Estimation of the error distributions of precipitation produced by convective parametrization schemes

By RONALD M. ERRICO¹*, DAVID J. STENSRUD² and KEVIN D. RAEDER¹

¹The National Center for Atmospheric Research, USA
²National Severe Storms Laboratory, USA

(Received 25 January 2001; revised 27 June 2001)

SUMMARY

If a parametrization scheme for convective precipitation is to be used for assimilating observations of precipitation using a statistically based technique, then statistics of the errors produced by that scheme are required. These are the errors produced by the scheme’s formulation itself, not counting any errors in the scheme’s input. Such errors are extremely difficult to estimate, but examination of differences produced by various suitable schemes can yield qualitative descriptions of such errors. Here, hourly accumulated convective precipitation fields produced from six different versions of a short-term forecast model are compared. The versions have identical initial and boundary conditions, but vary in the schemes used for either the convection or the planetary boundary layer, or both. The distribution of differences, or differences in logarithms of accumulations, between corresponding precipitating grid points for pairs of forecasts are examined using a simple binning technique. When the convection schemes differ, results reveal that if either a log-normal or normal distribution is a better characterization of the distributions, it is the log-normal one. The standard deviations of these logarithmic distributions correspond to different schemes at identical grid points producing values differing by factors of 2 or more. A large proportion of grid points that have non-zero hourly accumulations using one model version may have no accumulation using another version. For most pairs of forecasts examined, however, grid points having larger values of accumulation for one scheme tend to have a smaller fraction of values having no accumulation in the other scheme. These results suggest that the finite probability that the model produces no precipitation when the corresponding, true atmospheric state does, should be considered in the statistical description of the model errors and that, because of the large standard deviation of model errors as well as large possible errors of hourly precipitation observations, the quantitative usefulness of assimilating such observations may be very limited.

KEYWORDS: Convective precipitation Data assimilation Model error

1. INTRODUCTION

In the problem of atmospheric-data assimilation, some observations are only indirectly related to the fields being analysed. One example is a satellite observation of radiance. It can be related to atmospheric temperature, moisture, and chemical constituent profiles using a radiative-transfer model (Eyre 1989). Another example is an observation of precipitation rate or accumulation that can be related to fundamental atmospheric fields using a model that includes precipitation processes (Krishnamurti et al. 1993)

Modern atmospheric-data assimilation has its foundation in statistical estimation theory derived from Bayes’ theorem (e.g. Lorenc 1986; Talagrand 1997; Cohn 1997). When observations are only indirectly related to the assimilation fields and the relational

* Corresponding author: NCAR, PO Box 3000 Boulder CO, 80307-3000, USA.
e-mail: rmerico@ucar.edu
† NCAR is sponsored by the US National Science Foundation.
model can have non-negligible error, then the statistics of that model's error should be considered (Tarantola 1987). In meteorology, this error is often called the error of representativeness when the model is a spatial interpolation or integration model, but otherwise it is sometimes called the error of either the forward model or observation operator (Eyre 1989). In the assimilation of precipitation, this model error is likely to be non-negligible and therefore its statistics should be considered (Errico et al. 2000).

It is extremely difficult to directly estimate the error statistics of precipitation models. This error is the one caused by the model itself and not due to any errors in its input (e.g. input profiles of temperature, specific humidity, and winds). Since neither perfect input profiles nor perfect verifying observations of precipitation exist, the desired error estimates cannot be accurately and easily estimated by direct means. They can be less accurately estimated indirectly, however, by either examining innovation statistics (e.g. Hollingsworth and Lönnberg 1986) that include observation errors or by simulating an assimilation system (e.g. Daley and Mayer (1986), but for different purposes).

In a three-dimensional variational (3D-Var) framework, where convective precipitation rate rather than accumulation is most naturally assimilated, the appropriate forward model is a moist convection model, such as a convective parametrization scheme. In a four-dimensional variational (4D-Var) framework, however, the forward model should incorporate dynamics and other (e.g. planetary boundary layer) processes that can affect the structures of fields that subsequently affect the convection scheme. In this 4D-Var case, the forward-model error includes that of the entire forecast model.

Here we use a simplified simulation approach to estimate precipitation model error statistics, making use of a dataset produced in part by Weiss and Stensrud (2000) for a different but related purpose. Their purpose was to examine the possible uses of ensembles of short-range numerical forecasts for predicting severe weather events. They examined forecast similarities and differences produced by different configurations of the same basic forecast model, all having identical initial and boundary conditions.

