

# Impact of Ground-based GPS PW and MM5-3DVar Background Error Statistics on Forecast of a Convective Case

Y.-R. Guo<sup>1</sup>, H. Kusaka<sup>2</sup>, D. M. Barker<sup>1</sup>, Y.-H. Kuo<sup>1</sup>, and A. Crook<sup>1</sup>

<sup>1</sup>National Center for Atmospheric Research, CO, USA

<sup>2</sup>Central Research Institute of Electric Power Industry, Abiko, Japan

## Abstract

In order to improve the forecast of convection, the MM5 3D-Var (3-Dimensional Variational Data Assimilation) system is applied to initialize the mesoscale nonhydrostatic model, MM5, for a convective case on 12 to 13 June 2002.

The BES (Background Error Statistics) is the key component used in a 3D-Var system, which helps to extract the useful information from the observations and to make the final analysis balanced. In this study, an approach is proposed to construct a case-dependent BES file for a convective case with a relatively inexpensive computing cost.

By using MM5 3D-Var system and MM5 forecast model, a series of the 3D-Var experiments with the default BES and the newly constructed BES are carried out. The results showed that the assimilation of the conventional and GPS (Global Positioning System) PW (Precipitable Water) data with the new BES improved the forecast of convection. The real-time GPS PW itself also has a positive impact on the forecast of convection.

## 1. Introduction

In the United States, ground-based GPS PW measurements are available hourly in near real-time (about 50 min after the observation time). To exploit the information from this new type of observation, the MM5 3DVAR system (Barker et al. 2004), which is a newly developed analysis tool, is used to assimilate GPS PW measurements along with other conventional data. A convective case (12 June 2002) was chosen to assess the impact of the data assimilation with MM5 3D-Var system on short-range forecasts.

To extract useful information from the GPS PW measurements to improve the mesoscale convection forecast is still a challenging task because the 3DVAR system must be carefully tuned with the Background Error Statistics (BES). In previous work of assimilating the ground-based GPS measurements with variational approach (Nakamura et al. 2004; Koizumi and Sato 2004; Vedel and Huang 2004), the authors all felt that the BES used may have some problems for their experiments.

To exploit the information from the high temporal resolution GPS PW data, a short time window (3-h or 1-h) 3DVAR cycling run should also be implemented.

In this study, a series of numerical experiments were conducted to assess the impact of GPS PW data, 3DVAR assimilation system, and the assimilation strategy.

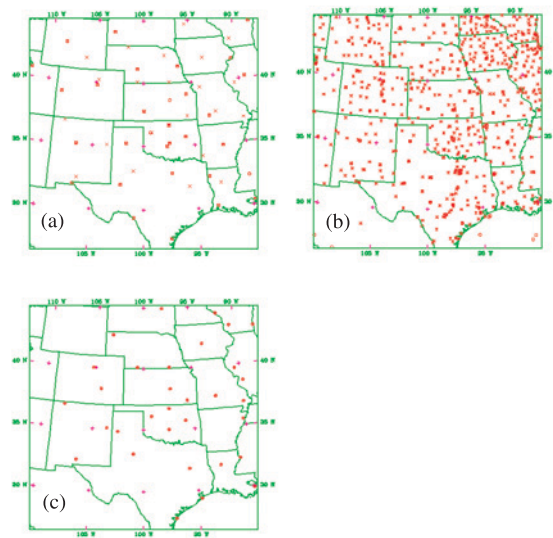


Fig. 1. Distribution of the observations assimilated at 1200 UTC 12 June 2002. (a) upper air data (sound and pilot); (b) surface data (synop, ship, and metar); and (c) GPS PW data

## 2. Data to be assimilated

The data assimilated in this study are conventional observations including upper air (sounding and pilot) data (Fig. 1a), surface (synop, ship, and metar) data (Fig. 1b), and the ground-based GPS PW data (Fig. 1c).

The hourly GPS PW data are provided by UCAR COSMIC Project, in near real-time mode. In this study, we assimilated GPS PW with its errors, usually 1.5 to 3.0 mm, estimated by the data provider. The observed GPS PW has a high correlation (0.964) with the MM5 simulation, but with a small moist bias of about  $-1.2$  mm for large values of PW.

## 3. Convective case on 12–13 June 2002

At 2200 UTC 12 June 2002, a convective line extended from western Oklahoma to the Texas panhandle. Two hours later, the squall line was well developed from southeast Kansas to the Texas panhandle (Fig. 2a). The maximum rainfall amount was 114.4 mm, located at Oklahoma-Kansas border. Then the squall line moved southeastward gradually and finally dissipated at around 1000 UTC 13 June.

