WESTERN REGION TECHNICAL ATTACHMENT
NO. 99-14
JULY 27, 1999

THE ETA DATA ASSIMILATION SYSTEM

James A. Nelson, Jr. - NWSFO-SLC/WRH-SSD

Introduction

As with all models, solutions generated by model equations are highly dependent on the data which goes into the model. Thus, the initialization procedure should be of great importance to the forecaster in determining the potential use of the model output. Assimilation packages are developed in order to find the best fit of both observational data and the model first-guess forecast. This Technical Attachment (TA) will detail the past, current, and future developments of the ETA Data Assimilation System (EDAS). A future TA will deal with the past, current, and future developments of the Global Data Assimilation System (GDAS).

Assimilation Procedures for the ETA

The EDAS is based upon the GDAS, Parrish and Derber (1992) and Derber et al. (1991). Currently, EDAS is comprised of the variational mathematical technique known as a 3-Dimensional Variational Analysis (3DVAR, Parrish et al. 1996). It was developed in order to produce an analysis and first-guess more consistent with the forecast model than the GDAS. 3DVAR is a modified version of the same analysis scheme used in the GDAS, known as the Spectral Statistical Interpolation (SSI). As mentioned above, this will be detailed in a later TA.

Prior to the changes made to the Eta in February of 1998, the EDAS provided the first-guess for its analysis at 00Z and 12Z using the Regional Optimal Interpolation Scheme (OI, DiMego, 1988). The EDAS was run as a 12-hour pre-forecast data assimilation, updated with analyzed (by an OI on Eta surfaces) corrections every 3-hours. A first-guess was provided by the GDAS using all available data. This first-guess then was applied to the Eta coordinate system. The original analysis converted data from spectral space (off of the Aviation (AVN) model grid) to the Eta model native grid and interpolated vertically to Eta coordinate surfaces. This adjusted guess was then interpolated to each observation location and the observed increments were computed. A multi-variate OI analysis of the observed increments was then performed on the Eta model grid and used to modify and update the "first-guess". This curtailed the need for any unnecessary interpolations. The only variables in the model which were updated during the OI analysis were temperature

(T), the u- and v-components of the wind field (u,v), the specific humidity (q), and the pressure at the model terrain level (p*). Lateral boundary conditions for both the assimilation and model forecast were obtained directly from the AVN. Boundary tendencies calculated from the EDAS every six hours from the 12-hour old AVN forecast.

The Meso-Eta, on the other hand, computed boundary tendencies every three hours from the current AVN forecast. In the current system, boundary tendencies are produced every three hours from the previous, now run four times a day, AVN. Since the AVN is run every six hours now, tendencies computed from the 06Z and 18Z time will be used for the last 6 hours of the EDAS and for the 48-hour forecast. The 3-hour updates made use of all data from RAOB/pibal, surface, Profiler, ACARS and polar-orbiting satellite (thicknesses). The last 3-hour forecast from the EDAS was the first-guess for 00Z or 12Z analysis. This procedure was very similar to what is being done with the RDAS for the NGM since 1991, and what the RUC is also doing now. A first-guess that has been generated by the Eta forecast model, via the EDAS, is preferable to one based on a static analysis of a global spectral model-based first-guess because of the matching physics, the better resolution and a less "spin-up" (DiMego, 1998).

Problems inherent in the OI scheme concerning the treatment of surface data, soil moisture, the cloud model, radiation scheme, and surface albedo were detailed in Staudenmeier, 1996. In summary; 1) 86% of the surface data over Western Region was being excluded from the analysis, and 2) soil moisture, carbon dioxide, ozone, and surface albedo were initialized with climatological values. In order to alleviate some of these problems, some changes were made to the analysis scheme. One such, alleviation was to increase the amount of surface observations into the analysis. This was achieved by reaching data within 1000 meters of the model terrain. Data was weighted according to its distance from the model terrain.

