Initialization of Cloud Water Content in a Data Assimilation System

XIANG-YU HUANG

Danish Meteorological Institute, Copenhagen, Denmark

(Manuscript received 11 April 1995, in final form 14 April 1995)

ABSTRACT

Cloud water content (CWC) is not treated in most operational objective analyses and initialization schemes. When CWC is used as a prognostic variable in a forecast model, it is necessary to define this variable at the initial time. A commonly used method is to set the initial CWC to zero or use a forecast CWC field from the previous data-assimilation cycle (the first-guess field for the objective analysis) without any modification. The inconsistent treatment of CWC and other fields leads to an imbalance between the first-guess cloud water field and other analyzed fields (winds, temperature, humidity, and surface pressure). In this study, the diabatic digital-filtering initialization scheme is used to alleviate this imbalance. It is shown that an intermittent data assimilation system with this initialization scheme can produce a better cloud evolution, a shorter spinup time, and a removal of the initial shock in precipitation.

1. Introduction

Numerical weather prediction models that explicitly predict cloud water content (CWC) require this field to be specified at the initial time. However, CWC is not normally treated in operational objective analyses and initialization schemes. Therefore, it is necessary to devise some method of defining this variable in a consistent manner at the initial time. Among the earlier attempts, Kristjánsson (1992) has tried to establish an initial CWC field from satellite data and model generated data. One problem found by Kristjánsson (1992) is that the CWC field thus obtained is not in a balance with other model fields.

In a recent study by Huang and Sundqvist (1993), a realistic CWC field is obtained as a by-product of an initialization scheme: the diabatic digital-filtering initialization (DFI) scheme of Huang and Lynch (1993). They have shown that the DFI produced CWC field qualitatively agrees with satellite observations and has a reasonable magnitude compared with the corresponding 24-h forecast. They have furthermore argued that the spinup time can be reduced with the DFI-generated CWC compared to that of a forecast started from zero cloud. The study of Huang and Sundqvist (1993) is just a case study. Only a forecast model is involved to demonstrate the potential usefulness of DFI in constructing a CWC field from zero cloud. The conclusions from Huang and Sundqvist (1993) may be different if a data assimilation system is used.

Once the data assimilation cycles have started, cloud information is available from the previous cycle. Without quantitative cloud observations and any treatment in the analysis scheme, it is natural to take the CWC field from first guess instead of starting from zero cloud (Kristjánsson 1992). New observations often lead to changes in the analyzed fields (temperature, wind, humidity, and surface pressure) to the first guess through the analysis scheme. These changes could be large both in phase and amplitude. As an example, the first guess and the analysis of the mean sea level pressure at 0000 UTC 16 September 1994 are shown in Fig. 1 together with the difference between these two fields (the analysis increment). The maximum difference is as large as 5 hPa. If the CWC field obtained from the first guess remains unchanged after analysis and initialization, it could contain large errors compared with other *properly* analyzed fields. The most evident symptom of these errors is an initial shock in the precipitation and a dramatic decrease in the mean cloud water in the beginning of the forecast. Although the spinup time is shorter than that without initial cloud, the mean cloud water seems to have difficulties to reach its actual level within a 6-h assimilation cycle. A possible consequence is that no matter how long the data assimilation is run (with a fixed cycle length, e.g., 6 h), the cloud amount is always different from its actual value, as shown schematically in Fig. 2. The inconsistency between the CWC and other model variables may also lead to an imbalance between the mass and wind fields, giving rise to numerical noise in the following forecast.

With the help of the DFI scheme, the shock–spinup process could be finished before the end of a 6-h cycle (middle thicker curve in Fig. 2). This is achieved by the initialization procedure as indicated by the thin

---

Corresponding author address: Dr Xiang-Yu Huang, Danish Meteorological Institute, Lyngbyvej 100, DK-2100 Copenhagen Ø, Denmark.

