A SYNOPTIC EVALUATION OF THE NCEP ENSEMBLE

Zoltan Toth[1], Eugenia Kalnay, Steve Tracton, Richard Wobus[1] and Joe Irwin

National Centers for Environmental Prediction
NOAA/National Weather Service
Washington DC, USA

Summary: Ensemble forecasting has been operational at NCEP (formerly NMC) since December 1992. In March 1994, more ensemble forecast members were added. In the new configuration, 17 forecasts with the NCEP global model are run every day, out to 16 days lead time. Beyond the 3 control forecasts (a T126 and a T62 resolution control at 00Z and a control at 12Z), 14 perturbed forecasts are made at the reduced T62 resolution. Global products from the ensemble forecasts are available from NCEP via anonymous ftp.

The initial perturbation vectors are derived from seven independent breeding cycles, where the fast growing nonlinear perturbations grow freely apart from the periodic rescaling that keeps their magnitude compatible with the estimated uncertainty within the control analysis. The breeding process is an integral part of the extended–range forecasts, and the generation of the initial perturbations for the ensemble is done at no computational cost beyond that of running the forecasts.

A number of graphical forecast products derived from the ensemble are available to the users, including forecasters at the Hydrometeorological Prediction Center of NCEP. The products include the ensemble and cluster means, standard deviations, probabilities of different events. One of the most widely used products is the "spaghetti" diagram where a single map contains all 17 ensemble forecasts, as depicted by a selected contour level of a field, e.g. 5520 m at 500 hPa height or 50 m/s wind speed at the jet level.

With the aid of the above graphical displays and also by objective verification we have established that the ensemble can provide valuable information for both the short and extended range. In particular, the ensemble can indicate potential problems with the high resolution control that occur on rare occasions in the short range. Most of the time, the "cloud" of the ensemble encompasses the verification, thus providing a set of alternate possible scenarios beyond that of the control. Moreover, the ensemble provides a more consistent outlook for the future. While consecutive control forecasts verifying on a particular date may often display large "jumps" from one day to the next, the ensemble changes much less and its envelope of solutions typically remains unchanged. In addition, the ensemble extends the practical limit of weather forecasting by about a day. For example, significant new weather systems (blocking, extratropical cyclones, etc.) are usually detected by some ensemble members a day earlier than by the high resolution control. Similarly, the ensemble mean improves forecast skill by a day or more in the medium to extended range, with respect to the skill of the control. The ensemble is also useful in pointing out areas and times where the spread within the ensemble is high and consequently low skill can be expected, and conversely, those cases in which forecasters can make a confident extended range forecast because the low ensemble spread indicates high predictability. Another possible application of the ensemble is identifying potential model errors. A case of low ensemble spread with all forecasts verifying poorly may be an indication of model bias. The advantage of the ensemble approach is that it can potentially indicate a system bias even for a single case while studies using only a control forecast need to average many cases.

1. INTRODUCTION

Since the advent of weather forecasting it has been evident to both the forecasters and the public that the forecasts have only a limited success, i.e. their skill drops off with lead time, and varies from case to case. Though it was clear that often there was not enough data to adequately estimate the initial state of the atmosphere, much of the blame for forecast failures was attributed to the numerical weather prediction (NWP) models and methods for data assimilation. This resulted in a successful effort to improve these forecast tools.

Model errors, however, are not the only source of errors in the forecasts. Recent studies (e.g., Reynolds et al., 1994) indicate that most of the synoptic scale errors in the extratropics in global NWP models *are not due* to model deficiencies. This observation has important consequences for practical weather forecasting. Even though model errors can occasionally play a role, we have to look for other sources of errors.

1. General Sciences Corporation (Laurel, MD) at NCEP

The work of Lorenz (1963, 1965, 1969) and others have provided an understanding of what makes forecasts go wrong. The atmosphere is a chaotic system, which means that even if we had a perfect atmospheric model, the smallest errors in the initial conditions would make the forecasts fail within two weeks or so. This is because instabilities in the atmosphere will amplify even infinitesimally small initial differences between two model runs (or between a "perfect" model run and the true evolution of the atmosphere) till the two forecast states are as far apart as two randomly chosen states in the atmosphere.

In the early years of NWP, forecast errors due to simplified model formulations dominated the total error growth. The traditional perception that forecast errors are primarily due to model errors dates back to those early years. By now, however, models have become much more sophisticated and it is the errors that arise due to instabilities in the atmosphere (even in case of small initial errors) that dominate forecast errors. The recognition of this situation requires a major shift in the perception of NWP. The models and the data assimilation can still be improved but that alone will not solve all problems. The forecasts, no matter how good our models may become, will always fail due to the inherent finite time limit of predictability of the atmosphere (which may become a few days longer as atmospheric analyses are improved). However, with ensemble forecasting we can learn a lot about when, where and how the skill is lost in the forecasts. Since the rate at which forecasts lose their skill varies substantially from case to case, this is very important information that can expand significantly the usefulness of the forecasts.

Ensemble forecasting, which is the only practical way to learn about the skill of the forecasts in advance, was first introduced by Leith (1974) and Epstein (1969). It consists of (1) estimating the probability distribution of the true state of the atmosphere around the control analysis, then (2) sampling that probability distribution and (3) running the forecast model from all the sample points (perturbed analyses). If everything is done right, we end up with a number of possible forecast scenarios that are equally likely (except that solutions near the control are somewhat more likely) and that will encompass the true state of the atmosphere. If, on a particular day, the ensemble members have a large spread at, let's say, at 3 or 5–day lead time, this means that the forecasts are less reliable and hence the forecasts should be worded differently or even alternate scenarios should be mentioned.

The ensemble strategy will work only if the models are good enough that model related errors do not dominate the final error fields. This is one of the reasons the ensemble approach was introduced into operational practice only about 3 years ago, when the reliability of the models (and the availability of enough computer power) made it possible to run multiple of forecasts. The first two centers that entered this new area of NWP were the National Centers for Environmental Prediction (NCEP, formerly NMC; Tracton and Kalnay, 1993) and ECMWF (Buizza et al., 1993; Molteni et al., 1995). The two centers use almost the same horizontal model resolution for the perturbed ensemble forecasts, however, the estimation (and sampling) of the initial probability distribution of the atmosphere is different. At ECMWF, the singular vector (SV) approach is used (Buizza and Palmer, 1995), in which the fastest growing perturbations are determined for a 2–day period at the beginning of the forecast. At NCEP, the breeding method is used (Toth and Kalnay, 1993, 1996a, hereafter referred to as TK93, TK96). In this technique (see next section) perturbations that grew very fast during the analysis cycles leading to the initial time, and thus are likely to dominate analysis errors, are used as initial ensemble perturbations.

After NCEP and ECMWF started at the end of 1992, other centers have also considered implementing ensemble forecasting. At the Fleet Numerical Meteorological and Oceanohraphic Center an ensemble forecast system based on the breeding method has been quasi–operational since 1995 (M–A. Rennick, personal communication, 1995). At the Canadian Meteorological Center an approach based on multiple data assimilation cycles, which is an extension of the breeding method, is being prepared for operational implementation later this year (Houtekamer et al., 1996). At the South African Weather Service, an ensemble system, using the breeding method, is being prepared for evaluation and implementation (Tennant, 1995, personal communication).

NCMRWF in India is also considering the application of a breeding based ensemble system (Iyengar, 1995, personal communication). At the UK Meteorological Office, ensemble experiments have been performed by introducing ECMWF initial perturbations on either the UK Meteorological Office or ECMWF analysis (Harrison et al., 1995).

In the rest of this paper we will describe how the breeding method works in the practice of NCEP to generate initial ensemble perturbations, using as an example the Blizzard of 1993 (section 2). A short description of the post–processed information that is made available at NCEP and is distributed through anonymous ftp every day to the NWS forecasters and to the general user community is given in section 3. Then we present several synoptic examples to illustrate the benefits of using the ensemble forecasts instead of relying only on a control forecast (section 4). Section 5 is a short discussion of questions and plans for the use of the NCEP ensemble.

2. CONFIGURATION OF THE ENSEMBLE

Since March 1994 the NCEP ensemble has consisted of 17 forecasts every day, including at 00Z a T126 (MRF) and a T62 (started from truncated T126 analysis) resolution control forecast and at 12Z a T126 control (AVN) forecast (see Fig. 1). Note that the high resolution control forecasts (MRF and AVN) are truncated to T62 after 7 and 3 days respectively, to save computer time, since Tracton and Kalnay (1993) showed that most of the advantages of high resolution are derived from just the first few days of the forecast. The rest of the ensemble consists of 10 perturbed forecasts at 00Z and 4 at 12Z, all done at a T62 resolution.

The initial perturbations must be representative of possible analysis errors. Since forecast errors grow fast from the beginning, it is obvious that the analysis fields contain fast growing errors. These errors get into the analysis through a dynamical process, due to the use of model generated first guess fields (TK93). This is because the analysis in data void regions relies heavily on the first guess and therefore contain errors similar to short range forecast errors. Beyond these dynamically conditioned, fast growing errors, there are random errors that are a result of stochastic errors in observations and/or in analysis procedures. The random errors, however, generally will not dominate forecast error, since they typically do not grow (and initially may even decay) until they become dynamically organized into growing patterns a few days later.

In the breeding method we attempt to capture the fastest growing analysis errors that are most likely to be responsible for the error in the control forecast. Therefore the breeding cycle intends to mimic how the fastest growing errors in short range forecasts develop. In the operational ensemble system, the difference field between a positively and negatively perturbed 24–hour forecast from the previous day, valid at the time of the latest analysis, is used as an initial perturbation for today's ensemble (Fig. 2). Before adding (and subtracting) that perturbation onto the latest control analysis, it is first "scaled down" to a perturbation amplitude that is representative of actual analysis errors. So, for example, initial perturbations over the Pacific ocean where rawinsonde observations are sparse, have amplitudes three times larger than over the continental US, which is relatively well observed. Since the same perturbations are added with both positive and negative sign onto the control analysis, one breeding cycle (Fig. 2) generates perturbations for a pair of ensemble forecasts. Operationally, there are 7 independently run breeding cycles that differ only in that the very first initial perturbations in each cycle are independent. Since the bred perturbations are computed from the differences of the previous day's ensemble forecasts (at one day lead time), breeding is done globally at no extra computational cost. Over the course of 3–4 days after initiating the ensemble, the perturbations loosely converge to a subspace of perturbations that grow fastest in a sustainable manner. The bred perturbations are closely related to the Lyapunov vectors of the atmosphere (see e.g., Tsonis, 1992); they can be considered as a nonlinear extension of the linear Lyapunov vectors (TK96).

In regions where fast developments (e. g., cyclogenesis) take place and instability is high, the perturbations from different breeding cycles are found to converge. As the cyclone decays the

perturbations in the region considered become less correlated again in the different breeding cycles, till another system with strong instabilities develops. We will document this behavior through the example of the "storm of the century", a massive and very fast developing storm that hit much of the eastern US in March 1993. In Fig. 3 we present a series of 500 hPa streamfunction fields, leading to the development of the storm. One can see that there is little change on the large scales and with conventional analysis techniques no precursor of the cyclone can be seen earlier than 24–36 hrs before it actually appears on the 13th of March.

In Figs. 4a and 4b, we show the corresponding bred perturbations from two independently run breeding cycles (differences between control and perturbed nonlinear forecasts). We can see in Fig. 4a that there is a packet of instability (larger amplitude perturbations with a negative–positive–negative wave triplet) on the 8th of March in the eastern part of the Pacific. This wave packet then travels primarily to the east and by 12th of March it appears over the western Atlantic. The speed of the packet of instability is larger than the actual wind speed. In fact, if we follow the evolution of the perturbations in time we can notice that there is a very strong downstream development (and upstream decay) that makes the fast speed possible. The same can be seen in Fig. 4b.

