Impact of a Digital Filter as a Weak Constraint in MM5 4DVAR

Tae-Kwon Wee and Ying-Hwa Kuo
UCAR COSMIC Project
P.O. Box 3000, Boulder, CO 80307

1. Introduction

A weak constraint formulation of a digital filter, which enables the assimilation model to fit observations closely while preventing it from dynamic imbalance, is introduced into the MM5 4DVAR. The dynamic imbalance between mass and wind fields can be introduced from a variety of sources, including the insertion of observation through the assimilation system, observation error, model errors, and the initialization procedures. Such dynamic imbalance leads to the excitation of high frequency gravity waves that are not supported by available observations, and reduces the overall effectiveness of an assimilation system. By introducing a penalty cost function (PCF hereafter), which measures the activities of high frequency gravity wave modes, via a digital filter, the assimilation model can be forced to use slowly varying "real" modes (known as the "slow manifold") to fit its solution to the observations, instead of the "gravity" wave modes which degenerate rapidly with time. The advantages of the MM5 4DVAR system constrained with digital filter are examined in the assimilation of GPS radio occultation soundings for a cyclogenesis case over the Antarctic region, in the context of the Observing System Simulation Experiment (OSSE).

2. Digital filter and its weak constraint formulation

The filtering of a discrete function of time can be thought of as a time averaging and the associated filter weights determine what will be filtered from the source function. The weights also determine the effectiveness of the filtering and, hence, the proper choice of filter is very important. In this study, we use a low-pass filter, which suppresses high frequency gravity wave modes. Low-pass digital filtering has been widely used for the initialization of numerical weather prediction models to suppress high frequency waves (for example, Huang and Lynch 1993).

The filtering of high frequency waves from a model trajectory, which exceeds a threshold value (the so-called digital cutoff frequency, q_c), can be achieved by a discrete convolution of model state ($\mathbf{x}(t_k)$) with a weighting function (a_k) at each time step (t_k) over a specified time span (t_N),

$$\bar{\mathbf{x}}(t_{N/2}) = \sum_{k=0}^{N} a_k \mathbf{x}(t_k) . \qquad (1)$$

The filtered model state, $\bar{\mathbf{x}}$, is defined at the middle of the specified time span, ($t_{N/2}$), for use in a symmetric and non-recursive filter. In practice, the filtering of a finite Fourier series introduces spurious oscillations in the frequency response, which are known as Gibbs oscillations. To control these oscillations, some special window functions can be used. Following Lynch (1997), we use the Dolph-Chebyshev windowed filter, which can have similar properties to an optimal filter under appropriate choice of parameters. It should be borne in mind that the filter coefficients have to be normalized over a given time span (i.e., $b_k = a_k / \sum_{m=-M}^{M} a_m$) to keep the mean quantity of the targeted sequence from amplifying or damping.

Following Gustafsson (1992), Polavarapu et al. (2000), and Gauthier and Thépaut (2001), the PCF can be defined as the distance between the filtered state and the corresponding unfiltered state so that

$$J_{DF}(\mathbf{x}) = \frac{l}{2}\left\langle \mathbf{x}(t_{N/2}) - \bar{\mathbf{x}}(t_{N/2}), \mathbf{W}[\mathbf{x}(t_{N/2}) - \bar{\mathbf{x}}(t_{N/2})] \right\rangle , \qquad (2)$$

where $<,>$ is an inner product, and l is an adjustment parameter of PCF which determines

Corresponding author address:
Tae-Kwon Wee, P.O. Box 3000, Boulder, CO 80307; Email: wee@ucar.edu; phone: 303-497-2612; fax: 303-497-2610.

the degree to which the weak constraint will be enforced. $\mathbf{W}$ is a diagonal matrix which defines the weightings of each control variable. Basically, this matrix is used for the scaling of variables, which may vary significantly from one another in terms of absolute magnitude. The scaling is particularly important because different variables are closely related to one another as far as gravity wave modes are concerned (through what we call "polarization" relationship), although the variables (for example, temperature and moisture) show different time variations in magnitude. In that situation, the deviation of each variable from its own filtered state contains the same information, i.e. the high frequency fluctuation of gravity wave modes, regardless of the specific variable type. Thus, it is necessary to normalize the variables with proper scaling factors to ensure the uniform contribution of each variable to the net PCF. This scaling is also helpful for a faster convergence of minimization. Another important benefit is that, by introducing this matrix, one can choose which variable to include in the cost function (one may exclude moisture, for example), as well as to weight each variable's relative contribution to the cost function. It turns out that the error covariance matrix of the background field is a good choice as the scaling factor.

