Ensemble-based observation targeting for improving ozone prediction in Houston and the surrounding area

Naifang Bei[1], Fuqing Zhang[2], and John W. Nielsen-Gammon[3]

[1]Molina Center for Energy and the Environment, La Jolla, California

[2]Department of Meteorology, The Pennsylvanian State University, University Park, Pennsylvania

[3]Department of Atmospheric Sciences, Texas A&M University, College Station, Texas

Submitted to *Pure and Applied Geophysics* for publication

Revised, May 7, 2011

Corresponding author address: Dr. Fuqing Zhang, Department of Meteorology, The Pennsylvania State University, University Park, PA 16802

E-mail: fzhang@psu.edu

Abstract

This study examines the effectiveness of targeted meteorological observations for improving ozone prediction in Houston and the surrounding area based on perfect-model simulation experiments. Supplementary observations are targeted for the location that has the highest impact factor (maximum Kalman gain) estimated from an ensemble and is expected to minimize ozone forecast uncertainty at the verification time. It is found that the observational impact factor field varies with time and is sensitive to ensemble resolutions and physics parameterizations. The efficiency of observation targeting is further examined through assimilating observations in areas with different impact factors using an ensemble Kalman filter. It is found that the ensemble sensitivity analysis is capable of locating supplementary observations that may reduce meteorological and ozone forecast error, but not as effectively as expected.

Keywords: Ozone prediction; Data assimilation; Observation targeting; Observational impact factor

1. Introduction

The severity of air pollution situations is determined by a complicated interaction among three factors: the emissions to the atmosphere, chemical reactions, and meteorology (*Banta et al.*, 2005). However, except for accidental releases or spills, whether high-pollution concentrations form on a given day is dominated principally by meteorological processes, which determine the dilution or accumulation of the pollutant emissions and can also impact other key processes, such as chemical reaction rates. Air-quality forecasters depend upon numerical weather prediction (NWP) model output for guidance in formulating their forecasts, so the reliability of air-quality forecasts is related to the accuracy of the NWP models. It has been known that the quality of a numerical weather forecast is related to the quality of its initial condition. If the initial condition has large errors, or if it has moderate errors in regions where forecast errors grow fast, the relevant numerical weather forecast may not be accurate enough. The importance of the accurate representation of meteorological conditions for ozone predictions in Houston and the surrounding area has been demonstrated by *Zhang et al*. (2007) recently.

One possible approach for minimizing forecast errors from day to day or in particular situations is the use of an adaptive observation network (*Emanuel et al.,* 1995). The existing methods for adaptive observations include the singular vector technique (*Palmer et al.,* 1998; *Buizza and Montani,* 1999; *Bergot et al*., 1999; *Gelaro et al.*, 1999, 2000; *Bergot*, 2001), the quasilinear inverse approach (*Pu et al.*, 1997; *Pu and Kalnay*, 1999), gradient and sensitivity approaches (*Bergot et al.*, 1999; *Langland et al*., 1999; *Baker and Daley*, 2000), ensemble spread techniques (*Lorenz and Emanuel*, 1998; *Hansen and Smith*, 2000; *Morss*, 1998; *Morss et al*.,

2001), the ensemble transform technique (*Bishop and Toth*, 1999; *Szunyogh et al.*, 1999), and the ensemble transform Kalman filter (*Bishop et al*., 2001; *Majumdar et al.*, 2001； *Hamill and Snyder*, 2002; *Liu and Kalnay* 2008; *Liu et al*. 2009). *Kang* (2009) and *Kang et al.* (2011) applied the LETKF for estimating surface CO2 fluxes and obtained promising results. *Hamill and Snyder* (2002) demonstrated the application of an algorithm to select the optimal adaptive observation location using the background-error statistics from an ensemble Kalman filter coupled to a quasigeostrophic model. They underscored the importance of accurate estimates of the background-error covariance matrix through using different data assimilation schemes. Their work focused on testing adaptive observation strategies for improving analyses but not forecasts. *Majumdar et al*. (2006) evaluated the similarities and differences among five types of adaptive sampling guidance for tropical cyclones that occurred during the Atlantic hurricane season of 2004. They found that the guidance using the same adaptive sampling technique with different numerical models was often similar while the guidance using the two main techniques usually differed significantly.

Data assimilation has been used operationally for meteorological modeling and prediction. It is also useful as an inverse modeling technique for diagnosing pollutant emission source locations and strengths (i.e. parameter estimation) (*Chang et al.*, 1997; *Elbern et al.,* 2000; *Mendoza-Dominguez and Russell*, 2001) and for identifying locations (in time and space) for field observation networks and adaptive observations (*Daescu and Carmichael*, 2003). However, much of these data assimilation works have focused on variational data assimilation techniques.

The ensemble-based Kalman filter (EnKF) is an alternative data assimilation approach

which has been applied in a number of studies (*Evensen*, 2003) since it was first introduced by *Evensen* (1994). It has also been used to improve air quality modeling (such as *van Loon et al*., 2000; *Hanea et al*. 2004; *Heemink and Segers,* 2002). Its advantages include flow-dependent background error covariance, ease of implementation, and its use of a fully nonlinear model. The EnKF system used here is the same as that employed in *Zhang et al*. (2006) and *Meng and Zhang* (2007). It is a square root EnKF with 20 ensemble members that uses covariance relaxation (*Zhang et al*. 2004, their Eq. (5) where a = 0.5) to inflate the background error covariance. More recent development and applications of this system can be found in *Meng and Zhang* (2008a, b), *Zhang et al*. (2009), *Hu et al*. (2010), *Zhang et al.* (2011) and *Weng et al*. (2011) while a comprehensive review of mesoscale applications of the EnKF is presented in *Meng and Zhang* (2011). *Stuart et al*. (2007) studied the use of ensemble-based Kalman filtering of chemical observations for constraining meteorological uncertainties and for selecting target observation locations following the method proposed by *Hamill and Snyder* (2002). They further demonstrated the potential usefulness of the above method for locating promising adaptive observations in a predictive model. However, they did not test the strategy through assimilating the additional observations at the selected locations. *Liu and Kalnay* (2008) proposed an ensemble sensitivity method, which was conducted within an ensemble Kalman filter, to measure observation impact on the reduction of forecast errors due to assimilation of observations. *Liu et al*. (2009) further investigated the analysis sensitivity, which is proportional to the analysis error and anti-correlated with the observation error. The purpose of this study is to apply and extend the method developed by *Hamill and Snyder* (2002) to determine the meteorological targeted

observation locations for improving ozone prediction in Houston and the surrounding area using a real-world NWP model (MM5) and photochemical model (CMAQ/Models3). In addition, the efficiency of the targeted observations is investigated using EnKF data assimilation under the perfect-model assumption.

A high-ozone event that occurred on 30 August during the Texas Air Quality Study of 2000 (TexAQS2000) is chosen for study. An analysis of the meteorological and ozone conditions on this day is provided by *Banta et al.* (2005), and mesoscale simulations of the meteorology and photochemistry on this day have been reported by *Bao et al.* (2005), *Zhang et al.* (2007), and *Cheng and Byun* (2008a, 2008b). The day featured westerly and northwesterly winds in the morning that died down by early afternoon. During mid- to late-afternoon, bay and sea breezes developed along the immediate coastline, and hourly-averaged ozone levels of nearly 200 ppbv were recorded between Galveston Bay and downtown Houston. Airborne lidar observations and model simulations indicate that very high concentrations of ozone were present over Galveston Bay, and photochemical model simulations also produce high ozone concentrations over the nearshore Gulf of Mexico east and southeast of the Houston-Galveston area. Skies were generally clear, and no deep convection or precipitation was present.

The initial and boundary conditions used in the meteorological model are interpolated from the Eta model's 3-hourly gridded analysis for the Global Energy and Water Cycle Experiment (GEMEX) Continental-Scale International Project (GCIP). Since the GCIP analysis for the initial conditions has already incorporated the conventional observation data from the current fixed network, the targeted observation locations we mention here and after are actually single

sounding observations embedded within a generic synoptic-scale observing network. In addition, the method for selecting targeted observation locations is not specific to Houston and its surrounding area.

The methodology for choosing targeted observation locations is given in Section 2. Section 3 describes observation targeting results and their sensitivity to the data treatment, resolutions and planetary boundary layer (PBL) parameterization schemes. Section 4 investigates the effectiveness of targeted observations, and the summary and discussion are provided in Section 5.

2. Methodology for choosing targeted observation locations

In this study, we try to explore the targeted meteorological observation locations for improving ozone prediction through both meteorological and photochemical ensemble forecasts.

2.1 Model Configurations

The Pennsylvania State University-National Center for Atmospheric Research fifth-generation nonhydrostatic mesoscale model (MM5) version 3 (*Dudhia* 1993) is used to run the meteorological simulations. Horizontal grid spacings of 12-km and 4-km are used in the coarse domain (D1) and fine domain (D2) (Fig. 1), with one-way nesting. There are 43 vertical layers in the terrain-following coordinate system, with the model top at 50 hPa and vertical spacing smallest within the boundary layer. The MRF boundary layer parameterization scheme (*Hong and Pan*, 1996) and the simple ice microphysical scheme (*Dudhia*, 1993) are used for

both the 12-km and 4-km domain. The cumulus scheme of *Grell* (1993) with the shallow cumulus option is used only for the 12-km domain while convection in the 4-km domain is fully explicit. The MM5 simulations use the 24 land-use categories created from the 30-second USGS global land cover data, but a separate land-surface model or urban canopy parameterization are not used.

The MM5 simulations consist of 21 ensemble members initialized at 0000 UTC 30 August 2000 (1800 LST 29 August 2000) and integrated for 24 h. The ensemble members are initialized with the climatological ensemble initialization method of *Aksoy et al*. (2006) and Zhang et al. (2007) in which dynamically consistent initial and boundary conditions are statistically sampled from the Eta model's 3-hourly gridded (40-km) analyses for the Global Energy and Water Cycle Experiment (GEWEX) Continental-Scale International Project (GCIP). Departures from the initial and boundary condition ensemble means are scaled down by 20% to reduce the ensemble spread over the Houston area to 0.4-0.6 m s^{-1} for horizontal wind components and 0.7-0.8 K for temperature. The scaled perturbations are added to the unperturbed initial and boundary conditions from the GCIP analyses for the model simulation period beginning at 0000 UTC 30 August which are used for the 12-km domain ensemble simulation.