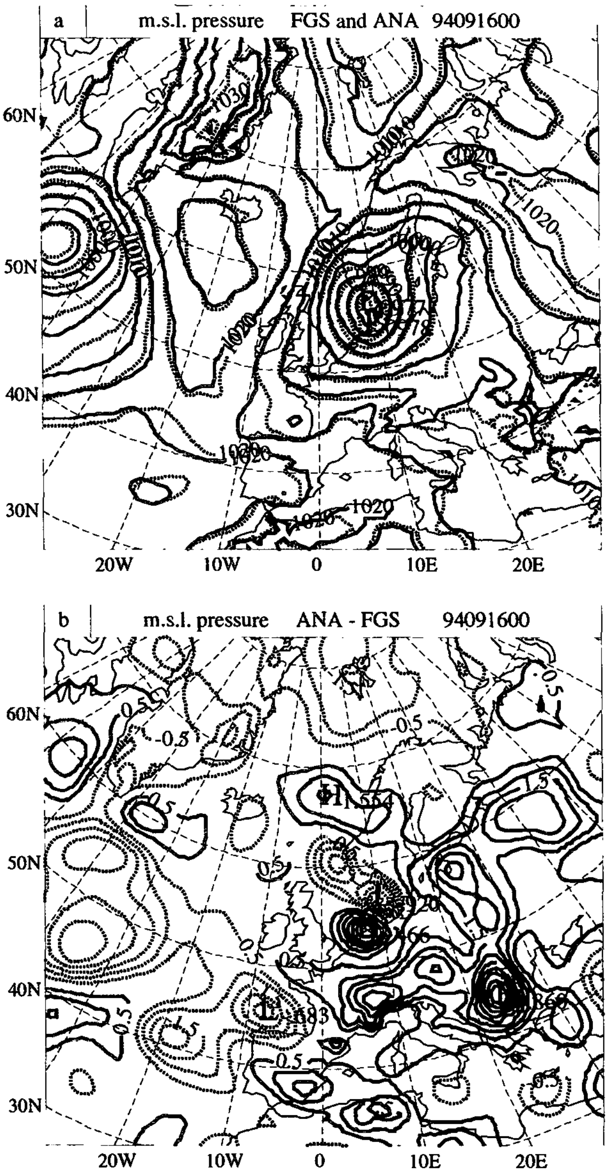

FIG. 1. (a) The first guess (dash line) and the analysis (full line) of the mean sea level pressure at 0000 UTC 16 September 1994 in the experiment NMIEXP. The contour interval is 5 hPa. (b) The analysis increment of the mean sea level pressure at 0000 UTC 16 September 1994 (positive is full line, negative is dash line, line is suppressed for zero). The contour interval is 0.5 hPa.

dashed line in the middle of Fig. 2. The digital-filtering initialization scheme requires first an adiabatic integration of the forecast model backward in time and then a diabatic integration forward in time. The initialized fields are obtained by applying a digital filter over the time series produced by the diabatic integration (Huang and Lynch 1993). As schematically shown in Fig. 2, the cloud amount remains on the same level during the adiabatic backward integration because only the advection of cloud is involved for CWC. The aforementioned imbalance also leads to an initial shock in precipitation in the beginning of the diabatic forward integration, but this happens during the initialization procedure (not in the forecast). The spinup process takes place before the forecast starts. The filtered initial state exhibits a balance between cloud and other fields (Huang and Sundqvist 1993). There is no initial shock in precipitation. The cloud amount in the initialized model state could not reach the normal level (Fig. 2), partly because the filtering process takes into account of the whole diabatic integration, including the spinup process. Therefore, after the initialization, there is still a short spinup period in the forecast, but this period is shorter than that with unchanged first-guess cloud as initial conditions. With the DFI scheme, the cloud amount in data assimilation cycles could reach the normal level within the 6-h cycle (Fig. 2).

In this study, a data assimilation system is used to test the initialization of CWC as discussed above in a preoperational environment.

2. The data assimilation system

The data assimilation system used here is a comprehensive numerical weather prediction system developed in a joint Nordic–Dutch–Irish research project (Machenhauer 1988; Gustafsson 1993). While a more detailed description can be found in Huang et al. (1994), a brief overview is given here. The system is

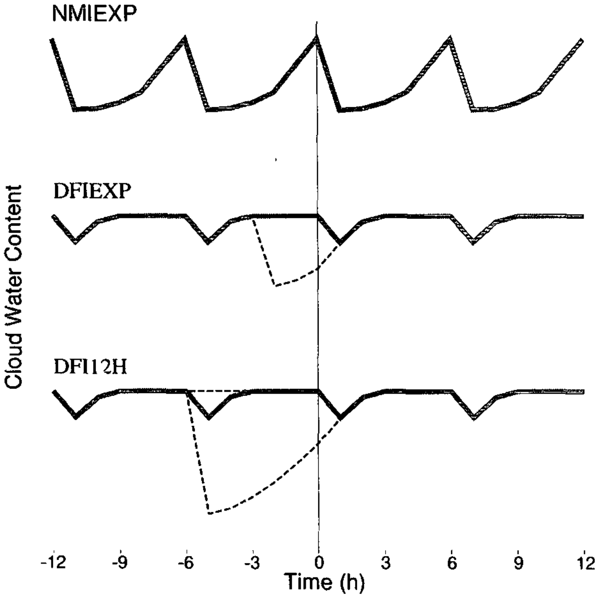

FIG. 2. Schematic diagram showing the qualitative evolution of mean cloud water C for NMIEXP (upper curve) and DFIEXP (middle curve). The lower curve (to be discussed later in section 5) is for DFI12H, which has the same setup as DFIEXP except that the filter span is 12 h in this experiment. The thin dashed lines show the qualitative evolution of C during the initialization process.

an intermittent data assimilation system including objective analysis, initialization, and forecast.

a. Analysis

In the objective analysis, the observation window covers a 6-h span around the analysis time (0000, 0600, 1200, and 1800 UTC). The first-guess field is the 6-h forecast from the previous data assimilation cycle. Three-dimensional multivariate statistical interpolation is used for wind, geopotential, and surface pressure. Three-dimensional univariate statistical interpolation is used for relative humidity. The cloud water content has not been incorporated in the analysis in any form. It is taken from the first guess without any change.

b. Initialization

In the initialization step, two schemes are used. The experiments to be discussed in this study only differ from each other in the initialization scheme involved and will therefore be named after the initialization schemes.