As shown in equation (2), PCF is not related to the observational part of the cost function, and consequently its minimization can be performed without the use of observations. The constrained 4DVAR acts as an initialization method, though the computational cost is high if used for this purpose, alone.

The gradient of PCF can be expressed as in Gauthier and Thépaut (2001):

$$\nabla_{\mathbf{x}} J_{DF}(\mathbf{x}_o) = \sum_{k=0}^{N} m_k \mathbf{M}'^{*}(t_o, t_k)\mathbf{W}[\mathbf{x}(t_{N/2}) - \bar{\mathbf{x}}(t_{N/2})], \quad (3)$$

where

$$m_k = \begin{cases} -b_k & if \quad k \neq N/2 \\ 1-b_k & if \quad k = N/2 \end{cases}, \quad (4)$$

and $\mathbf{M}'^{*}(t_o, t_k)$ is the adjoint of a model operator $\mathbf{M}(t_o, t_k)$. An interesting feature of adding digital filter as a weak constraint to the minimization process is that it functions in a recursive manner, although the filter itself is non-recursive. The filtered state changes at every iteration, as the 4DVAR alters the initial condition (and, consequently, the subsequent forecast within the assimilation window). Therefore, the minimization routine aims to approach a different target state in order to reduce the PCF at each iteration, which is different from observational cost function and background cost function that have a fixed target in the minimization process.

The efficiency of the constrained 4DVAR depends on the response of the filter employed and, therefore, the design of an improved digital filter is highly desirable. In this study, two efficient methods have been developed. It is possible to define multiple points of initialized states within the window, because the nearly optimal response of the Dolph-Chebyshev windowed filter enables us to use a shorter time span (Huang and Lynch 1993; Lynch et al. 1997). However, its direct application requires significant additional resources (disk space or memory) to keep track of the initialized states and the corresponding adjoint forcing terms. This problem can be remedied by applying the filter once again to the sequence of the filtered states:

$$\bar{\bar{\mathbf{x}}}(t_0) = \sum_{k=-\frac{M}{2}}^{+\frac{M}{2}} b_k \bar{\mathbf{x}}(t_k) = \sum_{k=-\frac{M}{2}}^{+\frac{M}{2}} b_k \left(\sum_{l=-\frac{M}{2}}^{+\frac{M}{2}} b_l \mathbf{x}(t_{k+l}) \right), \quad (5)$$

where $\bar{\bar{\mathbf{x}}}$ represents the initialized state through the double applications of the filter. The filtered states are not related sequentially, for the use of a non-recursive filter, so that rearranging the filter weights in (5) leads to,

$$\bar{\bar{\mathbf{x}}}(t_0) = \sum_{k=-M}^{+M} g_k \mathbf{x}(t_k) = \sum_{k=-M}^{+M} \left(\sum_{l=-\frac{M}{2}}^{+\frac{M}{2}} b_l b_{k-l} \right) \mathbf{x}(t_k), \quad (6)$$

where b_l is defined only for $\left[-\frac{M}{2}, +\frac{M}{2}\right]$ and zero elsewhere. Thus, the double application of the filter shown in (6) does not require additional resources; and it produces the same net result as the use of the multiple points of initialized state, as shown in (5).