One can also notice in Figs 4a and 4b that there is a second surge of perturbation energy that reaches the eastern half of the Pacific on the 10th of March. This wave packet, however, does not follow the previous one that traveled to the East, but rather takes a southerly course and plunges into the Gulf of Mexico, where the actual storm was formed. In Fig. 5 we zoom into the perturbations in time and space, just before the development of the cyclone. As we can see, the perturbation amplitude more than tripled in a 12 hour period ending 06Z on the 13th of March in the Gulf of Mexico, where the storm actually developed. This shows that the bred perturbations can grow extremely fast in the presence of strong instability. In addition, these perturbations are often related to actual forecast errors. In this case, for example, forecasts verified on March 13, 12Z with a lead time of 12, 36, ..., 132 hours (not shown) all had a similar error pattern with a maximum near the location of the negative perturbation center in the Gulf of Mexico in Fig. 5 (12Z panel). It is important to note that the same error was found even at 12 hour lead time (Fig. 6), suggesting that a 6–hour forecast valid at the same time, used in the analysis as a first guess, would also have a similar error pattern. And since in that area there are not enough observations to "correct" the first guess error, the analysis at 12Z should have had a similar error pattern.

The fact that the bred perturbations may appear as errors in the first guess and the analysis gives the justification for their use in ensemble forecasting. A successful ensemble scheme requires perturbations that are possible analysis errors. An objective comparison of the bred perturbations and analysis/first guess fields confirms the presence of bred perturbations within the analysis errors (Kalnay and Toth, 1994).

3. NCEP ENSEMBLE FORECAST PRODUCTS

An ensemble of 17 forecasts out to 16 days is a lot of data to examine and to archive. To reduce the volume, each day at 00Z and 12Z we create user files that contain forecast fields (all lead times and all forecast runs) for 17 different selected variables. The variables (Table 1) were chosen with different forecast applications in mind, including extratropical prediction, aviation forecasts, tropical predictions, hurricane forecasting, marine prediction, etc. The data are available in grib format at an anonymous ftp site (nic.fb4.noaa.gov, /pub/ens).

In addition to these gridded fields, postprocessed graphical information is also provided in GEMPAK metafile format to the general user community, along with a user's manual (nic.fb4.noaa.gov, /pub/nadata/meta/model/ens). These same products are used in the operational practice at the Hydrometeorological Predictions Center (HPC) of NCEP for the preparation of medium–range forecast guidance. Of all the graphical products available to the forecasters, the most heavily used is the "spaghetti" diagram (Fig. 7), where a single selected

contour of a variable is plotted for each ensemble member. This, in fact, is probably the simplest way one can display information directly from all individual forecasts on one panel. The advantage, from a synoptic point of view, is that the position of smaller scale features can be seen in each ensemble run. In the example of Fig. 7, we can see a large uncertainty at 4.5 day lead time regarding the position and amplitude of a trough over the eastern part of the US. As one can see, the control forecast is not necessarily in the middle of the pack, in this case it was on the side of the distribution that did not verify well. The most frequently looked at variable is the 500 hPa height, but other fields such as 850 hPa temperature, 1000/500 hPa thickness, mean sea level pressure or wind speeds at 850 and 250 hPa heights are also used. One should be aware that in areas of small gradients (like over a flat ridge) large differences among the contour lines do not necessarily mean substantial differences among the solutions. To avoid such problems we find it best if the spaghetti diagram is viewed alongside a full field map for the control forecast (or ensemble mean).

The ensemble mean field is also available to the forecasters. This field offers a forecast that is, on average, better than the control (or any other ensemble member) forecast in terms of RMS errors. The small scale features whose position is uncertain (but which nevertheless will most likely still appear somewhere in the verification) are naturally filtered out from the ensemble mean, so that this information is most useful for the extended range (days 6–8 and beyond). For the medium range, subsets of the ensemble (clusters, that are formed of ensemble members with similar solutions based on an objective algorithm) are more useful (Fig. 8). The two leading clusters shown in the example of Fig. 8 clearly identify two distinct scenarios. Note that in this case none of the three controls was classified in cluster 1, which verified very well over eastern US. The spread of the ensemble around the ensemble mean (Fig. 9) also indicated to the forecasters the large uncertainty in the position of the trough over the Eastern US.

Another important guidance is the map of probability. For example, the probability that the 850 hPa temperature will be below (or above) a certain contour level can be displayed in a map format, by counting how many of the 17–member ensemble runs had such a temperature (not shown). Similarly, the probability that the 500/1000 hPa thickness value will be above 60 m is displayed (Fig. 10). For the selected case, high probabilities are given for the Western part of the country and low over the Norhteast but in between the probability values reflect the uncertainty in the position of the trough over the Eastern US. Height tendencies can also be viewed in a probabilistic manner (Fig. 11), indicating a strong likelihood of height falls around the Ohio valley.

The set of products available from NCEP keeps growing through the interaction with the forecasters. Recently we added maps in which the high and low surface pressure centers are marked for each ensemble member; probabilistic quantitative precipitation forecasts, probabilistic "meteograms" (weather parameters at certain locations) and other products will also become available in the near future.

4. SYNOPTIC EXAMPLES

In this section, we show examples of the everyday use of the ensemble, pointing out the potential benefits the ensemble can offer (in addition to using only a control forecast), through a few selected synoptic examples. Perhaps the most important benefit the ensemble can offer is that it helps to distinguish between cases when a meaningful forecast can or cannot be made with confidence. This becomes obvious if we compare, for example, Fig. 12 and Fig. 13a. These two 10.5 day forecasts initiated only three weeks apart but their skill levels are very different. The case from April (low skill) is a good example of pure chaos, while the March case (high skill) ensemble is much more orderly, suggesting to the forecaster a relatively high degree of predictability even at this extended lead time.

Note also that in both cases the verification line, for most of the domain, lies within the cloud of the ensemble. This is true in general: only in 15% of the cases the ensemble fails to give reliable information in this respect, so that the forecaster can be fairly confident that the verification will fall

within (or close to) the envelope of the ensemble. The fact that with initial perturbation amplitudes that are representative of analysis uncertainties, our ensemble encompasses truth most of the time, confirms the assertion made in the Introduction that the extratropical forecast errors are due primarily to initial value uncertainty and not to model deficiencies. This point is further corroborated by the observation that in most cases when the ECMWF operational control forecast and the NCEP control have largely different solutions, the NCEP ensemble encompasses the ECMWF control as well. In the example of Fig. 14, the ECMWF control forecast had a trough over the West Coast of the US that was positioned substantially further west with a different tilt than that of the MRF control. The NCEP ensemble, however, incorporated not only the ECMWF control but also the verification line.

The fact that the verification typically falls within the ensemble also implies that the ensemble is a good indicator of skill not only in time but also in the spatial domain. In Fig. 13a, for example, one can see that in areas of small spread (West Pacific, mid–Atlantic) the control forecast (and most of the perturbed forecasts) is almost perfect while in areas of large disagreement (e. g., East Pacific) the control forecast can have larger errors.

Another feature of the ensemble welcomed by synopticians, and which can have great impact on how forecasts are made, is the fact that the ensemble is much more consistent in time. If we compare today's ensemble valid for the weekend and then compare it with yesterday's ensemble valid at the same time chances are we will find that they are quite similar, even if the corresponding control forecasts were very different. Fig. 13 gives one example of this feature of the ensemble, showing that 10.5, 6.5 and 3.5 days long ensemble forecasts that are remarkably similar, despite substantial shifts in the behavior of the control forecast. The change in the ensemble with decreasing lead time is primarily a reduction of the spread and not a shift in the position of the envelope of the ensemble.

Since time consistency in the forecasts is an important point for practicing synopticians, we present another example where the control forecasts from consecutive days suggested largely different scenarios over the east Pacific and western US (Fig. 15). The ensemble in this critical situation, again, displayed a much desired consistency in time. Note that when the controls lie north of the verification (Fig. 15a, 9.5 days lead time) the ensemble indicates that a deep trough in the SW is a possibility, whereas next day (Fig. 15b, 8.5 days lead time), when the controls take on this solution, much of the ensemble lies NE of the control, with the envelop basically unchanged. If the spread is large in the ensemble (as in our example here) the ensemble, of course, does not tell the forecaster which is the solution that will verify best. But at least it gives a range of solutions that will most likely encompass truth. Situations like this, indeed, call for some kind of probabilistic approach in forecasting.

Looking at our last example (Fig. 15b) one may notice that in areas where most ensemble members were on one side of the control (e.g., NE of the control over SW US or South of it over the Eastern US) the verification also tends to lie on the same side, i. e., on the side where the majority of the ensemble members are. So on average, it is worth adjusting the control somewhat into the direction of the bulk of the ensemble. There is a caveat there, though. As described in section 2, all perturbed forecasts within the ensemble are made at a resolution that is half of the resolution of the operational control forecasts (MRF or AVN). So in cases where the resolution may have a strong impact (like in certain synoptic situations around mountains) the low resolution ensemble may be handicapped with respect to the MRF or AVN controls. In the example of Fig. 16 the T62 resolution control is dramatically different from the MRF scenario over the SW and there is only one ensemble member that has a solution similar to the MRF. Note also that the AVN control, that is run at the high T126 resolution only for the first three days, is still similar to the MRF in the SW. This, and similar examples suggest that increasing the resolution of the ensemble for the first 3 days or so of the integration may enhance the utility of the ensemble in these situations.

As mentioned earlier and documented elsewhere (TK93, TK96) the ensemble mean, on average, verifies considerably better than the control forecast in the medium and extended ranges. This

amounts to extending the limit of useful forecast skill by a day or so. The advantage of the ensemble in this respect is not restricted to the ensemble mean, though. In difficult forecast situations where the control forecast is rather poor we notice that some ensemble members will capture the right solution a day in advance of the control forecast. Some NWP models have become quite good in reproducing blocking frequencies in a climatological sense. The prediction of individual blocking events, though, is still a challenge. This is because blocking may occur through the nonlinear interaction of several dynamic developments that need to be analyzed and predicted fairly well. In the example of Fig. 17a, the 5.5 day lead time control missed the blocking in the NW Atlantic. There was one ensemble member, however, that indicated the possibility of this development. The control forecast initiated the next day (Fig. 17b) was quite accurate, as were many members of the ensemble. Due to the often highly nonlinear nature of blocking development a much larger ensemble would be needed to reliably indicate possible blocking development at longer lead times.

So far our examples came from the medium and extended range. But on occasion the global ensemble can give the forecaster crucial information on short range as well. In Fig. 18b we show an example where the ensemble indicated an unusually high uncertainty in the 36 hour forecast. The control forecast obviously did not capture the fast developing storm that appeared over Canada while some of the ensemble members were close to the verification. The same storm had a strong impact over the NE US two days later (Fig. 18c). The control gave a similarly poor forecast at this lead time, which is not surprising given its earlier failure. The ensemble captured not only the NE storm but also a second storm affecting the western part of Canada. Concerning the NE storm, note how strongly nonlinear the distribution of the ensemble members is with respect to the control. It should be remembered that at the initial time the perturbations are symmetrically distributed around the control (half on one and another half on the other side of it). At three and a half day lead time (Fig. 18c) all perturbations are South of the control, indicating that if there is an error in the control initial condition that will most likely result in a (possibly large) underforecast of the storm in question. Given the high degree of nonlinearity and the large uncertainty in the ensemble the forecaster could have included in his/her prediction the possibility of a strong storm development over the extreme NE. Another lesson from this example is that a disappointing failure of the control forecast should not necessarily be blamed on the model. Small initial uncertainties, in areas of large instabilities, can amplify even within a couple of days to differences within the ensemble that can explain why the forecast fails.