The output from the meteorological ensemble simulations is used to drive a 21-member photochemical ensemble simulation using the EPA photochemical model CMAQ/Model-3 (*Byun and Ching* 1999) with the CBIV gas-phase chemical mechanism (*EPA*, 2003). The CMAQ model uses the same two horizontal nests as the MM5 simulations but employs just 21 vertical layers with the lowest three levels at approximately 21, 64 and 106 m. Anthropogenic and

biogenic emissions are directly downloaded from the online emission inventory of the Texas Commission on Environmental Quality (TCEQ) (ftp.tceq.state.tx.us) and converted to CMAQ-ready emission files. Initial chemistry conditions are obtained from a 24-hour spinup run at 12-km grid spacing initialized at 00Z 29 Aug 2000. Further details regarding the model configurations are found in *Zhang et al*. (2007).

2.2 Targeted observations sensitivity assessment

The method adopted for determining targeted observation locations is similar to that developed by *Hamill and Snyder* (2002), which follows closely from the theory of *Berlinear et al*. (1999). Its algorithm is mathematically identical to the ETKF of *Bishop et al*. (2001) and is also closely related to the ensemble sensitivity analysis in *Zhang* (2005), *Hawblitzel et al*. (2007), *Torn and Hakim* (2008), and *Sippel and Zhang* (2008, 2010). In this method, the norm used for the total decrease in the model uncertainty is the sum over all state variables of the individual differences in variances, or the trace of $P^b - P^a$. $P^b - P^a$ is written as (*Hamill and Snyder*, 2002):

$$P^b - P^a = P^b H^T \left(H P^b H^T + R \right)^{-1} H P^b \tag{1}$$

where,

$$P^b = \left\langle \left(x^t - x^b \right) \left(x^t - x^b \right)^T \right\rangle \tag{2}$$

$$P^a = \left\langle \left(x^t - x^a \right) \left(x^t - x^a \right)^T \right\rangle \tag{3}$$

P^a and P^b are the post- and pre-analysis background error covariance matrices, respectively. H is the operator matrix relating the model state variables to the observational variables, superscript ^{T}

means matrix transpose, and R is the observational error covariance matrix, which includes both the instrument error and the representation error. x^a is the m-dimensional analyzed state vector, x^b is the background state, and x^t is the true state, which is associated with the observations through the following formula:

$$Y_0 = Hx^t + \varepsilon, \quad \varepsilon \sim N(0,R) \tag{4}$$

Because the true state is not known, here we estimate the background error covariance using the ensemble covariance matrices around the ensemble mean. Thus P^b in (2) can be estimated by

$$P^b = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i^b - \overline{x^b}\right)\left(x_i^b - \overline{x^b}\right)^T = X^b(X^b)^T \tag{5}$$

where the subscript i denotes ensemble members, n represents the ensemble size. Substituting (5) to (1), we obtain

$$P^b - P^a = X^b\left(HX^b\right)^T\left[HX^b\left(HX^b\right)^T + R\right]^{-1}HX^b\left(X^b\right)^T \tag{6}$$

In this study, we intend to choose the locations for additional observations with the goal of minimizing the forecast-error variance. It is necessary to compare the forecasts from the initial condition with and without additional observations. If we mark quantities relating to these two forecasts by superscripts $f|a$ and $f|b$, respectively, the change in the forecast-error variance due to the additional observations is tr($P^{f|a} - P^{f|b}$). When the analysis errors are not too large, we have

$$P^{f|b} - P^{f|a} \sim M\,(P^b - P^a)M^T.$$

Here M is the linearization of the nonlinear forecast operator M (*Bishop et al.,* 2001; *Majumdar et al*., 2001). From (6), we can get

$$P^{f|a} - P^{f|b} \approx MX^b \left(HX^b\right)^T \left[HX^b \left(HX^b\right)^T + R\right]^{-1} HX^b \left(MX^b\right)^T \qquad (7)$$

Considering the ensemble forecasts from the background state, $x_i^{f|b} = M(x_i^b)$ for $i = 1, \ldots, n$. MX^b in (7) can be replaced by $X^{f|b}$ with the same accuracy, and HX^b can be replaced by Y, which represents the observation located at the model grid point, so (7) becomes

$$P^{f|a} - P^{f|b} \approx \left(X^{f|b} \cdot Y^T\right)\left[Y \cdot Y^T + R\right]^{-1}\left[Y \cdot \left(X^{f|b}\right)^T\right] \qquad (8)$$

where $X^{f|b}$ represents model state variables and Y represents observation variables. The symbol "·" means covariance between two variables. In this particular study, $X^{f|b}$ represents photochemical model output variables (such as the surface ozone concentration) or meteorological model output variables (such as u, v, and T), and Y represents meteorological observation variables at model grids (such as u, v and T at all vertical layers). R represents observation error variances of u, v and T. The observational error variances at vertical layers used in this study are interpolated from those provided by *Parrish and Derber* [1992], and are assumed to be uncorrelated between vertical levels and different variables (Table 1).

The trace of ($P^{f|b} - P^{f|a}$) can be computed for each additional observation location candidate (each H). Like *Stuart et al*. (2007), we define the trace of ($P^{f|b} - P^{f|a}$) normalized by its domain-wide maximum value as the observational impact factor. The locations with the maximum observational impact factor value are expected to be where additional observations can best improve ozone prediction or meteorological prediction, depending on which model variables are included in (8).

If we consider u, v, and T in an entire vertical column as one sounding observation and the surface O_3 concentration ([O_3]) as the model state variable (or the forecast variable), the above

formula can be used to calculate the observational impact factor of sounding observations on the surface [O_3]. Figure 1 shows the model domains (D1 and D2) for CMAQ/Models3, the domain (D_target) used for examining the effect of the targeted observations, the Houston area (innermost domain), and the locations (cross points) at which observational sounding data is extracted for assimilation. A total of 121 sensitivity experiments (one for each sounding) are conducted to examine the impact of each individual sounding as compared to the estimated impact derived from the ensemble sensitivity.

With the climatological ensemble initialization scheme employed here, the background error covariance is independent of information regarding previous observation locations. In practice, the background errors would be expected to be larger in places where observations are relatively sparse, such as over the Gulf of Mexico in this case, but such observation-sensitive background error information may not always be available. All other things being equal, the observation sensitivity would be higher than analyzed here where the background error covariance is large and smaller where the background error covariance is small.

3. Diagnosed Targeting Sensitivities

The ensemble mean forecast of ozone, temperature and wind at 21 UTC (15 LST, the approximate time of maximum ozone) is shown in Figure 2a for the 12-km run over D1 and Figure 2c for the 4-km run over D2. The highest values of ozone are located over Galveston Bay and the adjacent Gulf of Mexico, but a tongue of high ozone concentrations extends westward from Galveston Bay toward downtown Houston. The winds have weakened and shifted from an

offshore land breeze to an onshore bay/sea breeze. The ensemble mean ozone at 21 UTC is 120 ppb in the Houston area and is as high as over 150 ppb over Galveston Bay.

Figures 2b and 2d show the simulations of ozone, temperature, and wind from a separate model run (the "truth" run) that is the source of the simulated observations for the experiments described in Section 4. Because this is a single deterministic run, the ozone field has a more detailed structure than in the ensemble mean. In the context of this experiment, improvements in the model forecasts are only possible in places where the ensemble mean (Fig. 2a, c) and the truth run (Fig. 2b, d) disagree at 21 UTC.

Using the method described in Section 2, we compare the observational impact factors derived from different treatments of forecast variables (Houston urban ozone, domain-wide ozone, and meteorological variables), observation data (sounding vs. single layer), different model resolutions (coarse vs. fine resolutions), and different PBL schemes (single- vs. multi-scheme). Multi-scheme hereafter means ensembles with 5 different PBL schemes: MRF PBL (*Hong and Pan*, 1996), High-resolution Blackadar PBL (*Blackadar*, 1979), Burk-Thompson PBL (*Burk and Thompson*, 1989), Eta PBL (*Janjic*, 1990; 1994), and Gayno-Seaman PBL (*Ballard et al*., 1991; *Shafran et al*. 2000).

The spatial and temporal evolution of the observational impact factor for sounding observations on the peak time surface [O_3] averaged over Houston area (the inner box) is shown in Figure 3, along with the ensemble mean of the surface wind and temperature over the simulation period. Initially (00 UTC or 18 LST, Fig. 3a), the largest observational impact factor is found over land near the coast, where the ensemble mean of the wind direction is almost

perpendicular to the coastline and the gradient of the ensemble mean of the temperature has its maximum value. This shows that the location and intensity of the previous day's sea-breeze front has an important correlation in the model simulations with the formation of the next day's urban [O_3]. Another large impact factor area is located in the northeast corner of the domain. At 06 UTC (0 LST, Fig. 3b), the largest observational impact values are farther inland, though still collocated with the stronger winds and remnant temperature gradient of the sea breeze. At 12 UTC (6 LST, Fig. 3c), the greatest impact area is located near and upstream of Houston. Finally, at 18 UTC (12 LST, Fig. 3d), the area of the largest observational impact factors are generally smaller than before, though an area with large impact factor persists along the coast east and southeast of Houston. This area perhaps reflects the importance of the timing and magnitude of the local sea breeze on the ozone concentrations in the Houston area.

Figure 4 shows the observational impact factor for 21 UTC ozone concentrations throughout domain D1. The sensitivities at 00 UTC (Fig. 4a) are quite similar to those for Houston area ozone only (Fig. 3a), suggesting that observational influences on the ozone forecast this far in advance are fairly large-scale and not specific to the ozone in any particular region. As the verification time approaches, differences between the domain-wide impacts and Houston area impacts become larger. In general, the domain-wide ozone is more sensitive to sounding observations in the northern portion of the domain.

If the targeted sounding is to be assimilated by a sequential data assimilation system that ingests observations of u, v and T at each layer separately, the covariances and correlations between the observations on different layer may not properly be taken into account in the

calculation of the observational impact factor. When the multiple observations within a single sounding are redundant for analysis purposes, the impact factor for the multiple "single-layer" observations can be estimated as the average of the observational impact factor over each layer. Figure 5 shows the average observational impact factor for single-layer observations on the peak time surface [O_3] over D1. In general, the distribution of the observational impact factor is consistent with that for the sounding observations treated as a whole (Fig. 4), while the impacts themselves are smaller. However, the average single-layer impact factor is much smaller than the full-sounding impact factor in the coastal areas, especially at later assimilation times.

The large difference between single-layer and full-sounding impact factors in coastal areas is probably due to the importance of the sea breeze circulation for determining ozone concentrations. Only a small portion of a coastal or offshore sounding will sample the sea breeze flow or return circulation, and the rest of the observations aloft will be of relatively little value. In such a circumstance, averaging of impact factors from the various levels in the sounding would underestimate the impact of the sounding observation. In contrast, farther inland, an observation from anywhere in the lower troposphere would be sufficient to improve the model's analysis of the large-scale flow opposing the sea breeze and controlling its inland penetration. Here, averaging comes closer to the impact factor computed from a full sounding. This demonstrates that the impact of the multiple data points within a sounding should be estimated collectively rather than as a set of independent observations.