If, a priori, one version of the model cannot be claimed to be better than another, but all can be claimed as reasonably describing atmospheric behaviour, then we can alternately consider each distinct model forecast as representing 'truth' fields and any differences between it and other forecasts as 'errors.' Although there appears to be no evidence that one model version is consistently much better than another (Wang and Seaman 1997; Stensrud et al. 2000), it is unclear whether any version describes real precipitation well enough for our purposes. Even so, examining these differences is a good starting point for estimating required error statistics for 4D-Var. This is analogous to the approach originally adapted at the National Centers for Environmental Prediction (NCEP) and adopted at most other operational centres for estimating forecast-error covariances by comparing lagged forecasts: it is known to yield incorrect results under some conditions but nevertheless is a useful starting point (Parrish and Derber 1992).

Stensrud et al. (2000) have already examined many statistics of the precipitation differences between their different model simulations. They find that when 6 h precipitation totals are compared, none of the model configurations are consistently better than all the others. Yet the divergence between model solutions is dramatic when sensible weather parameters, such as precipitation, are considered. The differences they obtain are large, particularly with regard to accumulated precipitation, even on short (less than 6 h) time-scales.

Here, our focus is on questions raised in Errico et al. (2000). We want to estimate properties of the distribution of precipitation differences; in particular shapes and statistics of the distributions of differences obtained at corresponding locations and times in the forecasts. Is a log-normal distribution a better characterization of those
Figure 1. Precipitation accumulated over the first six hours' forecast using versions 1–6 of MM5 (labelled (a)–(f), respectively). See text for details of software. The contour interval is 1/3 cm.

differences than a normal (i.e. Gaussian) one? What are the standard deviations that describe those distributions? The relevance of these questions is discussed in greater detail in Errico et al. (2000).

In section 2 we describe the models and their datasets used in this study. In section 3 we describe our procedure for determining and examining the distributions. Results are presented in sections 4 and 5, followed by conclusions in section 6.

2. Model versions

The models used to produce the datasets examined are all versions of the mesoscale model developed jointly by the Pennsylvania State University and the National Center
for Atmospheric Research designated MM5 (Dudhia 1993; Grell et al. 1994). All versions are nested with an outer grid resolution of 96 km and an inner one of 32 km. The inner domain is shown in Fig. 1. Prognostic fields are defined on 23 atmospheric levels. The precipitation examined here is only that produced on the finer-resolution grid.

The MM5 software is designed to easily change physical parametrization schemes. The six model versions used to generate the forecasts are some combinations of three different convection schemes and three different planetary boundary layer (PBL) schemes. The convective schemes used include versions of the quasi-equilibrium schemes developed by Betts and Miller (1986) and Grell (1993), and a scheme developed by Kain and Fritsch (1990) specifically for mesoscale models. The Grell (GR) and Kain–Fritsch (KF) schemes both include the effects of convective downdraughts, while the Betts–Miller (BM) scheme does not. The PBL schemes used include the non-local closure schemes developed by Blackadar (1979) and Hong and Pan (1996), and a local 1.5-order closure scheme developed by Burk and Thompson (1989), designated as BK, HP and BT, respectively. The particular combinations applied in the six model versions used here are listed in Table 1.

Other choices of physical and numerical schemes common to all the forecasts are cloud formation, short-wave and long-wave radiation, grid-resolved microphysics parametrization, and an upper radiative boundary condition as in Stensrud et al. (2000). Forecasts are initialized at 0000 UTC 3 May 1999 and 18 May 1999 using the corresponding NCEP Eta model (Black 1994) initial condition at 25 hPa vertical increments. Boundary conditions are calculated from the Eta model forecasts at 12 h time intervals where the boundaries of the outer grid are relaxed towards the temporally interpolated Eta model forecast values. Model output is examined only during the first 12 h of the forecasts, although attention is focused on the shorter time-scales (particularly 3 or 6 h) currently used or proposed for precipitation assimilation (e.g. Hou et al. 2000).

Since the model physics is different in distinct model versions, although all forecasts for each case begin from identical initial conditions, the model fields are altered after the first time step in the places where the altered physics operates (i.e. at moist convective points or within the PBL). Thereafter, model dynamics as well as altered model physics act to further change the forecasts. In other words, after the first time step, it is not just the convection or PBL schemes themselves acting at a given time that change the precipitation, although all alterations are ultimately due to using different convection or PBL schemes.