Earlier in this paper we emphasized that our global models have become reliable enough that the forecast quality generally does not hinge on model deficiencies. Nevertheless the models are still far from perfect. Interestingly, potential problems with the model can be identified through the use of the ensemble in a much more efficient way than using only a control forecast. Through the traditional approach, a large number of control forecasts issued in similar synoptic situations would be needed to identify systematic errors with the model. We need many cases because the "signal", i.e., systematic model error, is often smaller than the errors that are due to fast developing instabilities. Running an ensemble can actually cut down the cost of this process. If in areas of small ensemble spread we see that the verification lies far outside the ensemble cloud at short lead times, we can suspect that there may be a model problem. In these areas there is little sensitivity to initial conditions yet all forecasts fail. So it is probably not the analysis but rather the model that is responsible for these errors. We can find an example for this situation in Fig. 18c. Over the West Coast of America the control and all ensemble members are far away from the verification. Note that the error here is not connected to a developing synoptic feature like the two waves discussed earlier further in the east but rather appears as a larger scale bias which perhaps may be due to some inaccurate parametrization in the model.

The ensemble approach can not only help to improve the model by identifying potential systematic errors but can also point to areas where additional observations may help to improve the *analysis*. In Fig. 18a we see a very short range ensemble forecast. When comparing the 12 hour lead time ensemble (Fig. 18a) to the 36 hour ensemble (Fig. 18b) it is obvious that the large uncertainty

appearing at 36 hour lead time came from an area upstream, north of Lake Winnipeg. Fig. 18a actually suggests that the analysis at 12Z 1995/03/14 would be quite uncertain in this area because the huge uncertainty in the first guess short range forecast. And since the ensemble forecasts indicate that at later lead times (see Fig. 18c) the forecast uncertainty becomes even greater, in a situation like this it would be ideal to send deployable observing platforms (such as manned or unmanned aircraft) to take additional observations in the area of high spread indicated in Fig. 18a. If such extra observations had been used for the 12Z analysis, they would certainly have reduced the uncertainty in the initial condition, and the forecast started at 12Z would become more reliable. The strategy described here is called "targeted observations" (see, e. g., Snyder, 1996) and it has the advantage that it can pinpoint areas where extra observations can make a critical contribution to increasing the skill of our forecasts (Kalnay et al, 1996).

5. DISCUSSION

Most of our discussion has so far focused on medium range predictions. The concept of ensemble forecasting, however, applies to any temporal and spatial scale. At NCEP, experiments are going on in the area of regional ensemble forecasting (Brooks et al., 1995). A successful short range ensemble would be an essential component in the National Weather Service's efforts toward initiating probabilistic quantitative precipitation forecasting. Plans are also prepared for running the coupled ocean–atmosphere model in an ensemble mode, using the breeding method (Toth and Kalnay, 1996b), for making the seasonal forecasts more reliable.

Regarding the use of the NCEP global ensemble, forecasters at the Hydrometeorological Prediction Center, the Climate Prediction Center and many NWS field offices have already adopted the ensemble as a very useful tool in their preparation of medium and extended range predictions. Nevertheless, the ensemble system can be further improved in several ways. Currently we are fine tuning the way we set the initial perturbation amplitudes in the ensemble, which have been perhaps too large in the summer. We also plan to increase the horizontal resolution for at least some of the perturbed forecasts for the first three days. Work is also underway to include more postprocessed products, many of them in probabilistic terms, related more closely to sensible weather parameters like precipitation, low level temperature and wind. In the use of the ensemble, the most difficult task will continue to be that of the forecasters who need to master a new perspective to weather forecasting and be able to interpret a larger amount of information than ever before. We truly believe, and indications are already clear, that this will lead to a genuine increase in the overall usefulness of weather forecasts.

REFERENCES:

Brooks, H. E., M. S. Tracton, D. J. Stensrud, G. DiMego, and Z. Toth, 1995: Short–Range Ensemble Forecasting (SREF): Report from a workshop. *Bull. Amer. Meteorol. Soc.*, **76**, 1617–1624.

Buizza, R., 1994: Sensitivity of optimal unstable structures. *Q. J. R. Meteorol. Soc.*, **120**, 429–451.

Buizza, R., and T. Palmer, 1995: The singular vector structure of the atmospheric general circulation. *J. Atmos. Science,* **52**, 1434–1456.

Buizza, R., J. Tribbia, F. Molteni, and T. Palmer, 1993: Computation of optimal unstable structures for a numerical weather prediction model. *Tellus*, **45A**, 388–407.

Epstein, E. S., 1969: Stochastic dynamic prediction. *Tellus,* **21**, 739–759.

Harrison, M. S. J., D. S. Richardson, K. Robertson, and A. Woodcock, 1995: Medium–range ensembles using both the ECMWF T63 and Unified models – An initial report. Technical Report No. 153, UK Met. Office. [Available from: Forecasting Research Division, Meteorological Office, London Road, Bracknell, Berkshire RG12 2SZ, UK.]

Houtekamer, P. L., L. Lefaivre, J. Derome, H. Ritchie, and H. L. Mitchell, 1996: A system simulation approach to ensemble prediction. *Mon. Wea. Rev.*, in press.

Kalnay, E., and Z. Toth, 1994: Removing growing errors in the analysis. Proceedings of the Tenth Conference on Numerical Weather Prediction, July 18–22, 1994, Portland, OR. AMS, p. 212–215.

Kalnay, E., Z. Toth, Z–X. Pu and S. Lord, 1996: Targeting weather observations to locations where they are most needed. WGNE Reasearch Activities in Atmospheric and Oceanic modeling, WMO–CAS/WCRP–JSC publication.

Leith, C. E., 1974: Theoretical skill of Monte Carlo forecasts. *Mon. Wea. Rev.*, **102**, 409–418.

Lorenz, E. N., 1963: Deterministic non–periodic flow. *J. Atmos. Sci.*, **20**, 130–141.

Lorenz, E. N., 1965: A study of the predictability of a 28–variable atmospheric model. *Tellus*, **17**, 321–333.

Lorenz, E. N., 1969: The predictability of a flow which possesses many scales of motion. *Tellus*, **21**, 289–307.

Molteni, F., R. Buizza, T. N. Palmer, and T. Petroliagis, 1995: The ECMWF ensemble system: Methodology and validation. *Q. J. R. Meteorol. Soc.*, in press.

Reynolds, C. A., P. J. Webster, and E. Kalnay, 1994: Random error growth in NMC's global forecasts. *Mon. Wea. Rev.*, **122**, 1281–1305.

Snyder, C., 1996: Summary of an Informal Workshop on Adaptive Observations and FASTEX. *Bull. Amer. Meteorol. Soc.*, in press.

Toth, Z., and Kalnay, E., 1993: Ensemble Forecasting at the NMC: The generation of perturbations. *Bull. Amer. Meteorol. Soc.*, **74**, 2317–2330.

Toth, Z., and Kalnay, E., 1996a: Ensemble forecasting at NCEP and the breeding method. *Mon. Wea. Rev.*, under review.

Toth, Z., and E. Kalnay, 1996b: Climate ensemble forecasts: How to create them? *Idojaras*, in press.

Tracton, M. S. and E. Kalnay, 1993: Ensemble forecasting at NMC: Operational implementation. *Wea. Forecasting*, **8**, 379–398.

VARIABLE	HEIGHT in hPa (or m)
Geopotential height (Gph)	1000
Gph	700
Gph	500
Gph	250
Winds (U,V)	10 m
U,V	850
U,V	500
U,V	250
Temperature (T)	2 m
T	850
Relative humidity	700
Mean sea level pressure	
Total accumulated precipitation	

TABLE 1. The list of variables that are available for the global ensemble from NCEP through anonymous ftp (nic.fb4.noaa.gov, /pub/ens).

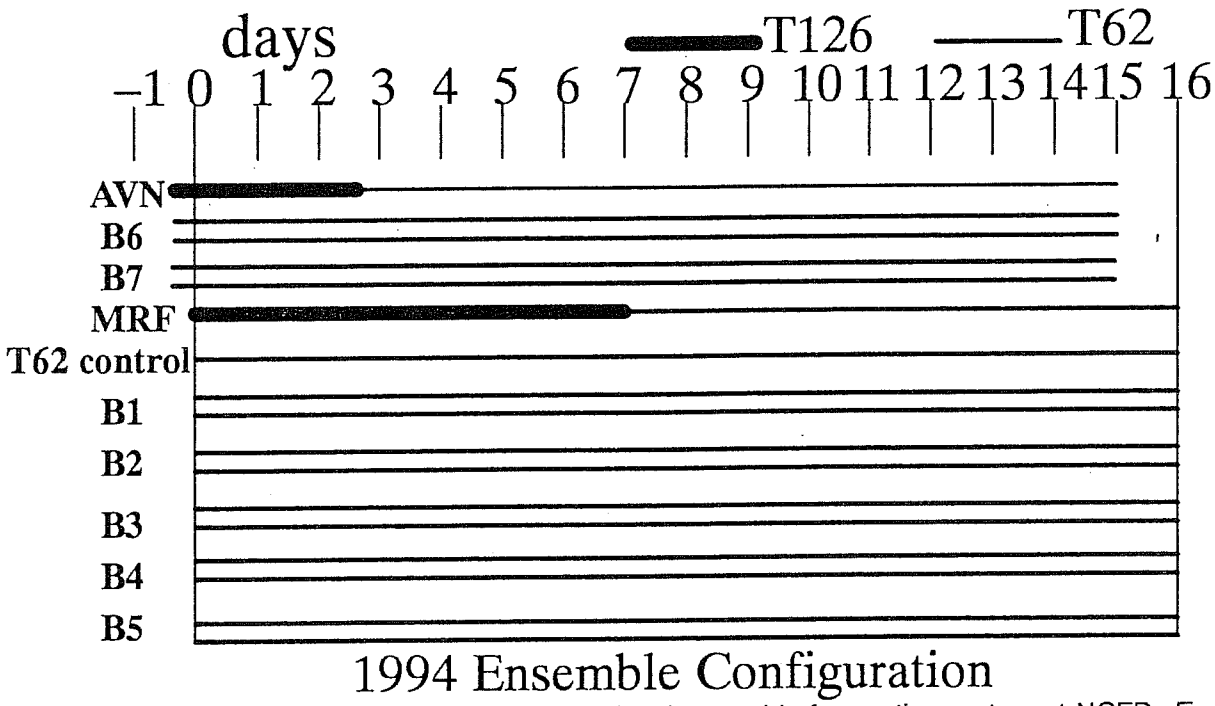

Fig. 1.: Schematic of the configuration of the operational ensemble forecasting system at NCEP. Each horizontal line represents a numerical forecast. High resolution, T126 forecasts are marked with heavy lines while the other forecasts are run at T62 resolution. Note that at 00Z there are two control forecasts, one started at T126 resolution and then truncated to T62 at day 7, while the other started at a T62 truncated resolution. At 12Z, the high resolution control is truncated after 3 days of integration. Pairs of perturbed forecasts based on the breeding method are marked as B1–B7. (From Toth and Kalnay, 1996.)

