ASSIMILATION OF PRECIPITATION DATA USING A COMPLEX 4DVAR DATA ASSIMILATION SYSTEM

D Zupanski and M Zupanski
NCEP/UCAR Visiting Research Program
Washington, D. C., U. S. A.

Summary: We use the regional NCEP's ETA/4DVAR data assimilation system to assess the effects of assimilation of accumulated hourly raingage precipitation observations. A special attention is paid to the time scale of the impact of the assimilated precipitation data on the future precipitation forecast. Experimental results demonstrate a potential of the 4DVAR method to outperform the optimal interpolation method.

1. INTRODUCTION

Assimilation of precipitation data is a challenging task for any data assimilation technique. In the case of optimal interpolation (OI) and nudging techniques a major difficulty is in defining a correct relationship between the observed quantity (e. g., precipitation) and the analyzed variables (e. g., surface pressure, temperature, wind, humidity). In Kalman filtering and four–dimensional variational (4DVAR) data assimilation methods this relationship is naturally given by the forecast model (used either as a strong, or as a weak constraint, e. g., Sasaki, 1970) and poses no problem in theory. In practice however, a big obstacle for both methods is in dealing with nonlinear and discontinuous physical processes (e. g., cumulus convection). These problems need to be solved when developing the necessary tools for the 4DVAR and Kalman filtering techniques: linearized (tangent linear and adjoint) models.

Recently, it was demonstrated that it is possible to overcome the difficulties caused by the discontinuous physical processes and successfully assimilate precipitation data by the 4DVAR data assimilation technique (Zupanski and Mesinger, 1995; Zou and Kuo, 1996; Tsuyuki, 1996). Here, for the first time, we use a complex state–of–the–art 4DVAR data assimilation system to assess the effects of assimilation of hourly raingage precipitation observations.

2. DATA ASSIMILATION SYSTEM

We use the National Centers for Environmental Prediction's NCEP/ETA 4DVAR regional data assimilation system, described in Zupanski and Zupanski (1996). It includes the following important features:

- weak constraint formulation of the forecast model (D Zupanski, 1997);
- non–diagonal, non–separable in space, background and model error covariances (Zupanski and Zupanski 1995);
- full–physics NCEP/ETA forecast model (Mesinger et al., 1988; Janjic, 1990, 1994; Janjic et al., 1995; Black, 1994) with tangent hyperbolic smoothing of discontinuities in the cumulus convection parameterization (Zupanski and Mesinger, 1995);
- corresponding diabatic adjoint model (excluding radiation processes);
- digital filter gravity wave penalty term (Zupanski and Zupanski, 1995);
- limited memory quasi–Newton minimization algorithm (Shanno, 1978; 1985) with an on–line, adaptive preconditioning (M Zupanski, 1993, 1996).

In this study we use the forecast model of a coarser resolution (80 km / 17 layers) than the present operational NCEP/ETA model (48km / 38 layers). Because of some recent enhancement of the operational model (e. g., OSU soil model, explicit cloud water scheme), the two models are somewhat different. The 4DVAR system has capability to assimilate all observations presently used in the operational, OI based regional data assimilation system (e. g., radiosonde, dropsonde, pibal wind profiles, SATEM, AIREP, ASDAR, ACARS, synop, ship, buoy data). We use the term "conventional" for these observations. In addition, we developed an algorithm to assimilate hourly raingage precipitation data, available at observational locations over the continental U. S. So far, the system has undergone extensive testing in different synoptic situations using conventional data. The results were very promising, clearly indicating an improvement as compared to the OI data assimilation results (e. g., Zupanski and Zupanski, 1995, 1996). Also, the 4DVAR experiments with assimilating 24–hourly accumulated raingage precipitation observations showed an improved precipitation forecast, as compared to the standard OI results (Zupanski and Mesinger, 1995).

3. THE MINIMIZATION PROBLEM

As in the case of the assimilation of conventional observations, we pose the problem as to find the minimum of the following functional:

$$
\begin{aligned}
J = & \frac{1}{2}\sum_{n=1}^{N} \left(y_n - y_n^{obs}\right)^T R^{-1} \left(y_n - y_n^{obs}\right) \\
& + \frac{1}{2}\sum_{k=1}^{M} \sum_{m=1}^{M} \Phi_m^T \; Q_{m,k}^{-1} \Phi_k + \frac{1}{2} \left(x_0 - x_b\right)^T B^{-1} \left(x_0 - x_b\right) \\
& + \left[x_0 - F\left(x_0\right)\right]^T W^{-1} \left[x_0 - F\left(x_0\right)\right] ,
\end{aligned}
\tag{1}
$$

where the weak constraint is defined by the forecast model G including the model error correction term Φ_m and a postprocessing operator H which transforms a state vector x from the model space to a vector y in the observational space, that is:

$$x_m = G\left(x_{m-1}\right) + \Phi_m \tag{2}$$

$$y_n = H_n(x_m) \; . \tag{3}$$

In the equations (1)-(3) the superscript T stands for a transpose, obs denotes observations, b a background value, the index n defines observational times and indexes m and k model time steps. Matrices R, B and Q are the observational, background and model error covariances, respectively. The matrix W defines the weight given to the digital filter gravity wave penalty term, where the digital filter operator is defined by F. As in the previous experiments (e. g., D Zupanski, 1997) the model error term includes both the systematic and the random error parts. Also, we use a definition for the model error covariance derived from the background error covariance matrix, as before. The control variable (z) includes the initial conditions x_0 and model random error r_N, that is:

$$z = \left(x_0,\ r_N \right) . \tag{4}$$

In the case of assimilating hourly precipitation data the observational part of the functional (first term in (1)) includes the following term:

$$\frac{1}{2} \sum_{n=1}^{N} \left(C(\boldsymbol{P}_n) - \boldsymbol{P}_n{}^{obs} \right)^T \boldsymbol{R}_p^{-1} \left(C(\boldsymbol{P}_n) - \boldsymbol{P}_n{}^{obs} \right) , \tag{5}$$

where $\boldsymbol{P}_n$ and $\boldsymbol{P}_n{}^{obs}$ are the forecasted and the observed hourly precipitation amounts, respectively. C is the observational operator (a simple collocation is used in our experiments) and $\boldsymbol{R}_p$ is the observational error (co)variance for precipitation data. The following observed variables are included in the definition of the functional: surface height (derived from the surface pressure observations), temperature, humidity, wind, and accumulated hourly precipitation. We use a diagonal matrix to define observational error covariances. The observational errors depend on the observational type and the pressure level. They vary in the range of 1–2.5 K for temperature, 10–20 m for surface height, 1–10m/s for wind observations. The hourly precipitation observations error is defined to be approximately 1 mm. With the particular weights chosen, the observational part of the functional is typically of the order of 10^5, while the model error and the background terms (second and third term in (1)) are an order of magnitude smaller. The gravity wave penalty term (fourth term in (1)) is approximately two orders of magnitude smaller than the observational term.

4. EXPERIMENTAL DESIGN

We assimilate the conventional observations every three hours and the precipitation observations every hour over the 12–h data assimilation period. The time resolution of the random error term is 3 hours (i. e., four random error terms during the data assimilation period). The minimization process is terminated after 10 iterations, which is approximately within the distance of 1–5% to the exact minimum of the functional (for more explanation, see M Zupanski, 1996). The reason for terminating the minimization process before the exact minimum was reached was primarily due to economy. In addition to this, the effect of further decrease of the functional on the forecast results was marginal, so the need for more iterations was not obvious.