### 3. Examination Method

For each model version (denoted by the index \( m \) in Table 1) and synoptic case, the precipitation accumulated over the 1-hour periods ending at each forecast hour \( t_k \) at each grid point \( j \) will be considered. These values will be denoted as \( R(j, m, t_k) \). The explicit notation \((j, m, t_k)\) will be omitted when confusion should not arise, i.e. the simpler notation \( R \) will be used if sufficient.
Some values of $R$ are extremely small. Such values can greatly skew some distributions, especially those of logarithms of ratios of $R$ for pairs of model versions, as when comparing a near-zero value with a large one. For this reason, values of $R$ smaller than 0.01 mm are treated as non-precipitating points, just as those that are zero-valued (MM5 sometimes actually records negative accumulations in its lateral-boundary sponge regions, and these too are treated as 0). In the remainder of this paper, it is only these adjusted values of $R$ that are considered. Filtering more values by restricting them to be larger than 0.1 mm had the effect of truncating the small-value tail of some logarithmic distributions. Allowing non-zero values less than 0.01 mm produced little change in the results, however, except for increasing the standard deviation of the logarithmic distributions by allowing some outlying, extremely small values to be considered. Although current quantitative precipitation observing systems cannot distinguish between distinct values less than a few tenths of millimetres, these small values are part of the distributions produced by the models, and along with zero-values that may be produced, they should be considered when formulating a statistical description of the model errors (see e.g. Errico et al. 2000).

For each $m$, $t_k$, and case, the mean $\overline{R}(m, t_k)$ and standard deviation $\sigma_R(m, t_k)$ of $R$ are determined. The average is performed over the number of points $N(m, t_k)$ for which $R > 0$ (considered after the very small values have been set to 0) rather than over the number of grid points in the full model domain. For each $R$, an $L = \ln(R)$ is determined, along with the corresponding means $\overline{L}(m, t_k)$ (so that $\overline{L} = \ln \overline{R}$) and standard deviations $\sigma_L(m, t_k)$, in the same way as for each distinct set of $R$ considered. The explicit notation $(m, t_k)$ will be omitted when confusion should not arise.

For each case, results produced by each member within each distinct pair of versions $(m_1, m_2)$ are compared. The ordering of the two members of each pair affects the signs of the differences examined, and therefore the means of the distributions of those differences. Standard deviations of the distributions are not affected, however, nor are the shapes of the distributions (aside from a reflection about the mean value), and therefore only one ordering for each pair need be considered.

For each $j, t_j, (m_1, m_2)$, and case for which both $R(j, m_1, t_k) > 0$ and $R(j, m_2, t_k) > 0$, differences

$$\Delta R(j, m_1, m_2, t_k) = R(j, m_1, t_k) - R(j, m_2, t_k)$$

are computed. The corresponding means $\overline{\Delta R}(m_1, m_2, t_k)$ and standard deviations $\sigma_{\Delta R}(m_1, m_2, t_k)$ are computed for each distinct set of values considered. The averaging is performed only over the number $N_{\Delta}(m_1, m_2, t_k)$ of common precipitating points for the pair, case and hour. Analogously,

$$\Delta L(j, m_1, m_2, t_k) = L(j, m_1, t_k) - L(j, m_2, t_k)$$

are computed with corresponding means $\overline{\Delta L}(m_1, m_2, t_k)$ and standard deviations $\sigma_{\Delta L}(m_1, m_2, t_k)$. As earlier, the notation $(m_1, m_2, t_k)$ will be omitted when possible.

Note that

$$\Delta L(j, m_1, m_2, t_k) = \ln[R(j, m_1, t_k)/R(j, m_2, t_k)].$$

A distribution of $\Delta L$ is not the same as one of $R(j, m_1, t_k)/R(j, m_2, t_k)$, but a mean or standard deviation of the former can be expressed in terms of an equivalent ratio $R_1/R_2$ by exponentiating the logarithmic values. This is how all such values will be presented here: e.g. instead of referring to 'differences of logarithms greater than in 2' we will refer to 'ratios greater than 2'. Note that when expressed this way, changing the order of $m_1$ and $m_2$ yields reciprocals of the original ratios.
The sums

\[ S(j, m, t_k) = \sum_{t=1}^{k} R(j, m, t) \]  

are also determined. The differences

\[ \Delta S(j, m_1, m_2, t_k) = S(j, m_1, t_k) - S(j, m_2, t_k) \]

are then computed, along with corresponding logarithms of \( S \). The means and standard deviations of all these distributions are also computed.

