P1.13 ASSIMILATION OF GROUND-BASED GPS PWV WITH A 3DVAR SYSTEM FOR AN IHOP CASE

Y.-R. Guo[1], H. Kusaka[2], D. M. Barker[1], Y.-H. Kuo[1], and A. Crook[1]

[1] National Center for Atmospheric Research
P. O. Box 3000, Boulder, CO 80307

[2] Central Research Institute of Electric Power Industry
Japan

1. Introduction

In the United States, the ground-based GPS PWV (precipitable water vapor) measurements are available hourly in near real-time (about 50 min after the observation time) from three networks: Suominet, FLSnet, and CORSnet. There are a total of 199 ground-based GPS sites from these three networks, but usually about 100 sites provide reports at a given time (see Fig. 1). To exploit the information from this new type of observations, the MM5 3DVAR system (Barker *et al*. 2004), which is a newly developed analysis tool, is used to assimilate the GPS PWV measurement along with other conventional data. A convective case (12 June 2002) during the IHOP (International H_2O Project) IOP was chosen to assess the impact of the ground-based GPS PWV measurements on short-range forecasts.

However, to extract useful information from the GPS PWV measurements to improve the mesoscale convection forecast is still a challenging task because the *Eta* analysis, which is used as the background fields in 3DVAR, has already assimilated a variety of the conventional and non-conventional observations. Moreover, the 3DVAR system must be carefully tuned with the Background Error Statistics (BES), and the lateral boundary condition must be carefully treated.

To exploit the information from the high temporal resolution GPS PWV data, a short time window (3-h or 1-h) 3DVAR cycling run should also be implemented. However, the use of high-frequency observations can excite spurious gravity waves, which must be overcome with certain techniques, such as the IAU (Incremental Analysis Update – Bloom et al, 1996).

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

2. Ground-based GPS PWV data

The hourly GPS PWV data are provided by UCAR COSMIC Project, starting from 21 April 2002 in near real-time mode. In addition to the GPS PWV and its observation error, the wet-delay, dry-delay, K-factor, as well as the surface pressure, temperature and relative humidity at the site, are also provided by UCAR COSMIC. In this study, we assimilated GPS PWV with its errors estimated by the data provider.

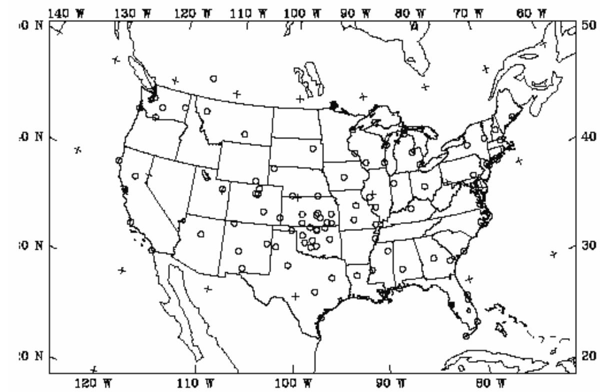

Fig. 1. A total of 102 GPS PWV Observations are available at 0000 UTC 13 June 2002 over the United States.

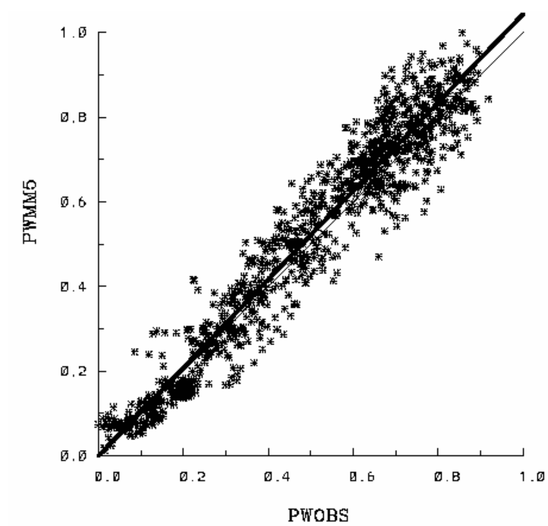

Fig. 2. The scatter diagram between GPS PWV observations and MM5 simulation. The coordinates have been scaled: the origin (0.0) is equivalent to PWV= 2.1 mm, and the value of 1.0 is equivalent to PWV = 57.8 mm. The linear regression is the thick solid line. The total number of samples is 1013, and the correlation coefficient = 0.964.

