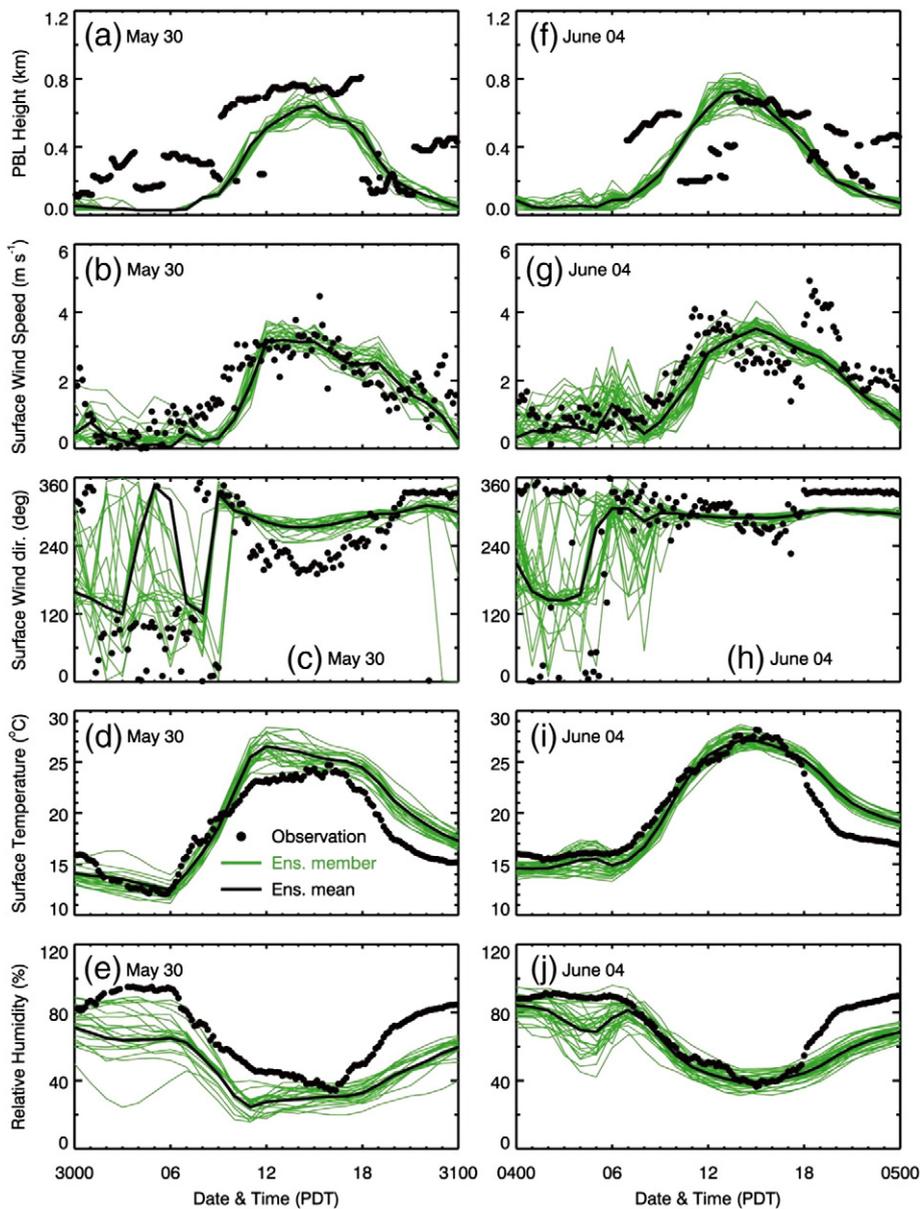
Fig. 3. Temporal evolutions of the ensemble mean and the spread of the meteorological fields at PQM on May 30 and June 04, 2010. (a) and (f) PBL height, (b) and (g) surface wind speed, (c) and (h) surface wind direction, (d) and (i) surface temperature, (e) and (j) surface relative humidity. Black dots: observations; Green lines: ensemble members; Black line: ensemble mean.
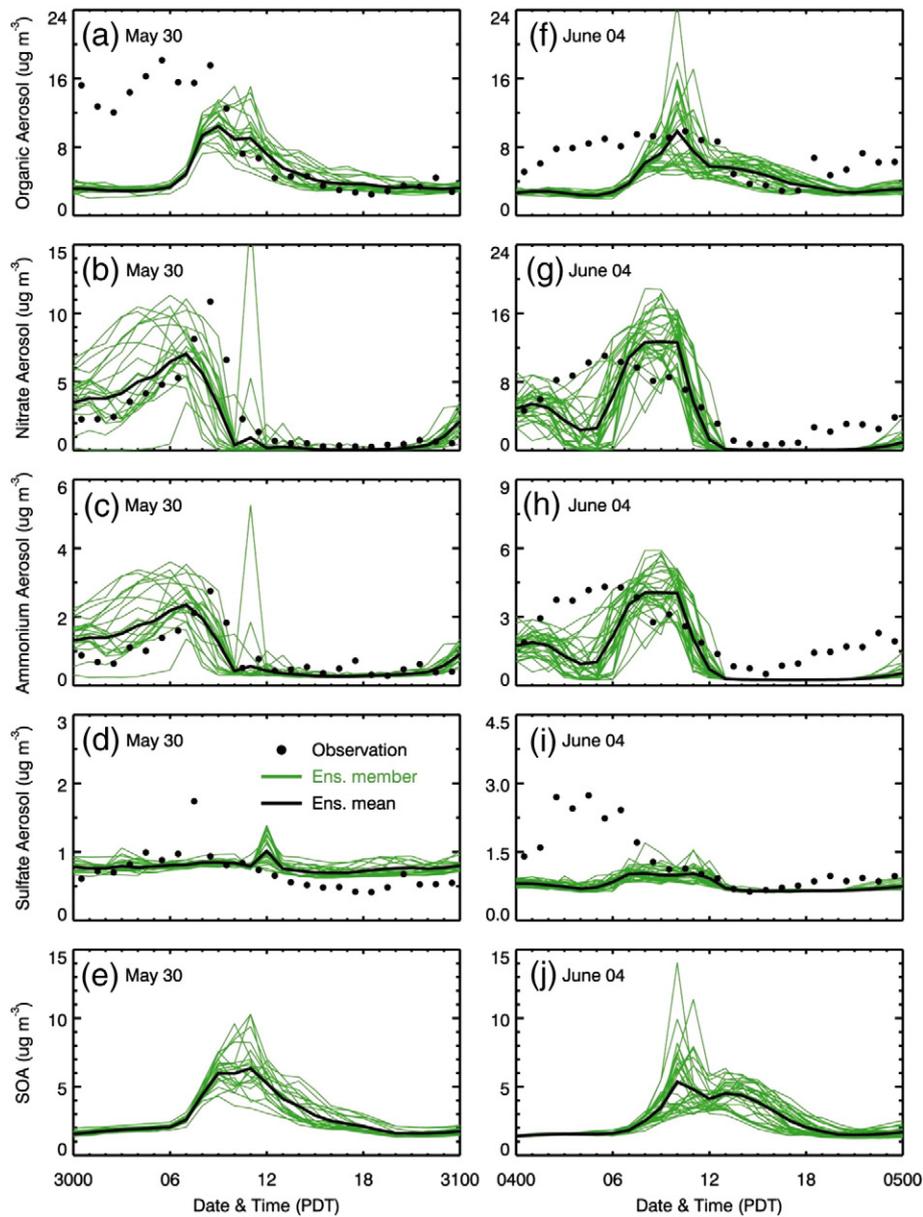
Fig. 4. Same as Fig. 3, but for aerosols at PQM. (a) and (f) organic aerosol, (b) and (g) nitrate aerosol, (c) and (h) ammonium aerosol, (d) and (i) sulfate aerosol, and (e) and (j) SOA.