3DVAR

Another change was to introduce the new analysis scheme known as 3DVAR. This scheme will allow the model to cycle data on itself. Cycling means the capability to use the previous EDAS to initialize the next. This has allowed many of the small-scale models, like the soil model, the cloud model, and the radiation model, to have initial values from the previous run, rather than having to initialize with zero information and spin up to reality. As mentioned earlier, the 3DVAR is a modified version of the same analysis scheme used in the GDAS. The major differences between the two analysis schemes are:

1) background error statistics are simulated in grid space instead of model space
2) no fast and slow variables
3) balance is maintained through a weak constraint on the thermal wind
4) primary analysis variables are at ob locations, not model points,

The fourth difference is achieved when one understands how the mathematical solution to this problem involves the solution of a matrix problem of dimension equal to the number of observations. The OI procedure makes an approximation to model points from a few nearby observation to solve the matrix. Hence, it has only been possible to use observations directly related to model variables. The 3DVAR method makes no local approximation. Therefore, two procedures are needed to solve for the matrix. First, the matrix is not computed directly, but is represented by a sequence of simple operations. Second, the problem is solved by iteration, using a technique known as the conjugate gradient method. Hence, using 3DVAR technique dictates that observations no longer need to be the same as model variables. It is only necessary to have a procedure which can compute a simulated observation. This is meant to minimize the distance between observations and the model first-guess. 3DVAR analyzes wind and specific humidity, like OI. However, OI inferred temperature from the analysis of thickness through height, 3DVAR analyzes both temperature and height. This allows the analysis scheme to assimilate isolated temperature or wind data. Thereby, the 3DVAR system will allow the use of many new data sources and better support the utilization of existing data. Some examples of high density data are: GOES radiances, GOES PW, other satellite derived cloud products, ACARS, VAD wind profiles, GOES PW, surface winds over land, SSMI sea-surface winds, mesonets, NEXRAD radial velocities (Parrish et al. 1996), cloud maps, hourly rain gage data(Lin et al. 1998), vegetation and soil type, etc.

Current data that is used by EDAS with 3DVAR:
- rawindsonde mass and wind
- pibal winds
- dropwindsondes
- wind profilers
- surface land temperature and moisture
- oceanic surface data (ships and buoys)
- aircraft winds
- satellite cloud-drift winds
- oceanic TOVS thickness retrievals
- ACARS temperature data
- surface winds over land
- VAD winds from NEXRAD
- SSM/I oceanic surface winds
- tropical cyclone bogus data

Performances of the 3DVAR versus the OI are located at:
http://www.nws.noaa.gov/om/tpb/447body.html

Currently, the EDAS can be run in a continuous cycling mode. The EDAS can be run either as full cycling or partial cycling. Full cycling incorporates winds, temperature, moisture, soil moisture, soil temperature, cloud water, and TKE. Partial cycling incorporate soil, cloud and TKE parameters, but atmospheric state variables are obtained from the GDAS.

In February 1998, the partial cycling mode was used at the NCEP. Since the EDAS/Eta model had lower precipitation biases than the GDAS, the cycled EDAS resulted in improved soil moisture. This led to better simulations of surface processes. The reason for this improvement was due to the increase in soil layers to four in the Eta model. The previous two layer soil model resulted in too long of a drying period. Tests showed that lower and mid tropospheric errors were smaller for the Eta-32 with EDAS than the Eta-48 with OI. The Eta-32 EDAS showed no improvement at the 250mb level. This was believed to be due to the fact that the only data available at off-times is ACARS temperature data for the mass calculations. Mixed results were experienced for surface variables.

In June 1998, the EDAS was set to full cycling, and degraded analysis of surface and lower tropospheric data (especially moisture) was noticed. It was determined by the NCEP that the vertical correlation error length scale for moisture and the background error co-variances were both too low. This would result in a bias towards the model's first-guess, preventing the accurate portrayal of low-level moisture fields (Rogers et al. 1998). In addition, an error in the original code was found that excluded all surface data.

Changes were implemented on 3 November 1998, after testing to problems detailed above were run. After these changes were made, a problem was noted with the skill of the analysis and forecast over the eastern Pacific as compared to the NGM and AVN. NCEP determined that no apparent data source issues seemed to explain the poor performance of the Eta.