We perform both data assimilation and forecast (up to 48–h) experiments in two different synoptic situations: 00 UTC 08 May 1995 - 12 UTC 10 May 1995 (CASE 1) and 00 UTC 5 Aug 1995 - 12 UTC 7 Aug 1995 (CASE 2). In both situations, considerable amounts of precipitation (up 20–30 mm/hour) were observed in the pre–forecast period over large areas in the continental U. S. (where the precipitation observations are available). Therefore, we expect that assimilating this precipitation data should have a noticeable effect on the 4DVAR data assimilation results. The results of the following three experiments are compared:

(1) 4DVAR_NOPREC, where the initial conditions for the forecast model are defined by applying the 4DVAR method, but using only the conventional observations.

(2) 4DVAR_PREC, where the 4DVAR method was applied to assimilate both the conventional and the precipitation data.

(3) OI_NOPREC experiment, where the OI analyses are created as in the operational Eta Data Assimilation System (EDAS, Rogers et al,, 1995), except for a different space resolution. As in the operational system, the precipitation data are not used. This experiment is used as a control, in order to evaluate the potential of the 4DVAR method to outperform the operational data assimilation method.

5. RESULTS AND DISCUSSION

5.1 Data assimilation experiments

As mentioned in the introduction, a serious problem related to the inclusion of physical parameterizations into the forecast and adjoint models is the presence of nonlinearities and discontinuities. This may have a consequence in slowing down the convergence of the functional, or, in some cases, even in the divergence of the functional (e. g., Verlinde and Cotton, 1993). These problems may become even more pronounced when assimilating the precipitation observations, due to small scale nonlinearities present in the data. Therefore, it is important to pay attention to the convergence of the minimization. In Fig. 1 we present the value of the total functional (eq. (1)) as a function of the iteration number obtained in the experiments 4DVAR_NOPREC and 4DVAR_PREC. The initial analysis is defined as in our earlier experiments (Zupanski and Zupanski, 1996) using previous 4DVAR data assimilation results. (This, so called cycled application of 4DVAR, provides considerably better initial guess analysis than the non–cycled 4DVAR and reduces the number of necessary iterations.) As we can see in Fig. 1, the convergence is relatively fast in both experiments, indicating no serious nonlinearity/discontinuity problems. The functional value gets more reduced in the experiment 4DVAR_PREC due to a better fit to the precipitation data. Other terms of the functional are approximately of the same magnitude.

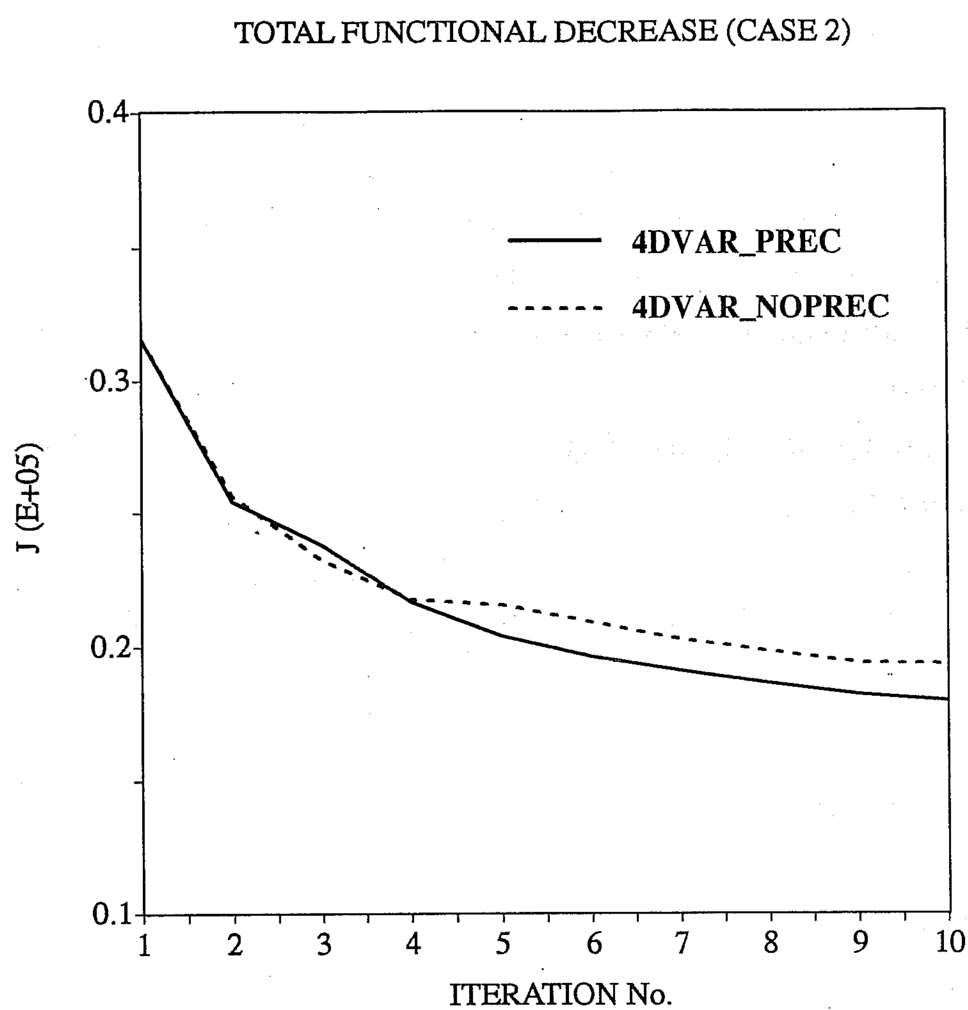

Fig. 1 Total functional decrease (including fitting to precipitation observations) as a function of iteration number calculated in the experiments 4DVAR_PREC and 4DVAR_NOPREC. The assimilation from 00 UTC 05 Aug 1995 to 12 UTC 05 Aug 1995 (CASE 2) is presented.

Another important issue related to the assimilation of precipitation data is the effect of the assimilated precipitation observations on the precipitation forecast fields. To examine this effect we present the observed hourly precipitation at the end of 12–h data assimilation period for CASE 2 in Fig. 2. The corresponding precipitation forecast fields obtained in the experiments 4DVAR_PREC and 4DVAR_NOPREC are given in Figs. 3 and 4, respectively.