Since the model resolution contributes to the quality of numerical weather forecast, we investigate the observational impact factor derived from the high resolution (4-km) ensemble

simulations. Figure 6 shows the observational impact factor for sounding observations on the peak time surface [O_3] within domain 2 calculated from 4-km ensemble simulations (D2 in Fig. 1) at different times. The evolutions of the observational impact factor in the 4-km run is chiefly consistent with those from the 12-km run, however, there exist more fine structures along the coastline, especially at 06 and 18 UTC (00 LST and 12 LST), indicating the significant contributions of the local circulation in the coastal area to the formation of the peak time surface [O_3].

Ozone prediction is also sensitive to various PBL schemes used in the meteorological model on an urban scale (*Mao et al*., 2007, *Bei et al*. 2010). We have compared the observational impact factors calculated from the ensemble simulations using a single-PBL scheme versus multi-PBL schemes, respectively (Figure 7). They exhibit similar basic patterns (compare to Fig. 6), but with large differences in the details of location and magnitude, especially along the coastline. It shows that the targeting depends upon details of the forecast system used in the data assimilation scheme, so we cannot use one technique for identifying target locations and another technique for assimilating the additional data. It also further shows that model errors besides model resolution should also be considered in the targeted observation network design.

4. Examination of the Improvement from Targeted Observations

The most interesting question is whether the assimilation of an additional observation actually reduces the forecast error in line with the calculated impact factor. We have conducted a group of experiments in an initial attempt to address this question. In each experiment, a synthetic

(fake) observational sounding extracted from the truth run (a perfect-model identical twin experiment) has been assimilated. Random uncorrelated noise of a particular magnitude (see Table 1) is added to each observation. The synthetic soundings are extracted from the places indicated by crosses in Fig. 1. Each synthetic sounding is assimilated using a standard Ensemble Kalman Filter algorithm, without covariance localization, inflation of member deviations, or hybridization. The EnKF used in the current study follows closely that of *Snyder and Zhang* (2003). As in the standard Kalman filter,

$$x^a = x^f + K(y - Hx^f),$$

where x^f represents the prior estimate or first guess, x^a is the posterior estimate or analysis, **y** is the observation vector, H is the observation operator that returns observed variables given the state, and K is the so-called Kalman gain matrix defined as

$$K = P^f H^T (HP^f H^T + R)^{-1},$$

where P^f and R represent the background and observational error covariance, respectively. In the EnKF, the flow-dependent P^f is estimated through an ensemble of short-range forecasts. Observations are taken sequentially with observations errors assumed to be uncorrelated. Further background on the EnKF refer to *Snyder and Zhang* (2003) and references therein.

The assimilation is performed at 00 h. For each experiment, we calculate the actual reduction in the root mean squared error (with respect to the truth run) of the surface wind speed (hereafter "rms-SPD") and the surface [O_3] (hereafter "rms-O_3") according to:

$$rms = \left\| \bar{x}^{f|b} - x^t \right\| - \left\| \bar{x}^{f|a} - x^t \right\| \quad (9)$$

We also identify the square root of the observational impact factor with the expected reduction in

the root mean squared error. The actual reduction in the rms-SPD over D1 and the actual reduction in the rms-O_3 over the Houston area are calculated and compared to their expected reduction.

Figure 8 shows the scatter plot of the expected and actual reduction in the rms-SPD at 21 UTC (15 LST) when the observations were assimilated at 00 UTC (18 LST). The rms-SPD at 21UTC is reduced when any of the synthetic soundings are assimilated at the model initial time. In general, a larger expected reduction in rms-SPD is associated with a larger actual reduction in rms-SPD. However, the expected reduction is always overestimated. The correlation is only 0.13, suggesting the value of the impact factor diagnosis is limited, at least in this case. One possible reason is that the background-error covariances provided to the EnKF may not be perfect since the ensemble size is limited. Also, the initial ensemble perturbations, despite being dynamically consistent, are derived directly from climatological uncertainties in this location, which may not have sufficient flow dependent statistics at the initial time. In addition, the actual error reduction (through EnKF assimilation of the synthetic soundings) also depends on the magnitude of the error (without assimilation) while the predicted impact factor may not.

Figure 9 shows the horizontal distributions of the expected and actual reduction in the rms-SPD. Both expected and actual reductions in rms-SPD are largest for synthetic soundings ingested from southwest and northeast of Houston. The actual reduction is much smaller than the expected reduction when fake soundings are assimilated in the eastern margin of the domain.

Reductions in meteorological errors should lead to reductions in ozone errors. However, ozone errors, particularly in the Houston area, should depend partly on the initial configuration

and transport of ozone and its precursors and partly on the interaction of the meteorological conditions with emissions during model integration. The latter process is not explicitly considered in the estimation of expected error.

A scatter plot of the expected and actual reduction in the rms-O_3 over the Houston area is shown in Figure 10. The rms-O_3 over the Houston area is reduced by assimilating any of the fake soundings extracted from the truth run. The average actual ozone impact is similar to the average expected ozone impact. The correlation between the expected and actual reduction in rms-O_3 is 0.23, which is still low but is slightly higher than for rms-SPD. For the low correlation between the expected and actual reduction in rms-O3, one possible reason is the non-linear response of the ozone concentration uncertainty to the meteorological field uncertainty (*Zhang et al*. 2007); another is the lack of consideration of ozone emissions mentioned earlier. On the other hand, the better estimate of the magnitude of the impact factor for ozone than for meteorological variables may be due to a greater dependence of O_3 on conditions throughout the forecast period, including the time at which the impact factor estimate is made, since O_3 at verification time is sensitive to transport and dilution that takes place beginning at analysis time and continuing throughout the forecast period.

The horizontal distributions of the expected and actual reduction in the rms-O_3 (Figure 11) show that both the expected and actual reduction in rms-O_3 are largest for synthetic soundings ingested from west and southwest of Houston. The actual reduction tends to be much smaller than the expected reduction when synthetic soundings are assimilated near certain boundaries, particularly the northern, eastern, and southwestern margins of the domain.

5. Summary and Discussion

An ensemble-based method for determining targeted observation locations developed by *Hamill and Snyder* (2002) has been extended and applied for improving ozone predictions in Houston and the surrounding area. In this method, an observational impact factor has been calculated from an ensemble run and employed to select the targeted observation locations. The locations with larger observational impact factors are expected to be where supplementary observations can better reduce the forecast error. The improvement brought by the targeted observations is further examined through an Ensemble Kalman Filter assimilation of synthetic observations.

The computational cost of estimating the observational impact factor as proposed here is only slightly more than the cost of performing the set of ensemble forecasts on which the calculation is based. Calculation of actual impact factor through assimilation of test observations requires one EnKF analysis and set of ensemble forecasts for each test observation, which rapidly becomes prohibitively expensive as the number of test observations increases.

The observational impact factor for meteorological observations on the peak time surface [O_3] evolves with time. Initially, the larger impact factor is mainly located at the landside of the coastline and the northeast portion of the domain. The larger impact factor area near the coastline moves to the north and then northeast at later assimilation times with the increase of south and southwest wind. Close to verification time, the observational impact factor along the coastline increases. This shows that the observational impact factor field is changed with the evolution of the wind circulation.

The observational impact factor from considering sounding data for different layers separately on the peak time surface [O_3] is different from that for the sounding observations over the entire column as a whole, indicating that the targeting strategies should be based on proper consideration of the impact of the sounding in its entirety on the forecast.

The observational impact factor derived from ensemble simulations with different resolution basically has the same pattern except for some minor differences in the distribution. Ensembles created with different PBL schemes can also produce differences in the detailed distribution and magnitude of the calculated observational impact factor fields, especially along the coastline, which indicates the model error due to the PBL scheme should also be included in the observation targeting. It also suggests we should not use one technique for identifying target locations and another technique for assimilating the additional data.

Through assimilating synthetic observational soundings extracted from the truth run using an Ensemble Kalman Filter, the surface wind speed forecasts are improved in the root mean square sense when any of the synthetic observational soundings is assimilated at the model initial time. In general, larger observational impact factors are associated with larger reductions in rms-SPD. However, at the peak ozone time (21 UTC), the correlation between the expected and actual reduction in the rms-SPD is low (0.13). The lack of agreement between expected and actual observation impacts could be produced by the data assimilation scheme, without using the localization in EnKF, the noise of the observation data, sampling ensemble with a limited ensemble size, and the limit of meteorological predictability (*Morss and Emanuel,* 2002; Zhang et al. 2007). The correlation between the expected and actual reduction in rms-O_3 is also low but

better than of rms-SPD, which can be partly explained by the non-linear response of the ozone concentration uncertainty to the meteorological field uncertainty (*Zhang et al*., 2007). Other reasons (such as lateral boundary influence) may also affect the results.

Acknowledgments: The authors are grateful to Yonghui Weng and Zhiyong Meng for help on the EnKF implementation. This research was started when both NB and FZ were employed at the Texas A&M University. This research is partially funded by GTRI/HARC Grant Project No. H24-2003 and NSF grant ATM-084065.

References

Aksoy, A., F. Zhang, and J. W. Nielsen-Gammon, 2006: Ensemble-based simultaneous state and parameter estimation with MM5, *Geophys. Res. Lett.*, **33**, L12801, doi:10.1029/2006GL026186.

Baker, N. L., and R. Daley, 2000: Observation and background adjoint sensitivity in the adaptive observation-targeting problem. *Quart. J. Roy. Meteor. Soc.,* **126,** 1431-1454.

Ballard, S. P., B. W. Golding, and R. N. B. Smith, 1991: Mesoscale model experimental forecasts of the Haar of northeast Scotland. *Mon. Wea. Rev.*, **119**, 2107-2133.

Banta, R. M., C. J. Senef, J. Nielsen-Gammon, L. S. Darby, T. B. Ryerson, R. J. Alvarez, S. P. Sandberg, E. J. Williams, and M. Trainer, 2005: A bad air day in Houston. Bull. *Amer. Meteor. Soc*., **86**, 657-659.

Bao, J.-W., S. A. Michelson, S. A. McKeen, and G. A. Grell, 2005: Meteorological evaluation of a weather-chemistry forecasting model using observations from the TEXas AQS 2000 field experiment. *J. Geophys. Res.*, **110**, D21105, doi:10.1029/2004JD005024.

Bei, N., Lei, W., Zavala, M., and Molina, L. T., 2010: Ozone predictabilities due to meteorological uncertainties in Mexico City Basin using ensemble forecasts, *Atmos. Chem. Phys.*, 10, 6295-6309.

Bergot, T., 2001: Influence of the assimilation scheme on the efficiency of adaptive observations. *Quart. J. Roy. Meteor. Soc.,* **127,** 635-660.