1) DFIEXP

In this experiment, the diabatic digital-filtering initialization scheme is used in the initialization step. A detailed description of the scheme is given in Huang and Lynch (1993). The parameters are chosen as in Huang and Sundqvist (1993) with Lanczos window, 6-h cutoff period, and 6-h filter span. The initial model state is obtained by filtering a model trajectory. Two model integrations are needed. First the model is integrated adiabatically backward in time from the analysis. During this integration, all model variables, including humidity and CWC, are involved, but the diffusion and physical processes are disabled. The advection of humidity and CWC may need special treatment, for example, the upstream scheme used for CWC needs a modification (to a ''downstream scheme'') for the backward integration due to numerical stability considerations. Using the model state at the end of the adiabatic integration as the initial condition, a full model is integrated forward in time. This integration provides the digital filter with a model trajectory centered around the analysis time. All model fields are filtered to yield the initialized model state. In this study, only one filter is used with the parameters given above. Other filters and/or filter parameters may also be used. The digital filter with the chosen parameters is used for the convenience of comparisons with earlier studies (Huang and Lynch 1993; Huang and Sundqvist 1993). The filter is designed to remove high-frequency oscillations from a time series. The amplitude of the oscillations with the cutoff period is about halved by the filter. The filter span controls the quality of the filtering. The parameters chosen here are found to give satisfactory filtering results (Lynch and Huang 1992).

2) NMIEXP

This experiment is performed as a reference. The nonlinear normal-mode initialization (NMI) scheme is used in the initialization step. The scheme is implicit. Only the adiabatic model is used during the initialization. Four modes are initialized. Two iterations are performed. The choice of the above parameters is based upon the operational setup currently used at the Danish Meteorological Institute. A diabatic NMI scheme with all modes initialized and six iterations could have a significant impact on the model noise, but not much on the precipitation rate due to the particular formulation of the diabatic NMI scheme that does not modify humidity and CWC fields (Huang et al. 1994).

c. Forecast

The forecast model is a primitive equation model with horizontal wind components, temperature, specific humidity, surface pressure, and cloud water content as prognostic variables. It has a rotated latitude–longitude horizontal grid; a hybrid sigma-pressure vertical coordinate; a second-order accuracy in finite-difference scheme; and a leapfrog, semi-implicit scheme with Asselin time filtering. It contains a comprehensive physics parameterization package including the cloud water scheme of Sundqvist et al. (1989). The European Centre for Medium-Range Weather Forecasts (EMCWF) forecasts available at the time of the analysis are chosen as the boundary data. The model is run with 110 × 100 × 16 grid points and 0.51° horizontal resolution. The time step is 5 min.

3. Results

The data assimilation cycle is run for a 5-day period from 0000 UTC 13 September 1994 to 0000 UTC 18 September 1994. The first analysis uses the ECMWF forecast as the first guess. The model domain used in data assimilation experiments is shown in Fig. 1.

a. Initial shock and spinup

One of the major motivations of this study is to show the impact of DFI on the spinup of CWC and related fields. The mean cloud water C (g m^{-2}) and the mean total precipitation rate R (mm day^{-1}) are shown for the 5-day data assimilation experiments in Fig. 3 and Fig. 4, respectively. Mean cloud water C is defined as

$$C = \frac{1}{IJ}\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}(\mathrm{CWC})_{ijk}\frac{\Delta p_k}{g},$$

where $(\mathrm{CWC})_{ijk}$ is the cloud water content at a model grid, Δp_k is the pressure difference between vertical levels, and g is the gravity. Note that C is not a simple average of $(\mathrm{CWC})_{i,j,k}$. The summation in the vertical includes the air density, ρ: $\Delta p_k/g = \rho_k \Delta z_k$ through the hydrostatic assumption. In order to show the charac-