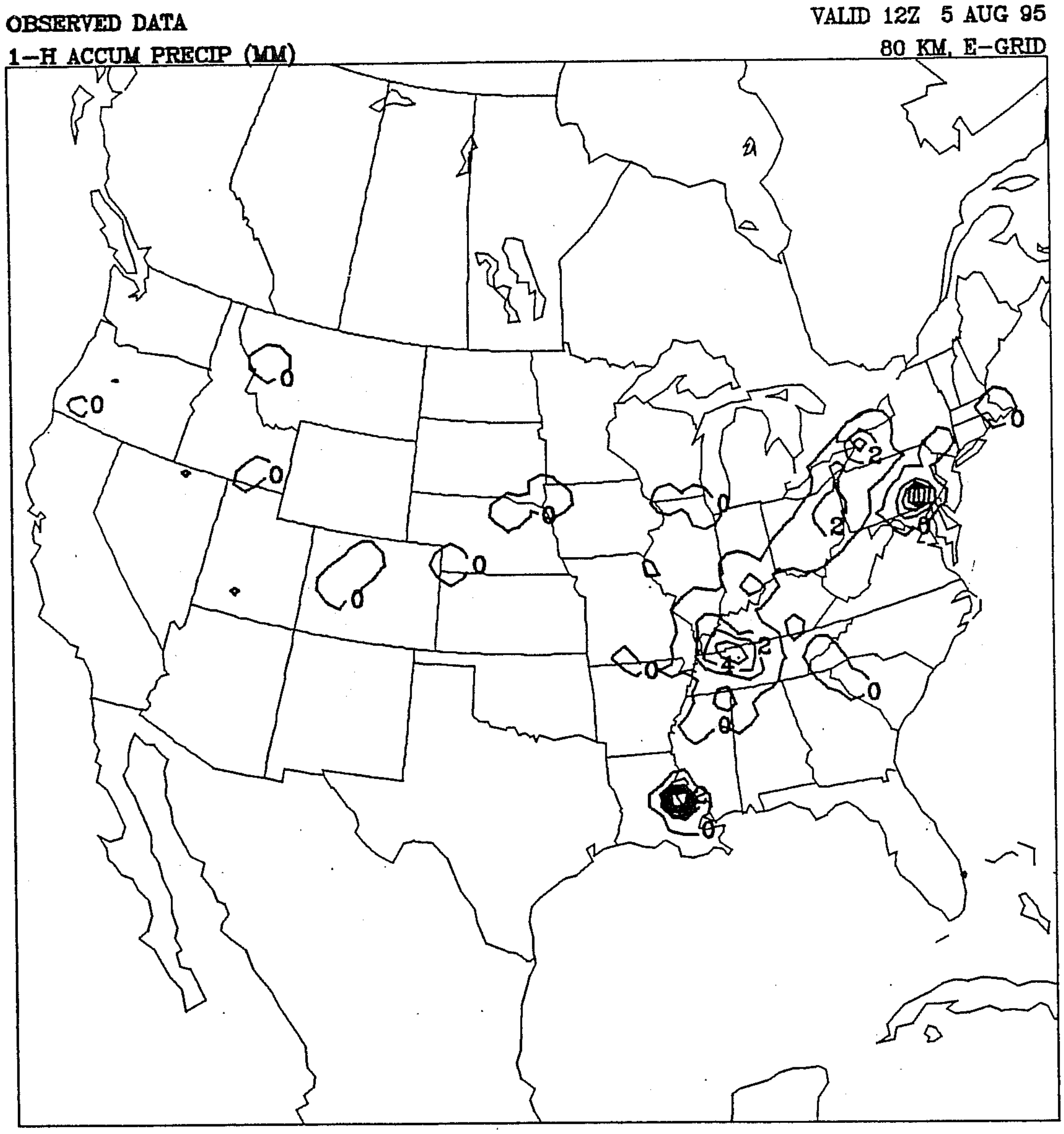

Fig. 2 Observed hourly accumulated precipitation, available over continental U. S., valid 12 UTC 05 Aug 1995. Contouring interval is 2 mm.

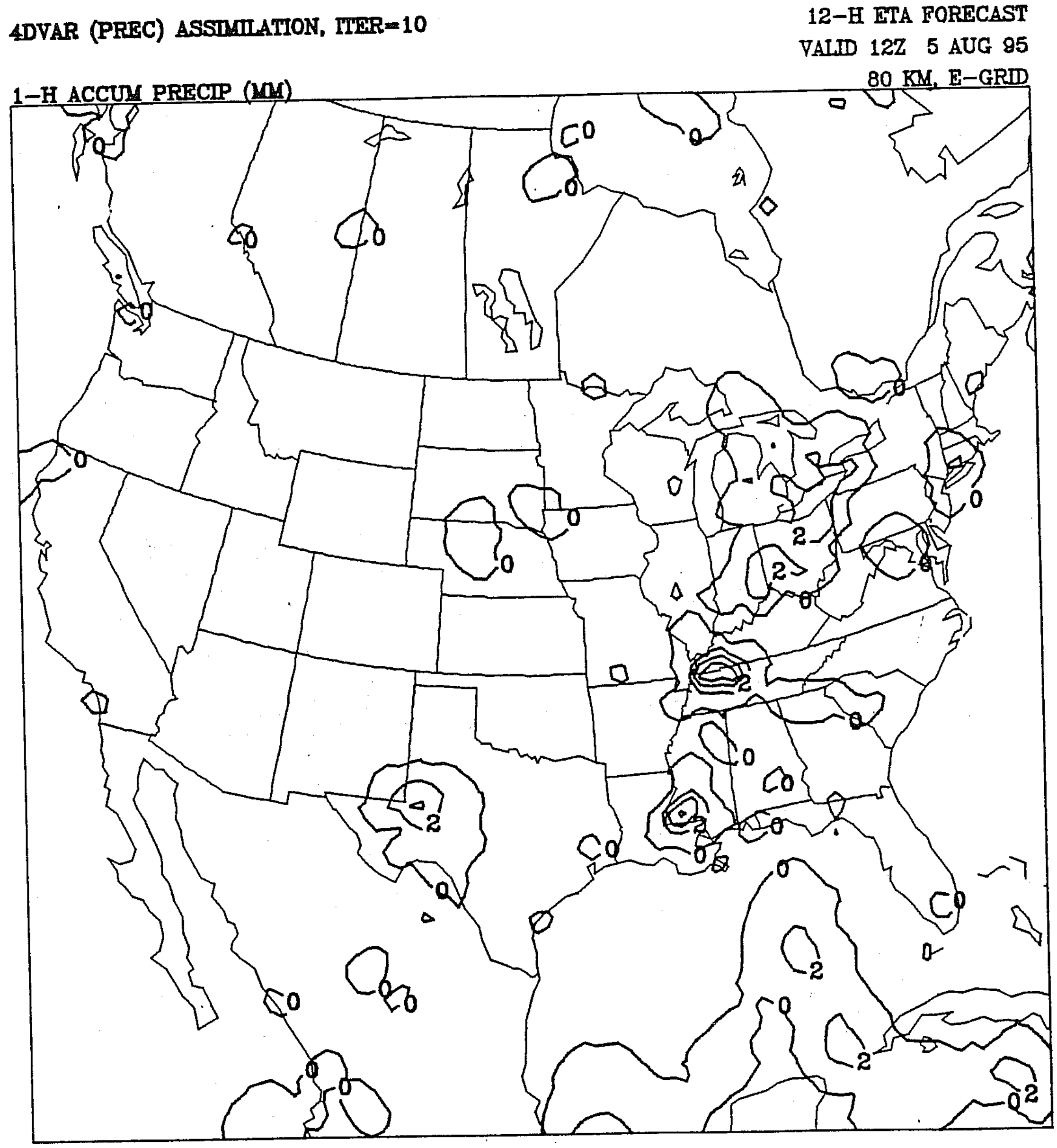

Fig. 3 A 12–h forecast of the accumulated hourly precipitation, valid 12 UTC 05 Aug 1995. It is obtained at the end of data assimilation period in the experiment 4DVAR_PREC after 10 iterations of the minimization. Contouring interval is 2 mm.