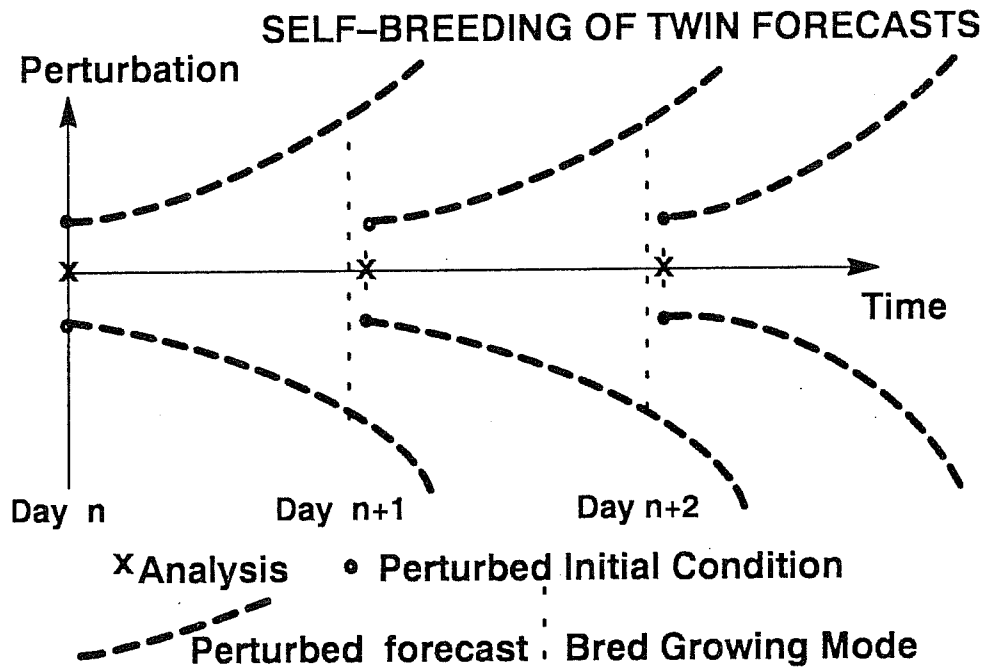

Fig. 2: Schematic of a self contained breeding pair of ensemble forecasts. Note that breeding is part of the extended ensemble forecasts and that the creation of efficient initial ensemble perturbations requires no additional computing resources beyond that needed to run the forecasts themselves. (From Toth and Kalnay, 1996.)

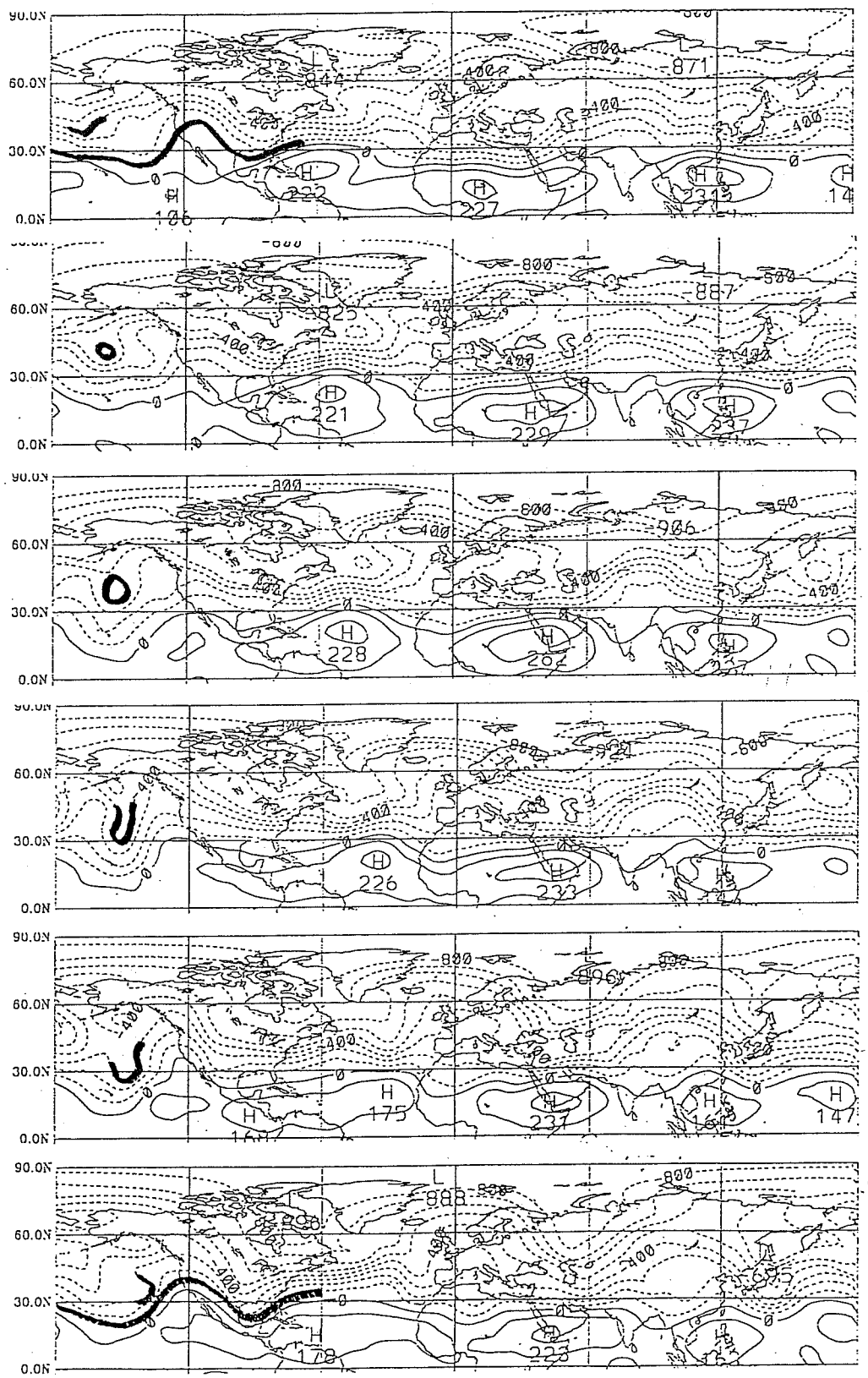

Fig. 3: 500 hPa streamfunction field (labels sclaed by 10^{-5}) for 1993 March 8–13, 00Z (from top to bottom).

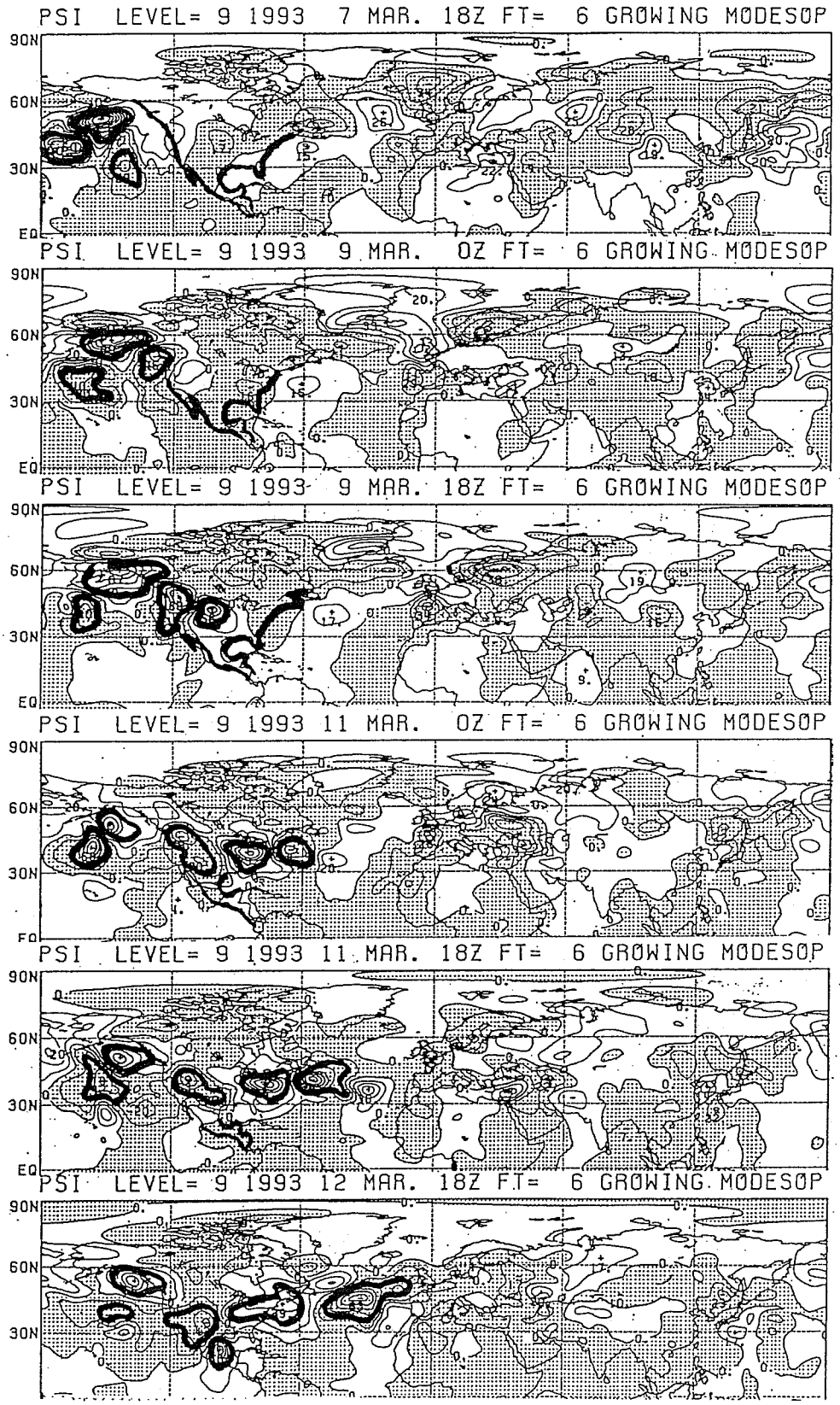

Fig. 4a: As in Fig. 3 except for operational bred perturbations (labels scaled by 10^{-4}). For 1993/03/09 and 11, the valid time is 06Z .