underestimated the observation during nighttime. The ensemble mean could not reproduce the observed wind direction changes in the morning and late afternoon during the transition of sea breeze, and the ensemble spread is very large in the morning. On May 30, most of the ensemble members substantially underestimated the observed surface winds between 0800 and 1000 PDT, causing the delay of the occurrence of the sea breeze and an overestimation of the surface temperature after

1000 PDT. On June 04, better predictions of the wind speed in the morning also improved the ensemble predictions of the surface temperature. However, the ensemble mean underestimated the observed relative humidity, especially during the nighttime on both days.

Fig. 4 presents the temporal evolutions of the ensemble mean and the spread of the organic and inorganic aerosol concentrations and the observations at PQM on the two days. The nitrate (Fig. 4b and
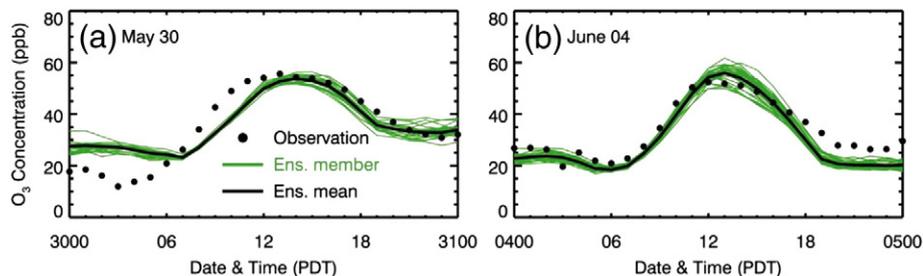

Fig. 5. Same as Fig. 3, but only for O_3 averaged over the sites along the coastal region.
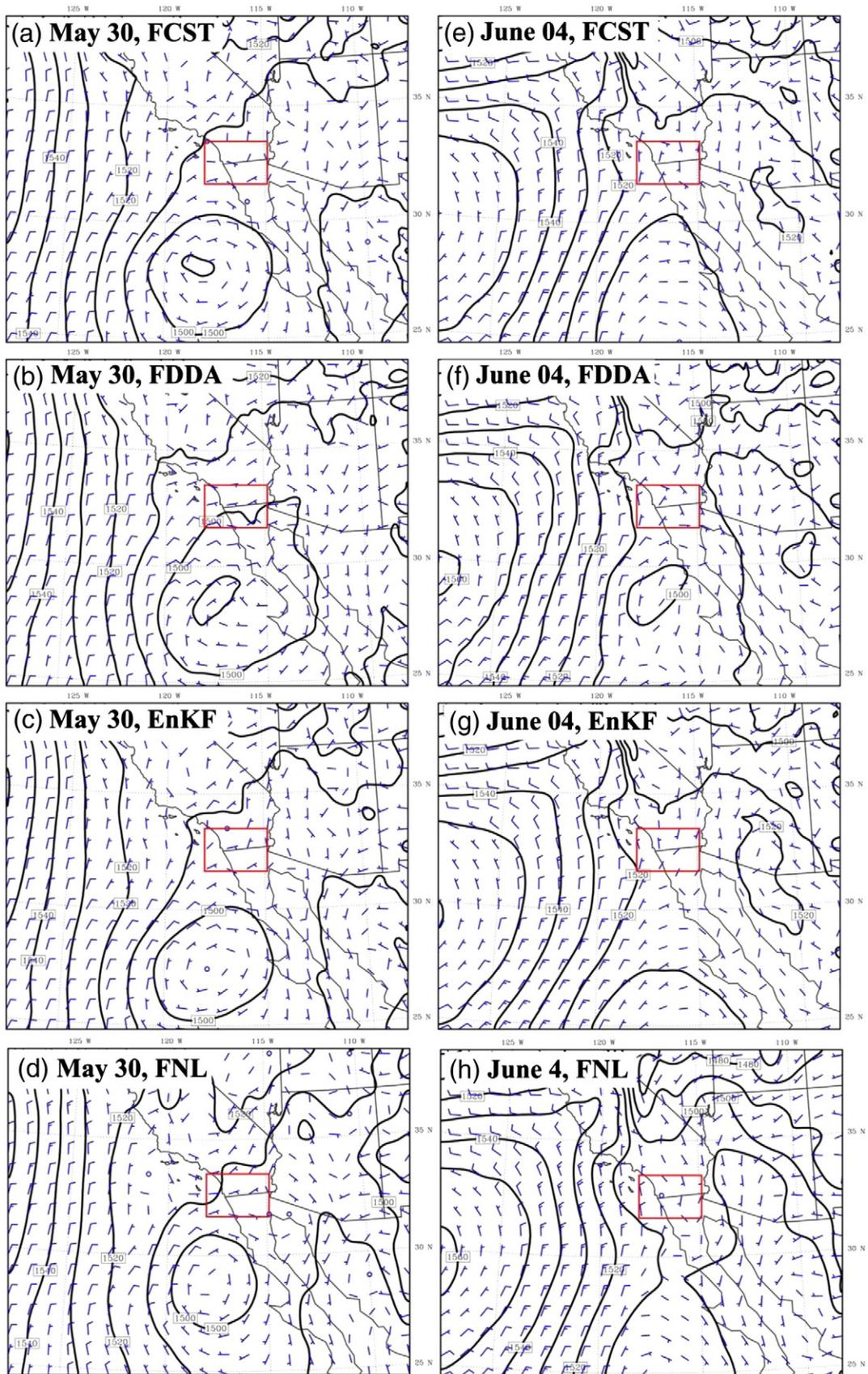
Fig. 6. Simulated 850 hPa geopotential heights and winds at 11 PDT by (a) and (e) FCST, (b) and (f) FDDA, (c) and (g) EnKF, and (d) and (h) FNL on May 30 and June 04, 2010.