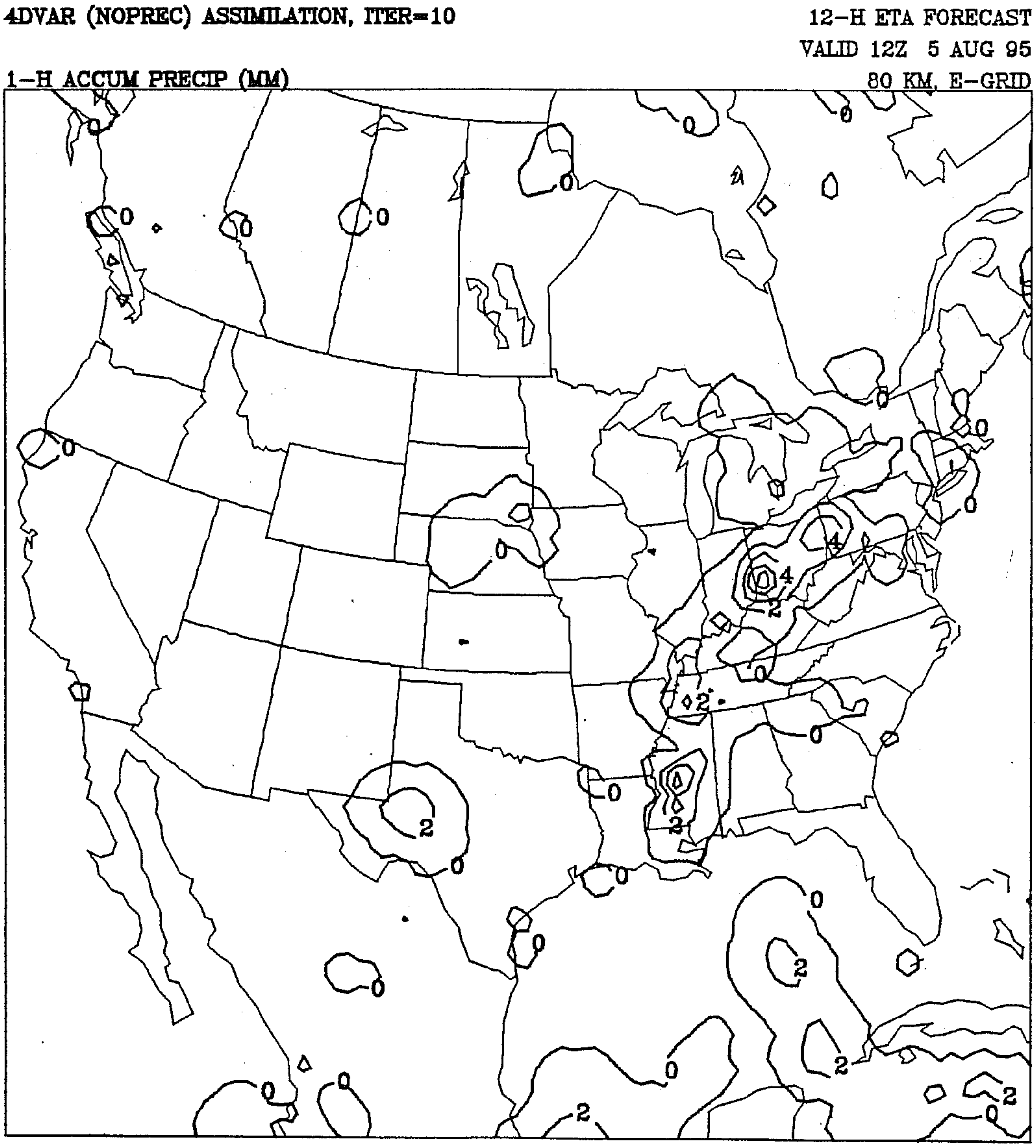

Fig. 4 As in Fig. 3, but for the experiment 4DVAR_NOPREC.

By comparing Figs. 2, 3 and 4 we can immediately see that the assimilation of the precipitation observations results in a considerable improvement of the precipitation forecast. For example, the centers of high precipitation amounts over Tennessee and Mississippi are in much better agreement with the observations in the experiment 4DVAR_PREC then in the experiment 4DVAR_NOPREC. Similar improvements are obtained in CASE 1, as well. It is important to note that the forecast model is capable to realistically produce some very localized maximums of high precipitations (it is even more pronounced in CASE 1, figure not presented). Therefore, we conclude that the fine scale precipitation data can be correctly assimilated even with the coarse resolution 4DVAR data assimilation system. Another important issue is the capability of the free forecast (i. e., forecast after data assimilation), to maintain these structures throughout the forecast period. In other words, we would like to address the question of how long these improvements will last in the future.

5.2 Forecast experiments

The answer to the question of the time scale of the impact of the precipitation data is not an easy task. Some previous experiments (e. g., Zou and Kuo, 1996) suggest that the effect of the precipitation data might be short lasting (3–6 hours) into the future. Similar results are also obtained in the experiments with nudging precipitation data. The forecast results we present here suggest that in some cases, such as CASE 1, the positive effect of precipitation assimilation is very quickly lost after the data assimilation period. In some other cases, such as CASE 2 presented here, the positive effect of assimilating precipitation data may be long lasting into the future.

In the next two figures we present the RMS errors of the accumulated hourly precipitation forecast as functions of the forecast time, calculated against the hourly precipitation observations. In Fig. 5 the results of the experiments 4DVAR_PREC and 4DVAR_NOPREC, CASE 1, are presented. Fig. 6 we shows the results of the same two experiments, but in CASE 2. As we can see, the forecast was considerably improved in both synoptic cases by assimilating the precipitation data during the data assimilation period (from –12h to 0h). In CASE 1 this improvement has quickly diminished in the free forecast period. In CASE 2 however, a small positive effect of the precipitation data remains throughout the 48–h forecast period.