Once the mean \( \overline{X} \) and standard deviation \( \sigma_X \) of the set of values \( X \) are determined for any distribution considered (e.g. for \( X \equiv \Delta R \)), then 21 ranges of values of \( X \) (called bins) are defined. Each bin spans \( 0.3\sigma_X \). The centre bin includes all values for which \( \overline{X} - 0.15\sigma_X < X \leq \overline{X} + 0.15\sigma_X \). The first and last bins include all values less than \(-3\sigma_X \) and greater than \( 3\sigma_X \), respectively. Once the numbers of grid points whose \( X \) fall within each bin are determined, the percentage of grid-point values within each bin with respect to the total number in all the bins is computed. In order to increase sample sizes when distributions for \( \Delta R \) or \( \Delta L \) are presented, values computed for the hours \( t_k, t_k - 1, \) and \( t_k - 2 \) are considered together, approximately tripling the numbers of independent values binned. For such binned sets, the parameters \( \overline{X} \) and \( \sigma_X \) used to define the bins are computed over the combined set of 3 h.

All plots of distributions include three curves. One is the distribution of values of \( R \) (or \( \Delta R \)) for some set considered. Another is that for the corresponding distribution of logarithmic values (i.e. either \( L \) or \( \Delta L \)). The third is the curve for a standard normal distribution (i.e. a Gaussian having mean 0 and standard deviation 1). All distributions are presented with an abscissa labelled in terms of standard deviations about a mean translated to 0; i.e. the distributions are re-scaled so that all are in their corresponding standard forms for easy comparison and presentation on a single diagram. Values are plotted as continuous functions rather than histograms, although they are computed as the latter.

4. Case Descriptions

The first synoptic case (denoted case 1) examined is the 12-hour period beginning 00 UTC 18 May 1999. This period was chosen because there was much convective precipitation over the inner model domain.

The convective precipitation accumulated over the first six hours with each model version for case 1 appears in Fig. 1. There are two primary areas of convection: one over and near Michigan, the other over and near Louisiana. Versions having the same convection scheme but different PBL schemes produce very similar 6-hour accumulated precipitation, but different convection schemes produce notably different results. For this case, the GR scheme produces less precipitation than the others, the BM scheme produces larger maximum values within the two primary regions, and the KF scheme additionally produces precipitation in a third region over and adjacent to Florida.

The domain-total precipitation

\[ \Sigma(m, t_k) = \sum_j R(j, m, t_k) \]  

accumulated during each 1-hour period for each model version for case 1 appears in Fig. 2(a). Note that model versions 4–5 (those using the KF scheme) experience a spin-down during the first six hours, so that at hour 6, those \( \Sigma \) are approximately 1/3 of the
Figure 2. Values of (a) total precipitation $\Sigma$, (b) number of precipitating points $N$, (c) area-mean precipitation $R$, (d) standard deviation of precipitation $\sigma_R$, (e) exponential of area-mean logarithm of precipitation $\exp(L)$, and (f) exponential of the standard deviation of the logarithm of precipitation $\exp(\sigma_L)$, for each hour and model version (indicated on figures) for case 1, computed as described in the text.
values at hour 1. The other model versions experience a weak spin-up during the first four hours, except for version 6 (the GR scheme) whose hourly accumulations vary little with time. Since most of the precipitation occurs over land or just off-shore, diurnal effects should be expected, so that spin-up or spin-down should not necessarily be considered a problem here, but the diurnal responses among the versions differ greatly.

Values of $N(m, t_k)$ and $\overline{R}(m, t_k)$ for case 1 for each $m$ and $t_k$ appear in Fig. 2(b)–(c), respectively. At all hours for all versions, $N > 200$, with the largest $N = 911$ (the latter is 12% of the grid points in the model’s fine-mesh domain). For $m = 4$ and 5, $N$ declines by about 40% between hours 2 and 6 but $\overline{R}$ declines by a factor of 2, implying that the strength of precipitation is declining with time for the versions using the KF scheme. For the GR-scheme version, $N$ increases between hours 1 and 10 by a factor of 3, but $\overline{R}$ declines, again implying that the strength of precipitation declines with time. For the BM versions ($m = 1, 2, 3$), there is less systematic variation in $N$, but it is such that the strength of precipitation may be characterized as peaking at hour 6. At hours 8–12, $\overline{R}$ is similar for all model versions ($0.5 \text{ mm} < \overline{R} < 0.9 \text{ mm}$) except for the GR version ($\overline{R} < 0.25 \text{ mm}$).