Bergot, T., G. Hello, A. Joly, and S. Malardel, 1999: Adaptive observations: A feasibility study. *Mon. Wea. Rev.,* **127,** 743-765.

Berliner, L. M., Z.-Q. Lu, and C. Snyder, 1999: Statistical design for adaptive weather observations. *J. Atmos. Sci.,* **56,** 2536–2552.

Bishop, C. H., and Z. Toth, 1999: Ensemble transformation and adaptive observations. *J. Atmos. Sci.,* **56,** 1748-1765.

Bishop, C. H., B. J. Etherton, and S. Majumjar, 2001: Adaptive sampling with the ensemble transform Kalman filter. Part 1: Theoretical aspects. *Mon. Wea. Rev.,* **129,** 420-436.

Blackadar, A.K., 1979: High resolution models of the planetary boundary layer, *Adv. Environ. Sci. and Eng.*, **1**, 50-85.

Buizza, R., and A. Montani, 1999: Targeting observations using singular vectors. *J. Atmos. Sci.,***56,** 2965-2985.

Burgers, G., P. J. van Leeuwen, and G. Evensen, 1998: Analysis scheme in the ensemble Kalman filter. *Mon. Wea. Rev.,* **126,** 1719–1724.

Burk, S. D., and W. T. Thompson, 1989: A vertically nested regional numerical prediction model with second-order closure physics. *Mon. Wea. Rev.*, **117**, 2305-2324.

Byun, D.W. and J.K.S. Ching, ed., (1999), Science Algorithms of the EPA Models-3 Community Multi-scale Air Quality (CMAQ) Modeling System, EPA Report, EPA/600/R-99/030, NERL, Research Triangle Park, NC.

Chang, M.E., D. E. Hartley, C. Cardelino, D. Haas-Laursen, and W. L. Change, 1997: On using inverse methods for resolving emissions with large spatial in homogeneities. *J. Geophys. Res.,* **102***,* 16023-16036.

Cheng, F.-Y., and D. W. Byun, 2008a: Application of high resolution land use and land cover data for atmospheric modeling in the Houston-Galveston metropolitan area, Part I: Meteorological simulation results. *Atmos. Env.*, **42**, 7795-7811, doi:10.1016/j.atmosenv.2008.04.055.

Cheng, F.-Y., and D. W. Byun, 2008b: Application of high resolution land use and land cover data for atmospheric modeling in the Houston-Galveston metropolitan area, Part II: Air quality simulation results. *Atmos. Env.*, **42**, 4853-4869,doi:10.1016/j.atmosenv.2008.02.059.

Daescu, D. N., and G. R. Carmichael, 2003: An adjoint sensitivity method for the adaptive location of the observations in air quality modeling. *J. Atmos. Sci.,* **60***,* 434-450.

Dudhia, J., 1993: A nonhydrostatic version of the Penn State-NCAR Mesoscale Model: Validation tests and simulation of an Atlantic cyclone and cold front. *Mon. Wea. Rev*., **121**, 1493-1513.

Elbern, H., H. Schmidt, O. Talagrand, and A. Ebel, 2000: 4-D variational data assimilation with an adjoint air quality model for emission analysis. *Environmental Modelling & Software, **15**,* 539-548.

Emanuel, K. A., and Coauthors, 1995: Report of the first prospectus development team of the U.S. Weather Research Program to NOAA and the NSF. *Bull. Amer. Meteor. Soc.,* **76,** 1194-1208.

Gelaro, R., R. Langland, G. D. Rohaly, and T. E. Rosmond, 1999: An assessment of the singular vector approach to target observations using the FASTEX data set. *Quart. J. Roy. Meteor. Soc.,* **125,** 3299-3327.

Gelaro, R., C. A. Reynolds, R. H. Langland, and G. D. Rohaly, 2000: A predictability study using geostationary satellite wind observations during NORPEX. *Mon. Wea. Rev.,* **128,** 3789-3807.

Grell, G. A., 1993: Prognostic evaluation of assumptions used by cumulus parameterizations. *Mon. Wea. Rev.*, **121**, 764-787.

Environmental Protection Agency (EPA) (2003), Models-3 Community Multiscale Air Quality (CMAQ) model, version 4.3, report, Research Triangle Park, N. C. (Available at http://www.cmascenter.org/modelclear. shtml).

Evensen G., 1994: Sequential data assimilation with a non-linear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.*, **99**, 10143-10162.

Evensen G., 2003: The Ensemble Kalman Filter: theoretical formulation and practical implementation. *Ocean Dynamics*, 53, 343-367.

Hamill, T. M. and C. Snyder, 2002: Using improved background-error covariances from an ensemble Kalman filter for adaptive observations. *Mon. Wea. Rev.,* **130***,* 1552 -1572.

Hanea, R. G., G. J. M. Velders, and A. Heemink, 2004: Data assimilation of groundlevel ozone in Europe with a Kalman filter and chemistry transport model. *J. Geophys. Res.,* **109**, D10302, doi:10.1029/2003JD004283.

Hansen, J. A., and L. A. Smith, 2000: The role of operational constraints in selecting supplementary observations. *J. Atmos. Sci.,* **57,** 2859-2871.

Hawblitzel, D. P., F. Zhang, Z. Meng, and C. A. Davis, 2007: Probabilistic evaluation of the dynamics and predictability of the mesoscale convective vortex of 10-13 June 2003. *Mon. Wea. Rev.,* **135**, 1544-1563.

Heemink, A. W., and A. J. Segers, 2002: Modeling and prediction of environmental data in space and time using Kalman filtering. *Stochastic Environmental Research and Risk Assessment,* **16***,* 255-240.

Hong, S.-Y. and H.-L. Pan, 1996: Nonlocal boundary layer vertical diffusion in a medium-range forecast model, *Mon. Wea. Rev.*, **124**, 2322-2339.

Hu, X., F. Zhang**,** and J. W. Nielsen-Gammon, 2010: Ensemble-Based Simultaneous State and Parameter Estimation for Treatment of Mesoscale Model Error: A Real-data Study. *Geophys. Res. Lett.*, **37**, L08802.

Janjic, Z. I., 1990: The step-mountain coordinate: physical package. Mon. Wea. Rev., **118**, 1429-1443.

Janjic, Z. I., 1994: The step-mountain eta coordinate model: further developments of the convection, viscous sublayer and turbulence closure schemes. Mon. Wea. Rev., **122**, 927-945.

Kang, J., 2009: Carbon Cycle Data Assimilation Using a Coupled Atmosphere-Vegetation Model and the Local Ensemble Transform Kalman Filter, Ph. D Thesis, University of Maryland.

Kang, J.-S., E. Kalnay, J. Liu, I. Fung, T. Miyoshi, and K. Ide, 2011: "Variable localization" in an Ensemble Kalman Filter: application to the carbon cycle data assimilation. J. Geophys. Res.,doi:10.1029/2010JD014673, in press.

Langland, R., R. Gelaro, G. D. Rohaly, and M. A. Shapiro, 1999: Target observations in FASTEX: Adjoint based targeting procedures and data impact experiments in IOP/7 and IOP/8. *Quart. J. Roy. Meteor. Soc.,* **125,** 3241-3270.

Liu, J., and E. Kalnay, 2008: Estimating observation impact without adjoint model in an ensemble Kalman filter. *Quart. J. Roy. Meteor. Soc.,* **134**, 1327-1335.

Liu, J., E. Kalnay, T. Miyoshi, and C. Cardinali, 2009: Analysis sensitivity calculation in an ensemble Kalman filter, *Quart. J. Roy. Meteor. Soc.,* **135**, 1842-1851.

Lorenz, E. N. and K. A. Emanuel, 1998: Optimal sites for supplementary observations: Simulation with a small model. *J. Atmos. Sci.,* **55,** 399-414.

Majumdar, S. J., C. H. Bishop, I. Szunyogh, and Z. Toth, 2001: Can an Ensemble Transform Kalman Filter predict the reduction in forecast error variance produced by targeted observations? *Quart. J. Roy. Meteor. Soc.,* **127,** 2803-2820.

Majumdar, S. J., S. D. Aberson, C. H. Bishop, R. Buizza, M. S. Peng and C. A. Reynolds, 2006: A comparison of Adaptive Oberving Guidance for Atlantic Tropical Cyclones. *Mon. Wea. Rev.*, **134**, 2354-2372.

Mao Q., L. L. Gautney, T. M. Cook, M. E. Jacobs, S. N. Smith, and J. J. Kelsoe: 2006, Numerical experiments on MM5-CMAQ sensitivity to various PBL schemes. *Atmos. Environ.,* **40**, 3092-3110.

Meng, Z., and F. Zhang, 2007: Tests of an ensemble Kalman filter for mesoscale and regional-scale data assimilation. Part II: Imperfect Model Experiments. Mon. Wea. Rev., **135**, 1403-1423.

Meng, Z., and F. Zhang, 2008a: 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. Wea. Rev.*, **136**, 522-540.

Meng, Z., and F. Zhang, 2008b: Tests of an Ensemble Kalman Filter for Mesoscale and Regional-Scale Data Assimilation. Part IV: Comparison with 3DVar in a Month-Long Experiment. *Mon. Wea. Rev.*, **136**, 3671-3682.

Meng, Z., and F. Zhang, 2011: Limited-area ensemble-based data assimilation. *Mon. Wea. Rev.*, in press.

Mendoza-Dominguez, A. and A. G. Russell, 2001: Estimation of emission adjustments from the application of four-dimensional data assimilation to photochemical air quality modeling. *Atmos. Environ.,* **35,** 2879-2894.

Morss, R. E., 1998: Adaptive observations: Idealized sampling strategies for improving numerical weather prediction. Ph.D. dissertation, Massachusetts Institute of Technology, 225 pp. [Available from UMI Dissertation Services, P. O. Box 1346, 300 N. Zeeb Rd., Ann Arbor, MI, 48106-1346.]

Morss, R. E., and K. A. Emanuel, 2002: Influence of added observations on analysis and forecast errors: Results from idealized systems. *Quart. J. Roy. Meteor. Soc.,* **128,** 285-322.

Morss, R. E., and K. A. Emanuel, and C. Snyder, 2001: Idealized adaptive observation strategies

for improving numerical weather prediction. *J. Atmos. Sci.,* **58,** 210-234.

Palmer, T. N., R. Gelaro, J. Barkmeijer, and R. Buizza, 1998: Singular vectors, metrics, and adaptive observations. *J. Atmos. Sci.,* **55,** 633-653.

Parrish S. and J. Derber, 1992: The national meteorological center's spectral statistical-interpolation analysis system, *Mon. Wea. Rev.*, **120**, 1747-1763.

Pu, Z.-X., and E. Kalnay, 1999: Targeting observations with the quasilinear inverse and adjoint NCEP global models: Performance during FASTEX. *Quart. J. Roy. Meteor. Soc.,* **125,** 3329-3337.