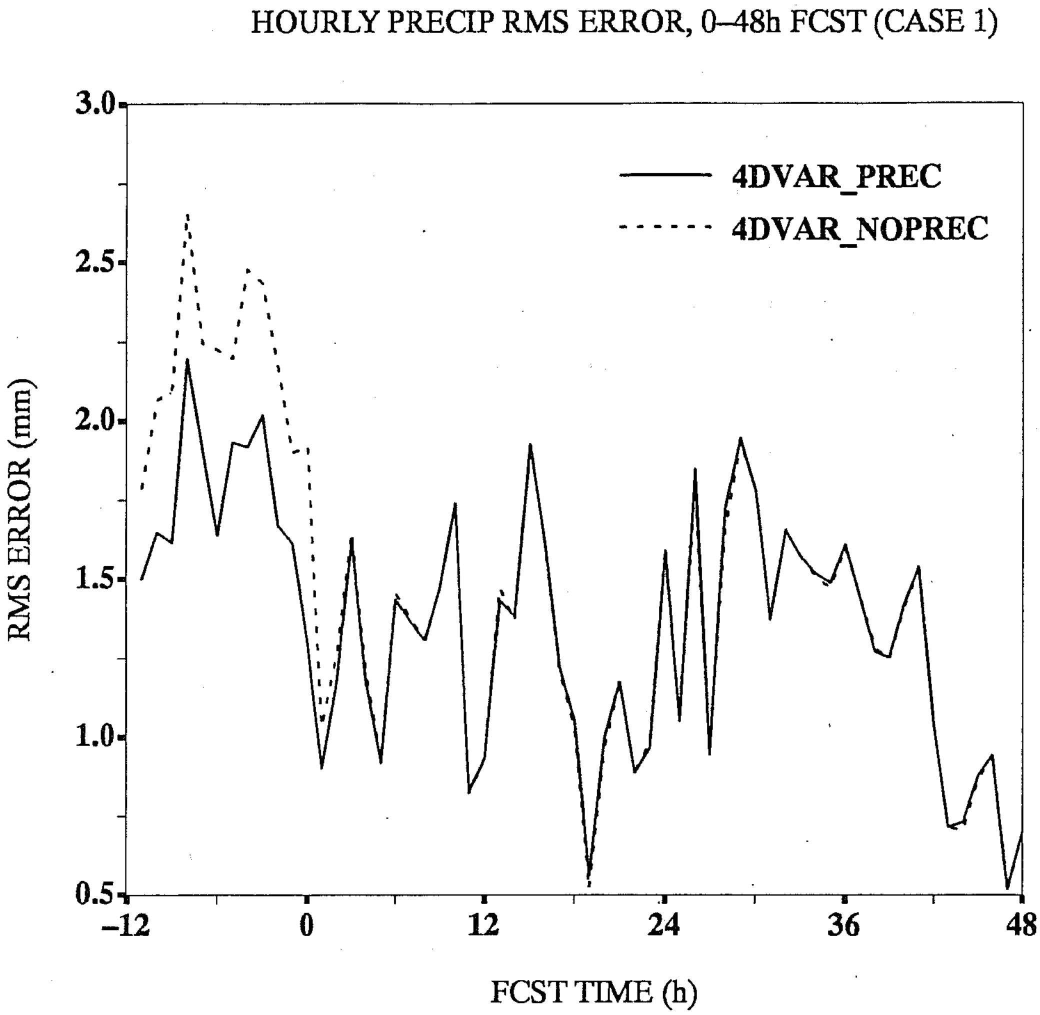

Fig. 5 Hourly precipitation RMS error, calculated against observations in the experiments 4DVAR_PREC (solid line) and 4DVAR_NOPREC (dashed line) for the forecast period from –12h to 48h (CASE 2).

Even though the effect of assimilation of the precipitation data may seem small in the presented figures, it may not be insignificant. A reason for small differences between the two experiments may be in the measure chosen. It seems that the RMS measure is not as sensitive to the small changes in the precipitation fields as some other normalized measures (e. g., equitable threat score). This may have a consequence that in the future we need to consider some alternative definitions of the precipitation part of the functional J using some more sensitive measure for the precipitation forecast error. Looking at the precipitation forecast fields, one can see that the improved precipitation fields in the pre–forecast period may have an impact on the improved precipitation forecast 48–h later. In Fig. 7 we present the observed hourly precipitation field valid 12 UTC 7 Aug 1995, used as verification.

Figs. 8 and 9 show the 48-h forecasts of hourly precipitation, valid at the same time, obtained by the experiments 4DVAR_PREC and 4DVAR_NOPREC, respectively. As indicated in the data assimilation results (Figs. 2, 3 and 4) the precipitation patterns over Mississippi and Tennessee were considerably more realistic in the 4DVAR_PREC experiment then the 4DVAR_NOPREC experiment. Here we notice a more realistic precipitation patterns over Tennessee and Louisiana in the experiment 4DVAR_PREC, which may be due to a successful assimilation of precipitation data in these areas. Also, the precipitation in Pennsylvania is better predicted in the 4DVAR_PREC experiment (a maximum of over 2 mm in the 4DVAR_NOPREC experiment was not observed).

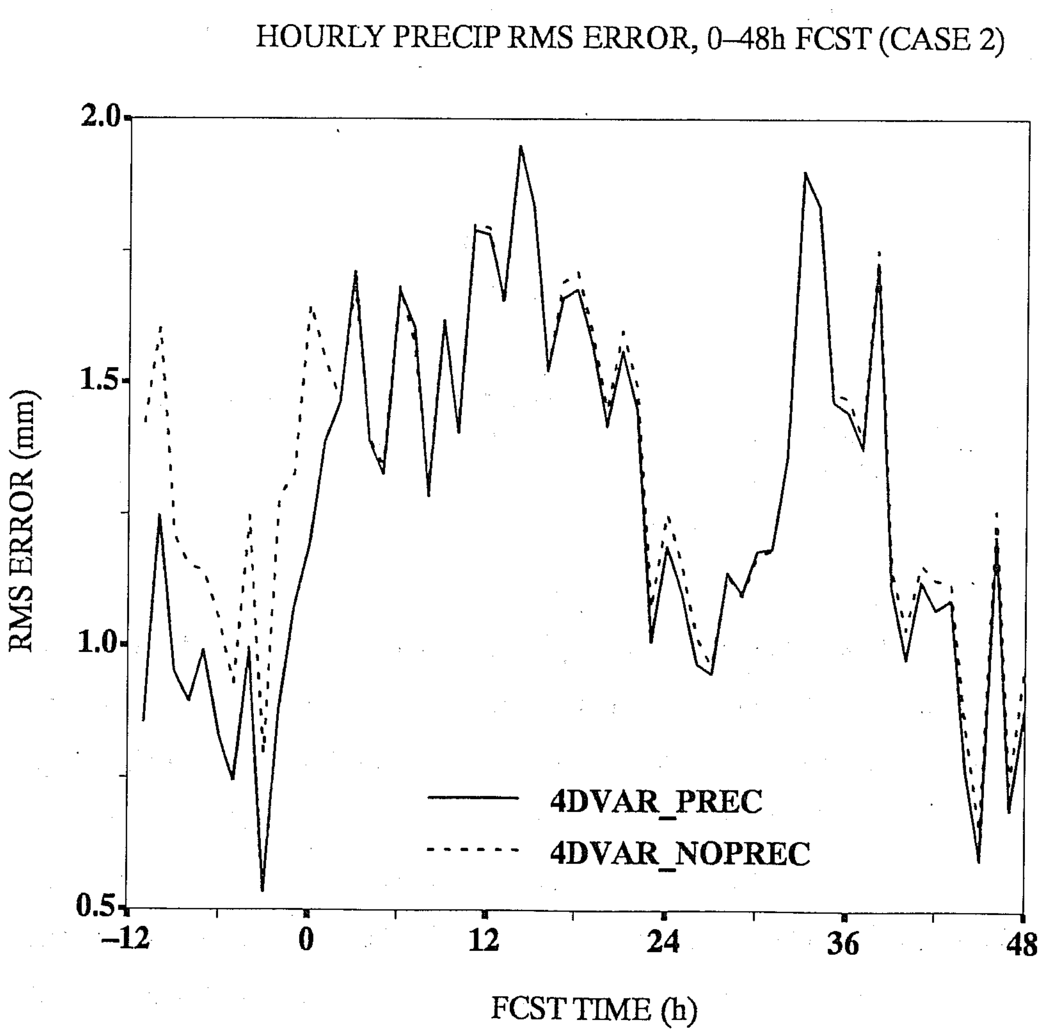

Fig. 6 As in Fig 5, but for CASE 2.