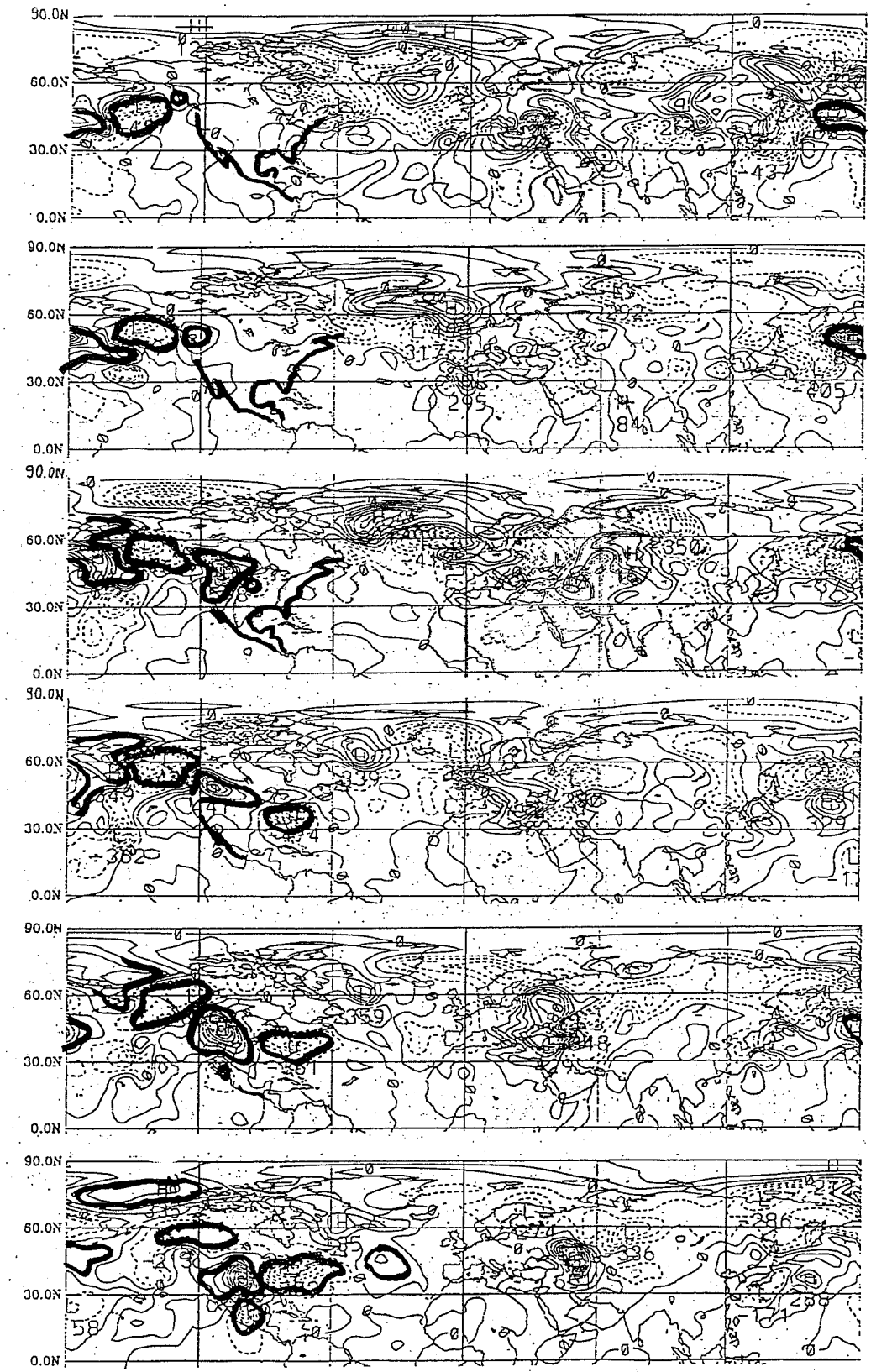

Fig. 4b: As in Fig. 4a except for perturbations from an independently run parallel breeding cycle (started with a different initial perturbaiton; all maps are for 00Z).

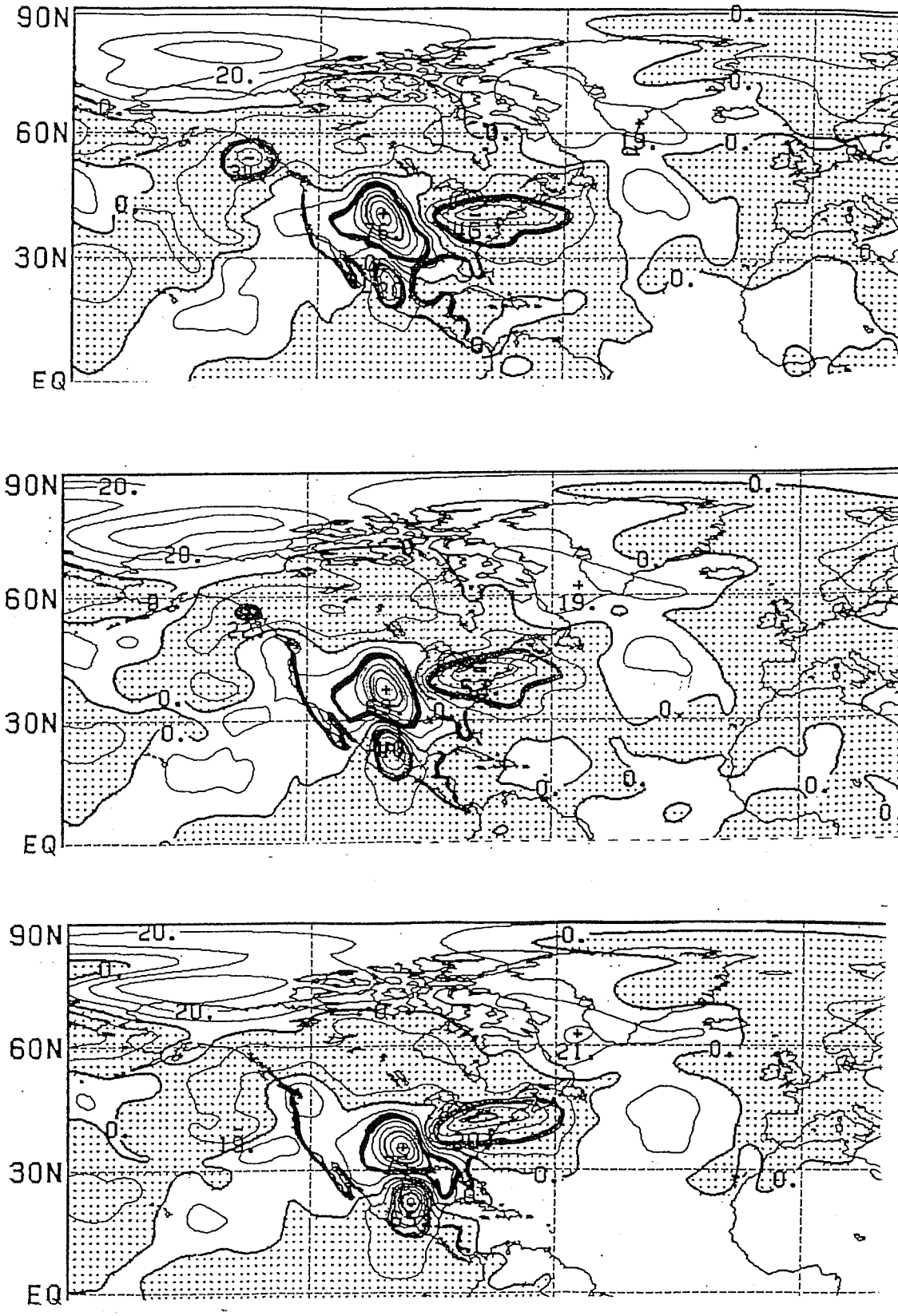

Fig. 5a–c: As Fig. 4a, except for 6–hr intervals (from top to bottom) between 1993/03/12/18Z and 1993/03/13/06Z.

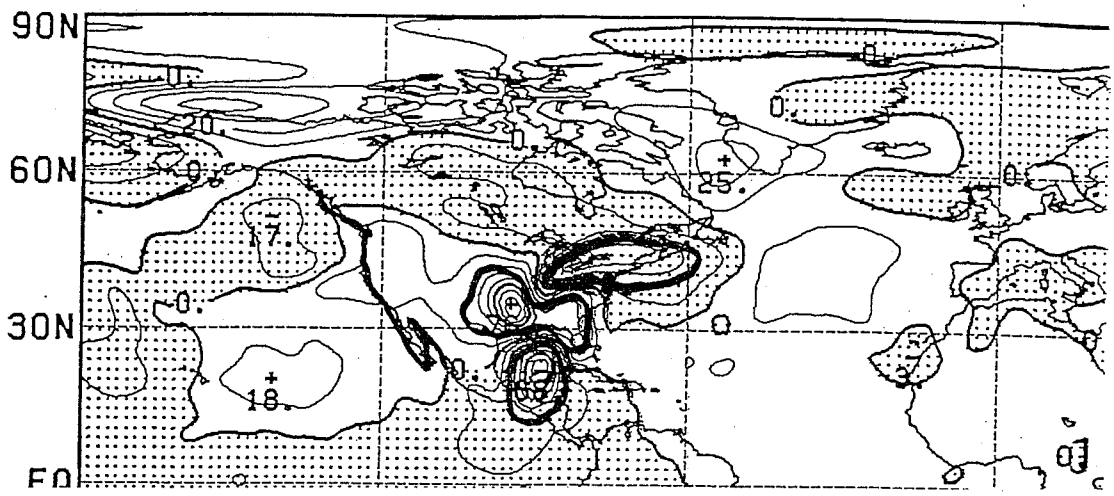

Fig. 5d: As Fig. 5a–c, except for 1993/03/13/12Z.

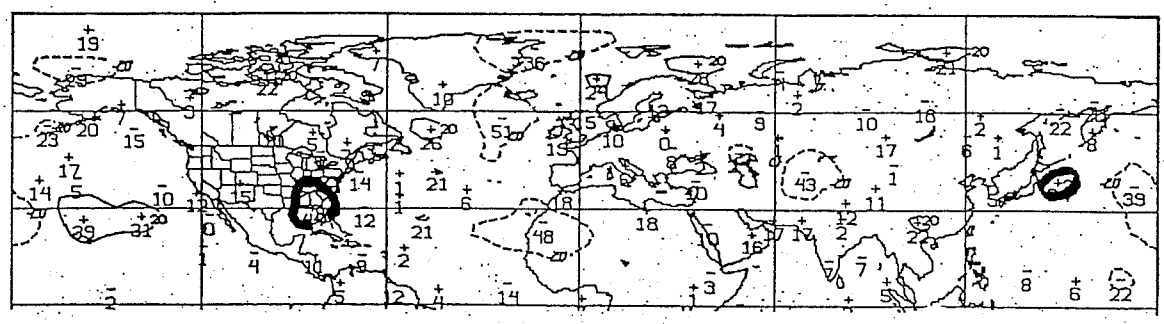

Fig. 6: 500 hPa height error field for operational 12–hour forecast started 1993/03/13/00Z.

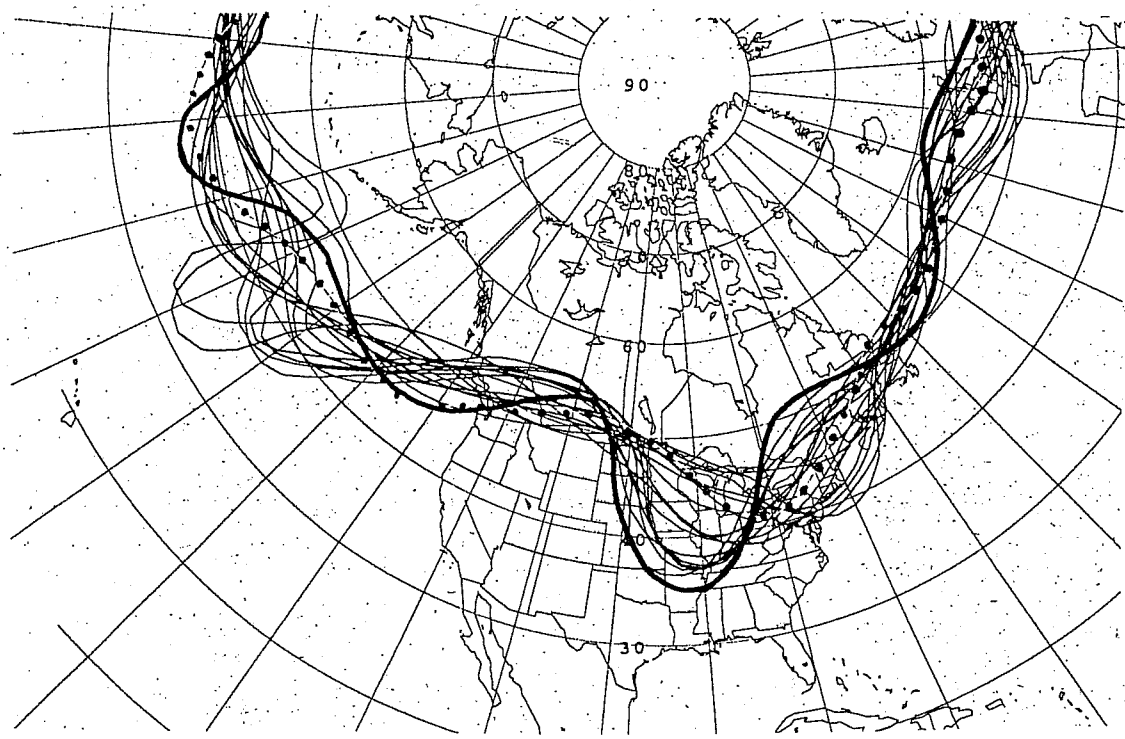

Fig. 7: 5640 m contour line of 500 hPa height field from all 17 108–hour lead time ensemble forecast members verifying 1995/10/20/12Z. The dotted line marks the high resolution control forecast (MRF) and the heavy solid line is the verifying analysis.