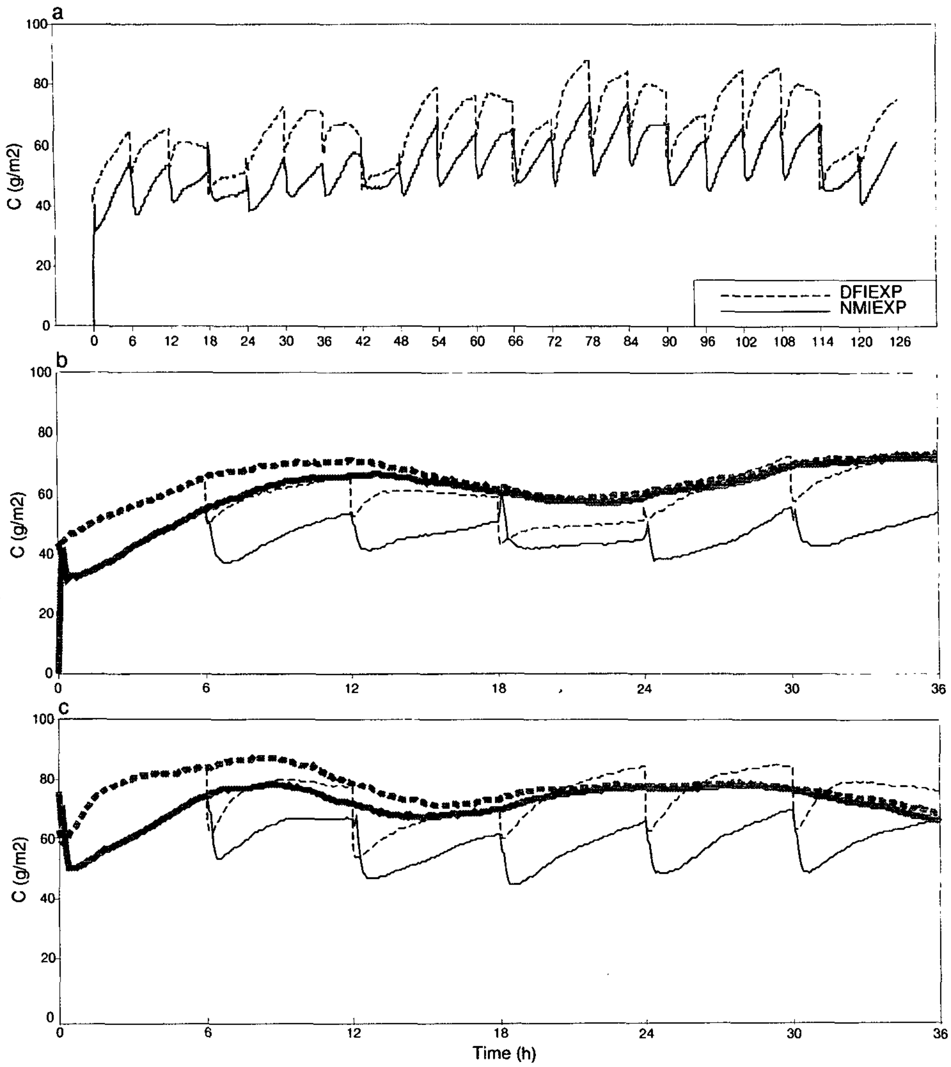

FIG. 3. Vertically integrated cloud water averaged over the model domain C (g m^{-2}) as a function of time (h). The results for NMIEXP are shown by full lines and DFIEXP by dash lines. (a) The 5-day data assimilation experiments. (b) The first 36-h forecasts (thicker lines) together with the data assimilation results in the same period (thin lines). (c) The 36-h forecasts started at 0000 UTC 16 September 1994 (thicker lines) together with the data assimilation results in the same period (thin lines).

teristics of the spinup, 36-h forecasts from 0000 UTC 13 September 1994 (the very first cycle) and that from 0000 UTC 16 September 1994 (well into the data assimilation cycle, the thirteenth cycle) are also shown in the figure by thicker lines. The results for NMIEXP are shown by full lines and DFIEXP by dash lines.

The first cycle of the data assimilation experiment needs some special attention. The analysis in the first cycle uses the ECMWF forecast as the first guess, which has no cloud. Since the NMI scheme does not include CWC, the first forecast in NMIEXP started without CWC. It takes more than 6 h for C to reach a reasonable amount. Even longer time is needed for R to spin up to an acceptable intensity. Cloud water C increases in time with a rapid jump on the first time step; R also increases in time with an initial shock on the second time step.

The DFI scheme takes the full model into consideration and produces a CWC field during the forward integration of DFI. The first forecast in DFIEXP starts from the DFI-generated CWC. The improvements by using the initialized CWC can be noticed from the figure: a shorter spinup time (about 3 h) for both C and R, no initial shock in R. It is interesting to notice that R in the two forecasts converges after 18-h integration.

From the second cycle, the data assimilation system has the first guess from its own forecast. The