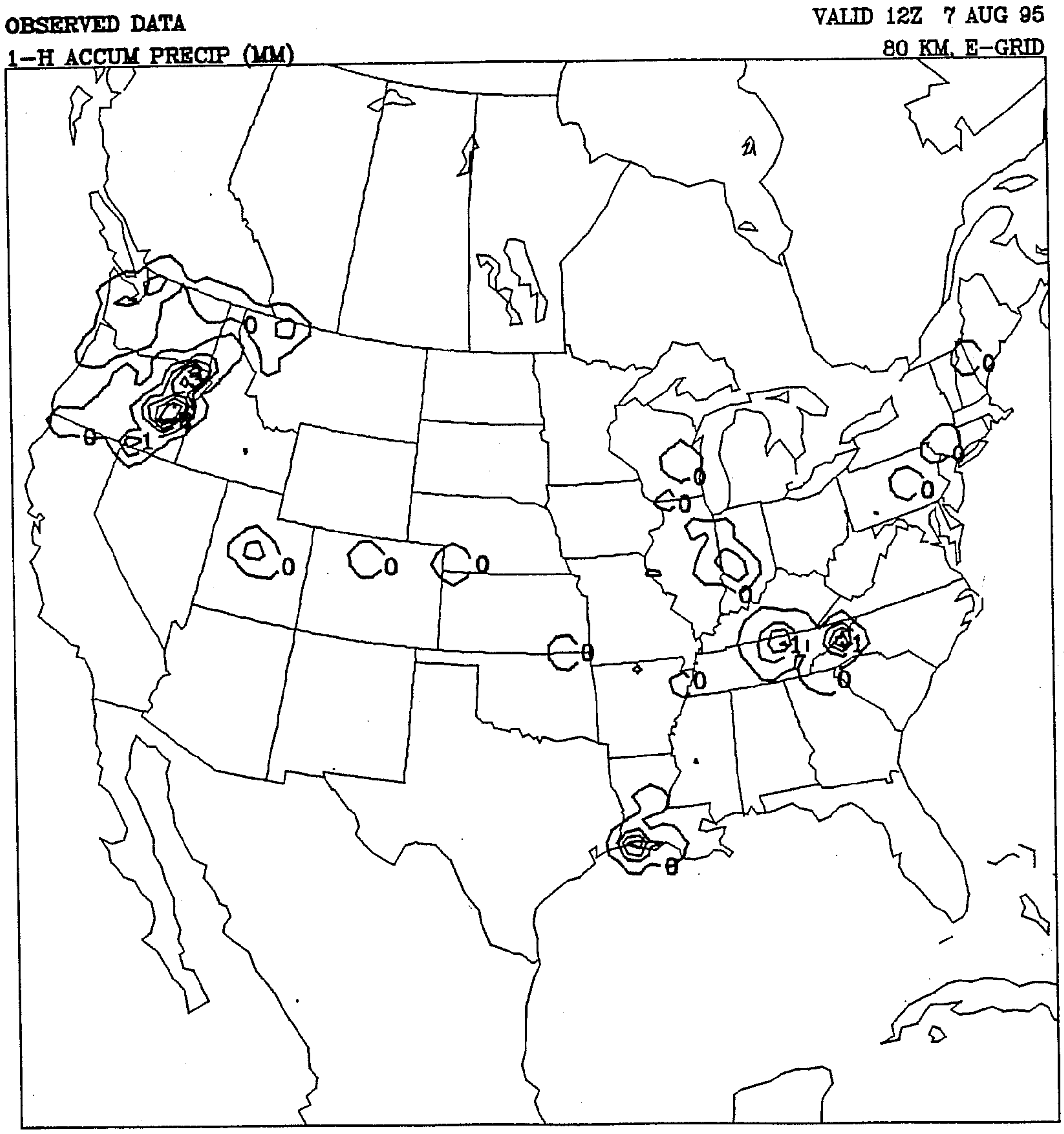

Fig. 7 Observed accumulated hourly precipitation, valid 12 UTC 07 Aug 1995 (CASE 2). Contouring interval is 1 mm.

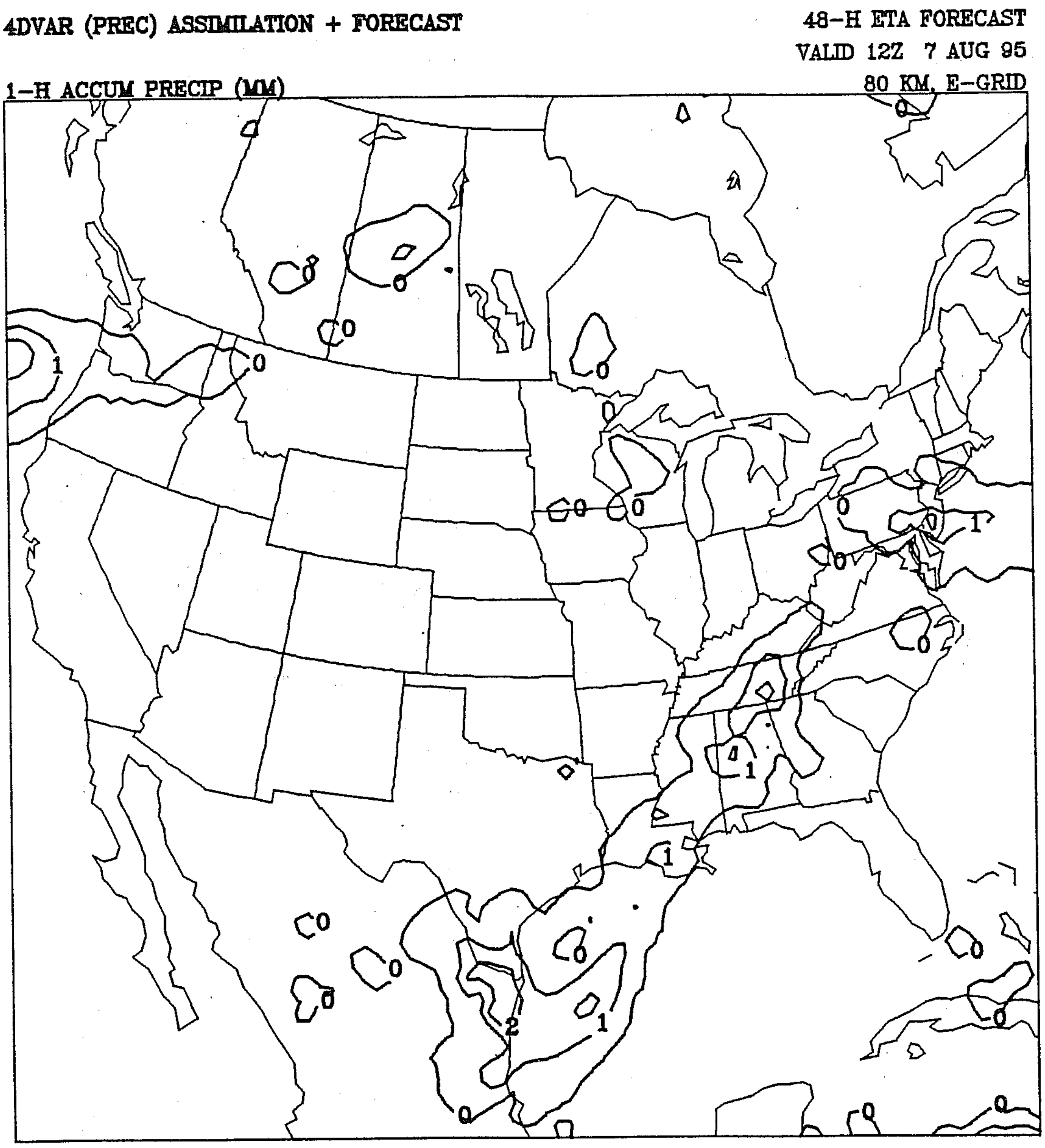

Fig. 8 A 48–h forecast of the accumulated hourly precipitation, obtained in the experiment 4DVAR_PREC, valid 12 UTC 07 Aug 1995. Contouring interval is 1 mm.

