THREE DIMENSIONAL DATA ASSIMILATION FOR HIGH RESOLUTION WEATHER AND POLLUTION FORECASTING IN THE LOS ANGELES BASIN

Michael D. McAtee, James F. Drake, and Leslie O. Belsma
The Aerospace Corporation, El Segundo, California

1. BACKGROUND

Air quality predictions for the Los Angeles basin pose particular challenges for the weather forecasts that support them. In addition to transport by the regional scale winds, the forecast must accurately capture the sea and land breeze circulation, and the convectively driven upslope circulation in the mountains surrounding the basin, which can inject pollutants into the strong stable layer aloft, where they can travel far, return to the surface, and contribute to air pollution episodes (Liu and Turco, 1995). A high-resolution weather model is a necessary component of air quality modeling in order to adequately represent the small scale phenomena which are significant to air pollution in this region.

The accuracy of predictions from the fine-scale, limited-area models generally used to support air quality applications can depend on several factors. Some of more important factors include the nature of the weather phenomenon one is attempting to predict, the model's ability to correctly represent physical processes (e.g., boundary layer turbulence, radiation), the quality of boundary conditions employed, and the initial conditions supplied to the model. The quality of the model initial conditions, in turn depends critically on the number, distribution, and quality of atmospheric observations, as well as the methods used to analyze them. Relatively few observations go into an operational forecast in comparison to the number of forecast quantities. Maximizing the use of all available observations, even if these observations are not direct measures of the model variables, is thus very important when initializing a weather model. The work described below focused on improving fine-scale, real-time predictions from the MM5 over the LA basin through the optimal assimilation of space based and local observations. In Section 2 we describe the observational and other sources of initialization data used in the system. A brief description of the 3-Dimensonal Variational Analysis (3DVAR) system used to assimilate the observational data is given in Section 3. This is followed in Section 4 by a description of the MM5 configuration and the data assimilation cycle used. In Section 5 we describe the model verification software developed for this system. Finally in Section 6 we discuss our plans to improve the system.

2. INITIALIZATION DATA

The system assimilates observational data obtained from local sources as well as from the worldwide observational database of the Air Force Weather Agency (AFWA) in Omaha, Nebraska. The AFWA data includes surface and rawinsonde observations; aircraft reports (AIREP); cloud-drift winds from geostationary satellites; and precipitable water and surface wind speed over oceans from the Defense Meteorological Satellite Program (DMSP) Special Sensor Microwave/Imager (SSM/I). Additional data are pulled from servers of the surface networks of the South Coast Air Quality Management District of California (SCAQMD) and the Remote Automated Weather Stations (RAWS) of the Bureau of

Land Management (BLM), as well as the boundary layer profiler (BLP) network of NOAA's Forecast Systems Laboratory (FSL). Figure 1 gives examples of some of the numerous observations that were assimilated in the inner model domain at 00 UTC on July 10 2003. The examples given are intended to highlight data that are not typically used in meteorological data assimilation for air quality applications.

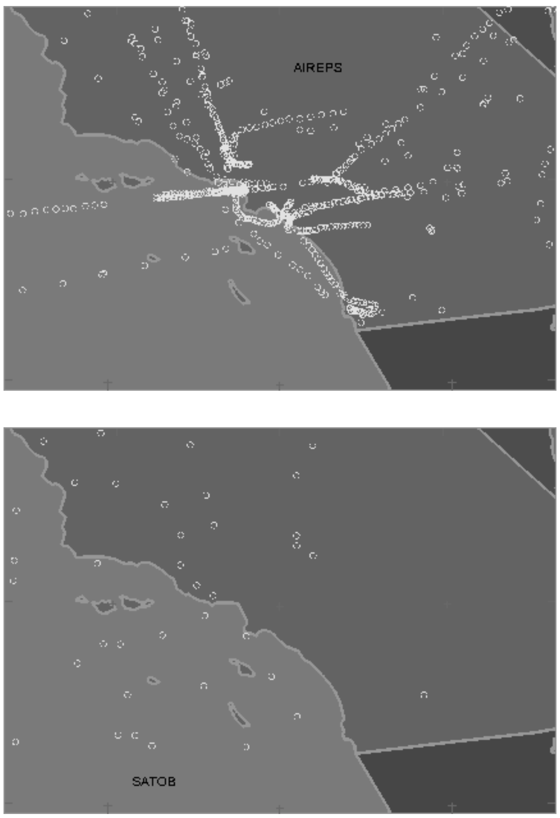

Figure 1. Inner Model Domain AIREP and SATOB Observations for 00Z 10 Jul 03

To make an acceptable forecast over a domain with significant ocean area, MM5 needs the sea surface temperature (SST) field to be specified reasonably well. Daily operational global Navy SST fields are obtained from AFWA to specify a SST field that does not change during the forecast period.