g) and ammonium (Fig. 4c and i) aerosols had the largest spread in the morning on both days, which was possibly due to the large uncertainties in surface temperature and humidity simulations (Fig. 3d,e,i, and j) because their formations are sensitive to temperature and humidity. The uncertainties of predicted surface winds in the morning also possibly constitute an important factor for the spreads of nitrate and ammonium aerosols through transporting their precursors. The organic aerosol predictions showed large spread between 0900 and 1200 PDT on both days, which was caused by the uncertainties of the SOA formation (Fig. 4e and j), the same as the results of Bei et al. (2012a). On the two days, the ensemble mean did not produce the observed high organic aerosol concentrations in the early morning hours when the primary organic aerosols are dominant, so the uncertainties of the emission inventory might contribute significantly to the underestimation of organic aerosols. Because the SOA observations at PQM are not available, only the simulated temporal evolutions of the ensemble mean and the spread of the SOA concentrations are shown in Fig. 4e and j. On May 30, the simulated SOA concentration reaches its peak between 0900 and 1200 PDT, with the ensemble mean of around 5–6 $\mu g\ m^{-3}$ and the ensemble spread of up to 1.7 $\mu g\ m^{-3}$. On June 04, the simulated peak SOA concentration occurs between 1000 and 1400 PDT, with the ensemble mean of around 4–5 $\mu g\ m^{-3}$ and the ensemble spread of up to 2.6 $\mu g\ m^{-3}$. In addition, the ensemble forecasts were not able to capture the observed high peaks of sulfate aerosols that were influenced by multiple sources, including the inefficient formation through the gas phase reaction of SO_2 with OH, the oxidation of SO_2 in cloud droplets as well as direct emissions from power plants and industries. The sulfate

aerosol peak at PQM was primarily determined by the transport of the emissions from Rosarito power plant (a point source) in the coastal region in Tijuana. Therefore the bias of predicted wind fields in the morning might be the main reason for the underestimation of sulfate aerosol peaks. In general, the ensemble mean performed reasonably well in predicting the nitrate and ammonium aerosols, but the underestimation or overestimation still existed.

In general, the ensemble mean of the surface O_3 concentrations averaged over the coastal region was in good agreement with the observations (Fig. 5). The ensemble spread of peak O_3 concentrations (hereafter we refer to the O_3 peak time as 1200–1500 PDT) on June 04 was greater than that on May 30, which was likely caused by the large wind field uncertainties in the morning on June 04 (Fig. 3f). The ensemble spread of surface O_3 concentrations was less than 10 ppb, and also less than that of aerosols at PQM, which was due to the average of the spread of surface O_3 concentrations over many sites.

4.2. Influence of using EnKF on meteorological simulations

Fig. 6 shows the large-scale synoptic conditions at 1100 PDT simulated by FCST, FDDA, and EnKF on the selected two days. The simulated large-scale synoptic patterns have been modified due to the implementation of data assimilation. For example, at 1100 PDT on May 30, both the location and intensity of the simulated 850 hPa geo-potential height low center by EnKF (Fig. 6c) were in better agreement with the NCEP-FNL analysis (Fig. 6d) compared to those by FCST and FDDA (Fig. 6a and b), which in turn caused the simulated wind circulation over the
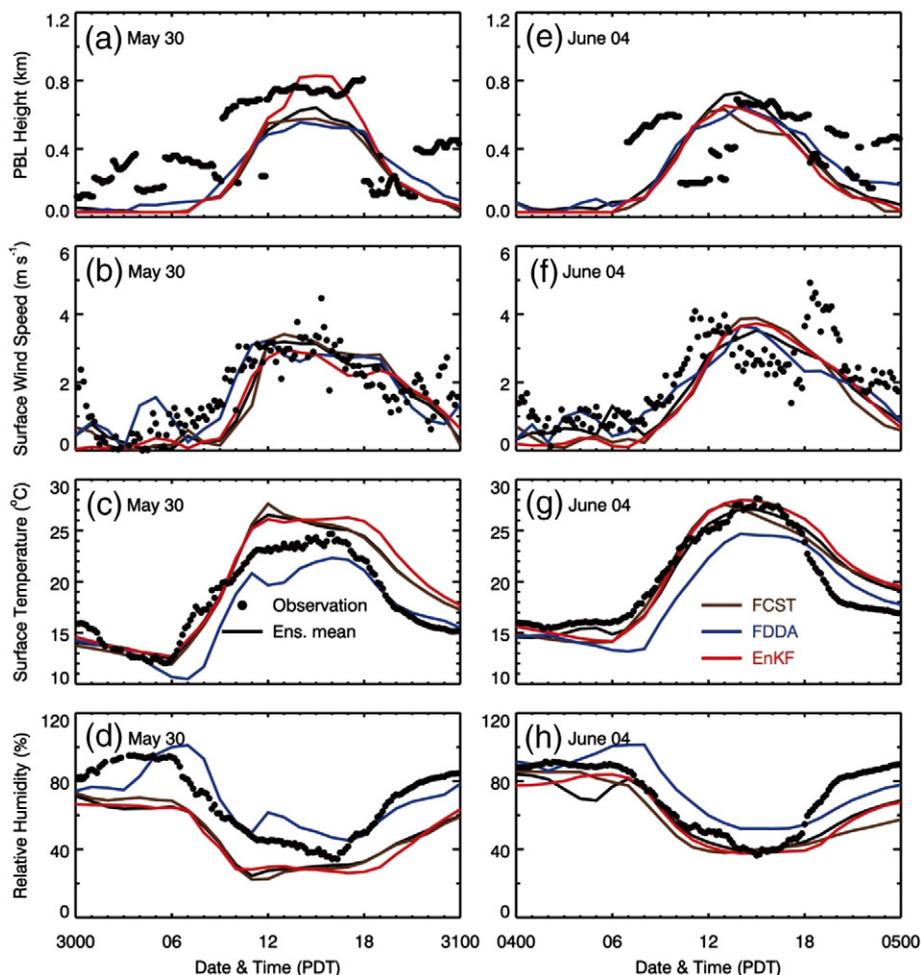

Fig. 7. Temporal evolutions of simulated and observed (a) and (e) PBL height, (b) and (f) surface wind speed, (c) and (g) surface temperature, (d) and (h) surface relative humidity at PQM on May 30 and June 04, 2010. Black dots: observations; black line: ensemble mean; blue line: FDDA; red line: EnKF; brown line: FCST.