Pu, Z.-X., E. Kalnay, J. Sela, and I. Szunyogh, 1997: Sensitivity of forecast errors to initial conditions with a quasi-inverse linear model. *Mon. Wea. Rev.,* **125,** 2479-2503.

Sippel, J. A., and F. Zhang, 2008: A probabilistic analysis of the dynamics and predictability of tropical cyclogenesis. *J. Atmos. Sci.,* **65,** 3440-3459.

Sippel, J. A., and F. Zhang, 2010: Factors Affecting the Predictability of Hurricane Humberto (2007). *J. Atmos. Sci.*, **67**, 1759-1778.

Shafran, P. C., N. L. Seaman, and G. N. Gayno, 2000: Evaluation of numerical predictions of boundary layer structure during the Lake Michigan ozone study, J. Appl. Meteor., **39**, 412-426, 2000.

Snyder, C. and F. Zhang, 2003: Assimilation of simulated Doppler radar observations with an ensemble Kalman filter. *Mon. Wea. Rev.*, **131**, 1663-1677.

Stuart, A. L., A. Aksoy, F. Zhang, and J. W. Nielsen-Gammon, 2007: Ensemble-based data assimilation and targeted observation of a chemical tracer in a sea breeze model. *Atmos. Environ*., **41**, 3082-3094.

Szunyogh, I., Z. Toth, K. A. Emanuel, C. H. Bishop, C. Snyder, R. E. Morss, J. Woolen, and T. Marchok, 1999: Ensemble–based targeting experiments during FASTEX: the effect of dropsonde data from the Lear jet. *Quart. J. Roy. Meteor. Soc.,* **125**, 3189-3218.

Torn, R. D., and G. J. Hakim, 2008: Ensemble-based Sensitivity Analysis. *Mon. Wea. Rev.*, **136**, 663-677.

Van Loon, M., P. J. H. Builtjes, A. J. Segers, 2000: data assimilation of ozone in the atmospheric transport chemistry model LOTUS. *Environ Modeling Software*, **15**, 603-609.

Weng, Y., M. Zhang, and F. Zhang, 2011: Advanced Data Assimilation for Cloud-Resolving Hurricane Initialization and Prediction. *Computing in Science and Engineering*, **13**, 40-49.

Zhang, F., 2005: Dynamics and Structure of Mesoscale Error Covariance of a Winter Cyclone Estimated through Short-Range Ensemble Forecasts. *Mon. Wea. Rev.*, **133**, 2876-2893.

Zhang, F., C. Snyder, and J. Sun, 2004: Impacts of initial estimate and observation availability on convective-scale data assimilation with an ensemble Kalman filter. *Mon. Wea. Rev.*,

132, 1238-1253.

Zhang, F., Z. Meng, and A. Aksoy, 2006: Tests of an ensemble Kalman filter for mesoscale and regional-scale data assimilation. Part I: Perfect Model Experiments. Mon. Wea. Rev., **134**, 722-736.

Zhang, F., N. Bei, J. W. Nielsen-Gammon, G. Li, R. Zhang, A. Stuart, and A. Aksoy, 2007: Impacts of meteorological uncertainties on ozone pollution predictability estimated through meteorological and photochemical ensemble forecasts. *J. Geophys. Res.*, **112**, D04304, doi:10.1029/2006JD007429.

Zhang, F.**,** Y. Weng, J. A. Sippel, Z. Meng, and C. H. Bishop, 2009: Cloud-resolving Hurricane Initialization and Prediction through Assimilation of Doppler Radar Observations with an Ensemble Kalman Filter. *Mon. Wea. Rev.,* **137**, 2105-2125.

Table 1. Observation error variances for temperature and u and v wind components.

Level	***Pressure (hPa)***	***T*** (K^2)	***u*** (m^2s^{-2})	***v*** (m^2s^{-2})
1	1010	3.24	1.96	1.96
2	1001	3.24	1.96	1.96
3	989	3.16	2.04	2.04
4	967	3.06	2.22	2.22
5	936	2.85	2.53	2.53
6	895	2.65	2.92	2.92
7	830	2.31	3.68	3.68
8	676	1.69	5.95	5.95
9	483	1.85	8.06	8.06
10	291	4.2	11.35	11.35
11	98	9.79	6.3	6.3

Figure Captions

Figure 1. Model domains used in photochemical ensemble simulations. The horizontal grid spacing of domain 1 (D1) and domain 2 (D2) is 12-km and 4-km, respectively. The box indicated by D_target denotes the domain used in Figs. 2, 3, 7, and 9. The innermost box denotes the domain used as the Houston area. Cross points indicate positions used to extract the profiles for assimilation. D1 grid points are numbered along the margins.

Figure 2. The ensemble mean forecast of ozone (ppb, shaded), temperature (Celsius, contoured) and wind (vectors) at 21 UTC (15 LST) in (a) D1 and (c) D2, and the simulations of ozone, temperature, and wind from the "truth" run in (b) D1 and (d) D2. Domain grid points are numbered along the margins.

Figure 3. The observational impact factor (shaded) for sounding observations on the peak time surface [O_3] (21 UTC) over the Houston area (the inner domain) along with the ensemble mean surface temperature (Celsius, contoured) and surface wind (vectors).

Figure 4. The observational impact factor (shaded) for sounding observations on the peak time surface [O_3] (21 UTC) over D1 along with the ensemble mean surface temperature (contoured) and surface wind (vectors) from 00 to 18 UTC. Inner domain indicates D2 and the innermost domain indicates the Houston area.

Figure 5. The observational impact factor (shaded) for single layer observations on the peak time surface [O_3] (21 UTC) over D1 along with the ensemble mean surface temperature (contoured) and surface wind (vectors) from 00 to 18 UTC. The inner domain indicates the Houston area.

Figure 6. Same as Fig. 4, but the observational impact factor (shaded) is derived from ensemble simulations in D2 (4-km run). The inner domain indicates the Houston area.

Figure 7. Same as Fig. 6, but the observational impact factor (shaded) is derived from ensemble simulations with multi-PBL schemes.

Figure 8. Scatter plot of the expected and actual reduction in the root mean squared error of surface wind speed (rms-SPD, m/s) over D1 at 21 UTC (15 LST). The solid (dotted) line denotes data fitted (ideal) line.

Figure 9. Horizontal distributions of (a) the expected reduction and (b) the actual reduction in rms-SPD (m/s) over D1 at ozone peak time through data assimilation at 00 UTC (6 LST).

Figure 10. Scatter plot of the expected and actual reduction in the root mean squared error of surface [O_3] (rms-O_3, ppb) over the Houston area at 21 UTC (15 LST). The solid (dotted) line denotes data fitted (ideal) line.

Figure 11. Horizontal distributions of (a) the expected reduction and (b) the actual reduction in rms-O_3 (ppb) over the Houston area at peak ozone time through data assimilation at 00 UTC (6 LST).

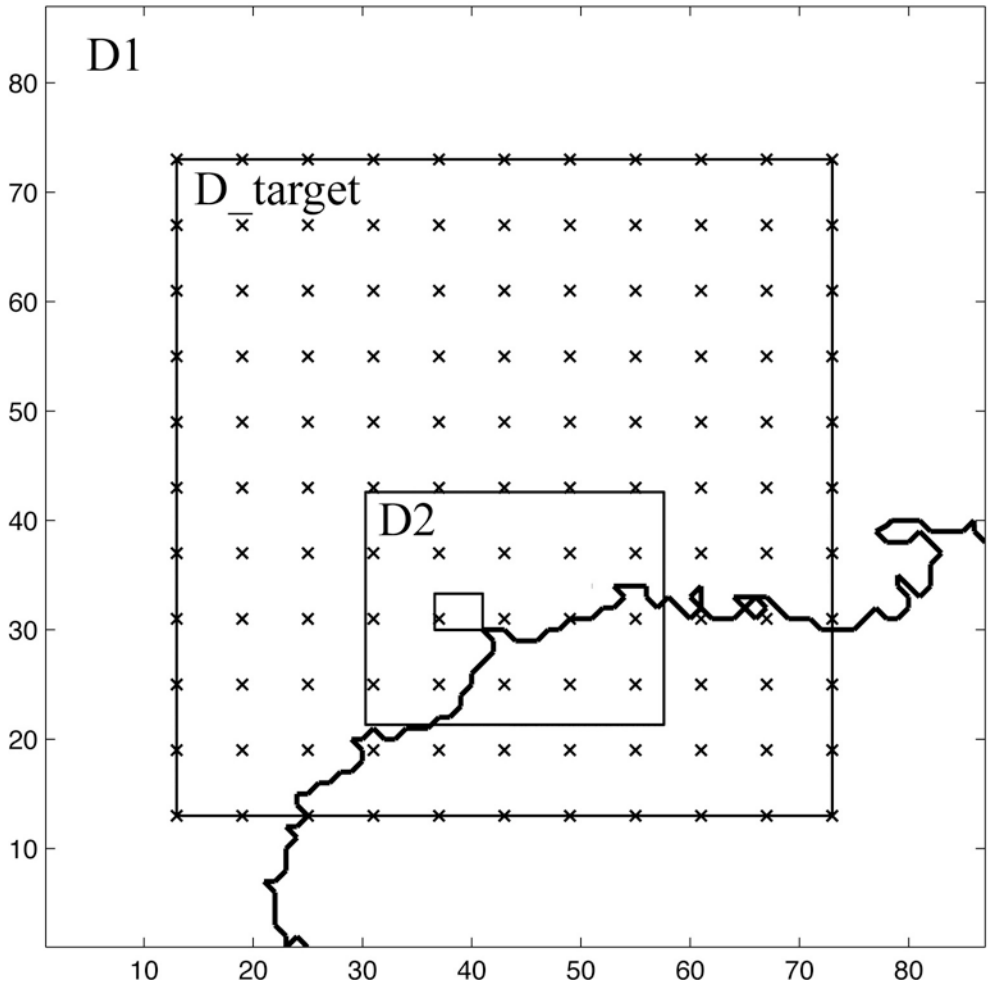

Figure 1. Model domains used in photochemical ensemble simulations. The horizontal grid spacing of domain 1 (D1) and domain 2 (D2) is 12-km and 4-km, respectively. The box indicated by D_target denotes the domain used in Figs. 2, 3, 7, and 9. The innermost box denotes the domain used as the Houston area. Cross points indicate positions used to extract the profiles for assimilation. D1 grid points are numbered along the margins.