The temporal fluctuations of $\sigma_R(m, t_k)$ (Fig. 2(d)) are similar to those for the corresponding $\overline{R}(m, t_k)$. As the strength of precipitation increases or decreases in the mean, the degree of deviation from the mean does likewise. At hours 8–12, $0.59 \text{ mm} < \sigma_R < 1.03 \text{ mm}$ except for the GR version for which $\sigma_R \approx 0.32 \text{ mm}$.

Values of $\exp(\overline{L}(m, t_k))$ and $\exp(\sigma_L(m, t_k))$ appear in Figs. 2(e) and (f), respectively. The former resembles $\overline{R}(m, t_k)$, except corresponding values are approximately half the size. Unlike $\sigma_R(m, t_k)$, however, $\exp(\sigma_L(m, t_k))$ varies relatively little with respect to $t_k$ and $m$, being approximately $4 \pm 1$, except for $m = 3$ at $t_k = 5$. The expression of standard deviation in terms of $L$ therefore seems more stable than in terms of $R$.

Generally $\sigma_R(m, t_k) > \overline{R}(m, t_k)$, implying that the distributions of precipitation values cannot be approximately Gaussian since $R$ is necessarily non-negative. Examples of the distributions of $R$ appear in Fig. 3. Specifically presented are the distributions of precipitation accumulated during the hour ending at $t_k = 3 \text{ h}$ for case 1, versions 1
and 4. For version 1, $\bar{R} \approx \sigma_R$; the distribution of $R$ peaks near $\bar{R} - 0.6\sigma_R$ and generally becomes smaller as $R$ increases further. The distribution of $L$ peaks near $0.6\sigma_L$. The distribution looks neither Gaussian nor log-normal. For version 4 on the other hand, the distribution of $R$ is very poorly approximated by a Gaussian throughout the range of $R$, and instead the log-normal approximation appears very good except for values $L > \bar{L} + 2\sigma_L$ where the $L$ distribution is zero. The result for version 1 is similar to those for versions 2 and 3; that for version 4 is similar to those for versions 5 and 6. At later times, those for versions 1–3 look more like the others: i.e. for all a log-normal approximation is better than a Gaussian one. As an example, the distributions for hour 9 of version 1 appear in Fig. 4(a).

The second synoptic case (denoted case 2) examined is the 12-hour period beginning 00 UTC 3 May 1999. This is the overnight period just prior to a tornado outbreak that devastated portions of Oklahoma and Kansas. This is a period with approximately one half of both the amount and the area extent of convective precipitation compared with case 1 (such that values of $\bar{R}$, however, are similar). The convective precipitation accumulated over the 6-hour forecasts produced with model versions 1 and 5 for case 2 appear in Fig. 5. These are the model versions that create the most different accumulations during this period.

Some of the results reported for case 2 are qualitatively similar to those for case 1. Notably, the temporal fluctuations of $\bar{R}$ and $\sigma_R$ are similar to each other. The 1-hour accumulations for versions 4–5 generally (but not monotonically) decay during the first nine hours as do their numbers of accumulation grid points. Notably different, however, is that for hours 4–6, the 1-hour area totalled accumulations for version 1 are one half or less than those of versions 2–3 (the other BM versions) and less than that of version 6. Also, at hour 12, $\bar{R}$ for $m = 5$ is twice that for $m = 4$ (the other KF version). Only examining two cases is therefore insufficient to fully and accurately characterize the differences in precipitation statistics produced for each scheme separately. The shapes of distributions of values for case 2, however, are qualitatively similar to the corresponding ones for case 1, except for being a bit noisier due to the smaller sample size ($N$ being smaller). As an example, the distributions for hour 3 of version 1, case 2, are presented in Fig. 4(b) for comparison with those in Fig. 3(a). Similarly, the gross shapes of
corresponding distributions of $\Delta R$ and $\Delta L$, as well as values of $\sigma_{\Delta L}$ produced for the two cases are similar, as discussed in the following section.

5. RESULTS REGARDING DIFFERENCES

(a) Same PBL but different convection schemes

Four pairs of experiments have the same PBL schemes, but different convection schemes. These are the pairs (1, 4), (1, 6), (5, 2), and (4, 6), denoted as set 1. All use the BK scheme, except (5, 2) which uses the BT scheme.