Table 1
Statistical comparison of simulated and measured hourly averaged meteorological parameters and aerosol concentrations at PQM site on May 30 and June 04, 2010.

Species	Method	May 30			June 04		
		MB	R^2	RMSE	MB	R^2	RMSE
PBL height (m)	EnsM	−170	0.66	220	−87	0.04	250
	FCST	−190	0.65	230	−140	0.01	270
	FDDA	−160	0.60	210	−78	0.04	220
	EnKF	−130	0.67	210	−130	0.04	260
Surface temperature (k)	EnsM	1.3	0.91	2.1	0.25	0.87	1.6
	FCST	1.2	0.91	2.2	0.21	0.85	1.8
	FDDA	−1.5	0.83	2.3	−1.56	0.78	2.6
	EnKF	1.7	0.89	2.6	0.67	0.87	2.0
Wind speed (m s^{-1})	EnsM	−0.30	0.77	0.62	−0.35	0.65	0.75
	FCST	−0.27	0.73	0.69	−0.49	0.60	0.96
	FDDA	−0.05	0.70	0.54	−0.32	0.59	0.76
	EnKF	−0.32	0.77	0.63	−0.47	0.65	0.90
Relative humidity (%)	EnsM	−21	0.91	22	−10	0.82	13
	FCST	−21	0.91	22	−12	0.69	16
	FDDA	1.1	0.71	11	4.3	0.67	12
	EnKF	−21	0.91	23	−10.7	0.81	13
Organics (μg m^{-3})	EnsM	−3.8	0.10	6.9	−2.5	0.11	3.4
	FCST	−4.3	0.11	6.9	−1.2	0.10	3.0
	FDDA	−5.2	0.12	8.0	−3.0	0.21	3.7
	EnKF	−4.0	0.09	7.0	−2.7	0.46	3.3
Nitrate (μg m^{-3})	EnsM	−0.36	0.56	1.9	−1.6	0.54	3.3
	FCST	−0.02	0.57	2.0	−0.34	0.70	2.4
	FDDA	1.0	0.72	2.5	−1.7	0.58	3.2
	EnKF	−0.08	0.73	1.5	−1.0	0.83	2.3
Ammonium (μg m^{-3})	EnsM	0.04	0.50	0.50	−0.97	0.39	1.4
	FCST	0.14	0.51	0.61	−0.61	0.54	1.1
	FDDA	0.44	0.66	0.93	−1.0	0.39	1.5
	EnKF	0.13	0.68	0.49	−0.81	0.72	1.1
Sulfate (μg m^{-3})	EnsM	0.07	0.28	0.26	−0.54	0.10	0.85
	FCST	0.06	0.18	0.25	−0.53	0.28	0.83
	FDDA	0.03	0.12	0.27	−0.55	0.09	0.88
	EnKF	0.09	0.10	0.28	−0.60	0.37	0.86

border region (inner box indicated by red color in Fig. 6) in EnKF to be more consistent with the NCEP-FNL analysis fields. On June 04, the simulated 850 hPa wind fields inside the border region by EnKF (Fig. 6g) also generally performed better than those by other two experiments (Fig. 6e–f).

The PBL height is an important parameter in air quality simulations. Fig. 7a and e illustrates the comparisons of the modeled hourly PBL heights with the observation at PQM on the selected two days. FCST generally underestimated the PBL heights, with the MB of −90 m and −140 m on May 30 and June 04, respectively (Table 1). EnKF performed best in modeling the PBL height on May 30, with the MB of −130 m, R^2 of 0.67, and RMSE of 210 m. On June 04, all the simulation methods failed to reproduce the variation of the PBL height with R^2 less than 0.1, due to frequent abrupt changes of the observations. For the predictions of the wind speed, surface temperature, and relative humidity, FDDA performed differently compared to FCST, EnKF, and EnsM. FDDA yielded poor diurnal variations of the wind speed with the R^2 of 0.70 and 0.59 on May 30 and June 04, respectively, less than those from the other three methods. In addition, FDDA underestimated the surface temperature with MB of less than −1.5 °C and overestimated the

relative humidity with the MB of greater than 1% on the two days, contrary to the results from the other three methods, indicating the strong predicted see breeze. In terms of R^2, EnKF yielded the best simulations of the surface wind speed, temperature, and relative humidity. Overall, the simulated meteorological parameters by EnKF are similar to those by EnsM, and showed better performance compared to FCST and FDDA (Fig. 7b–d, f–h).