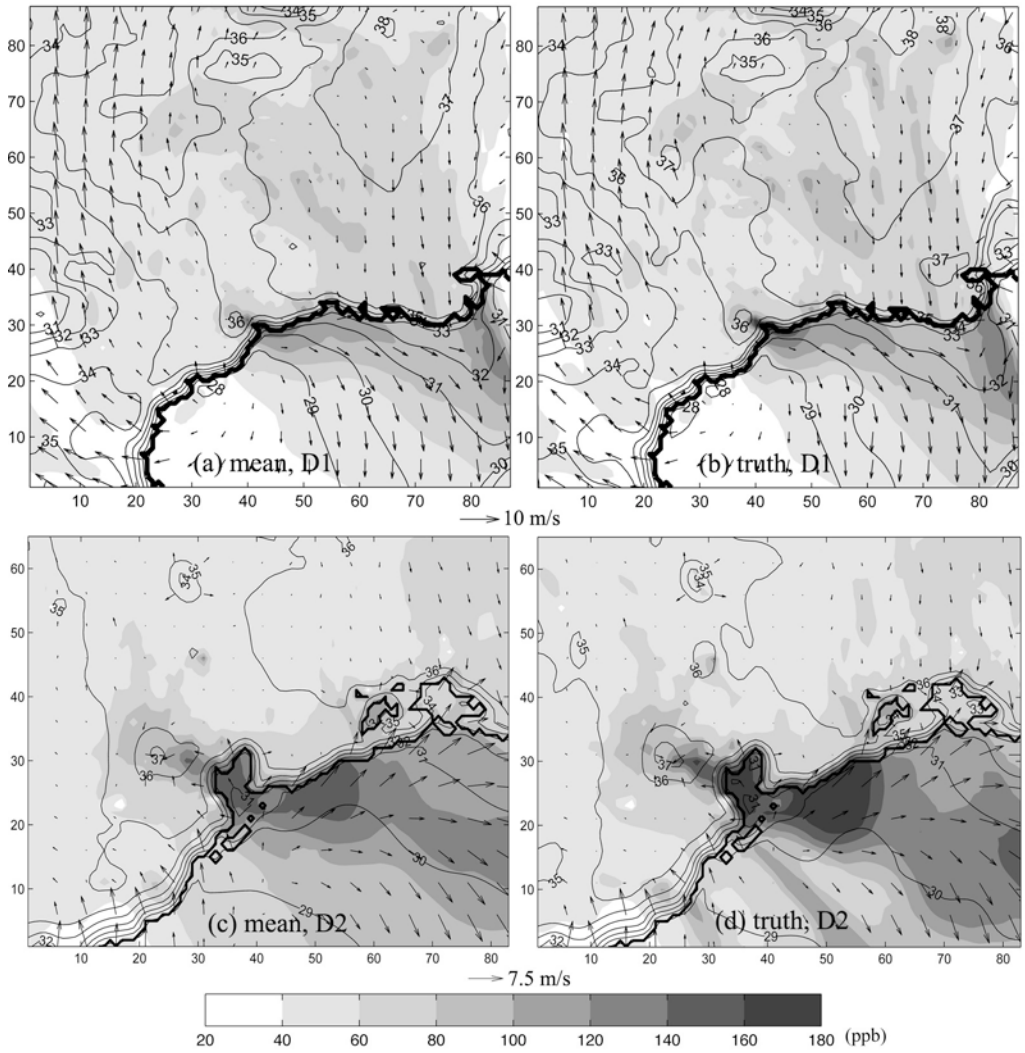

Figure 2. The ensemble mean forecast of ozone (ppb, shaded), temperature (Celsius, contoured) and wind (vectors) at 21 UTC (15 LST) in (a) D1 and (c) D2, and the simulations of ozone, temperature, and wind from the "truth" run in (b) D1 and (d) D2. Domain grid points are numbered along the margins.

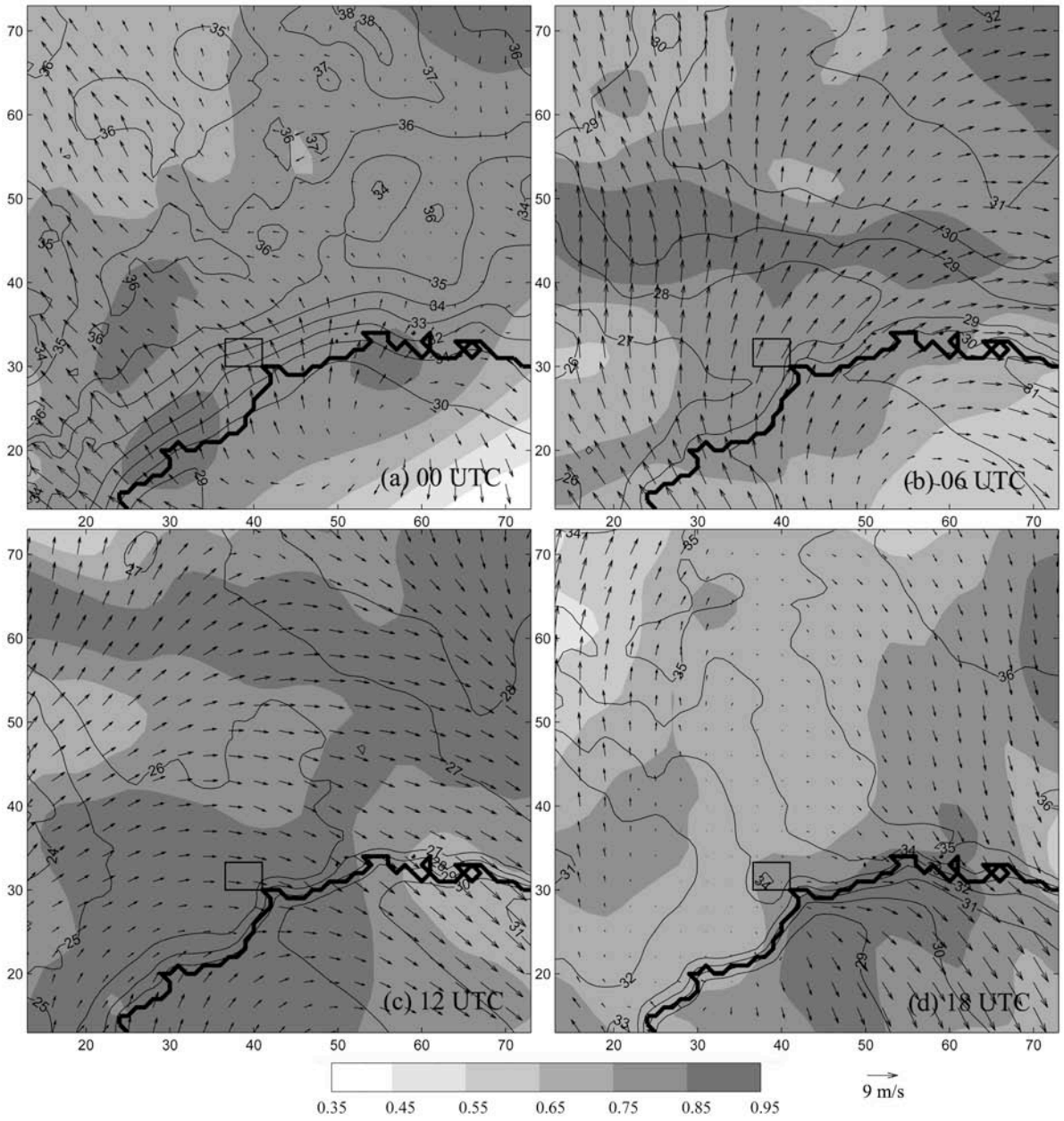

Figure 3. The observational impact factor (shaded) for sounding observations on the peak time surface [O_3] (21 UTC) over the Houston area (the inner domain) along with the ensemble mean surface temperature (Celsius, contoured) and surface wind (vectors).

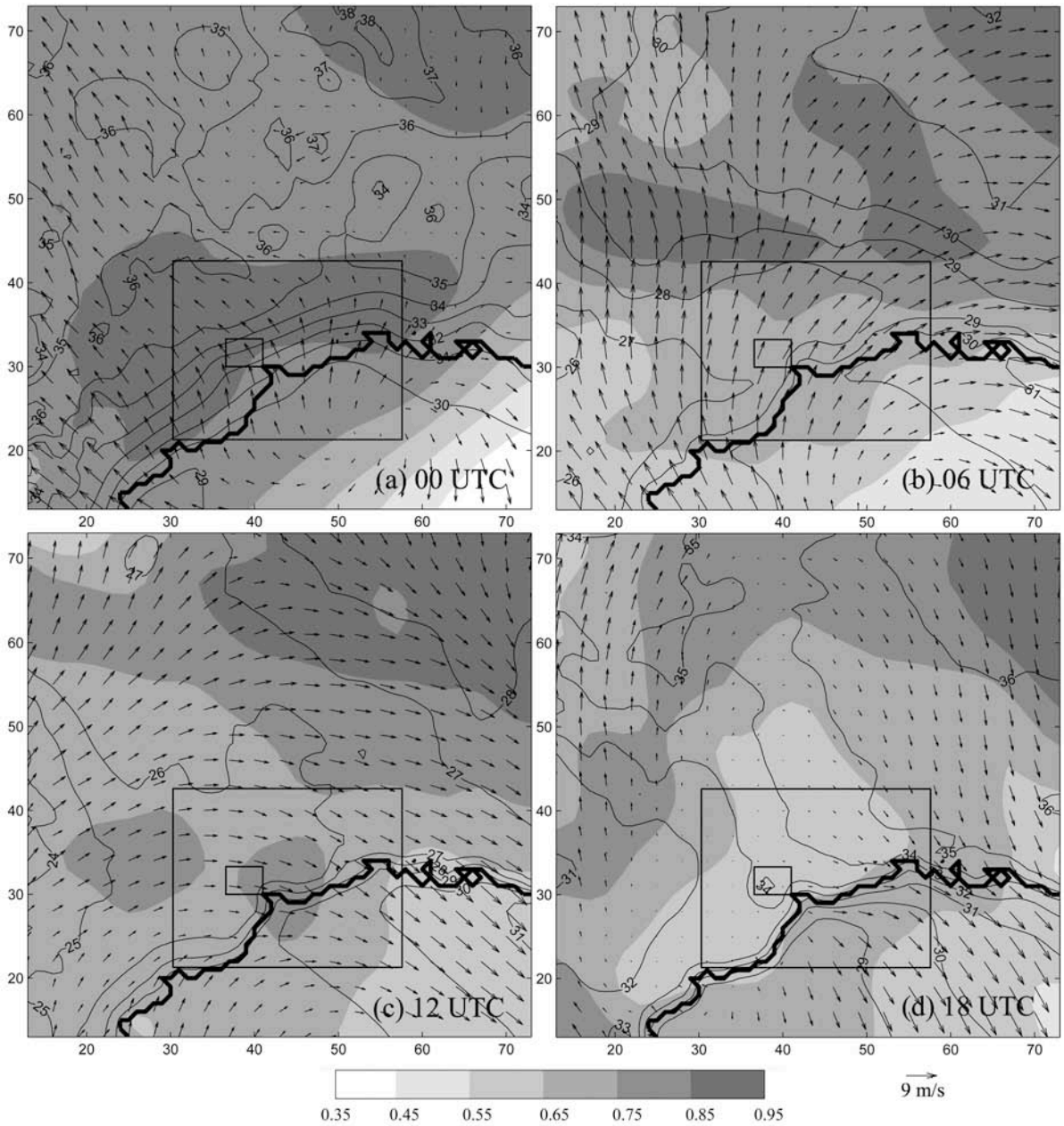

Figure 4. The observational impact factor (shaded) for sounding observations on the peak time surface [O_3] (21 UTC) over D1 along with the ensemble mean surface temperature (contoured) and surface wind (vectors) from 00 to 18 UTC. Inner domain indicates D2 and the innermost domain indicates the Houston area.