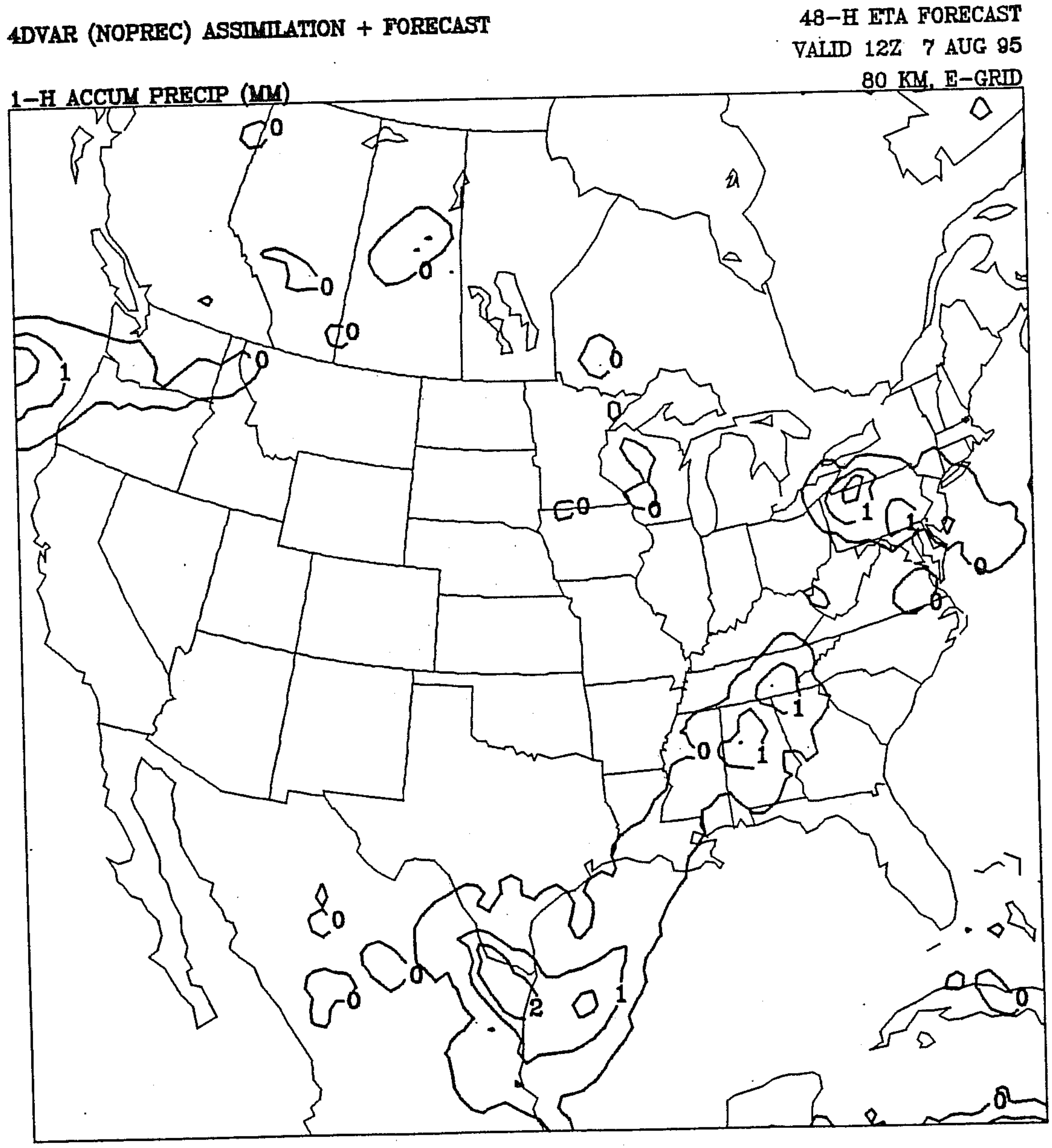

Fig. 9 As in Fig. 8, but for the experiment 4DVAR_NOPREC.

We also compare the results of the 4DVAR method with the OI method. The RMS errors of the hourly precipitation forecasts (up to 48h) obtained by the experiments 4DVAR_PREC and OI_NOPREC are plotted as functions of the forecast time in Figs. 10 (CASE 1) and 11 (CASE 2). As the figures show, the 4DVAR method can outperform the OI method by assimilating the precipitation data and considerably improving the precipitation forecast. The improved forecast is obtained throughout the 48h forecast period.

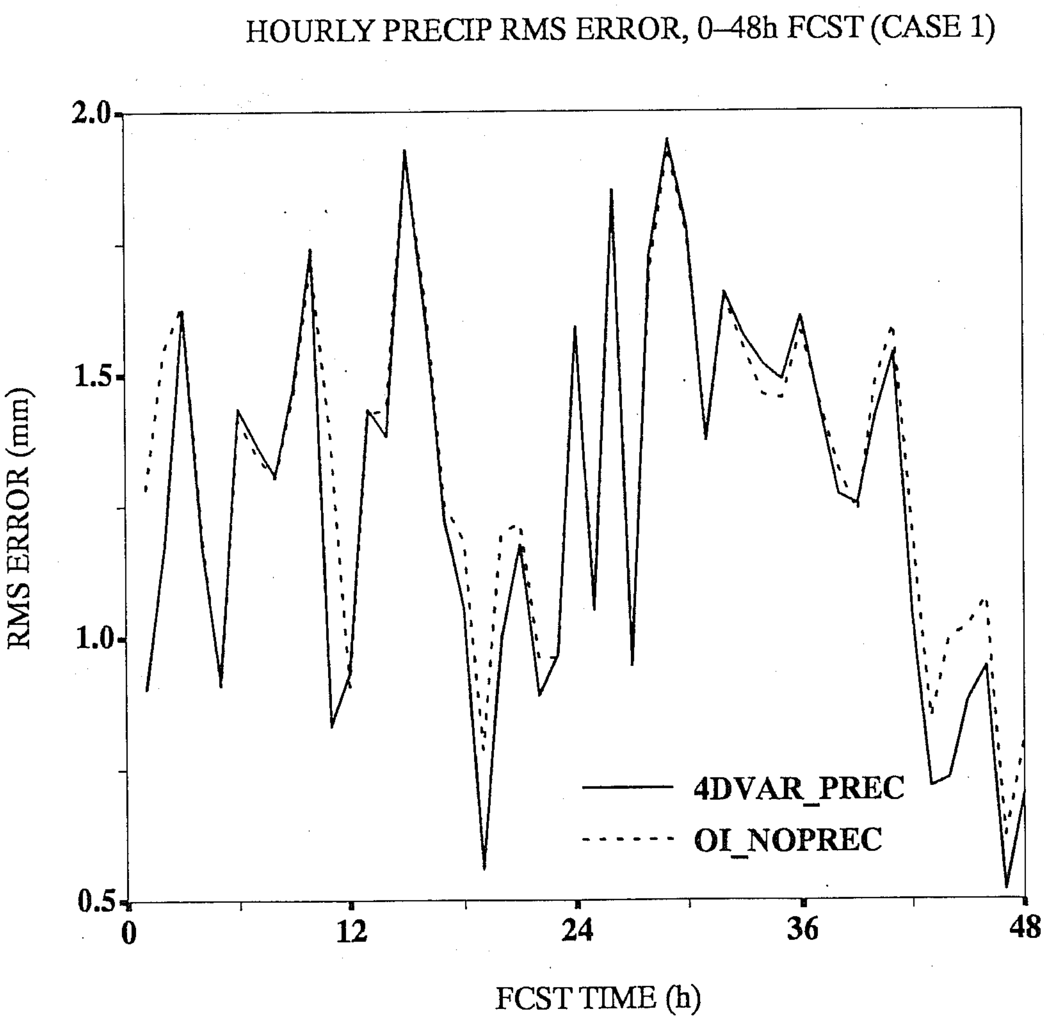

Fig.10 Hourly precipitation RMS error, calculated against observations in the experiments 4DVAR_PREC (solid line) and OI_NOPREC (dashed line) for the forecast period from 0h to 48h (CASE 1).