Corresponding author: Yong-Run Guo, Mesoscale and Microscale division, National Center for Atmospheric Research, P.O. Box 3000, Boulder, CO 80307, USA. E-mail: guo@ucar.edu.  
©2005, the Meteorological Society of Japan.

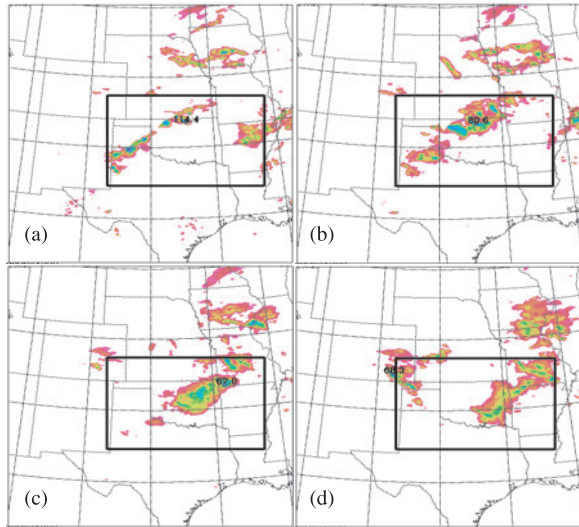


Fig. 2. 3-h accumulated precipitation derived from the National Stage-IV Precipitation Analysis (from NCEP). a) 2100 UTC 12 to 0000 UTC 13, b) 0000 to 0300 UTC 13, c) 0300 to 0600 UTC 13, and d) 0600 to 0900 UTC 13 June 2002. The color scale is referred to Fig. 4. The inner box is used for Equitable Threat Score calculation.

#### 4. Experiment design

Two sets of experiments were conducted: a Cold-start run and a Cycling-run (see Fig. 3 and Table 1). In order to find the optimal Background Error Statistics (BES) in the 3DVAR system for this case study, we first performed the Cold-start experiments. This was then followed by a set of Cycling-run experiments to exploit the information from the high temporal resolution GPS PW data.

For the Cold-start set, the *Eta* analysis at 1200 UTC 12 is used as the initial condition for the control experiment (CONTRL), and used as the first guess for other 3D-Var experiments. For Cycling experiments, the starting time is 0000 UTC 12 with the *Eta* analysis as the first guess, and then the 3-h MM5 model forecasts ending at 03, 06, 09, and 12Z, are used as the first guess for each of the cycles.

For MM5 model integration, the 3-h boundary conditions between 0000 to 1200 UTC 12 June are obtained from 3 hourly *Eta* analyses, and the boundary conditions between 1200 UTC 12 to 1200 UTC 13 June are from the 6 hourly *Eta* forecasts.

The model physics are exactly the same in all experiments, which include MRF-PBL, Dudhia radiation scheme, multiple soil layers, Kain-Fritsch-2 cumulus parameterization scheme, and Goddard mixed phase microphysics scheme with graupel (<http://www.mmm.ucar.edu/mm5/overview.html>). The model domain with a mesh size of 200x200x27 and 10-km grid distance is shown in Fig. 1. The integration time step is 30 seconds.

#### 5. Background error statistics

There are four important input datasets in a 3DVAR system: the background field, background error statistics, and observations and its error statistics. In the Cold-start experiments, the first guess, *Eta* analysis with 40-km resolution, has already blended the information from all

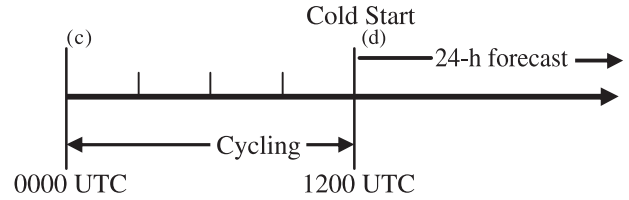


Fig. 3. Schematic diagram for experiment design.

Table 1. Experiment design.