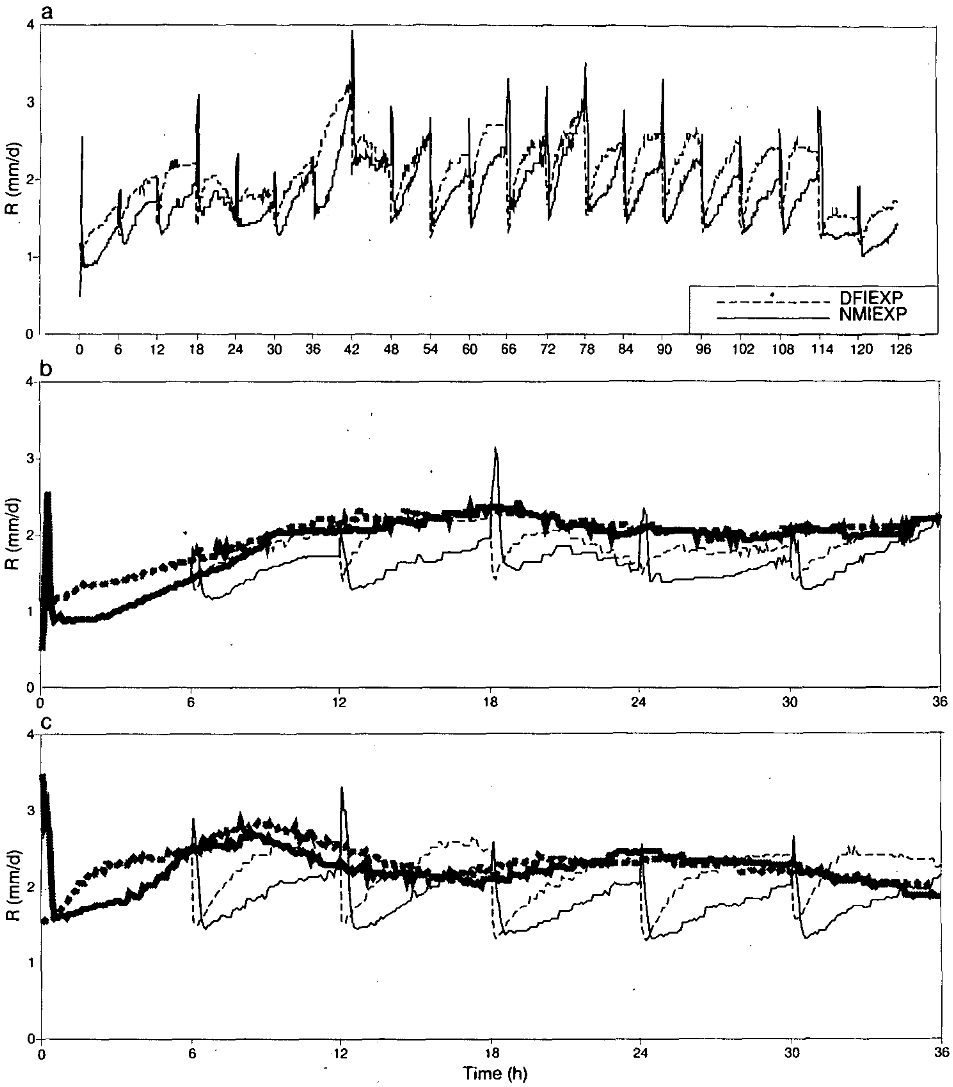

FIG. 4. Mean total precipitation rate R (mm day^{-1}) as a function of time (h). The results for NMIEXP are shown by full lines and DFIEXP by dash lines. (a) The 5-day data assimilation experiments. (b) The first 36-h forecasts (thicker lines) together with the data assimilation results in the same period (thin lines). (c) The 36-h forecasts started at 0000 UTC 16 September 1994 (thicker lines) together with the data assimilation results in the same period (thin lines).

conditions for initial CWC in NMIEXP have changed: a 6-h forecast CWC field from the previous cycle is used directly together with other fields, which are analyzed. The problem now comes from the imbalances between CWC and other fields. It is not difficult to find examples in which the analysis scheme makes large changes in analyzed fields to the first guess (e.g., in Fig. 1). The initialization scheme NMI adds on further changes in these fields. Only CWC remains unchanged from the first guess. This imbalance leads to a significant decrease of C and an initial shock in R. However, it may be argued that the CWC from the first-guess is better than nothing. Comparing C and R in later cycles and that from the first, it seems that the spinup time becomes shorter. The 36-h forecast in the thirteenth cycle can be seen as a representative run.

Both C and R evolve in a qualitatively similar manner in the rest of the cycles of DFIEXP as described above for the first cycle. The discussion made earlier on the first cycle can also be made here for other cycles, that is, the spinup time is shorter than that in NMIEXP and there is no initial shock in R. It seems evident that the DFI scheme has provided a balanced initial model state that leads to a smoother cloud evolution in the following forecast. Our wish to alleviate the imbalances due to the lack of consistent analyses of cloud is basically fulfilled.

b. Noise control

One of the purposes of using initialization schemes is to control the numerical noise caused by the imbalances between mass and wind fields introduced by the objective analysis. The mean absolute surface pressure tendency N is chosen to measure the global noise level, which is defined as

$$N = \frac{1}{IJ} \sum_{i=1}^{I} \sum_{j=1}^{J} \left| \frac{\partial p_s}{\partial t} \right|_{ij},$$

where p_s is the surface pressure and the summation is calculated over the whole model domain. In Fig. 5, the variation of N [hPa $(3\ \mathrm{h})^{-1}$] as a function of time is shown for the 5-day data assimilation experiments. The results for NMIEXP are shown by full lines and DFIEXP by dash lines. In order to show a "noise free" evolution of N, 36-h forecasts from the first and thirteenth cycle are also shown in Fig. 5.