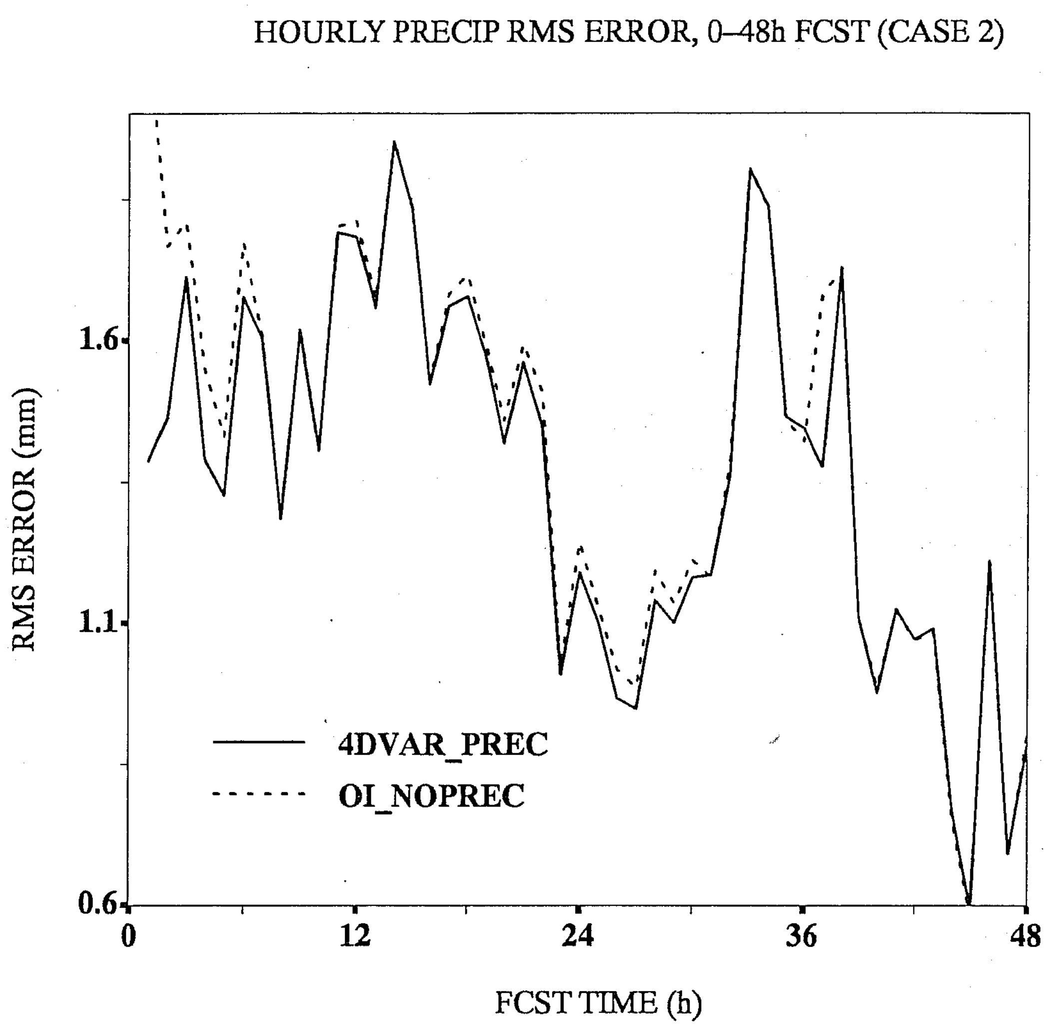

Fig.11 As in Fig. 10, but for CASE 2.

7. CONCLUSIONS

We used a complex state–of–the–art 4DVAR data assimilation system to assess the effects of assimilation of precipitation data. The experimental results showed that the precipitation data can be successefully assimilated by the 4DVAR method. The small scale features in the observed hourly accumulated precipiation data can be realistically retrieved, significantly improving the precipitation forecast during the data assimilation period.

The positive effect of assimilating the precipitation data may sometime last long into the future, improving the precipitation forecast up to 48 hours, as in our example in CASE 2. In some other cases, such as CASE 1 of this paper, the effect of precipitation data diminishes quickly after the data assimilation period. This variable time scale of the impact of precipitation data may be related to the type of the physical processes that played a major role in a specific synoptic situation. This phenomena needs to be examined further. In the future, we also need to consider some altenative definitions of the precipitation part of the functional J using some more sensitive measure for the precipitation forecast error then the RMS error (e. g., the equitable threat score).
By comparing the results of the 4DVAR method and OI method we demonstrated that the capability of the 4DVAR metod to assimilate precipitation data can contribute to further improve the NCEP's operational forecasts, providing the same model resolution is used in the 4DVAR method. This may be a limiting factor at present, due to a limited computer power, but achievable in the near future.

References

Black, T. L., 1994: The new NMC mesoscale eta model: Description and forecast examples. *Wea. Forecasting,* **9**, 265–278.

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 development of the convection, viscous sublayer and turbulence closure schemes. *Mon. Wea. Rev.*, **122**, 927–945.

Janjic, Z. I., F. Mesinger, and T. L. Black, 1995: The pressure–advection term and additive splitting in split–explicit models. *Quart. J. Roy. Meteor. Soc.*, **121**, 953–957.

Mesinger, F., Z. I. Janjic, S. Nickovic, D. Gavrilov, and D. Deaven, 1988: The step mountain coordinate: Model description and performance for cases of Alpine lee cyclogenesis and for a case of an Appalachian redevelpoment. *Mon. Wea. Rev.*, **116**, 1493–1518.

Rogers, E., D. G. Deaven, and G. J. DiMego, 1995: The regional analysis system for the operational "early" eta model: Original 80–km configuration and recent changes. *Wea. Forecasting,* **10**, 810–825.

Sasaki, Y., 1970: Some basic formalisms on numerical variational analysis. *Mon. Wea. Rev.*, **98**, 875–883.

Shanno, D. F., 1978: Conjugate gradient methods with inexact line search. *Math. Operations Res.*, **3**, 244–256.

Shanno, D. F., 1985: Globally convergent conjugate gradient algorithms. *Math. Program.*, **33**, 61–67.

Tsuyuki T. 1996: Variational data assimilation in the tropics using precipitation data. Part II: 3–D model. *Mon. Wea. Rev.*, in press.

Verlinde, J., and W. R. Cotton, 1993: Fitting microphysical observations of nonsteady convective clouds to a numerical model: An application of the adjoint technique of data assimilation to a kinematic model. *Mon. Wea. Rev.*, **121**, 2776–2793

Zou, X., and Y.–H. Kuo, 1996: Rainfall assimilation through an optimal control of initial and boundary conditions in a limited–area mesoscale model. *Mon. Wea. Rev.*, in press.

Zupanski, D., 1997: A general weak constraint applicable to operational 4DVAR data assimilation systems. *Mon. Wea. Rev.*, submitted.

Zupanski, D., and F. Mesinger, 1995: Four–dimensional variational assimilation of precipitation data. *Mon. Wea. Rev.*, **123**, 1112–1127.

Zupanski, M., 1993: A preconditioning algorithm for large scale minimization problems. *Tellus*, **45A**, 578–592.

Zupanski, M., 1996: A preconditioning algorithm for four–dimensional variational data assimilation. *Mon. Wea. Rev.*, **124**, 2562–2573.

Zupanski, M., and D. Zupanski 1995: Recent developments of the NMC's regional four–dimensional varational data assimilation system. *Proc. of the Second International Symposium on Assimilation of Observations in Meteorology and Oceanography*, Tokyo, Japan,WMO, 376–372.

Zupanski M., and D. Zupanski, 1996: A quasi–operational application of a regional four–dimensional variational data assimilation. *11 th Conference on Numerical Weather Prediction.* Norfolk, Virginia, WMO, 94–95.