One drawback in both the single and double applications of a digital filter is that the initialized states are defined only at the mid-point of the assimilation window. Prior to the mid-point, the model state can be viewed as a transient, uninitialized period. The ideal method would be to have the initialized point at the start of the window, but this is practically unfeasible for the

use of a symmetric filter. However, one can make the initialization point to be closer to the beginning of window by using a shorter time span than the assimilation window. In this study, we use two filters jointly; with 3-h and 6-h time spans centered at 1.5-h and 3-h, respectively, over a 6-h assimilation window. The response of a filter sharpens and approaches an ideal low-pass filter, as its time span increases. Thus, the filter with the longer time span works effectively through the 6-h assimilation window, while the filter with the shorter time span gives an additional filtering effect prior to 3-h. We found that this method, which uses two filters simultaneously, is considerably more effective than the single or double application of a filter. Because of the additive nature of the adjoint gradient, this method can be easily implemented by summing the gradients associated with each filter.

3. Experiment Design

All experiments shown in this study are conducted in the context of an OSSE and the details of the OSSE procedure are described in Kuo et al. (2002). As an OSSE, a high-resolution MM5 model with sophisticated physics (30-km grid spacing with 50 vertical layers) is initialized at 0000 UTC 13 October 1995, using the ECMWF global analysis data enhanced with available upper air and surface observations, and is integrated for 72 hours to 0000 UTC 16 October 1995. The case chosen for this research is a cyclogenesis event that took place over Western Antarctica, during the period of 13-16 October 1995. During this period, several cyclones developed and moved around the Antarctic, over the sub-Antarctic Ocean and coastal regions.

The simulation from the high-resolution experiment, known as the "NATURE" run, is assumed to be the "true" atmosphere and is used to simulate observations from a hypothetical observing system similar to COSMIC (Constellation Observing System for Meteorology, Ionosphere and Climate). The spatial and temporal distributions of the simulated radio occultations are based on a realistic set of orbit parameters of GPS and COSMIC satellites. Also, the representative observation errors of GPS/MET refractivity, as estimated by Kursinski et al. (1997) (e.g., 1% below 5 km and 0.2% up to 30 km), are added to the simulated observations.

These observations are then assimilated into a lower-resolution model with relatively inexpensive physics. The assimilation model has 120-km grid spacing and 20 vertical layers; it is representative of a typical operational data assimilation/forecast system. Although the simulated GPS radio occultation data are used in this study, the weak constraint of digital filter is general enough to be used, regardless of observation type.

The effectiveness of the assimilation, and the impact of the data on subsequent forecasts, can be assessed by comparing the results with the known "true" atmosphere (the NATURE run). In the case of real data assimilation, such evaluation is not a simple task because accurate analysis of the atmosphere at high temporal and spatial resolution is not available for a detailed verification. On the contrary, OSSE enables us to precisely estimate the analysis and forecast errors with the knowledge of the "true" atmosphere. Therefore, OSSE is particularly valuable in evaluating the performance of a digital filter as part of an assimilation system.

We performed a large number of experiments, and obtained results on the sensitivities of the

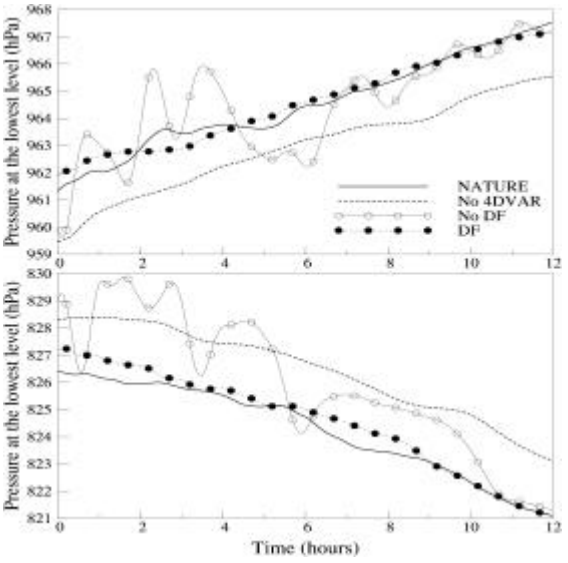

Figure 1: The traces of pressure at the lowest model level at two grid points.