Without initialization schemes, N is typically above 8 hPa $(3\ \mathrm{h})^{-1}$ at the start of the forecast (Lynch and Huang 1992; Huang and Lynch 1993). From Fig. 5, it can be seen that both initialization schemes control the noise effectively. It is also clear that DFI is better than NMI in the noise control according to this measure. The formulation of the DFI scheme and the consideration of all diabatic effects all contribute to the differences in the figure. Similar results can also be found in Huang et al. (1994), in which CWC is not included as a prognostic variable in the model formulation.

c. Forecast quality

In each data assimilation cycle, a 36-h forecast is run to assess the impact of initialization schemes on the forecast quality. To obtain an objective comparison, model fields are directly verified against observations of European radiosonde and synoptic stations (Hall 1987). The results are summarized in Fig. 6. Both bias (the lower pair of curves) and rms (the upper pair of curves in each panel) are shown as functions of forecast length. The averages are taken from all 20 forecasts in the 5-day data assimilation. As in the Danish Meteorological Institute operational setup, 12 parameters are chosen: geopotential height Z; temperature T; wind V at 250, 500, and 850 hPa; mean sea level pressure MSLP; 2-m temperature T02M; and 10-m wind V10M. The results for NMIEXP are shown by full lines and DFIEXP by dash lines.

The quality of analysis is given by bias and rms at time 0 h. Using observations as a reference, the objective analysis has hardly been affected by using different initialization schemes. Even the noticeable differences found in the 2-m temperature are only 0.3 K in bias and 0.2 K in rms. Further investigations may be needed to find out whether or not there is an inconsistency in the treatment of surface parameters in the DFI scheme. It can be concluded that the difference between NMIEXP and DFIEXP at analysis time is very small, especially when compared with the observation error (Daley 1991).

The quality of first guess is given by bias and rms at time 6 h. Here the differences between NMIEXP and DFIEXP are somewhat greater than those at the analysis time. For some parameters (e.g., 500-hPa temperature) NMIEXP is better than DFIEXP in producing a first-guess field, while for other parameters (e.g., 500-hPa geopotential height) DFIEXP is better. If the only concern is the quality of the data assimilation cycles, these differences have no significance since the analyses partly determined by these first-guess fields are of the same quality.

Considering the whole range of 36-h forecasts, it is evident that the DFI scheme works satisfactorily. The performance of DFI is at least as good as that of NMI. There are small differences in bias and rms between the two experiments and it seems, in general, that DFIEXP has a slightly better forecast skill than NMIEXP. In addition, it should be pointed out that the DFI scheme has a number of advantages over the NMI scheme, for example, better noise control and a quicker spinup time.

The verification of humidity and cloud has not been included in the observation verification package. In this sense, the observation verification scores discussed above are not complete. The humidity verification against observations is currently under testing. There are still difficulties in the understanding of humidity scores. The quantitative comparisons between cloud observations and model-simulated cloud fields are even more difficult (Raustein et al. 1990). Further efforts are needed in this direction.

4. Conclusions

Cloud water content is not treated in most operational objective analysis and initialization schemes. When CWC is used as a prognostic variable in a forecast model, it is necessary to define this variable at the initial time. A commonly used method is to set the initial CWC to zero or use a forecast CWC field from the previous data assimilation cycle (the first-guess field for the objective analysis) without any modification. The inconsistent treatment of CWC and other fields leads to an imbalance between the first-guess cloud water field and other analyzed fields (winds, temperature, humidity, and surface pressure).

At the start of the data assimilation cycles, there is no information about CWC and the system has to begin with no cloud. This leads to a spinup time typically longer than 12 h. Attempts have been made to introduce an initial CWC field from observations and/or other model variables. One of them is to use the diabatic digital-filtering initialization, which is able to create a CWC field as a "by-product" of the initialization process (Huang and Sundqvist 1993). It has been shown