4.3. Influence of using EnKF on surface O_3 simulations

The O_3 simulations driven by the three different meteorological fields (FCST, FDDA, and EnKF) along with EnsM were evaluated and compared with the measurements at the surface monitoring stations along the coastal area. Fig. 8 shows the diurnal cycle of observed and simulated surface O_3 concentrations averaged over the monitoring sites along the coastal region. On both days, FDDA substantially underestimated the surface O_3 concentrations during the peak time, which was consistent with the strong simulated sea breeze that transported the plume inland more efficiently. On May 30 during the O_3 peak time, FCST slightly overestimated the surface O_3
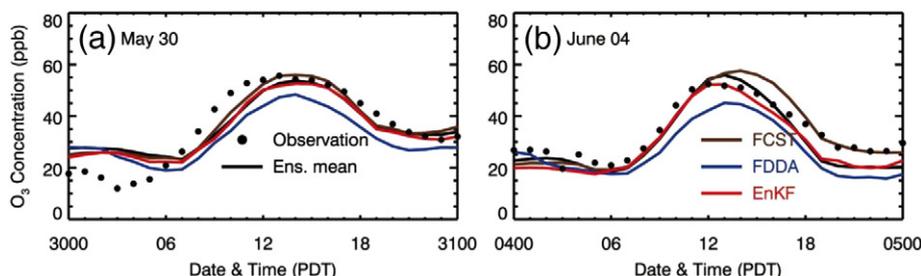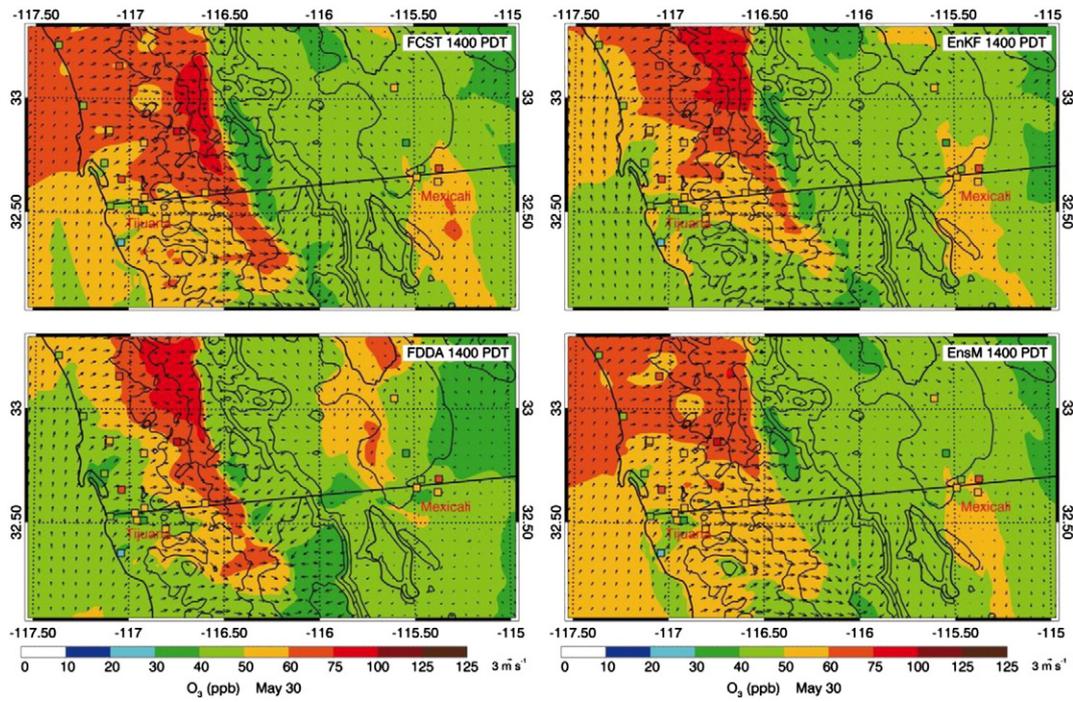

Fig. 8. Same as Fig. 7, but for O_3 averaged over the sites along the coastal region.

Fig. 9. Pattern comparison of simulated vs. observed near-surface O$_3$ at 1400 PDT on May 30. Colored squares: O$_3$ observations; color contour: O$_3$ simulations; black arrows: simulated surface winds.

concentrations, while EnKF and EnsM slightly underestimated the surface O$_3$ concentrations. On June 04, FCST considerably overestimated the surface O$_3$ concentrations in the afternoon, and the simulations by EnKF and EnSM were more consistent with the observations compared to FCST and FDDA during the peak time. Generally, the performance of EnKF was similar to EnsM, which is suggested to be an efficient method for reducing the meteorological uncertainties in simulations of CTMs.

Furthermore, the outflow plumes were also altered due to changes in meteorological fields caused by the data assimilation. Figs. 9 and 10 provide the horizontal distributions of O$_3$ along with the simulated wind fields at 1400 PDT on May 30 and June 04. The modified meteorological fields not only changed the horizontal distribution patterns of O$_3$ but also the magnitude of the maximum concentration. On May 30 (Fig. 9), FDDA produced the lowest surface O$_3$ concentrations along the coastal region, and underestimated the surface O$_3$ concentrations
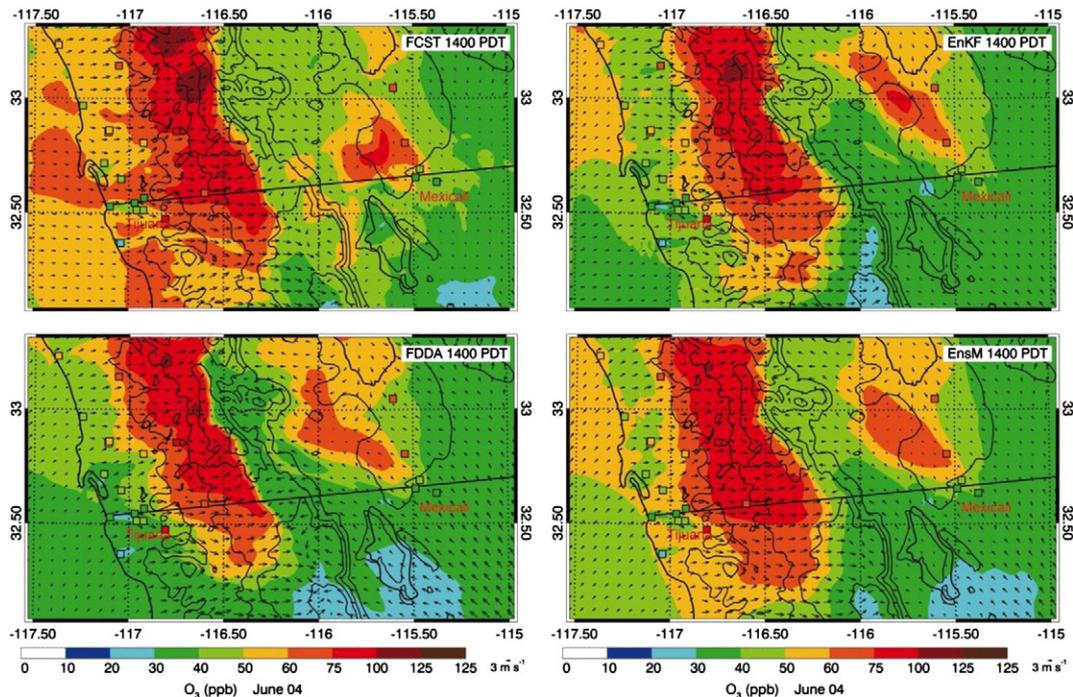
Fig. 10. Same as Fig. 9, but for June 04, 2010.

in and around the urban region of Tijuana, due to the simulated strong westerly winds by FDDA. FCST predicted the highest surface O_3 concentrations and overestimated the surface O_3 concentrations along the coastal region of San Diego, likely caused by the simulated weak sea breeze. Generally, the simulated surface O_3 concentration distributions by EnKF were in a good agreement with the observations, which can be explained by the simulated weaker convergence at the east of the coast and the stronger southerly wind along the coast by EnKF in comparison with FCST and FDDA. On June 04 (Fig. 10), FCST significantly overestimated the surface O_3 concentrations in and around the urban region of Tijuana, and predicted the largest area with the surface O_3 concentrations exceeding 60 ppb. The plume movement in FDDA was fast compared to the other three methods, particularly in the Tijuana region, causing the underestimation of the surface O_3 concentration in and around the urban region of Tijuana. The patterns of surface O_3 concentrations in EnKF were similar to those in EnsM on the both days, generally agreed well with the observed O_3 distributions.