Exp. Name	Data	3D-var/BES	mode
CONTRL			cold
3DOBES	Conventn+PW	default	cold
3DOBSL	Conventn+PW	default+SL tuned	cold
3DNBSL	Conventn+PW	New BES	cold
3DNBPW	PW only	New BES	cold
CTRL00			cycl
CYCLE0	Conventn+PW	New BES	cycl
CYCNPW	Convention	New BES	cycl

conventional upper air and surface observations as well as other remote sensing data available (SATOB, satellite radiances, etc.) through the NCEP EDAS (Rogers et al. 1996). In our experiments, in addition to the hourly GPS PW data, we only have the 12-h interval upper air and the hourly surface observations available from the NCAR archive. As seen below, it is difficult to improve upon the forecast initialized from the *Eta* analysis (CONTRL). To improve the performance of MM5 initialized from the MM5 3DVAR system, one thing we can do is to fine-tune the BES file.

As the default, we have an interpolated BES from a 210-km resolution global background error statistics file. When the 3DVAR is run with this default BES, the results (3DOBES) are worse than those from CONTRL (Fig. 5 and Table 2).

In MM5 3DVAR system, the BES is composed of (i) the eigenvector and eigenvalue derived from the vertical covariance matrix of background errors; (ii) the regression coefficients used in transforming the increments of the streamfunction and potential velocity to the increments of the balanced pressure; and (iii) the scale-lengths used in the recursive filter modeling of the horizontal correlation function.

##### 5.1 Eigenvector and eigenvalue

The eigenvector and eigenvalue included in the default BES may not represent the error variances and the correlation structure of our specific case, a 10-km resolution model and a convective event. Therefore, we use five 24-h forecasts which are started at 1200 UTC 10, 0000 and 1200 UTC 11, and 0000 and 1200 UTC 12 June, just prior to the initial time of our experiments, to derive the new eigenvector and eigenvalue, using the NMC method (Parrish and Derber 1992). There are two advantages with the new eigenvector and eigenvalue: (1) the error variances and the correlation structure will better represent the case studied here because the data are from the same model integration and close to the event; and (2) since only five 24-h forecasts (equivalent to an 120-h forecast) are needed in the calculation, this is computationally very cheap. This could even be done in real-time when the 3DVAR is implemented opera-

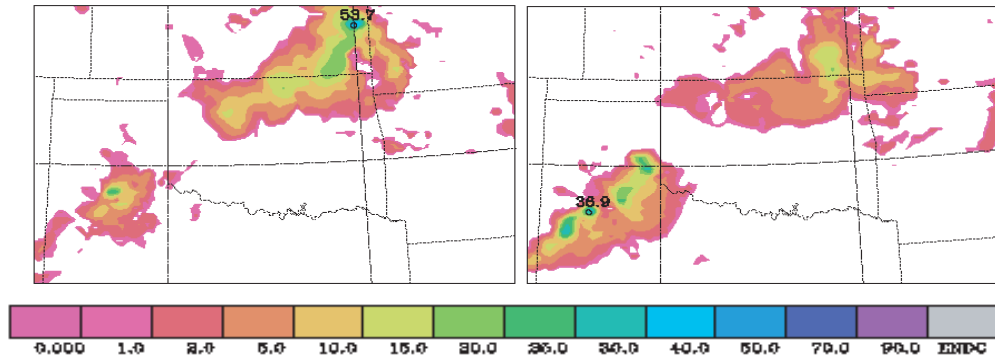


Fig. 4. Forecast of the 3-h accumulated precipitation ending at 0300 UTC 13 June within the box shown in Fig. 2 by a) Exp. 3DNBSL and b) Exp. 3DOBES.

tionally.

### 5.2 Regression coefficients for unbalanced pressure

In the current NMC-method code (see [http://www.mmm.ucar.edu/individual/guo/3DVAR\\_code\\_data.htm](http://www.mmm.ucar.edu/individual/guo/3DVAR_code_data.htm)), the regression coefficients are computed in a latitude-dependent (or a Y-direction-dependent) way to account for the latitudinal variation of the relationship between the wind and balanced pressure. However, since only five forecasts are used in the computation, the results may not be statistically stable, i.e., the regression coefficients are much different between the latitudes and produced latitude-strip-shape pressure increments. To avoid this problem, the domain-averaged regression coefficients are used in our experiments because the model domain is rather small.