3. THE DATA ASSIMILATION SYSTEM

The 3DVAR system used in this study is fully described by Barker et al (2004) so only a brief description is given below. We chose to use 3DVAR primarily because of its ability to assimilate a wide variety of observations especially those that are not direct measures of the model state variables (e.g., satellite data). 3DVAR categorizes observations by type, each with its own error statistics. Through the laws of physics, (linearized) observation operators relate the values of the model state variables at the analysis time to observed quantities. The goal of three-dimensional assimilation is to specify the model state variables at analysis time so as to minimize the difference between the analysis and observations and a prior estimate of the model state (background). The error statistics of the background field are known extrinsically from prior estimates. In the theory of 3DVAR there is a cost function $J(\mathbf{x})$ given by Equation 1, which is the sum of two terms: one (J^b) depends only on the background field, the other (J^o) depends only on the observations.

$$J(\mathbf{x}) = J^b + J^o = \frac{1}{2}(\mathbf{x}-\mathbf{x}^b)^T\mathbf{B}^{-1}(\mathbf{x}-\mathbf{x}^b) + \frac{1}{2}(\mathbf{y}-\mathbf{y}^o)^T(\mathbf{E}+\mathbf{F})^{-1}(\mathbf{y}-\mathbf{y}^o) \quad (1)$$

where:

$\mathbf{x}$ = a vector of the model variables at a given time (e.g., temperature values on a three-dimensional grid)

$$J^b = \frac{1}{2}(\mathbf{x}-\mathbf{x}^b)^T\mathbf{B}^{-1}(\mathbf{x}-\mathbf{x}^b)$$

$$J^o = \frac{1}{2}(\mathbf{y}-\mathbf{y}^o)^T(\mathbf{E}+\mathbf{F})^{-1}(\mathbf{y}-\mathbf{y}^o)$$

$\mathbf{x}^b$ = a vector of the model variables at a given time as given by the background field

$\mathbf{B}$ = the matrix of the covariance of error in $\mathbf{x}^b$

$\mathbf{y} = \mathbf{H}\mathbf{x}$

$\mathbf{H}$ = the observation operator

$\mathbf{y}^o$= the vector of observations

E = the (diagonal) matrix of observational (instrumental) error

F = the "representivity" error matrix, i.e. the error associated with the observation operator

The problem of assimilation reduces to finding the **x** that minimizes $J(\mathbf{x})$; the **x** that does this is denoted $\mathbf{x}^a$, for analysis. To make it practical to perform the minimization in the time available for analysis, the 3DVAR algorithm uses control variables that are linearly related to **x** by a nonsingular transformation rather than directly with **x** . After J is minimized in the space of control variables, the algorithm inverts the transformation to give $\mathbf{x}^a$.

4. FORECAST MODEL DESCRIPTION

The model system is run with two nested domains. Table 1 shows their specifications and Figure 2 shows where they are located. The terrain used is based on a 30 sec (0.9 km) global dataset.

Table 1. Model domain specifications for analysis and forecast

Domain	Grid Box Size (km)	Domain Dimensions (grid boxes)
D01	15	91 x 85
D02	5	121 x 91

Figure 2. Locations of the Model Domains

In a separate run for each domain, a multiprocessor Cray Research SV1 produces the analyses with a parallel version of 3DVAR. These fields are then used to initialize forecast runs of the MPP version of MM5 on the Cray. MM5 version 3.5 is configured as follows: one-way interaction at the boundaries between a parent domain and its child; cumulus parameterization (Grell, 1994) in just the outer domain; cloud radiative cooling; mixed phase ice physics; and a multilayer soil temperature model. Countergradient vertical transport within the planetary boundary layer (PBL) is considered to be significant. The Hong-Pan (1996) scheme, also known as the MRF scheme, is used because it is the only one among the PBL parameterizations offered by MM5 with this feature. MM5 is set to use 37 vertical (half-sigma) levels with the top of the model at 100 hPa.