4.4. Influence of using EnKF on aerosol simulations

Fig. 11 shows the temporal evolution of the simulated aerosol mass (including organic and inorganic mass) from FCST, FDDA, EnKF, and EnsM and the observation at the PQM site. On May 30 (Fig. 11a–d), none of the four methods produced good simulations of organic aerosols, with the *MB* raging from -5.2 to -3.8 μg m^{-3}, R^2 ranging from 0.09 to 0.12, and *RMSE* ranging from 6.9 to 8.0 μg m^{-3}, due to the substantial underestimation in the morning (Table 1). On June 04, only

EnKF yielded reasonable diurnal variations with the R^2 of 0.46, much more than the R^2 of the other three methods. The simulated nitrate and ammonium aerosols by EnKF were consistently better than all the other experiments on the both days, with the largest R^2 and smallest *RMSE*. The considerable improvements on nitrate and ammonium aerosol simulations by EnKF may be benefitted from the overall improvements of meteorological fields. For the simulations of sulfate aerosols, none of the methods performed well due to the failure to reproduce the morning peaks, which might be influenced by the large point source (Rosarito power plant) along the coastal region of Tijuana. In addition, the differences in predictabilities of May 30 and June 04 were likely caused by the different flow regimes since we have used the same model and set up for all selected days (e.g., Nuss and Miller, 2001; Zhang et al., 2007b).

Although the overall performance of EnKF in simulating aerosols was generally better than those of FCST and FDDA, the discrepancies between EnKF simulations and the measurements were still considerably large, indicating that, in case of reasonable emission inventory, there are still ample rooms for improvement either in the observation network, the data assimilation system, and/or in the forecast model. The EnKF has been proposed for observation system design (e.g., Bei et al., 2012b), and for simultaneous state and parameter estimation that can help in reducing the initial condition uncertainties as well as improving the forecast model itself (e.g., Aksoy et al. 2006; Hu et al., 2010a,b). More recent studies have also showed that further improvement in data assimilation accuracy may be achieved through hybrid and/or coupling the EnKF with either 3DVAR or 4DVAR (e.g., Zhang et al., 2009b, 2013;
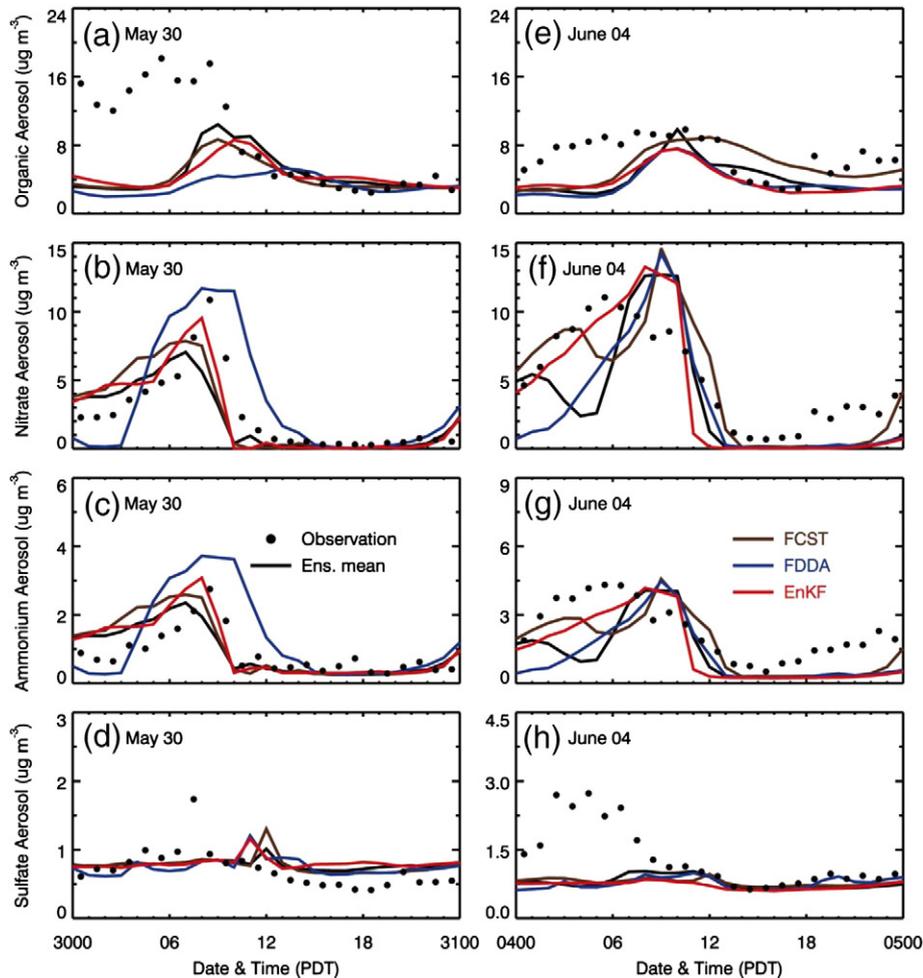

Fig. 11. Same as Fig. 7, but for aerosols at PQM. (a) and (e) organic aerosol, (b) and (f) nitrate aerosol, (c) and (g) ammonium aerosol, and (d) and (h) sulfate aerosol.

Zhang and Zhang, 2012). Future studies will investigate the potential of these advanced techniques in further improving the air quality modeling as in the current study.

5. Conclusions

In the present study, we have investigated the impact of using EnKF on the air quality simulations in the California–Mexico border region during Cal–Mex 2010 comparing with the reference forecast initialized with NCEP analysis data and the simulation with observation nudging. All the experiments were performed on two days (May 30 and June 04, 2010).

We first examined the uncertainties in O_3 and aerosol simulations in the Cal–Mex border region due to the meteorological initial uncertainties through ensemble simulations. The ensemble spread of O_3 averaged over the coast area was less than 10 ppb, similar to our previous results from Houston and Mexico City (Zhang et al., 2007a; Bei et al., 2010). The nitrate and ammonium aerosol were found to have large uncertainties during the morning on both days, which was attributed mostly to the large uncertainties in surface temperature and humidity simulations because of the high sensitivity to temperature and humidity during their formation. The organic aerosol simulations also showed large spreads between 0900 and 1200 PDT when the SOA became dominant. The discrepancies between the ensemble mean and observations still existed, particularly for organic and sulfate aerosols in the morning.