### 5.3 Scale-length used in recursive filter

The computation of the scale-lengths needs a large number of samples and is very expensive. Definitely, five forecasts are not sufficient to produce stable results. As Wu et al. (2002) found, the horizontal scales decrease when the resolution of forecast model is increased. Here we use the scale-lengths from the default (210-km resolution) BES in our experiments but with tuning factors. The tuning factors are obtained based on the “single OBS” tests. For the 10-km resolution model, we found tuning factors of 0.11, 0.11, 0.11, and 0.45 for the control variables, streamfunction, potential velocity, unbalanced pressure, and the specific humidity, respectively to work reasonably well. This means that the scale-lengths for streamfunction, potential velocity, and unbalanced pressure are 1/3 of those in the default BES and 2/3 for the specific humidity. The value of 0.11 equals  $(1/3)^2$ , and 0.45 equals  $(2/3)^2$ .

Based on the above consideration, a new BES file was constructed with (i) re-computed eigenvector and eigenvalues, (ii) domain-averaged re-computed regression coefficients, and (iii) the scale-lengths from the 210-km global BES with tuning factors of 0.11, 0.11, 0.11, and 0.45. Exp. 3DNBSL was performed with this new BES. To distinguish the effects of the tuning scale-length from the re-computed eigenvector, eigenvalues, and regression coefficients, another experiment (3DOBSL) was also carried out with only the scale-length tuning factors applied to the default BES.

## 6. Results

The main concern in this study is the convection

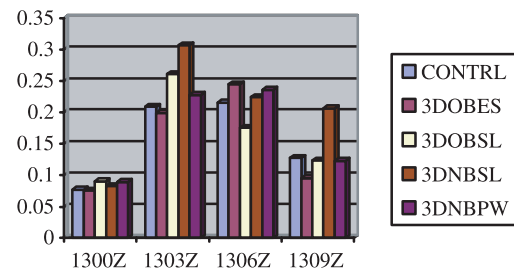


Fig. 5. The equitable threat scores of the 3-h accumulated precipitation forecast verified against the Stage IV precipitation analysis, for threshold = 5 mm and Cold-start 3DVAR experiments.

Table 2. ETS averaged over 4 periods for different thresholds for Cold-start 3DVAR experiments.

Exp.	1 mm	5 mm	10 mm
CONTRL	0.2377	0.1573	0.1067
3DOBES	0.2358	0.1538	0.0880
3DOBSL	0.2651	0.1626	0.0717
3DNBSL	<b>0.2749</b>	<b>0.2050</b>	<b>0.1399</b>
3DNBPW	0.2452	0.1691	0.1037

that occurred over Oklahoma-Kansas region between 2200 UTC 12 to 1000 UTC 13 June. Therefore, the equitable threat scores of precipitation forecast verified against the 3-h accumulated precipitation derived from the NCEP/OH Stage IV precipitation analysis are used to assess the impacts of the 3DVAR system with the different BES specification, the GPS PW assimilation, etc. on the convection forecast.

### 6.1 Cold-start experiments

The 3-h accumulated rainfall forecast ending at 0300 UTC 13 June for 3DOBES and 3DNBSL are shown in Fig. 4. Comparing them with Fig. 2b, 3DNBSL gave a better forecast of the precipitation than 3DOBES in shape and intensity. The equitable threat score (ETS) computed over the area of interest (the box shown in Fig. 2) may still be a useful and objective index for evaluation.

Figure 5 shows the ETS for different Cold-start



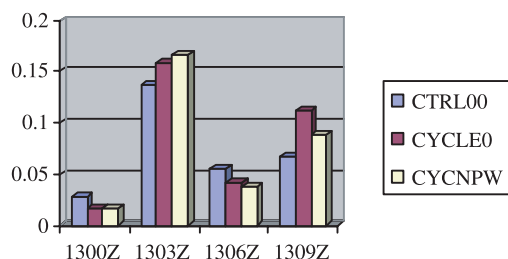


Fig. 6. The equitable threat scores of the 3-h accumulated precipitation forecast verified against the Stage IV precipitation analysis, for threshold = 5 mm and 3DVAR Cycling-run experiments.

Table 3. ETS averaged over four 3-h periods for different thresholds for 3DVAR Cycling-run experiments.

Exp.	1mm	5 mm	10 mm
CTRL00	0.1411	0.0720	0.0352
CYCLE0	<b>0.1747</b>	<b>0.0822</b>	<b>0.0489</b>
CYCNPW	0.1629	0.0773	0.0354

experiments with a threshold of 5 mm. Using the newly reconstructed BES, the scores from 3DNBSL are significantly higher than the other experiments in the 3-h periods ending at 0300 and 0900 UTC, and similar in other 2 periods. Table 2 summarizes the equitable threat scores averaged over the 4 periods for thresholds: 1, 5, and 10 mm. It is clear that 3DNBSL gives consistently higher scores, especially for heavy rain amounts with the thresholds of 5 and 10 mm.