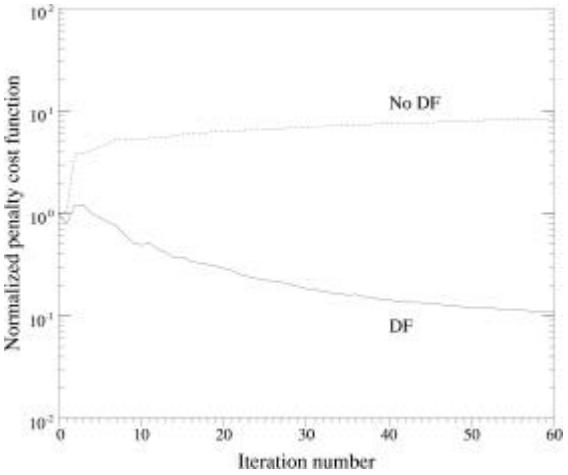

Figure 2: The normalized penalty cost function versus minimization iteration.

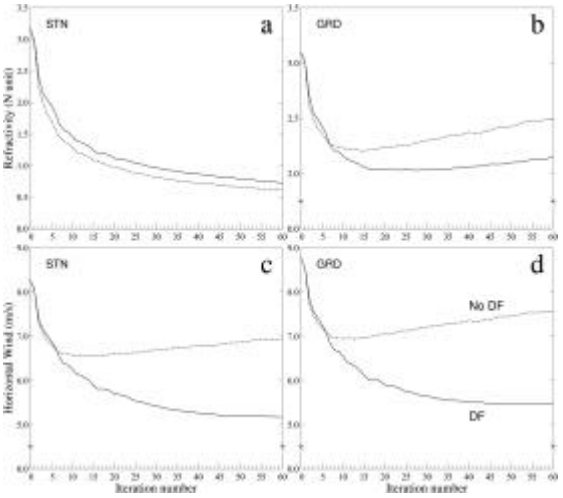

Figure 3: RMSE errors of refractivity (a) and horizontal wind (c), in observation space. (b) and (d) are the correspondences in model grid space.

digital filter parameters, observation types, observation errors, the degrees to which the constraint will be forced, and the choice of filtered variable, etc. Due to limitations of space, we will report only on the results of a few key experiments.

The first experiment is the "PERFECT" run which uses the perfect initial condition directly extracted from the grids of the NATURE run (at every 4th grid point). In general, this represents the best possible initial condition for a forecast model. The second experiment is the "No 4DVAR" run in which no GPS refractivity observation is assimilated. The initial condition of this experiment is a 12-h forecast from the NCEP/NCAR reanalysis data. The initial condition of "No 4DVAR" is also used as the background field of assimilation experiments.

The remainders are two assimilation experiments; one uses only the strong constraint ("No DF"), and the other has the additional weak constraint of digital filter ("DF"). The former uses basically the same method as those used in the previous GPS OSSE studies (Zou et al 1995, Kuo et al. 1997), with minor modifications.

4. Results

Figure 1 shows the pressure trace at the lowest model level at two selected grid points. The pressure traces taken from the NATURE and the 'No 4DVAR' evolve slowly with time, but the pressure trace in "No DF" shows the presence of spurious high frequency oscillations, with periods shorter than 3 hours. On the contrary, the pressure trace in "DF" does not contain such high frequency oscillations.

Figure 2 compares the normalized PCFs as a function of the iteration numbers in the minimization process. The PCF in "No DF" is defined in the same way as that in "DF", except that it is a diagnosed value provided only for the purpose of comparison. Thus, it is neither included in the total cost function nor used in the adjoint computation. The PCF in "No DF" continuously increases with iteration, and increases by nearly one order of magnitude after 60 iterations. On the contrary, the PCF in "DF" decreases steadily in the course of minimization, except in the initial few iterations. It is interesting to note that the PCFs in "DF" and "No DF" are heading in opposite directions, and end up differing by two orders of magnitude. By its own definition, the magnitude of the PCF is a measure of high frequency wave activity, although it is scaled by the error covariance of the background field. Therefore, the manifestation of high frequency oscillations, as seen in Figure 1, can be expected from the increased PCF in "No DF".