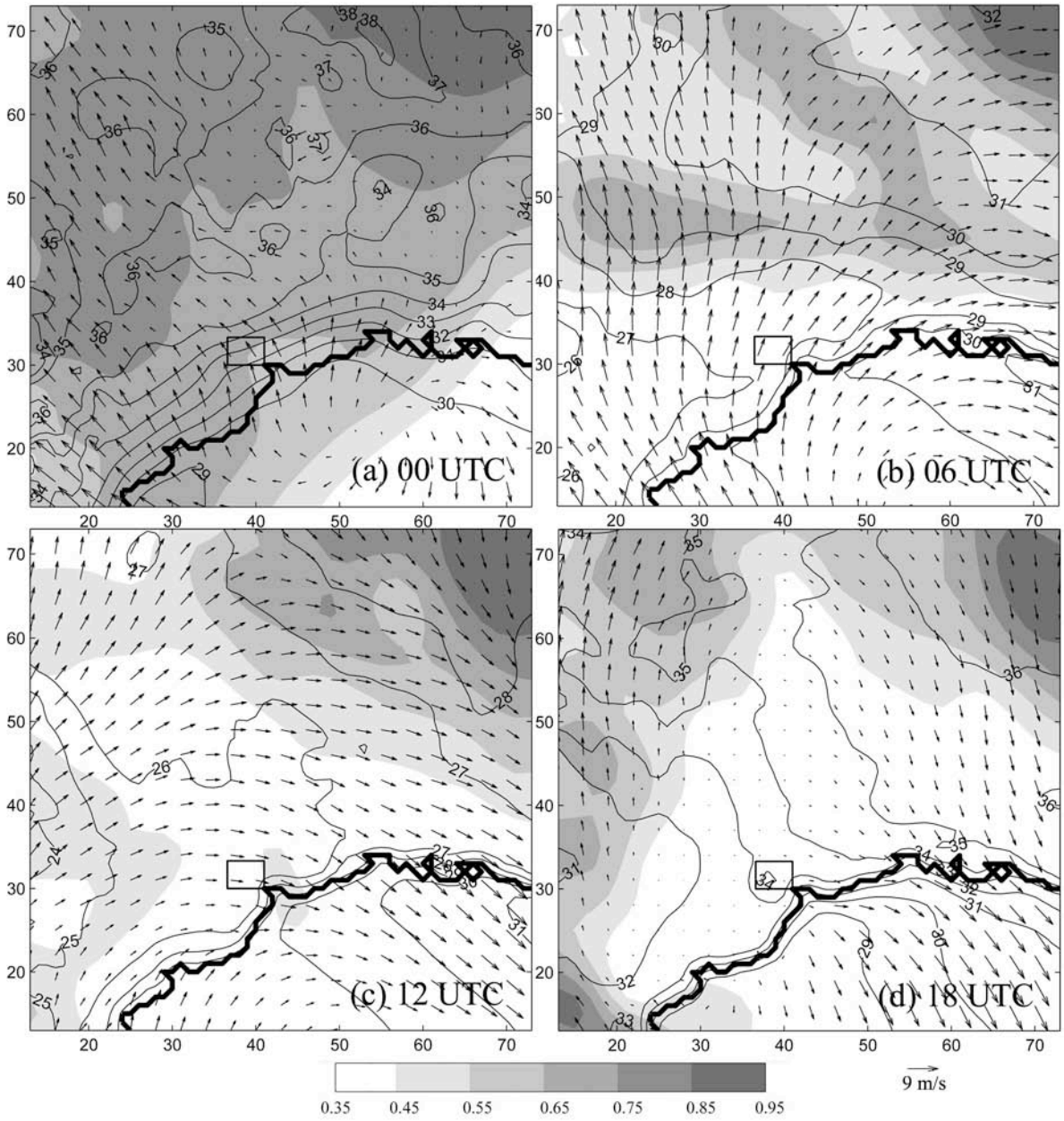

Figure 5. The observational impact factor (shaded) for single layer observations on the peak time surface [O_3] (21 UTC) over D1 along with the ensemble mean surface temperature (contoured) and surface wind (vectors) from 00 to 18 UTC. The inner domain indicates the Houston area.

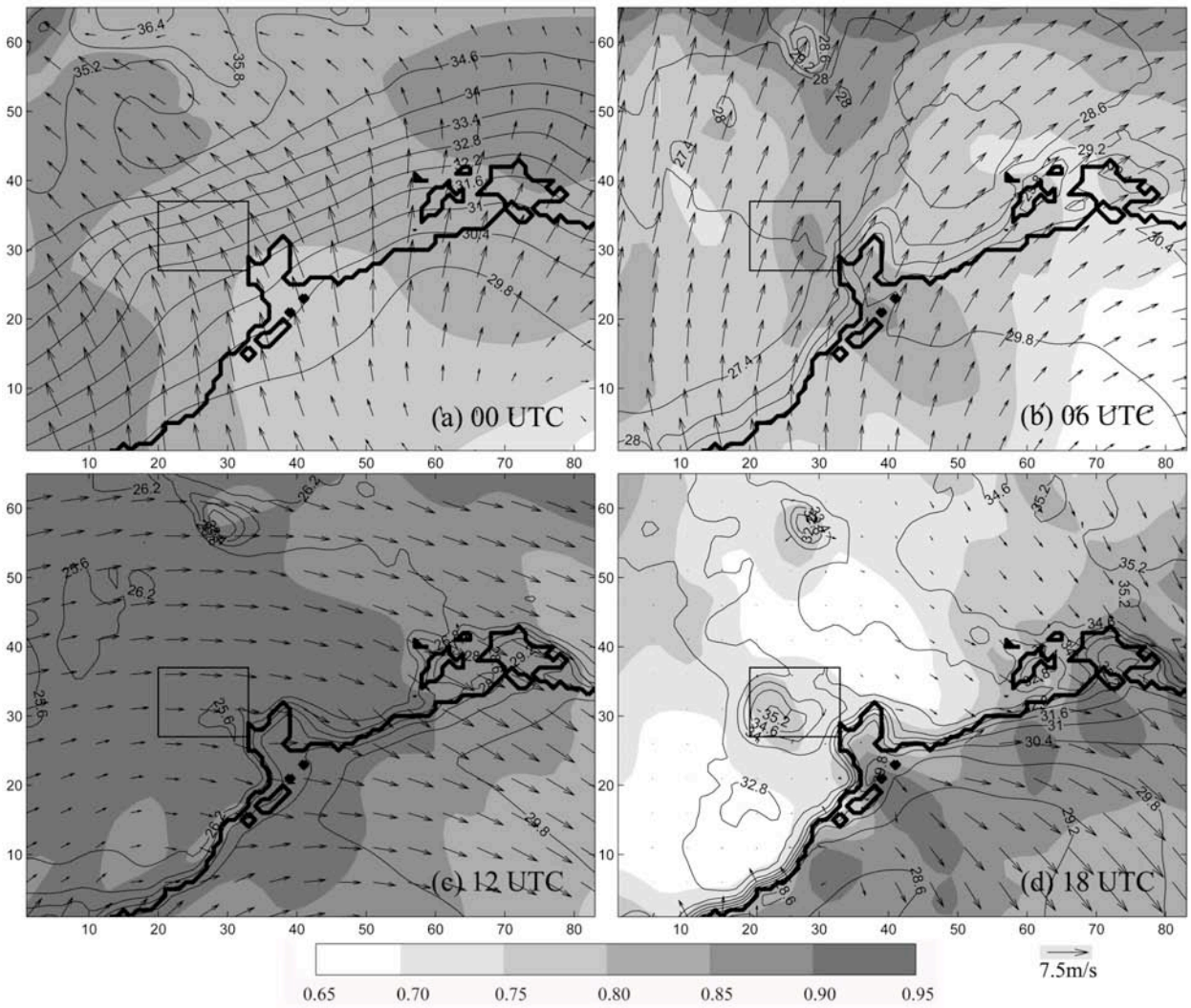

Figure 6. Same as Fig. 4, but the observational impact factor (shaded) is derived from ensemble simulations in D2 (4-km run). The inner domain indicates the Houston area.