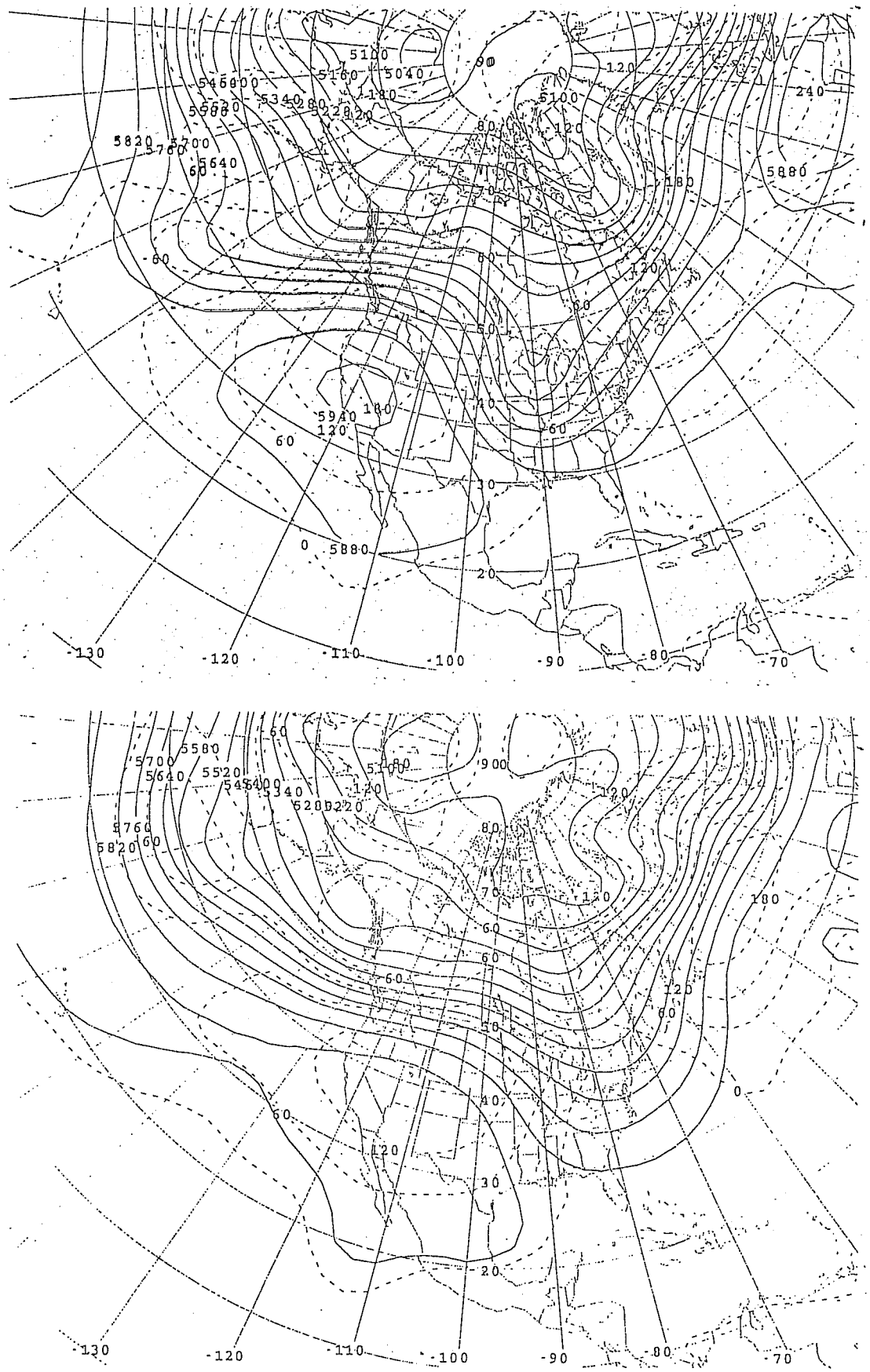

Fig. 8: Leading two 500 hPa height clusters (from top to bottom) for 108–hour lead time ensemble forecasts, valid at 95/10/20/12Z. Dashed lines indicate anomalies from climate mean.

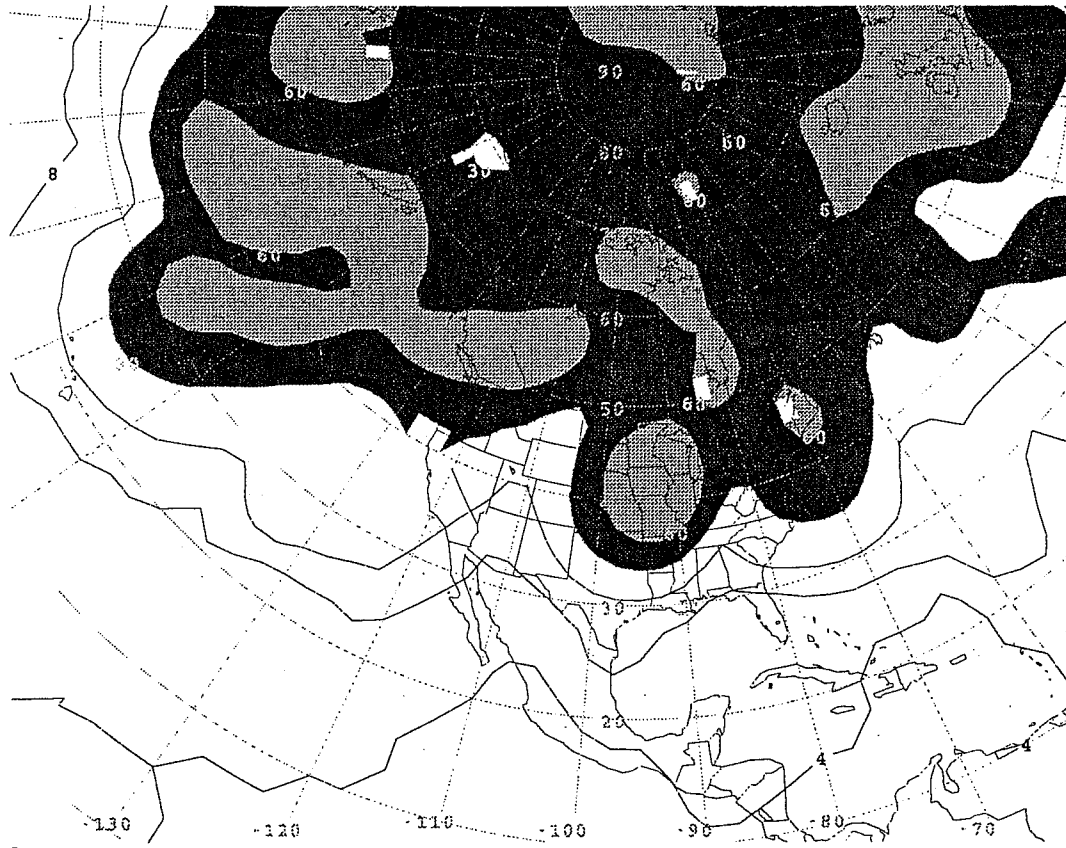

Fig. 9: Spread of ensemble members (m) around the ensemble mean at 500 hPa height for 108–hour lead time, valid at 95/10/20/12Z.

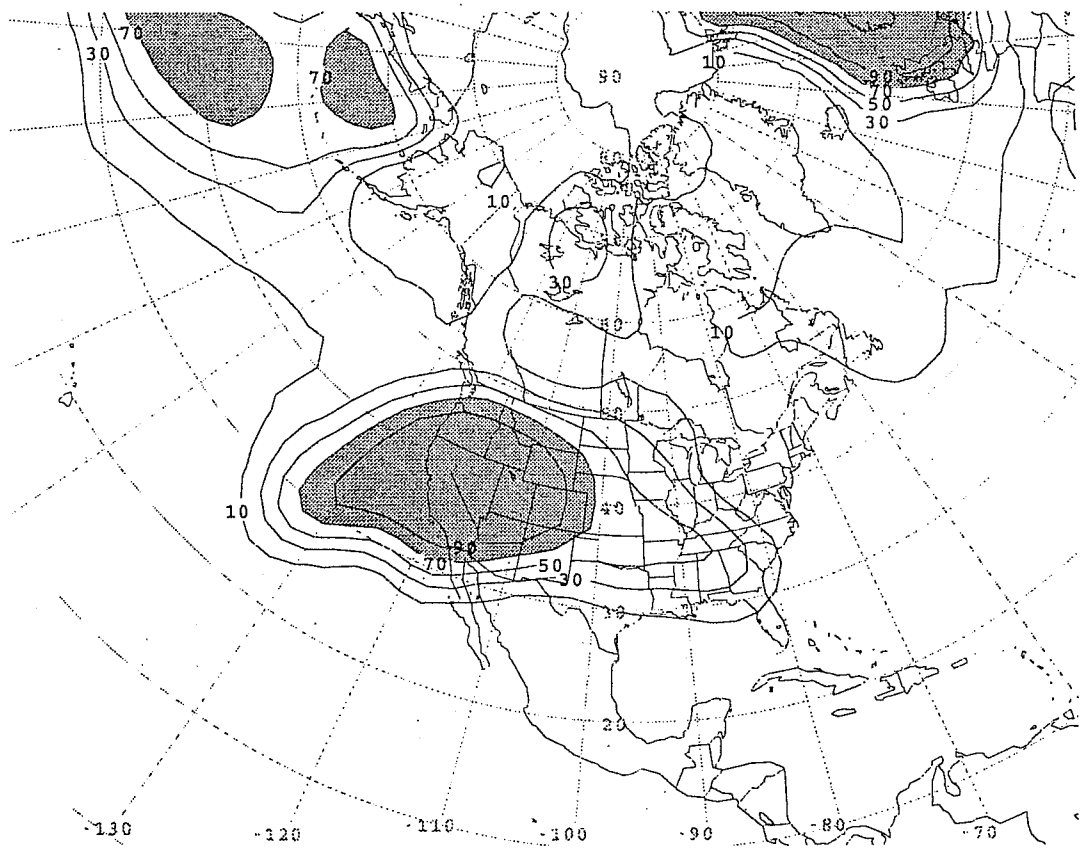

Fig. 10: Probability that 500/1000 hPa relative topography anomaly (from climate mean) will be above 60 m. Based on 120–hour ensemble forecasts valid at 95/10/21/00Z.