Values of $N_{\Delta}(m_1, m_2, t_k)$ for set 1 of case 1 are shown in Fig. 6(a). The values vary between 64 and 397 for different hours and pairs. The ratios of common 1-hour accumulation points with respect to the union of all such points in either forecast (not shown in figure) is between 0.1 and 0.4, indicating that many grid points have precipitation due to one scheme but not the other. For the distributions that will be presented, the consideration of three hours' worth of 1-hour accumulations renders the numbers of distinct values binned as between 412 and 1005.

The values of $\Delta R$ for set 1 of case 1 vary between approximately $\pm 1.2$ mm. The corresponding $\sigma_{\Delta R}$ appear in Fig. 6(b). These vary between 0.04 and 0.24 mm, and for pairs (1, 4), (1, 6) and (5, 2) the $\sigma_{\Delta R}$ increase by a factor of 4 or more during the first 5 or 6 hours. Values of $\exp(\overline{\Delta L})$ for this set (Fig. 6(c)) vary between approximately 0.25 for pair (6, 4) and 6.5 for pair (1, 6), in agreement with the precipitation for version 6 being much smaller than for the others. The corresponding $\exp(\sigma_{\Delta L})$ (Fig. 6(d)) vary between 2.3 and 8.3, indicating that many rather large ratios of pairs of corresponding values exist.

The distributions for the combined first 3 hours of $\Delta R$ and $\Delta L$ values for each pair in set 1, case 1, are presented in Fig. 7. Note that, for all pairs, the log-normal distribution appears to be as good as, or better than, a Gaussian one: i.e. the curve describing the binned log of ratios appears more like the reference normal curve than that for the binned differences. For all these pairs, the distributions of differences are more strongly peaked than a normal distribution, one or other tail is more populated, and values at either 1 or $-1$ standard deviations are greatly underpopulated. This description of the distributions applies at other times as well.

The distributions of $\Delta S$ for set 1, case 1, and their corresponding logarithms of ratios (not shown) appear very similar to those presented in Fig. 7. In particular, all
the distributions except that for (1, 6) appear much more log-normal than normal at all times. For \( t_k > 6 \) h, this statement is true even for (1, 6).

For case 2, since the numbers of points where convection occurs during any hour period are much smaller than for case 1, there are also many fewer common precipitating points for most pairs of set 1. The distribution of values therefore appears much noisier. For many pairs and hours for set 1, the \( \Delta L \) and \( \Delta R \) distributions appear indistinguishable as better approximations to a normal distribution, but for those comparisons for which one distribution does appear more normal, it is always the log-normal one. Values of \( \exp(\sigma_{\Delta L}) \) vary between 2.6 and 6.8 for different times and pairs, as presented in Fig. 8.

(b) Same convection but different PBL schemes

As a contrast to varying convection schemes, pairs of versions having only different PBL schemes were also examined. These are experiment pairs (1, 2), (1, 3), (3, 2) and
(5, 4), denoted as set 2. All use the BM convection, except the last that uses the KF scheme.

Values of $N_\Delta(m_1, m_2, t_k)$ for set 2 of case 1 are shown in Fig. 9(a). The numbers vary between approximately 200 and 850 for different hours and pairs. The ratios of the numbers of common precipitating points to the numbers of points precipitating in either experiment (not shown) is 0.9 initially, decaying to as low as 0.43 at hour 12. The similarities of precipitation regions is therefore much greater in these pairs of experiments than for those in set 1, but the similarities decrease as the forecasts have more time to diverge from one another. Between 750 and 2400 points are binned for each distribution shown.

The values of $\Delta R$ for set 2, case 1, vary between $\pm 0.3$ mm with no systematic variation in time. The corresponding $\sigma_{\Delta R}$ (Fig. 9(b)) generally increase with time, varying between 0.17 and 1 mm. This is approximately 1/3 of the values when only the convective scheme is varied instead. Values of $\exp(\Delta L)$ (not shown) vary between 0.7
Figure 8. As in Fig. 6(d), except for case 2.

Figure 9. Values of (a) the common number of precipitating points $N_\Delta$ and (b) the standard deviation of precipitation differences $\sigma_{\Delta R}$, for 1-hour periods ending at the indicated times for model version pairs in set 2, case 1. The curves are labelled by the two indices for each pair of versions compared.

and 1.3, with $\exp(\sigma_{\Delta L})$ approximately 1.4 at $t_k = 1$ h and increasing to approximately 3.0 at $t_k = 12$ h for all pairs of set 2.