Both routine and additional meteorological observations during the campaign have been assimilated into the meteorological model on the two selected days. The simulated large-scale synoptic patterns over the border region have been modified due to the implementation of the EnKF before the model initial time. The simulated wind circulation, temperature, and humidity fields in the border area by EnKF have been improved to some degree in comparison with FCST and FDDA against the observations. The simulated surface O_3 distributions on both days by EnKF were consistently better than FCST and FDDA compared to the measurement.

The simulated nitrate and ammonium aerosols by EnKF were in better agreement with the observations at PQM on both days. However, EnKF, FCST, FDDA, and EnsM could not produce good simulations of organic and sulfate aerosols in the morning, indicating that there are still substantial rooms for improvement in the current data assimilation system in case of the reasonable emission inventory, which could be observation, model, or data assimilation method itself.

Since this work focuses only on two days, comparisons between the simulations with and without data assimilations during a long period are also necessary. In addition, the observations used in this study were confined to soundings and wind profilers. Other available meteorological observations are also expected to be incorporated into the EnKF to further improve the air quality forecasts or simulations.

Acknowledgments

References

Aksoy A, Zhang F, Nielsen-Gammon JW. Ensemble-Based Simultaneous State and Parameter Estimation in a Two-Dimensional Sea-Breeze Model. Monthly Weather Review 2006;134:2951–70.

Beekmann M, Derognat C. Monte Carlo uncertainty analysis of a regional-scale chemistry model constrained by measurements from the Atmospheric Pollution Over the Paris Area (ESQUIF) campaign. J Geophys Res 2003;108:8559. http://dx.doi.org/10.1029/2003JD003391.

Bei N, de Foy B, Lei W, Zavala M, Molina LT. Using 3DVAR data assimilation system to improve ozone simulations in the Mexico City basin. Atmos Chem Phys 2008;8:7353–66.

Bei N, Lei W, Zavala M, Molina LT. Ozone predictabilities due to meteorological uncertainties in Mexico City Basin using ensemble forecasts. Atmos Chem Phys 2010;10: 6295–309.

Bei N, Li G, Molina LT. Uncertainties in SOA simulations due to meteorological uncertainties in Mexico City during MILAGRO-2006 field campaign. Atmos Chem Phys 2012a;12:11295–308. http://dx.doi.org/10.5194/acp-12-11295-2012.

Bei N, Zhang F, John WN. Ensemble-based observation targeting for improving ozone prediction in Houston and the surrounding area. Pure Appl Geophys 2012b;169(3): 539–54.

Bei N, Li G, Zavala M, Barrera H, Torres R, Grutter M, et al. Meteorological overview and plume transport patterns during Cal–Mex 2010. Atmos Environ 2013;70:477–89.

Bergin MS, Noblet GS, Petrini K, Dhieux JR, Milford JB, Harley RA. Formal uncertainty analysis of a Lagrangian photochemical air pollution model. Environ Sci Technol 1999;33: 1116–26.

Binkowski FS, Roselle SJ. Models-3 Community Multiscale Air Quality (CMAQ) model aerosol component: 1. Model description. J Geophys Res 2003;4183. http://dx.doi.org/10.1029/2001JD001409.

Curier RL, Timmermans R, Calabretta-Jongen S, Eskes H, Segers A, Swart D, et al. Improving ozone forecasts over Europe by synergistic use of the LOTOS-EUROS chemical transport model and in-situ measurements. Atmos Environ 2012;60: 217–26.

Dabberdt WF, Miller E. Uncertainty, ensembles and air quality dispersion modeling: applications and challenges. Atmos Environ 2000;34:4667–73.

Dabberdt WF, Carroll MA, Baumgardner D, Carmichael G, Cohen R, Dye T, et al. Meteorological research needs for improved air quality forecasting: report of the 11th prospectus development team of the U.S. Weather research program. Bull Am Meteorol Soc 2004;85:563–86.

Evensen G. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. J Geophys Res 1994;99:10143–62.

Fast JD. Mesoscale modeling and four-dimensional data assimilation in areas of highly complex terrain. J Appl Meteor 1995;34:2762–82.

Grell GA, Peckham SE, Schmitz R, McKeen SA, Wilczak J, Eder B. Fully coupled "online" chemistry within the WRF model. Atmos Environ 2005;39:6957–75.

Grell GA, Devenyi D. A generalized approach to parameterizing convection combining ensemble and data assimilation techniques. Geophys Res Lett 2002;29(14). http://dx.doi.org/10.1029/2002GL015311.

Hanna SR, Lu ZG, Frey HC, Wheeler N, Vukovich J, Arunachalam S, et al. Uncertainties in predicted ozone concentrations due to input uncertainties for the UAM-V photochemical grid model applied to the July 1995 OTAG domain. Atmos Environ 2001; 35:891–903.

Hong S-Y, Dudhia J, Chen S-H. A revised approach to ice microphysical processes for the bulk parameterization of clouds and precipitation. Mon Weather Rev 2004;132: 103–20.

Hu X-M, Nielsen-Gammon JW, Zhang F. Evaluation of three planetary boundary layer schemes in the WRF model. J Appl Meteorol Climatol 2010a;49:1831–44.

Hu X-M, Fuentes JD, Zhang F. Downward transport and modification of tropospheric ozone through moist convection. J Atmos Chem 2010b;65:13–35.

Janjic ZI. Nonsingular implementation of the Mellor-Yamada level 2.5 scheme in the NCEP Meso Model, NCEP Office Note, vol. 437; 2002 [61 pp.].

Kalnay E. Atmospheric modeling, data assimilation and predictability. New York: Cambridge Univ. Press; 2003 [341 pp.].

Li G, Lei W, Zavala M, Volkamer R, Dusanter S, Stevens P, et al. Impacts of HONO sources on the photochemistry in Mexico City during the MCMA-2006/MILAGO Campaign. Atmos Chem Phys 2010;10:6551–67.

Li G, Zavala M, Lei W, Tsimpidi AP, Karydis VA, Pandis SN, et al. Simulations of organic aerosol concentrations in Mexico City using the WRF-CHEM model during the MCMA-2006/MILAGRO campaign. Atmos Chem Phys 2011a;11:3789–809.

Li G, Bei N, Tie X, Molina LT. Aerosol effects on the photochemistry in Mexico City during MCMA-2006/MILAGRO campaign. Atmos Chem Phys 2011b;11:5169–82.