Another important benefit of digital filter is the treatment of the top boundary problem. The existing upper boundary conditions in MM5 have

not been well tested in high latitude regions, and it is uncertain whether these boundary conditions can deal with problems such as the reflection of gravity waves (e.g., Wei et al. 2001). Furthermore, the PCF in "No DF" increases with iteration, as shown in Fig. 2; consequently, the assimilation with strong constraint makes the upper boundary problem even worse, as more high frequency gravity waves are excited. We found that the constrained 4DVAR with digital filter can reduce this upper boundary problem, significantly, by directly filtering out the high frequency gravity waves at upper levels, and by suppressing the vertically propagating waves, through filtration, at the lower levels (not shown).

Figure 3 compares the root-mean-square errors (hereafter RMSE) of refractivity and horizontal winds in "No DF" and "DF", as a function of iteration numbers during the assimilation period. Two different sets of RMSE are used; one is computed in observational space and the other in model grid point space. The RMSE of refractivity evaluated in observational space is very similar to the observational part of the cost function. We found the RMSE of "DF" decreases slightly slower than that of "No DF" with the additional constraint (digital filter) imposed (Fig. 3a). On the other hand, the 'retrieved' variables in observational space, and all variables including refractivity in model space, at all grid points of assimilation model at one hour intervals, show the opposite behavior. In this case, the RMSEs in "DF" are smaller than those in "No DF" after 5 or 6 iterations. Thus, "No DF" fits the observed refractivity more closely than "DF", only in observational space, without accurately recovering the relationships among temperature, moisture and pressure; this causes an increase of dynamic imbalance with iteration. Because wind is not directly related to refractivity, the assimilation model can improve the wind field only through its own dynamics. Therefore, the degree of mutual adjustments between wind and mass fields determines the degree of improvement possible. In this regard, the reduction of imbalance in "DF" plays a crucial role in improving the wind field.