With the default BES (3DOBES), the scores are lower than those of CONTRL, initialized with *Eta* analysis. The default BES with the tuned scale-length factors (3DOBSL) gives the improved forecast skill for 1 and 5 mm but not for 10 mm thresholds. The GPS PW only data assimilation experiment with the new BES, also improved the forecast skill for 1 and 5 mm, and has comparable skill to CONTRL for 10 mm. The assimilation of both conventional and GPS PW data yields the best results.

### 6.2 3DVAR Cycling-run experiments

The 3DVAR Cycling-run experiments started from 0000 UTC 12 June, 12 hours earlier than the cold-start run. A 36-h forward model integration is also conducted as a benchmark.

Figure 6 shows that the 3DVAR Cycling-runs with new BES gives higher scores in the periods ending at 0300 and 0900 UTC 13 for 5 mm threshold, but slightly lower scores in the other two periods as compared with the control forward integration (CTRL00). Without GPS PW data assimilated, the scores, in general, are lower than with the GPS PW assimilated. This means that the GPS PW data have added value to this forecast of convection.

Table 3 shows the 4-period averaged scores for CTRL00, CYCLE0, and CYCNPW for thresholds of 1, 5, and 10 mm. The CYCLE0 gives the best results, followed by CYCNPW, and then CTRL00.

## 7. Discussion and conclusions

### 7.1 Discussion

Comparison of Table 3 with Table 2 shows that the 3DVAR Cycling-run has lower forecast skill than the Cold-start run. This may be caused by the reduced quality of the first guess (3-h MM5 forecasts) and lack of observations in the 12hr cycling period of 3DVAR (only surface and sondes, no radar, and satellite data as in EDAS). In the future, we will attempt (i) to obtain more observations, such as wind profiler data, satellite data, etc., and (ii) to make use of the next-generation meso-scale model, WRF, which should have an improved ability in simulating convection.

### 7.2 Conclusions

Some preliminary conclusions can be drawn from this initial 3DVAR study:

- (1) The newly constructed background error statistics are a key component of a 3DVAR system, which allows the 3DVAR system to more effectively extract useful information from conventional and GPS PW data, and to improve forecast of convection.
- (2) The proposed approach to construct the BES file is relatively inexpensive in terms of computing cost. It can be derived for the specific case. Only the scale-lengths need to be tuned subjectively for different model's resolution. In the future, such an approach may be implemented operationally.
- (3) With the newly constructed BES, GPS PW data assimilation has a positive and neutral impact on the forecast of this warm season convection. However, the cycling experiments conducted here did not show benefit over the cold-start run.

## Acknowledgements

This research was supported by Central Research Institute of Electric Power Industry, Japan.

## References

- Barker, D., Y.-R. Guo, W. Huang, A. J. Bourgeois, and Q. N. Xiao, 2004: A three-dimensional variational data assimilation system for MM5: Implementation and initial results. *Mon. Wea. Rev.*, **132**, 897–914.
- Koizumi, K., and Y. Sato, 2004: Impact of GPS and TMI precipitable water data on mesoscale numerical weather prediction model forecasts. *J. Meteor. Soc. Japan*, **82**(1B), 453–457.
- Nakamura, H., K. Koizumi, and N. Mannoji, 2004: Data assimilation of GPS precipitable water vapor into the JMA mesoscale numerical weather prediction model and its impact on rainfall forecasts. *J. Meteor. Soc. Japan*, **82**(1B), 441–452.
- Parrish, D. F., and J. C. Derber, 1992: The National Meteorological Center's spectral statistical-interpolation analysis system. *Mon. Wea. Rev.*, **120**, 1747–1763.
- Rogers, E., T. L. Black, D. G. Deaven, G. J. DiMego, Q. Zhao, M. Baldwin, N. W. Junker, and Y. Lin, 1996: Changes to the operational "Early" Eta analysis/forecast system at the National Center for Environmental Prediction. *Wea. Forecasting*, **11**, 391–413.
- Vedel, H., and X.-Y. Huang, 2004: Impact of ground based GPS data on numerical weather prediction. *J. Meteor. Soc. Japan*, **82**(1B), 459–472.
- Wu, W.-S., J. Purser, and D. E. Parrish, 2002: Three-dimensional variational analysis with spatial inhomogeneous covariances. *Mon. Wea. Rev.*, **130**, 2905–2916.