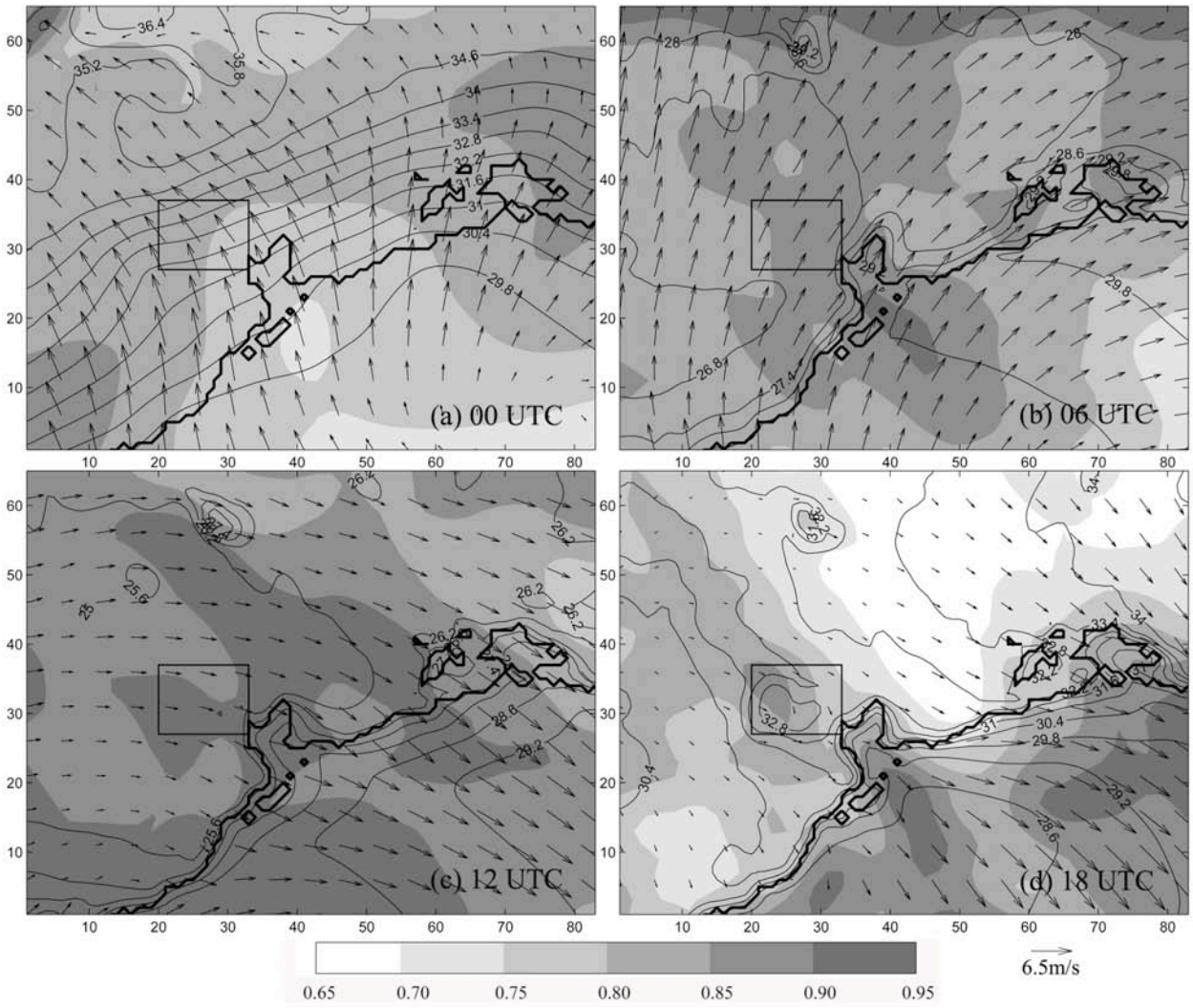

Figure 7. Same as Fig. 6, but the observational impact factor (shaded) is derived from ensemble simulations with multi-PBL schemes.

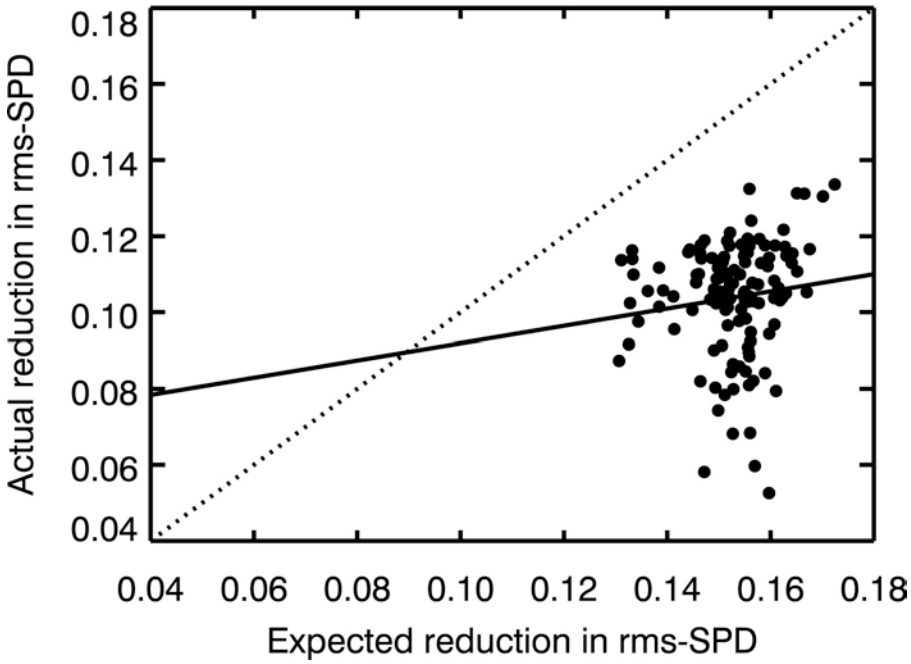

Figure 8. Scatter plot of the expected and actual reduction in the root mean squared error of surface wind speed (rms-SPD, m/s) over D1 at 21 UTC (15 LST). The solid (dotted) line denotes data fitted (ideal) line.

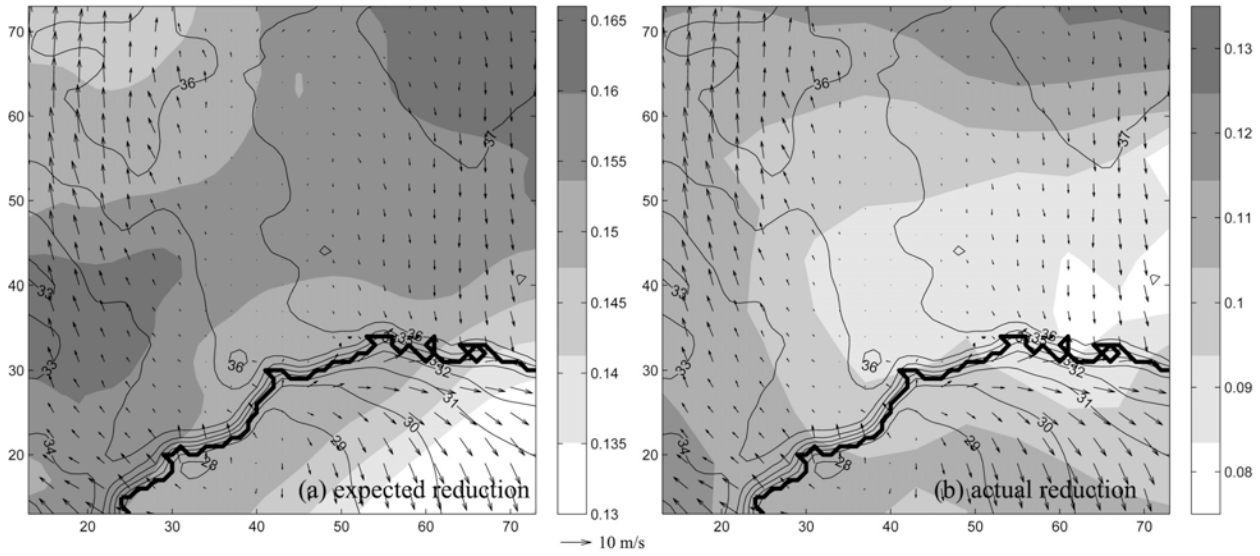

Figure 9. Horizontal distributions of (a) the expected reduction and (b) the actual reduction in rms-SPD (m/s) over D1 at ozone peak time through data assimilation at 00 UTC (6 LST).

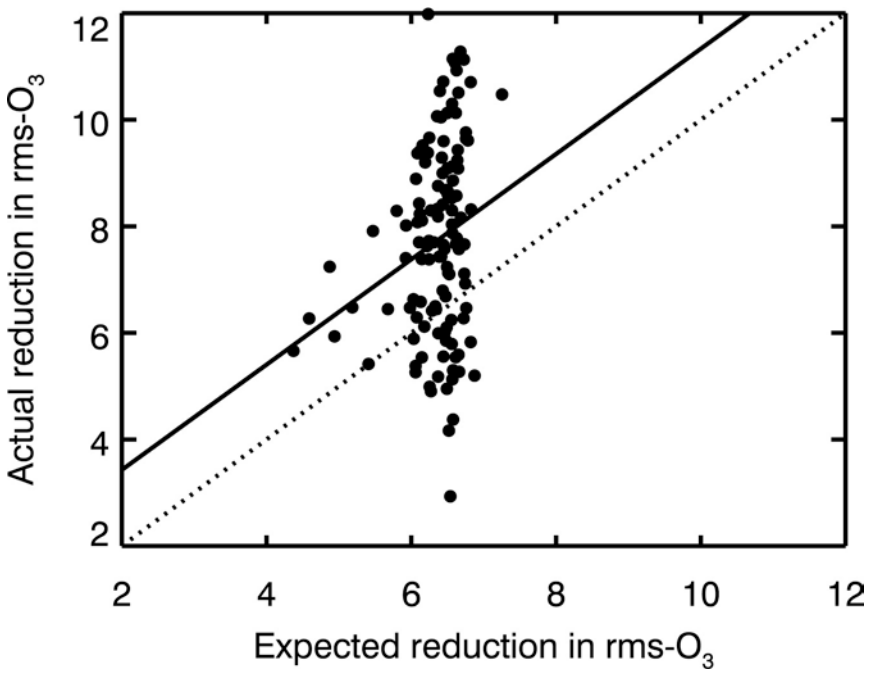

Figure 10. Scatter plot of the expected and actual reduction in the root mean squared error of surface [O_3] (rms-O_3, ppb) over the Houston area at 21 UTC (15 LST). The solid (dotted) line denotes data fitted (ideal) line.

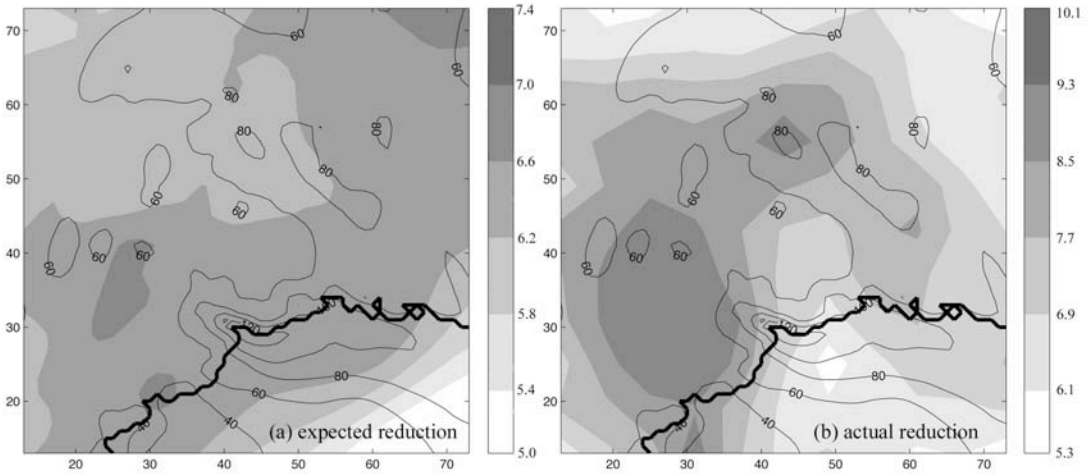

Figure 11. Horizontal distributions of (a) the expected reduction and (b) the actual reduction in rms-O_3 (ppb) over the Houston area at peak ozone time through data assimilation at 00 UTC (6 LST).