Figure 4 depicts the spectra of pressure error

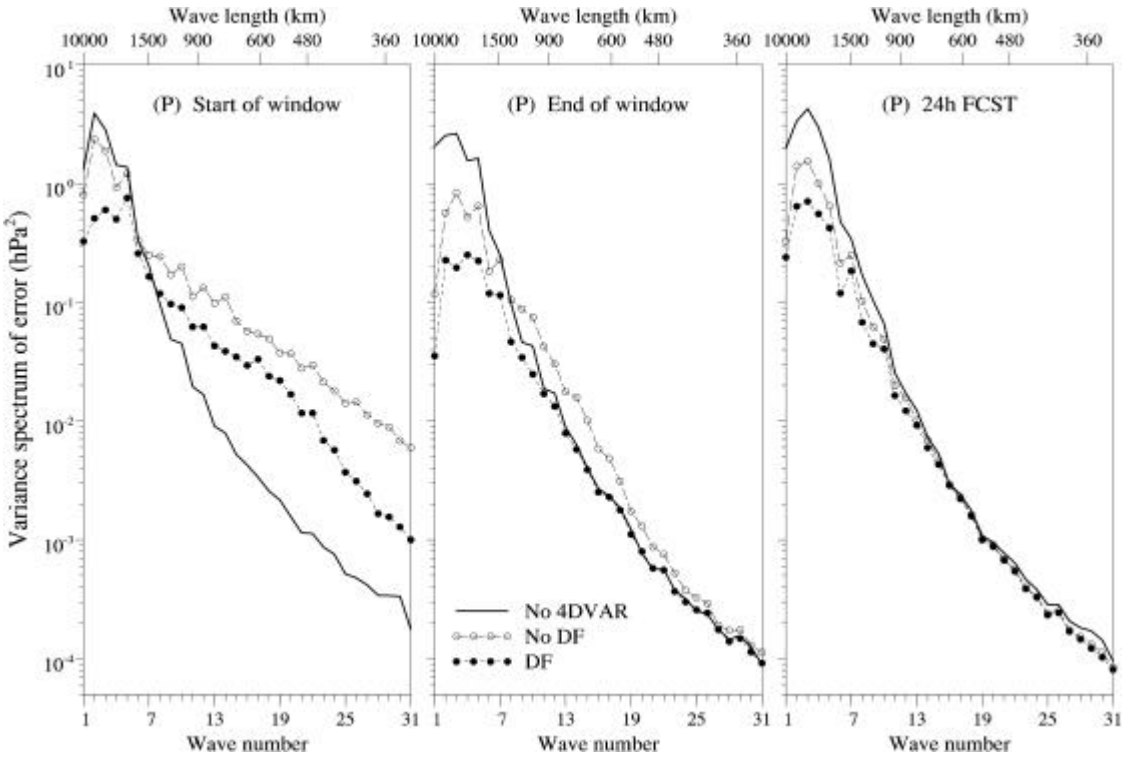

Figure 4: Variance spectra of pressure error at the start of the window (a), at the end of the window (b), and at the time of 24-h forecast (c).

variances at the start of the window, the end of the window, and at the 24-h forecast, respectively. Astonishingly, in the range of small spatial scales, where the wave number is greater than 7 and the wave length is shorter than 1440 km, both "No DF" and "DF" show larger errors than "No 4DVAR", at the start of the window. The assimilations only improve the large-scale features at the start time. But at the end of the window, the errors have decreased to a much lower level throughout the whole spectrum, and consequently the errors in "DF" are smaller than those in "No 4DVAR", at all spatial ranges. It is noteworthy that the error in "DF" is always smaller than that in "No DF". This suggests that "DF" utilizes the observation more effectively (by keeping the assimilation model from using high frequency gravity wave modes to fit observation), via the application of the weak constraint of digital filter. The spectra of the 24-h forecast show that both assimilation experiments contain errors smaller than "No 4DVAR", even in the high frequency range. This indicates that the model itself can partly recover the small-scale features, through the internal dynamics, once the assimilation recovers the large-scale features.

Figure 5 compares the error evolutions including the 36-h free forecast period. It clearly shows that the effectiveness of the constrained 4DVAR ("DF"), when compared to the unconstrained 4DVAR ("No DF"), even beyond the assimilation period. With the assimilation of GPS radio occultation data, "No DF" reduces its errors to about half of that in "No 4DVAR". With the use of digital filter weak constraint, "DF" further reduces the errors of "No DF" by half. It is also notable that the error in the specific humidify field in "No DF" at the start of window can be larger than that of "No 4DVAR" (Fig. 5b), but the errors in "DF" are more evenly distributed through the window. This happens because the 'optimal' initial condition through 4DVAR assimilation does not lead directly to the best possible description of the atmosphere at the start time, because 4DVAR only tries to obtain an initial condition that results in the minimum cost function under the constraints imposed. The closeness between the 'optimal' and 'improved' initial conditions is, of course, related to the accuracies of model and observation. This again confirms that the dynamic imbalance in 4DVAR