First, we perform a validation study for this new type of data with the MM5 model simulation, and to see if there is new information contained in this type of data.

Figure 2 is a scatter diagram between the observed GPS PWV and the model simulated PWV during a 24-h period from 12 UTC 12 to 12 UTC 13 June 2002 within our experimental model domain (Fig. 3).

From this figure, the observed GPS PWV has a high correlation with the MM5 simulation, but the model simulation has a small positive bias for the large values of PWV. The mean of GPS PWV is 29.83 mm, while that for MM5 is 31.04 mm. The standard deviations for GPS PWV and MM5 are 13.48 mm and 14.60 mm, respectively. For this case, the model may not gain too much from the ground-based GPS PWV data since the correlation coefficient is quite high already (0.964).

3. Convective case on 12-13 June 2002

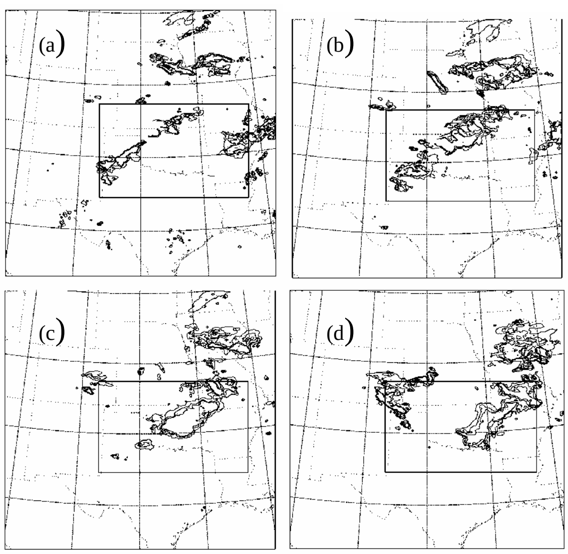

Fig. 3. 3-h accumulated precipitation derived from the National Stage-IV Precipitation Analysis (from NCEP). a) 2100 UTC 12 to 0000 UTC 13, b) 0000 to 0300 UTC 13, c) 0300 to 0600 UTC, and d) 0600 to 0900 UTC 13 June 2002. The counter level is 1, 5, and 10 mm.

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 Texas panhandle (Fig. 3a). 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.

4. Experiment design

Two sets of the experiments were conducted: a Cold-start run and Cycling-run. In order to find an 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 PWV data.

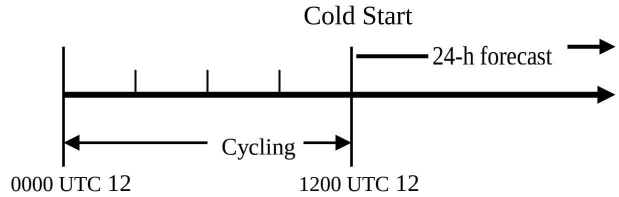

Fig. 4. Schematic diagram for experiment design

For the Cold-start set, the control Experiment (CONTRL) is a straightforward 24-h forecast with the *Eta* analysis as the initial condition at 1200 UTC 12.

Cold-start 3DVAR experiments:

3DOBES: Both conventional and GPS PWV were assimilated at 1200 UTC via 3DVAR with the default BES, followed by a 24-h forecast
3DOBSL: same as 3DOBES but the tuned scale-length.
3DNBSL: 3DVAR with the New computed BES and tuned scale-length from the default BES.
3DNBPW: same as 3DNBSL but only GPS PWV assimilated.

In all the 3DVAR experiments, the *Eta* analysis at 1200 UTC 12 was used as the background fields and the 6 hourly *Eta* forecasts provided the lateral boundary conditions for model forecasts.

3DVAR Cycling-run experiments:

CTRL00: 36-h forecast initialized form *Eta* analysis at 0000 UTC 12.
CYCLE0: 3-h 3DVAR cycling from 0000 UTC 12. The data assimilated include the upper-air data at 0000 and 1200 UTC 12, and the surface and GPS PWV data at 3-h interval from 0000 to 1200 UTC 12. The BES file is the same as in 3DNBSL.
CYCNPW: same as CYCLE0 but No GPS PWV data assimilated.