In order to improve the quality (e.g., fine-scale features) of the background field used by 3DVAR, the near real-time runs of 3DVAR and MM5 are employed in a three-part data assimilation cycle for each domain. Cycle (part) 1 is based at 00 UTC. For this cycle ETA model data at 40km grid spacing is used as a background field for 3DVAR. The analysis valid at 00 UTC is performed 6-7 hours after valid time to allow for the collection of all possible observations. Atmospheric measurements from satellites often have latencies of a few hours so waiting allows for the use of this data. MM5 is integrated to 6 hours for both domains using 3DVAR's analyses as initial conditions. Cycle 2, which is based at 06 UTC, uses the fine-scale 6-hour forecast for each domain from Cycle 1 as a background field for 3DVAR. Once again the analyses are performed 6-7 hours after 06 UTC to maximize amount of data used. As in Cycle 1 these analyses are used to initialize MM5 and a 6-hour forecast is produced. Cycle 3, the 12 UTC cycle, is the main forecast cycle. Because of the desire to deliver 24 hours of useful forecast data to air quality forecasters and due to the runtimes of 3DVAR and MM5, Cycle 3 is initiated at 14 UTC. MM5 is integrated out to 36 hours. Graphics for various forecast parameters are created and posted to a web site (http://www.aerospaceweather.com). The digital data in MM5 format is also posted to an ftp site where it is available to personnel from the SCAQMD. Additionally, the daily 36-hour forecasts from Cycle 3 are archived and are available on request. The system has been running routinely since the fall of 2003.

5. MODEL VERIFICATION

Verification software was developed to compare MM5 output to observations. The software interpolates MM5 predictions to the observation location and compares them with measured values, which are taken to be true. The program can compare temperature, relative humidity, dew point, mixing ratio, total precipitable water, wind speed, wind direction. It performs separate comparisons for sounding and surface-station data. For truth data, it reads all the data files from AFWA (surface and rawinsonde observations; AIREPs; GOES cloud-drift winds and DMSP SSMI precipitable water and surface wind speed over oceans), surface station data from RAWS, SCAQMD, and lidar and BLP profiles. Verification scatter plots by forecast hour with domain biases and RMSE's are posted to the web site as they become available. Individual times series of model forecasts and corresponding surface observations for a number of SCAQMD sites are also posted.

6. PLANS

Development of the 3DVAR/MM5 system for daily forecasts for the Los Angeles basin will continue in several ways. One is to establish a continual data assimilation cycle. Using this approach, one hopes to eliminate the loss of fine-scale flows that can occur when ETA analyses are used as a background for the domains. New data sources will be assimilated such as Quikscat ocean surface winds and radar data. Refined estimates of the background error fields will also be incorporated into the system.

Acknowledgments: We would like to acknowledge the cooperation of the AFWA, the BLM, the FSL, the NCEP, and the SCAQMD in making available to us the meteorological data we have used in this study.

REFERENCES

Barker, D. M., Huang, W., Guo, Y.-R., Bourgeois, A. J., Xiao, Q. N. 2004: A Three-Dimensional Variational Data Assimilation System for MM5: Implementation and Initial Results. *Monthly Weather Review*: Vol. 132, No. 4, pp. 897–914.

Grell, G.A., Dudhia, J., Stauffer, D.R. *A Description of the Fifth-Generation Penn State/NCAR Mesoscale Model (MM5)*. National Center for Atmospheric Research: Boulder, CO, 1994; NCAR Technical Note, NCAR/TN-398+STR.

Hong, S-Y, Pan, H-L. 1996: Nonlocal Boundary Layer Vertical Diffusion in a Medium-Range Forecast Model. *Monthly Weather Review*: Vol. 124, No. 10, pp. 2322–2339

Lu, R., Turco, R.P., 1995Air pollutant transport in a coastal environment—II. Three-dimensional simulations over Los Angeles basin *Atmos. Environ:* 29, 1499-1518.