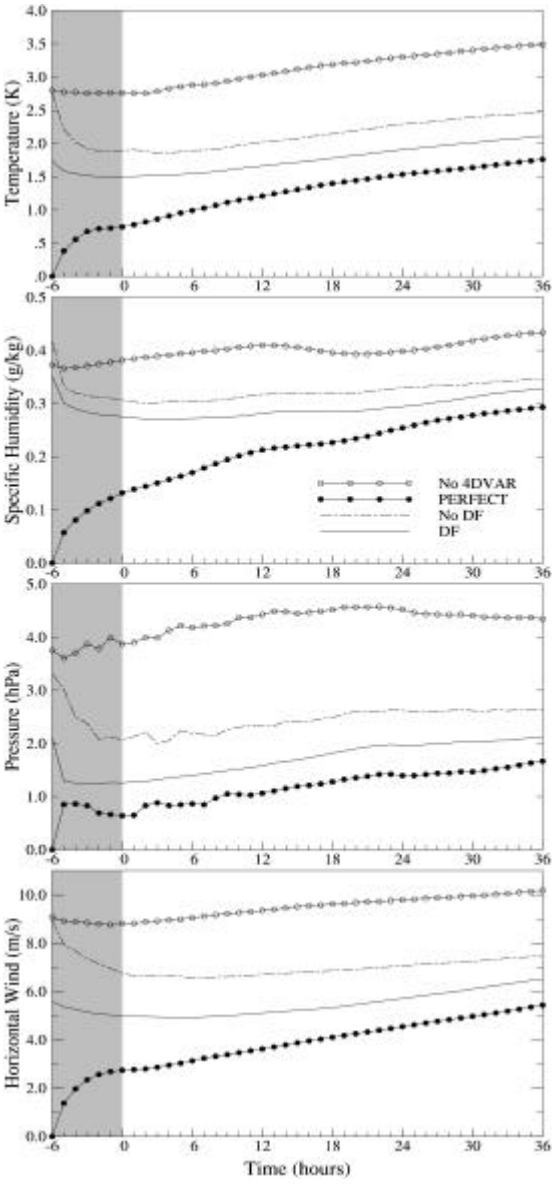

Figure 5: Root mean square errors with time. Shading represents the assimilation period.

can be effectively controlled through the weak constraint of a digital filter, and hence, the model state is dynamically more consistent through the assimilation window. This, in turn, enables the assimilation system to use observations more effectively and leads, ultimately, to a better description of the atmosphere.

Through a series of initialization experiments in which the minimum of imbalance is pursued without the use of observations, it is concluded

that the initialization (reduced imbalance) in itself does not improve the forecast skill; hence, it only helps the assimilation system to use the observations more effectively (not shown).

5. Conclusion

A weak constraint of digital filter is introduced into MM5 4DVAR. This method is computationally efficient, easy to implement, and flexible with regard to further modification (for example, the redesign of digital filter). The constrained 4DVAR is evaluated in the context of OSSE. It is shown that the method is very effective in reducing dynamic imbalance in the course of assimilation, and leads, ultimately, to more effective data assimilation and an improved forecast.

Acknowledgements

This research is funded by NASA, through Grant NAG 5-9518 to Ohio State University and the University Corporation for Atmospheric Research. We thank Kris Conrad for technical editing of this manuscript.

References

Gauthier, P., and J.-N. Thépaut, 2001: Impact of the digital filter as a weak constraint in the preoperational 4DVAR assimilation system of Météo-France. *Mon. Wea. Rev*., **129**, 2089-2102.

Gustafsson, N., 1992: Use of a digital filter as weak constraint in variational data assimilation. *Proc. ECMWF Workshop on Variational Assimilation with Special Emphasis on Three-dimensional Aspects*, Reading, United Kingdom, ECMWF, 327–338.

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

Kuo, Y.-H., X. Zou, and W. Huang, 1997: The impact of GPS data on the prediction of an extratropical cyclone: An observing system simulation experiment, *J. Dyn. Atmos. Ocean*, **27**, 439-470.

Kuo, Y.-H., T.-K. Wee, and D.-H. Bromwich, 2002: Potential impact of COSMIC GPS radio occultation data on regional weather analysis and prediction over the Antarctic, *Preprints, Antarctic NWP and Forecasting Workshop*, 4-5 June 2002, Boulder, Colorado, 33-42.

Kursinski, E. R., G. A. Hajj, J. T. Schofield, R. P. Linfield, and K. R. Hardy, 1997: Observing Earth's atmosphere with radio occultation measurements using the global positioning system, *J. Geophys. Res.*, **102**, 23,429-23,465.

Lynch, P., 1997: The Dolph–Chebyshev window: A simple optimal filter. *Mon. Wea. Rev*., **125**, 655–660.

Polavarapu, S., M. Tanguay, and L. Fillion, 2000: Four dimensional variational data assimilation with digital filter initialization. *Mon. Wea. Rev.*, **128**, 2491–2510.

Wei H., D.-H. Bromwich, L.-S. Bai, Y.-H. Kuo, and T.-K. Wee, 2001: A case study of the impact of the upper boundary condition in Polar MM5 simulations over Antarctica. *Preprints, The Ross Island Meteorology Experiment (RIME)*, 11-13 September 2001, Columbus, Ohio, BPRC Miscellaneous Series M-423, 83-89.

Zou, X., Y.-H. Kuo, and Y.-R. Guo, 1995: Assimilation of atmospheric radio refractivity using a nonhydrostatic mesoscale model. *Mon. Wea. Rev*., **123**, 2229-2249.