In all 3DVAR Cycling-run experiments, the *Eta* analysis at 0000 UTC 12 is used as the first guess. The 3-h boundary conditions between 0000 to 1200 UTC 12 are obtained from 3 hourly *Eta* analyses and the boundary conditions between 1200 UTC 12 to 0000 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. The model domain with a mesh size of 200x200x27 and 10-km grid distance is shown in Fig. 3. The integration time step is 30 seconds.

5. Background error statistics

There are three important input datasets in a 3DVAR system: *the first guess field, observations available, and the background error statistics*. In the Cold-start experiments, the first guess, *Eta* analysis with 40-km resolution, has already blended the information from all conventional upper air and surface observations as well as other remote sensing data available (SATOB, satellite radiances, etc.) through the NCEP EDAS (*Eta* data assimilation system). In our experiments, in addition to the hourly GPS PWV 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.4 and Table 1).

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 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, 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 online when the 3DVAR is implemented operationally.

5.2 Regression coefficients for unbalanced pressure

In the current NMC-method code, the regression coefficients are computed in a latitude-dependent (or a Y-direction-dependent) way to account for the inhomogeneity of the relationship between the wind and balanced pressure at different latitudes. 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 the 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 that 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 the tuning factors. The tuning factors are obtained based on the “single OBS” tests. For the 10-km resolution model, we found the 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 for the specific humidity is 2/3. 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 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 PWV assimilation, etc. on the convection forecast. The equitable threat scores are computed over the box shown in Fig.3.

6.1 Cold-start experiments

Figure 4 shows the equitable threat scores for different Cold-start experiments with the 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. For other thresholds, 1 mm and 10 mm, the results are similar. Table 1 summarizes the equitable threat scores averaged over the 4 periods for thresholds: 1, 5, and 10 mm. It is clear that 3DNBSL gives consistent higher scores, especially for heavy rain amounts with the thresholds of 5 and 10 mm.

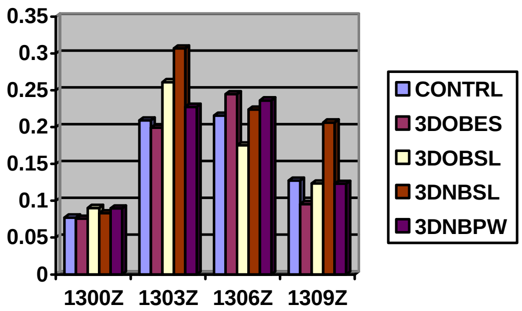

Fig. 4. The equitable threat scores of the 3-h accumulated precipitation forecast verified against the Stage IV precipitation analysis, for threshold = 5 mm and Cold-start 3DVAR experiments

Table 1. Equitable threat scores 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	**0.2749**	**0.2050**	**0.1399**
3DNBPW	0.2452	0.1691	0.1037

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 PWV only data assimilation with the new BES also produces the improved forecast skill for 1 and 5 mm, and has comparable skill to CONTRL for 10 mm. The assimilation of both conventional and GPS PWV data yields the best results. This suggests that the MM5 3DVAR system with the new BES can extract additional useful information from the observations, and improve upon the *Eta* analysis which has already made use of those conventional observations.

6.2 3DVAR Cycling-run experiments

The 3DVAR Cycling-run experiments started from 0000 UTC 12 June. To assess the impact of the 3DVAR and GPS PWV assimilation, we conducted a forward model integration starting from 0000 UTC 12 June as the benchmark. We did not directly compare the cycling-run results with those of the above Cold-start experiments because the first guess used in the Cold-start runs was from the *Eta* data assimilation system, which used different forecast model and observation dataset (more observations than those from NCAR archive).

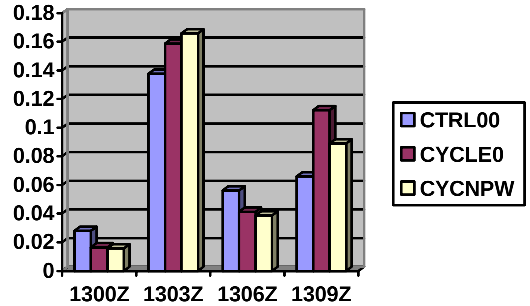

Fig. 5. The equitable threat scores of the 3-h accumulated precipitation forecast verified against the Stage IV precipitation analysis, for threshold = 5 mm and 3DVAR Cycling-run experiments.