According to analysis schemes, wind-(mass-)only observations will modify the mass(wind) field through a balance condition due to a thermal wind constraint (Parrish et al. 1996). The observations should set the balance where both exist. Balance problems occur in regions where there are predominately mass- or wind-only or single-level observations. This happens more readily over areas such as the eastern Pacific. This would lead to a lack of geostrophic balance between mass and wind analysis corrections away from the reference level. Therefore, the NCEP determined that the previous fix to make the analysis fit the observations better led to a weakened mass-wind balance.

In order to correct for these errors, a modification to the correlation lengths and geostrophic coupling of the mass and wind observations were needed. Correlation lengths determine the vertical and horizontal extent of the observation's influence on surrounding areas. These lengths were increased to allow for greater coverage over data-sparse regions. These modifications resulted in the following impacts to model output; http://www.nws.noaa.gov/om/tpb/3d-eta.htm

1) Temperature - initial values worsened, as expected, because of the weakened mass-wind balance constraint. But by the 48-hour forecast values had come back to slightly better results.

2) Vector wind - lower errors were recorded at all levels at all times.

3) Geopotential height - values were slightly worse than the control runs through all times and levels.

4) Relative humidity - all levels and times were similar to the control runs. Thereby, achieving the desired effect of fixing wind without hurting the relative humidity fix developed in November of 1998.

5) Precipitation - 10-20% increase in skill at all thresholds with a similar bias to the control runs.

Future

Currently, correlation lengths are static for all observations, meaning each observation is treated the same as to its vertical and horizontal influence. In the future, a varying correlation length is expected, where an observation's influence is determined by the relative strength of the mean flow's advective properties. The question lies in how to vary this field and how much background error to tax the values. Should it be varied isentropically or along wind.

Acknowledgments

The author would like to thank Dave Parrish (NCEP, Environmental Modeling Center, EMC) for his insight into how the EDAS works and the future direction of model assimilation.

References

DiMego, G. 1988: The National Meteorological Center Regional Analysis System. *Mon. Wea Rev.* **116**, pp, 977-1000.

Lin, Y. and coauthors., 1998: Assimilation of Real-Time, Multi-Sensor Hourly Precipitation Observations into the NCEP Eta Model. Preprints, 12th *AMS Conference on Numerical Weather Prediction.*

Mesinger, F., 1998: Comparison of Quantitative Precipitation Forecasts by the 48- and by the 29-km Eta Model: An Update and Possible Implications. Joint Preprints, *12th AMS Conference on Numerical Weather Prediction or 16th Conference on Weather Analysis and Prediction.*

Mittelstadt, J.,1998: The ETA-32 Model. *WR-Technical Attachment 98-03.*

Rogers, E., D. Parrish, Y. Lin and G.J.DiMego, 1996: The NCEP Eta Data Assimilation System: Tests with a Regional 3-D Variational Analysis and Continuous Cycling. Preprints, *11th AMS Conference on Numerical Weather Prediction.*

Rogers, E., D. Parrish, and G.J. DiMego, 1998: Data Assimilation Experiments with the Regional 3-D Variational Analysis at the National Centers for Environmental Prediction. Preprints, *12th AMS Conference on Numerical Weather Prediction.*

Rogers, E. and co-authors, 1997: Changes to the NCEP Operational "Early" Eta Analysis/ Forecast System. *NWS Technical Procedures Bulletin No. 447, NOAA/NWS, Washington, D.C.*

Rogers, E. and co-authors, 1999: Changes to the NCEP Operational Eta Analysis. *NWS Technical Procedures Bulletin No. 454, NOAA/NWS*, Washington, D.C.

Staudenmaier, M., 1996: The Initialization Procedure in the Meso Eta Model. *WR-Technical Attachment 96-30.*

Staudenmaier, M., and J. Mittelstadt, 1997: Results of the Western Region Evaluation of the Eta-10 Model. *WR-Technical Attachment 97-18.*

Zhao, Q., et.al.,1998: Assimilating Cloud and Precipitation Observations into the ETA Model to Improve Cloud and Precipitation Forecasts. Preprints, *12th AMS Conference on Numerical Weather Prediction.*