Li G, Lei W, Bei N, Molina LT. Contribution of garbage burning to chloride and PM$_{2.5}$ in Mexico City. Atmos Chem Phys 2012;12:8751–61. http://dx.doi.org/10.5194/acp-12-8751-2012.

Li G, Bei N, Zavala M, Molina LT. Ozone formation along the California–Mexican border region during Cal–Mex 2010 field campaign. Atmos Environ 2014;88:370–89.

Meng Z, Zhang F. Tests of an ensemble Kalman filter for mesoscale and regional-scale data assimilation. Part II: imperfect model experiments. Mon Weather Rev 2007;135: 1403–23.

Meng Z, Zhang F. Tests of an ensemble Kalman filter for mesoscale and regional-scale data assimilation. Part III: comparison with 3DVAR in a real-data case study. Mon Weather Rev 2008a;136:522–40.

Meng Z, Zhang F. Tests of an ensemble Kalman filter for mesoscale and regional-scale data assimilation. Part IV: comparison with 3DVar in a month-long experiment. Mon Weather Rev 2008b;136:3671–82.

Meng Z, Zhang F. Limited-area ensemble-based data assimilation. Mon Weather Rev 2011;139:2025–45.

Menut L. Adjoint modeling for atmospheric pollution process sensitivity at regional scale. J Geophys Res 2003;108:8562. http://dx.doi.org/10.1029/2002JD002549.

Molina LT, Kolb CE, de Foy B, Lamb BK, Bruce WH, Jimenez JL, et al. Air quality in North America's most populous city-overview of MCMA-2003 Campaign. Atmos Chem Phys 2007;7:2447–73.

Navon IM. Practical and theoretical aspects of adjoint parameter estimation and identifiability in meteorology and oceanography. Dyn Atmos Oceans 1998;27:55–79.

Nielsen-Gammon JW, Hu X-M, Zhang F, Pleim JE. Evaluation of planetary boundary layer scheme sensitivities for the purpose of parameter estimation. Mon Weather Rev 2010;138:3400–17.

Nuss WA, Miller DK. Mesoscale predictability under various synoptic regimes. Nonlinear Processes Geophys 2001;8:429–38. http://dx.doi.org/10.5194/npg-8-429-2001.

Ryerson TB, Andrews AE, Angevine WM, Bates TS, Brock CA, Cairns B, et al. The 2010 California Research at the Nexus of Air Quality and Climate Change (CalNex) field study. J Geophys Res Atmos 2013;118. http://dx.doi.org/10.1002/jgrd.50331.

Seaman NL. Meteorological modeling for air-quality assessments. Atmos Environ 2000; 34:2231–59.

Skamarock WC, Klemp JB, Dudhia J, Gill DO, Barker DM, Duda M, et al. A description of the advanced research WRF version 3, NCAR Technical Note, Boulder, CO; 2008.

Snyder C, Zhang F. Assimilation of simulated Doppler radar observations with an ensemble Kalman filter. Mon Weather Rev 2003;131:1663–77.

Stauffer DR, Seaman NL. Use of four-dimensional data assimilation in a limited-area mesoscale model—Part 1: experiments with synoptic-scale data. Mon Weather Rev 1990; 118:1250–77.

Stauffer DR, Seaman NL. Use of four-dimensional data assimilation in a limited-area mesoscale model—Part 2: effects of data assimilation within the planetary boundary layer. Mon Weather Rev 1991;119:734–54.

Stauffer DR, Seaman NL. Multiscale 4-dimensional data assimilation. J Appl Meteorol 1994;33:416–34.

Stuart AL, Aksoy A, Zhang F, Nielsen-Gammon JW. Ensemble-based data assimilation and targeted observation of a chemical tracer in a sea breeze model. Atmos Environ 2007; 41:3082–94.

Tang X, Zhu J, Wang ZF, Gbaguidi A. Improvement of ozone forecast over Beijing based on ensemble Kalman filter with simultaneous adjustment of initial conditions and emissions. Atmos Chem Phys 2011;11:12901–16.

Weng Y, Zhang F. Assimilating airborne Doppler radar observations with an ensemble Kalman filter for convection-permitting hurricane initialization and prediction: Katrina (2005). Mon Weather Rev 2012;140:841–59.

Weng Y, Zhang M, Zhang F. Advanced data assimilation for cloud-resolving hurricane initialization and prediction. Comput. Sci. Eng. 2011;13:240–9.

Zhang M, Zhang F. E4DVar: coupling an ensemble Kalman filter with four-dimensional variational data assimilation in a limited-area weather prediction model. Mon Weather Rev 2012;140:587–600.

Zhang F, Snyder C, Sun J. Impacts of initial estimate and observation availability on convective-scale data assimilation with an ensemble Kalman filter. Mon Weather Rev 2004;132:1238–53.

Zhang F, Meng Z, Aksoy A. Tests of an ensemble Kalman filter for mesoscale and regional-scale data assimilation. Part I: perfect model experiments. Mon Weather Rev 2006; 134:722–36.

Zhang F, Bei N, Nielsen-Gammon JW, Li G, Zhang R, Stuart AL, et al. Impacts of meteorological uncertainties on ozone pollution predictability estimated through meteorological and photochemical ensemble forecasts. J Geophys Res 2007a;112:D04304. http://dx.doi.org/10.1029/2006JD007429.

Zhang F, Bei N, Rotunno R, Snyder C, Epifanio CC. Mesoscale predictability of moist baroclinic waves. J Atmos Sci 2007b;64:3579–94.

Zhang F, Weng Y, Sippel JA, Meng Z, Bishop CH. Cloud-resolving hurricane initialization and prediction through assimilation of Doppler radar observations with an ensemble Kalman filter. Mon Weather Rev 2009a;137:2105–25.

Zhang F, Zhang M, Hansen JA. Coupling ensemble Kalman filter with four-dimensional variational data assimilation. Adv Atmos Sci 2009b;26:1–8.

Zhang M, Zhang F, Huang X, Zhang X. Intercomparison of an ensemble Kalman filter with three- and four-dimensional variational data assimilation methods in a limited-area model over the month of June 2003. Mon Weather Rev 2011a;139:566–72.

Zhang F, Weng Y, Gamache JF, Marks FD. Performance of convection-permitting hurricane initialization and prediction during 2008–2010 with ensemble data assimilation of inner-core airborne Doppler radar observations. Geophys Res Lett 2011b;38: L15810. http://dx.doi.org/10.1029/2011GL048469.

Zhang F, Zhang M, Poterjoy J. E3DVar: coupling an ensemble Kalman filter with three-dimensional variational data assimilation in a limited-area weather prediction model and comparison to E4DVar. Mon Weather Rev 2013;140:900–17.