Table 2. Equitable threat scores 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	**0.1747**	**0.0822**	**0.0489**
CYCNPW	0.1629	0.0773	0.0354

Figure 5 shows that the 3DVAR Cycling-runs 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 PWV data assimilated, the scores, in general, are lower than with the GPS PWV assimilated. This means that the GPS PWV data have the added values to this convection forecast. Table 2 shows the 4 periods 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.

Comparison of Table 2 with Table 1 shows that the 3DVAR Cycling-run has much lowers forecast skill than the Cold-start run. For the thresholds of 5 and 10 mm, all the Cycling experiments, CTRL00, CYCLE, and CYCNPW, have almost no skills in this convection forecast.

7. Discussion and conclusions

7.1, Discussion

One major problem in this study is that the 3DVAR Cycling-run experiments possessed very little skills in convection forecast. Actually we tried many different strategies to improve the 3DVAR Cycling run, such as using IAU (Incremental Analysis Update) technique, more frequent (1-h) cycling, etc., but with limited success. We realized that a 10-km resolution model might have limited skill in predicting such kind of fine scale convection (Fig. 3). However, at 1200 UTC 12 June, the 3DVAR Cold-start run and 3DVAR Cycling-run used the same observations, the same BES, and the same lateral and low boundary conditions during the ensuing 24-h forecast period. The only difference is the first guess field: Cold-start run used the *Eta* analysis, and Cycling-run used MM5 forecast through the previous four 3-h cycles. This suggests that the quality of the first guess field in the Cycling-run is worse than the *Eta* analysis. The lack of observations in the 12hr cycling period of 3DVAR (only surface and sondes) is likely to be a major source of error.

An assimilation system has three components: (i) observation database, (ii) analysis approach, and (iii) forecast model. Through the Cold-start experiments, the analysis approach, MM5 3DVAR, used in this study is shown to be capable of extracting additional useful information to improve short-range forecast. In addition to the lack of observations, the problems in Cycling-run experiments might come from a significant model error of a non-perfect forecast model (MM5), in particular in its ability to handle convection.

To exploit the information from the high temporal resolution GPS measurements, it is necessary to perform cycling 3DVAR or, better yet, the 4DVAR. 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 mesoscale model, WRF, which may have an improved ability in handling convection. We also plan to test the 3DVAR system over an extended (several week) period to provide statistically significant results. Such work is necessary to obtain sufficient information for a thorough tuning of the background error statistics.

7.2, Conclusions

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

(1) The newly constructed background error statistics is a key component of 3DVAR system, which allows the 3DVAR system to more effectively extract the useful information from the conventional and GPS PWV data, and to improve convection forecast.

(2) The proposed approach to construct the BES file is relatively inexpensive in terms of the computing cost. It is worthwhile to test it with other cases. In the future, such approach may be implemented online for operation. Use of observation minus forecast differences (Hollingsworth and Lonnberg, 1986) and/or variational techniques (Dezrosiers and Ivanov, 2001) would provide a more physically-based tuning of the BES, than the empirical case-dependent tuning performed here.

(3) GPS PWV data can be used to improve convection forecast. However, the best results are obtained when the GPS PWV data are assimilated together with other the conventional and remote sensing data.

8. Reference

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*, Accepted.

Bloom, S. C, Takacs, L.L, DA Silva, A. M., and Ledvina, D. , 1996: Data assimilationusing Incremental Analysis Updates. Mon. Wea. Rev., 124, 1256-1271

Desroziers, G., and S. Ivanov, 2001: Diagnosis and adaptive tuning of observation-error parameters in a variational assimilation. Quart. J. Roy. Meteor. Soc., 127, 1433-1452.

Hollingsworth, A. and P. Lönnberg, 1986: The statistical structure of short-range forecast errors as determined from radiosonde data. Part I: The wind field. *Tellus*, **38A**, 111-136.

Parrish, D. F., and J. C. Derber, 1992: The National Meteorological Center's spectral statistical-interpolation analysis system. *Mon. Wea. Rev*. **120**, 1747-1763.

Wu, W.-S., J. Purser, and D. E. Parrish, 2002: Three-Dimensional Variational Analysis with Spatial Inhomogeneous Covariances. *Mon. Wea. Rev*., **130**, 2905-2916.