The distributions for the first 3 hours of $\Delta R$ and $\Delta L$ values for pair (4, 5) of case 1 are presented in Fig. 10. The distributions appear similar when presented in this standard form. Neither appear normal: the percentage of values within 1/2 standard deviation of the mean is greater than for a normal (by more than 100%) and the percentages for $\Delta R$ or $\Delta L$ near $-1\sigma$ are about half or less of those for a normal. Similar statements apply to other pairs of set 2 at this and other times, for both cases 1 and 2.

(c) Different convective and PBL schemes

All the pairs of versions for which both convection and PBL schemes differed were also examined. Both their distributions and statistics were very similar to corresponding
pairs for which only convection schemes were varied (i.e. for which one of the PBL schemes is altered to yield a pair in set 1). This agrees with the result that varying only the convection schemes has greater effect than varying only the PBL schemes.

(d) Non-overlapped convection points

In Errico et al. (2000), for a simple stratiform precipitation model, it was assumed that if the model produces an estimated precipitation rate $r_e > 0$ given correct temperature and moisture input, then the probability that the true rate $r_t = 0$ is smaller for larger $r_e$. In other words, for larger $r_e$, it is less likely that the error is such that no precipitation should actually occur. We can test that hypothesis here for convective precipitation models by treating any one model version as truth.

All version pairs in sets 1 and 2 are investigated. We consider three consecutive hours of hourly accumulations ending at hours $t_k$. For each model version $m$, the number $M_\ell(m, t_k)$ of such values that fall in each of 10 bins (indexed by $\ell$) are determined. Unlike earlier, these bins are defined by choosing bin ranges such that the $M_\ell$ would all be identical if the $R$ were distributed log-normally, given its calculated $\bar{L}$ and $\sigma_L$ for the set of values of $R$ considered. For each $m$ and $t_k$, the index value $\ell$ is noted for each grid point $j$ for which $R_{j,m,t_k} \neq 0$. Then, for each $m$, $\ell$, and three hours ending at $t_k$, for each other version $n$, the number $M_{\ell_0}(m, n, t_k)$ of corresponding grid points for which $R_{j,n,t_k} = 0$ is determined, and the fraction $f_{\ell}(m, n, t_k) = M_{\ell_0}(m, n, t_k)/M_{\ell}(m, t_k)$ is calculated. Unlike for the examinations in previous sections, no filtering of small values of $R$ is applied here. Although the $M_\ell(m, t_k)$ are not all equal, for almost all $m$ and $t_k$ the bins are all well populated, which was our motivation for using the bin ranges we chose.

Examples of $f$ for pairs in set 1 of case 1 appear in Figs. 11(a)–(d) for 3 hours of values ending at hours 3 and 6. Since $f$ does depend on the ordering of the pair of versions considered, eight curves are presented for each time (four distinct pairs on one diagram, and the four with order reversed on another). The abscissa value corresponding to each bin is that which divides that bin into two parts having equal population,
assuming a log-normal distribution with mean $\bar{L}(m, t)$ and standard deviation $\sigma_L(m, t)$ for the sets of hours considered. The abscissa is labelled in units of $\sigma_L(m, t)$ about a mean translated to 0. Values are plotted as continuous curves, although they are in fact computed as for histograms.

At hour 3 (Figs. 11(a) and (b)), for most pairs there is a marked decrease of $f$ with increasing $L$, but for three pairs $f$ is a minimum near $L = \bar{L}$. For the latter pairs, $f$ for the bin of largest $L$ is similar to that for the bin of smallest $L$. For two of these pairs at hour 6 (Figs. 11(c) and (d)), the minimum of $f$ is still obtained at other than the bin of largest $L$. At later times (not shown), however, all pairs have the minimum $f$ in the bin of largest $L$.

In Errico et al. (2000), the equivalent of $f$ was assumed to be proportional to $\exp(-\beta R)$ with $\beta$ as a constant. The results in Fig. 9 suggest, however, that $f$ being linear in $\ln(R)$, with some bounding between 1 and 0, may be a better assumption.
6. Conclusions

First, distributions of grid-point values of hourly accumulated convective precipitation produced from 12-hour forecasts using each of six versions of a forecast model were examined using a simple binning technique. These distributions were better characterized as log-normal than normal. This result was expected, based on both investigations of observed precipitation (e.g. Kedem et al. 1990) and on theoretical arguments using simple models (e.g. Lopez 1977). The shapes of distributions produced by the model versions and cases investigated here are therefore consistent with other reported studies of real or simulated precipitation.