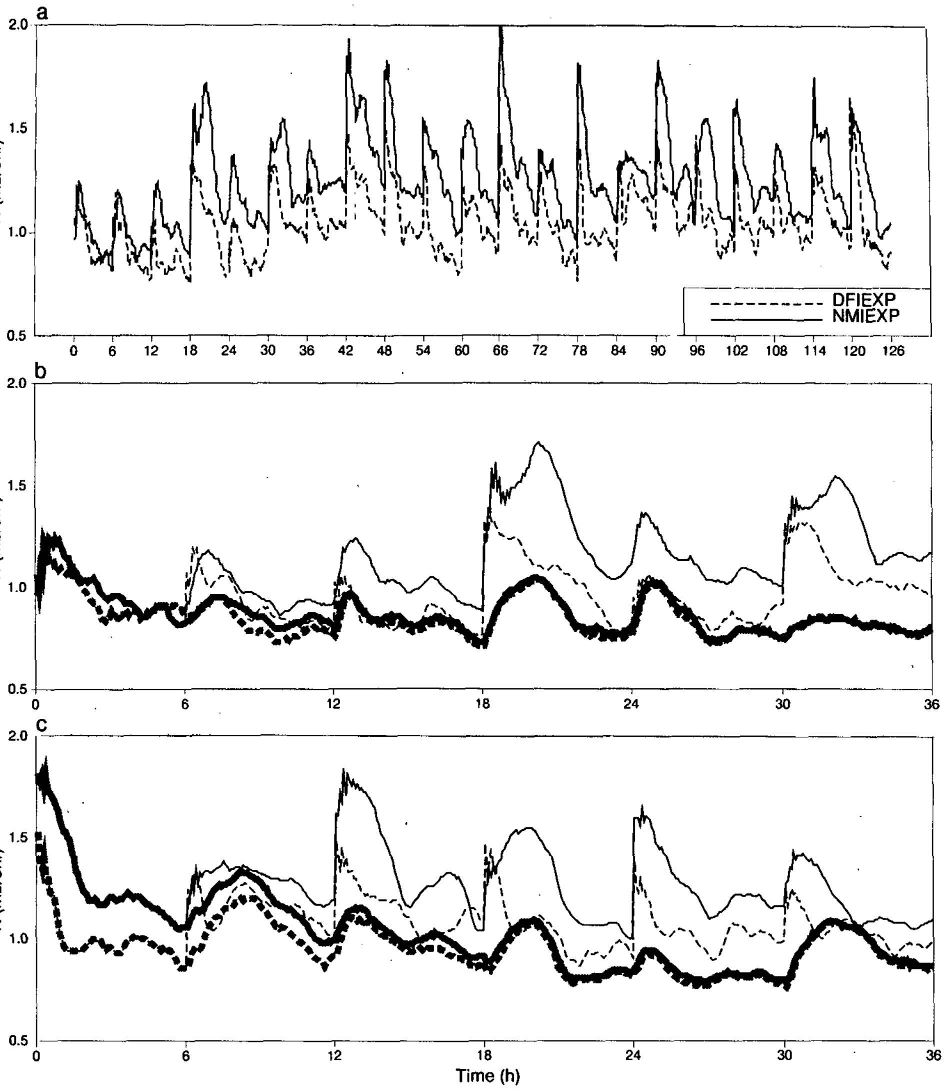

FIG. 5. Mean absolute surface pressure tendency N [hPa (3 h)$^{-1}$] as a function of time (h). The results for NMIEXP are shown by full lines and DFIEXP by dash lines. (a) The 5-day data assimilation experiments. (b) The first 36-h forecasts (thicker lines) together with the data assimilation results in the same period (thin lines). (c) The 36-h forecasts started at 0000 UTC 16 September 1994 (thicker lines) together with the data assimilation results in the same period (thin lines).

that the CWC field thus created has both a realistic spatial distribution, compared with satellite observations, and a quantitative magnitude, compared with that from a 24-h forecast.

During the data assimilation, a CWC field is available from the previous cycle (first guess). Using the first-guess CWC field directly as the initial field for the forecast, the spinup time is shorter than that with no initial cloud. However, the first-guess cloud field is not in a balance with other properly analyzed fields, causing an initial shock in the precipitation and a dramatic decrease in the mean cloud water of the model in the beginning of the forecast. The mean cloud water seems to be always in a shock–spinup process and below its actual value throughout the data assimilation period.

The spinup process could be finished within an assimilation cycle if the DFI scheme is used as the initialization step. The initialization procedure produces a balance between cloud and other fields leading to no initial shock in precipitation. There is still a short spinup period in the forecast because the filtering procedure takes into account of the whole evolution of the model including the spinup process during the initialization. However, the spinup period in the forecast following the DFI scheme is shorter than that without DFI. The cloud amount in the data assimilation period can reach its normal level.