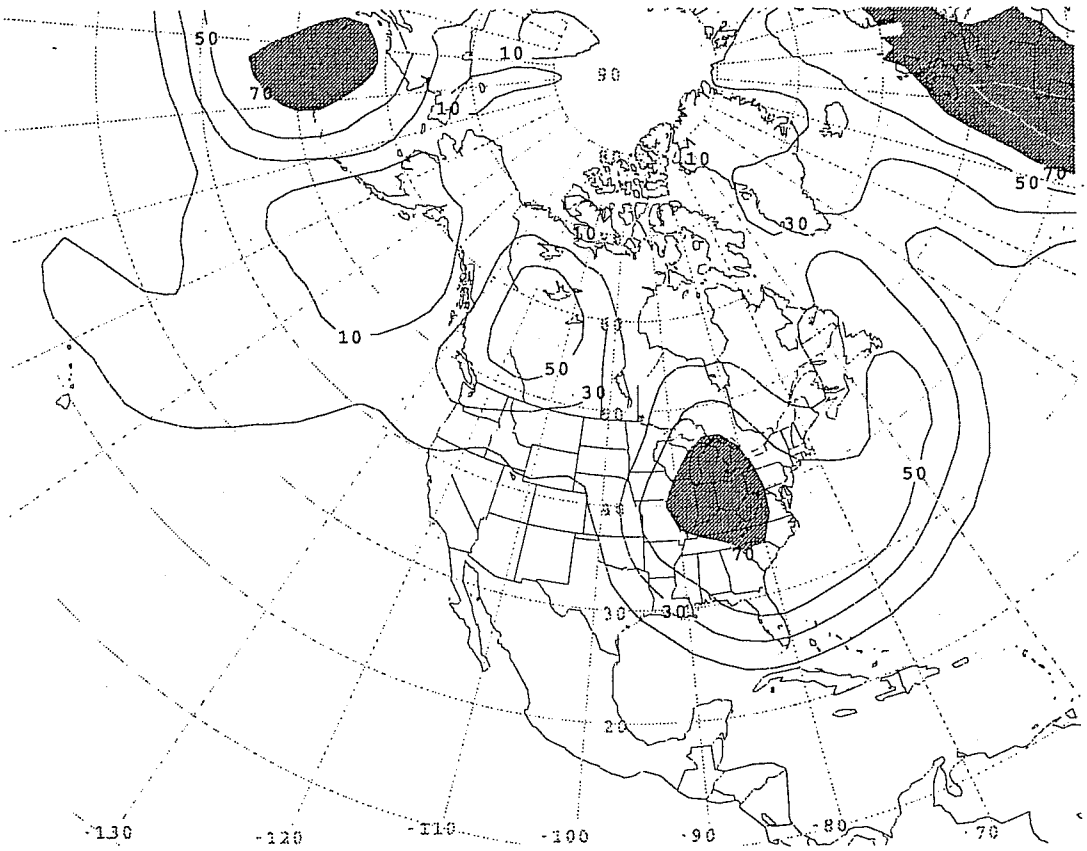

Fig. 11: Same as Fig. 10 except for probability of 500 hPa height tendency being below –30 m.

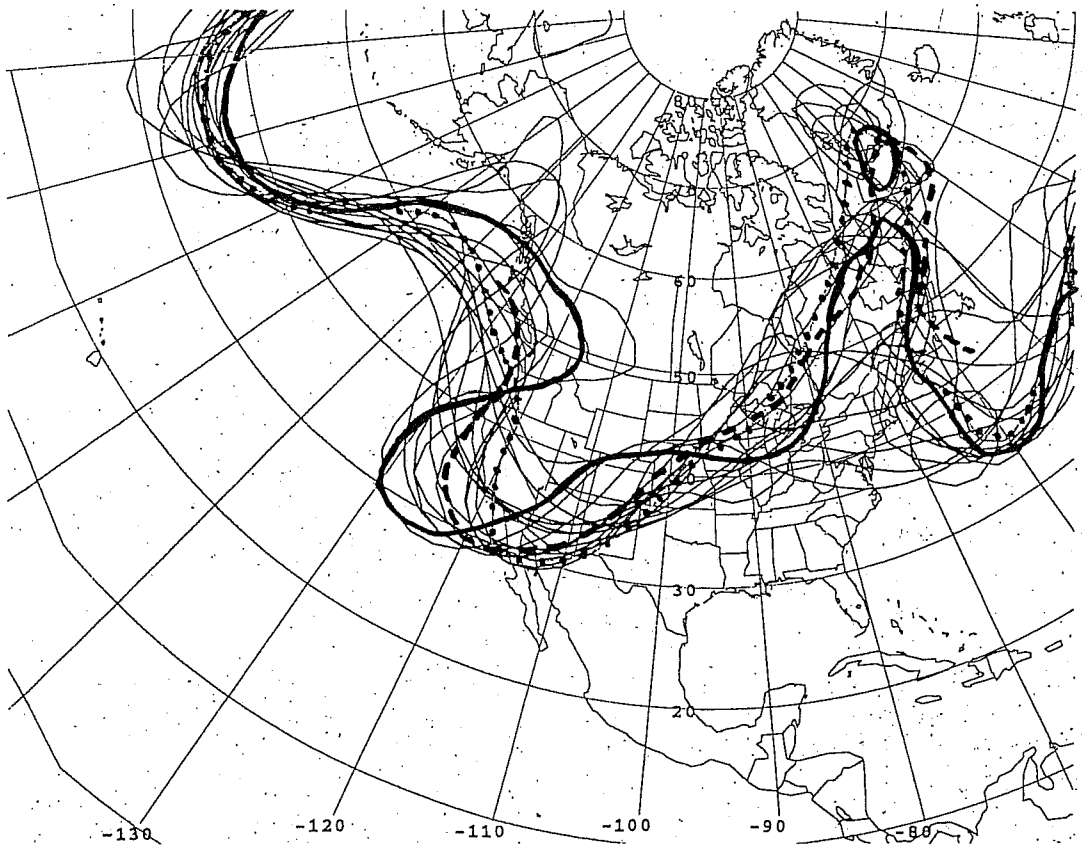

Fig. 14: Same as Fig. 7 except for 108–hour forecasts verifying on 95/05/10/12Z. Also included is the 120–hour ECMWF operational forecast valid at the same time (heavy dashed line).

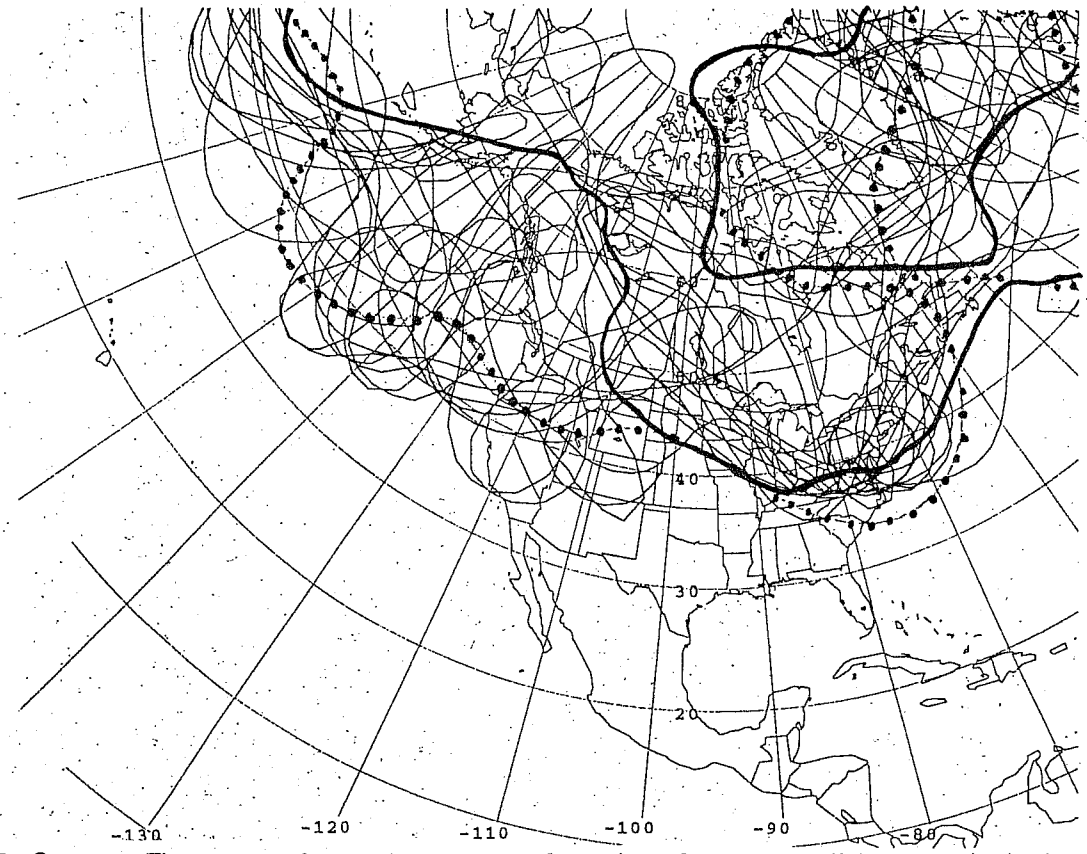

Fig. 12: Same as Fig. 7 except for 5520 m contour of 252–hour forecasts verifying on 95/04/25/12Z.

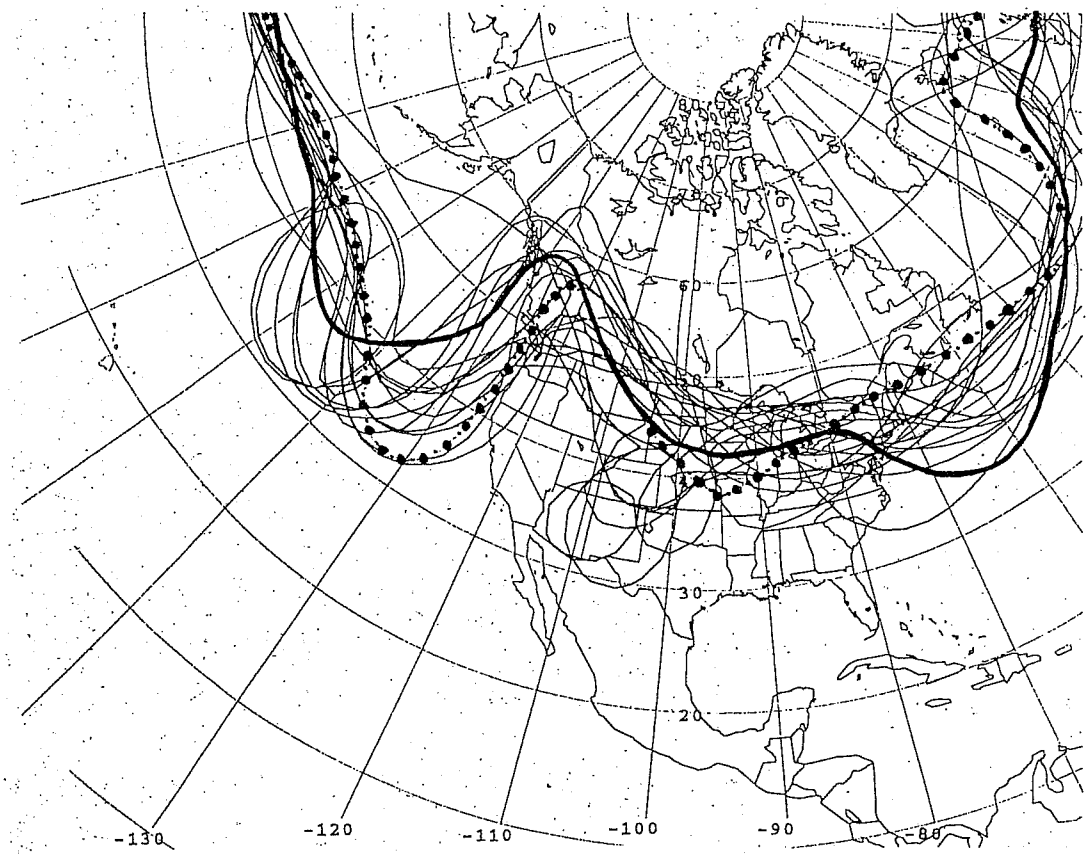

Fig. 13a: Same as Fig. 12 except for forecasts verifying on 95/04/03/12Z.