Attention was then focused on pairs of corresponding grid-point values of hourly accumulated convective precipitation of model versions having different convection schemes. At grid points and times where both versions had non-zero values of hourly accumulations, distributions of both differences of the pair of values and differences of their logarithms were examined. When one kind of distribution appeared more normal than the other for any time period or pair of versions, it was always the logarithmic one. This means that the distribution of the logarithms of ratios of corresponding values appears more normally distributed than that of the corresponding differences of values. The standard deviations of the logarithms of ratios of pairs are equivalent to factors of 2 or more, even for forecast durations as short as three hours.

When pairs of forecasts using the same convection schemes but different PBL schemes were examined in the same way, the distributions of differences appeared neither normal nor log-normal, being too sharply peaked near the distribution mean for both distributions. When both convection and PBL schemes were different, however, results were similar to those when only the convection schemes differed. In particular, the corresponding distributions had similar shapes, means, and standard deviations.

When a pair of model versions have different convective schemes, many grid points that experience hourly accumulations of convective precipitation in one forecast experience none in the other, and vice versa. For any hour, the ratio of the number of common convective grid points to the number of grid points for which either forecast has accumulations is less than 0.4. In contrast, when only PBL schemes differ, the ratio is near 0.9 for hour 1, but decreases to less than 0.7 or less at hour 6. For hour 6 and later, the larger a grid-point hourly accumulation is in one forecast, the more likely it is that the other forecast will also have a non-zero accumulation. At hour 3, however, this relationship is not observed for all pairs of model versions investigated.

The purpose of our study was not to characterize results produced by different convective and PBL schemes. Instead our intention was simply to use such differences to characterize model precipitation error, under the assumption that any one of the model versions can be considered 'truth' and differences from it therefore considered as 'errors'. This assumption relies on the fact that all of the model versions investigated have been claimed to produce useful, realistic simulations, and none has been demonstrated to be generally and significantly, statistically better than another in well-designed experiments. This assumption is likely not valid, at least for some versions investigated, although definitive evidence is lacking.

For the last reason, any quantitative results we report should not be used to specify the precipitation error statistics of an arbitrary convective parametrization scheme or forecast model. We consider it reasonable, however, to use our qualitative results for that purpose until some better estimation of model error is produced. So, for accumulations of 1 to several hours (up to 12 hours were investigated), it appears better to claim a log-normal error distribution than a normal one, and for such distributions, to use standard deviations equivalent to ratios of 2 or more. For forecast times of 6 h or longer, ratios
as large as 4 or more may even be more realistic. It is for the same reason that some more sophisticated techniques commonly employed by statisticians (e.g. Gilchrist 1984, chapter 11) to test fits to distributions were not used in this study. The binning method employed is sufficiently informative, and information about details, such as fits to the tails of distributions, is not very relevant for our intentions. (In data assimilation, the tails of distributions are often modelled by some quality-control procedure: e.g. as in Lorenc and Hammon 1988.)

We expected that our results would support the specification of log-normal errors in precipitation models, as was assumed in Errico et al. (2000) for a model of non-convective precipitation. What was very surprising, however, was how large the standard deviations of such distributions may be for present convection schemes. A standard deviation corresponding to a factor of 2 implies that 30% of the values produced by the scheme are expected to be greater than 2 times or less than one-half 'truth'. This value of 2 appears to be on the low side too.

For an observation to be valuable for data assimilation, there must be a reliable way to relate, or model, that observation to the fields being analysed. If that model has a significant probability of having large errors, the utility of the observation will be small. Furthermore, if the assimilation system ignores the fact that such errors are large, there is a likelihood the analyses it produces will be degraded by consideration of otherwise accurate observations. When the model errors and observation errors can both be very large, as is the case with present convective precipitation models and precipitation observation systems, the usefulness of such observations for data assimilation should be very carefully evaluated. It is therefore important to obtain good estimates of both model- and observation-error distributions for precipitation, and to assess their utility in well-designed, realistic or simulated, data-assimilation systems.

ACKNOWLEDGEMENTS

The authors acknowledge fruitful discussions with Luc Fillion regarding this study. This work was partially funded by the United States Office of Naval Research, grant N00014-00-C-0243, NASA contract W-18,077, Mod. 6, and the United States Weather Research Program funded at the National Center for Atmospheric Research.

REFERENCES


Black, T. L. 1994 The new NMC mesoscale eta model: Description and forecast experiments. Weather Forecasting, 2, 266–278

Blackadar, A. K. 1979 High resolution models of the planetary boundary layer. Advances in Environmental Science and Engineering, 1, 50–85