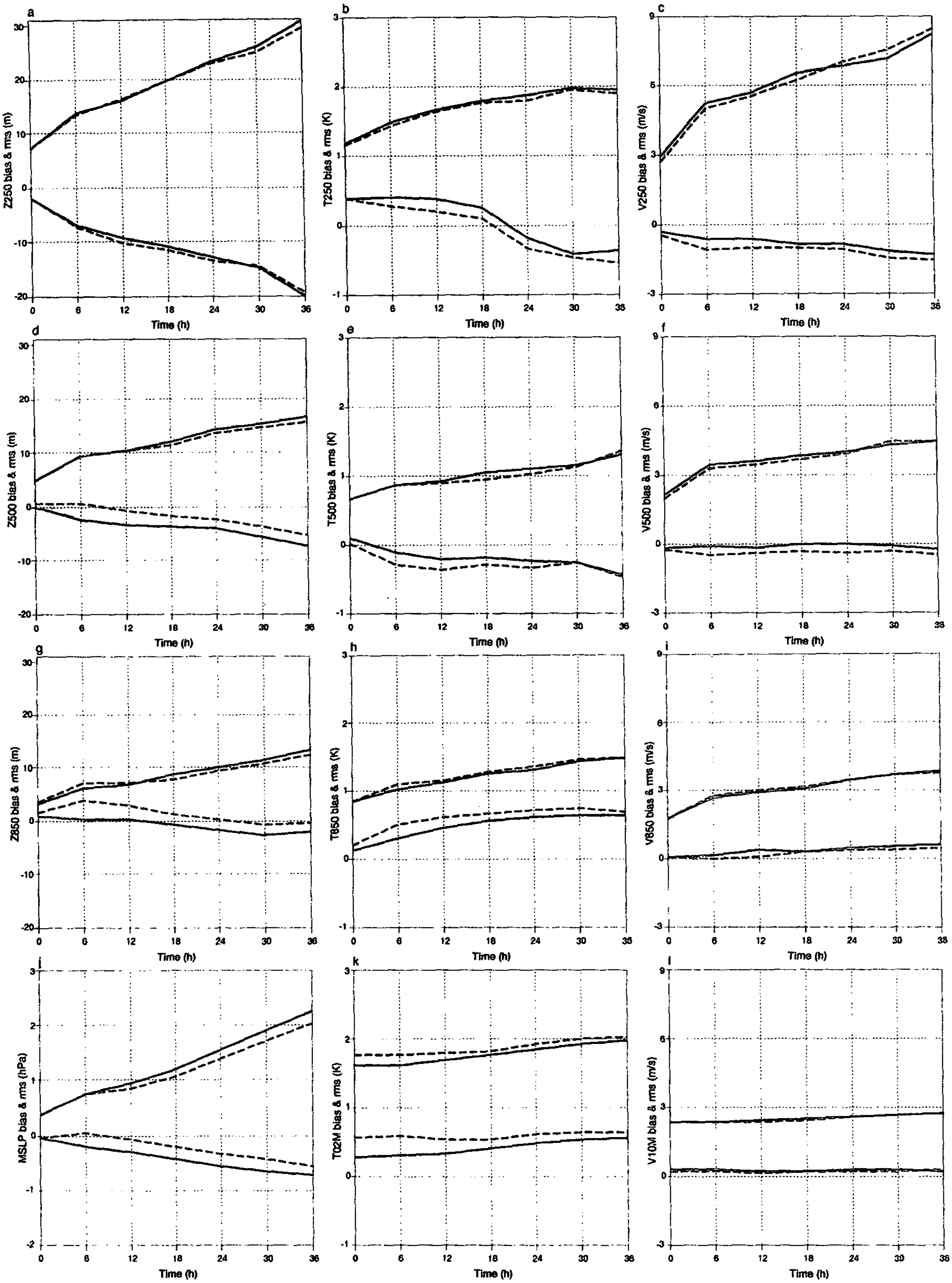

FIG. 6. Observation verifications (bias and rms) as functions of forecast length (h) for the 5-day data assimilation experiments. The results for NMIEXP are shown by full lines and DFIEXP by dash lines.

Experiments have also been conducted to shorten the spinup time even further. One idea is to use a longer filter span in the initialization scheme in the hope that the spinup process can be finished before the diabatic forward integration passes the analysis time ($t = 0$ in Fig. 2). However, with a longer filter span the adiabatic backward integration will result in a larger imbalance between cloud and other fields at the end of the integration. This in turn leads to a larger shock at the start of the diabatic forward integration (lower thin dashed curve in Fig. 2) and a longer spinup time. The evolution of cloud in the whole data assimilation period shows similar characteristics to that with a shorter filter span (compare the thicker curves in the middle and lower parts of Fig. 2).

In this study, no cloud observation is used in the data assimilation system, and yet the main motivation is to find a better initial cloud field for the forecast model. It is shown that the diabatic digital-filtering initialization scheme is able to alleviate the imbalances between the first-guess CWC and other analyzed fields, leading to a better cloud evolution, a short spinup time, and a removal of the initial shock in precipitation.

Acknowledgments. I would like to thank Erland Källén, Peter Lynch, and Hilding Sundqvist for discussions. I am also grateful to Jess U. Jørgensen for assistance with programs, and Julie Pietrzak and anonymous reviewers for comments on the manuscript.

REFERENCES

Daley, R., 1991: *Atmospheric Data Analysis.* Cambridge University Press, 457 pp.

Gustafsson, N., 1993: HIRLAM 2 final report. HIRLAM Tech. Rep. 9, 126 pp.

Hall, C., 1987: A common verification scheme for limited area models. *EWGLAM Newsletter,* **15.**

Huang, X.-Y., and P. Lynch, 1993: Diabatic digital filtering initialization: Application to the HIRLAM model. *Mon. Wea. Rev.,* **121,** 589–603.

——, and H. Sundqvist, 1993: Initialization of cloud water content and cloud cover for numerical weather prediction models. *Mon. Wea. Rev.,* **121,** 2719–2726.

——, A. Cederskov, and E. Källén, 1994: A comparison between digital filtering initialization and nonlinear normal mode initialization in a data assimilation system. *Mon. Wea. Rev.,* **122,** 1001–1015.

Kristjánsson, J. E., 1992: Initialization of cloud water in a weather prediction model. *Meteor. Atmos. Phys.,* **50,** 21–30.

Lynch, P., and X.-Y. Huang, 1992: Initialization of the HIRLAM model using a digital filter. *Mon. Wea. Rev.,* **120,** 1019–1034.

Machenhauer, B., 1988: HIRLAM final report. HIRLAM Tech. Rep. 5, 116 pp.

Raustein, E., H. Sundqvist, and K. B. Katsaros, 1990: Quantitative comparison between simulated cloudiness and clouds objectively derived from satellite data. *Tellus,* **43A,** 306–320.

Sundqvist, H., E. Berge, and J. E. Kristjánsson, 1989: Condensation and cloud parameterization studies with a mesoscale numerical weather prediction model. *Mon. Wea. Rev.,* **117,** 1641–1657.