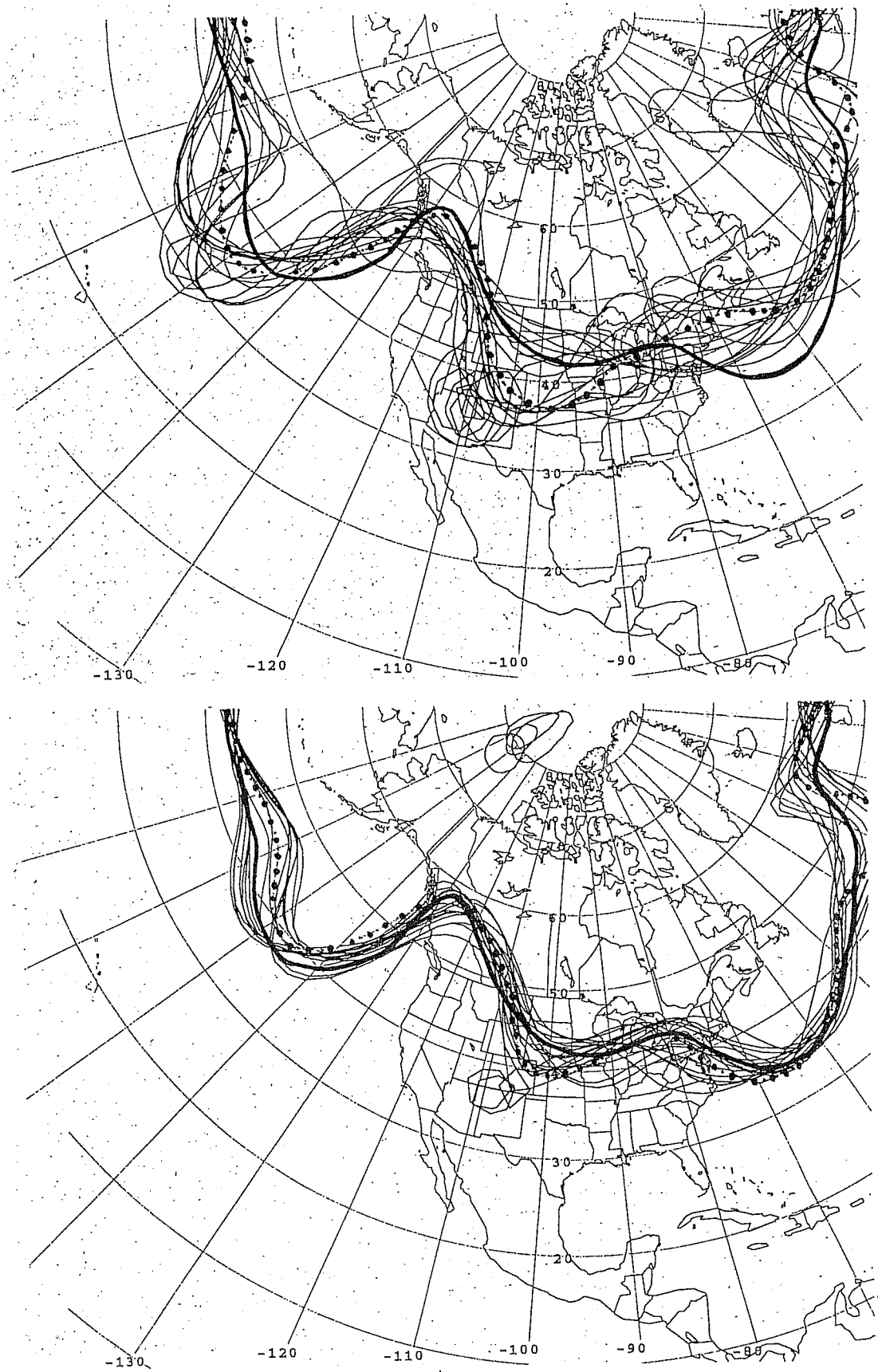

Fig. 13b,c: Same as Fig. 13a except for 156–hour (top) and 84–hour (bottom) ensemble forecasts verifying on 95/04/03/12Z .

Fig. 14 is placed after Fig. 11.

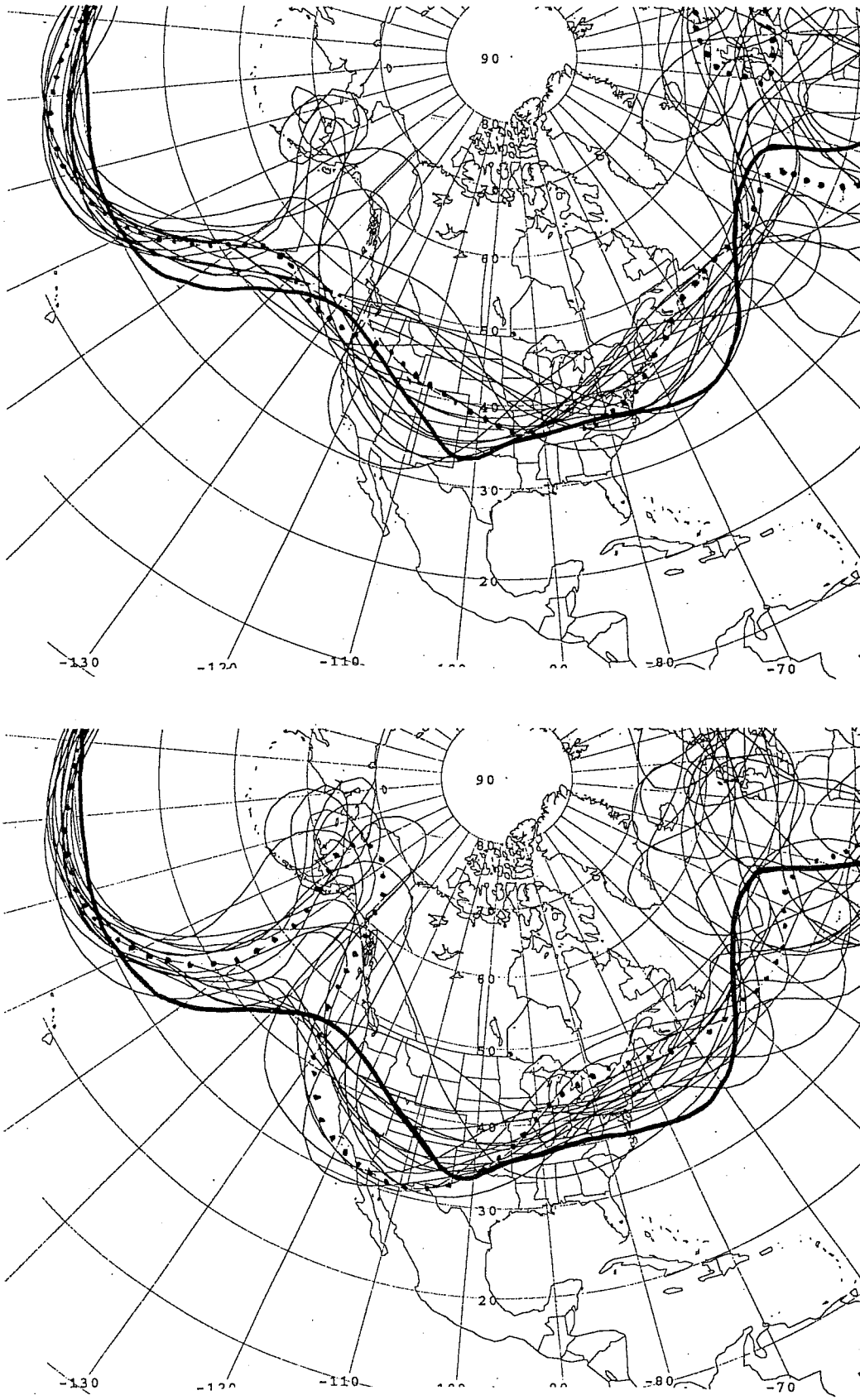

Fig. 15: Same as Fig. 7 except for 228–hour (a, top) and 204–hour (b, bottom) forecasts verifying on 95/12/08/12Z.

Fig. 16 is placed after Fig. 18.

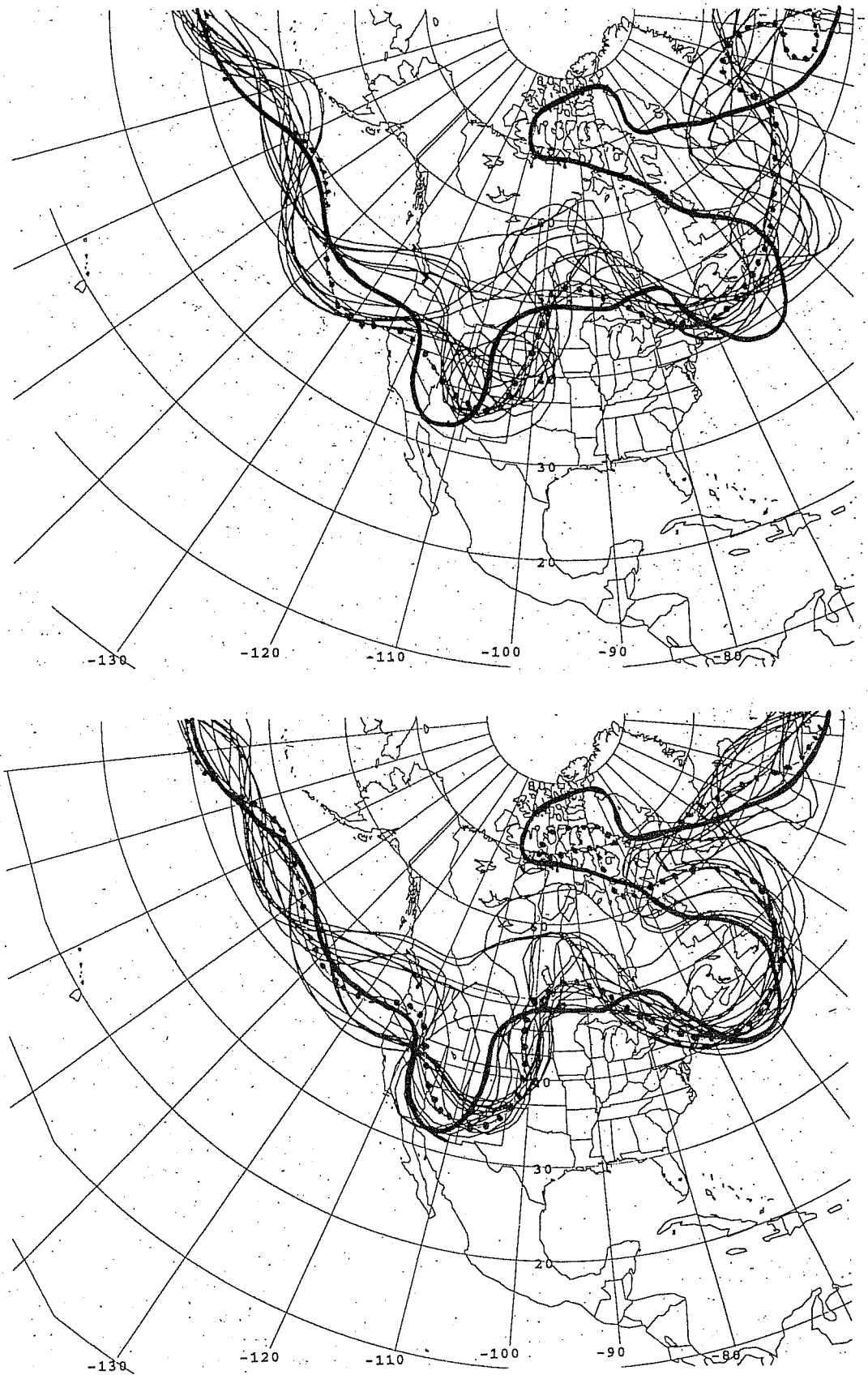

Fig. 17: Same as Fig. 12 except for 132–hour (a, top) and 108–hour (b, bottom) forecasts verifying on 95/04/17/12Z.

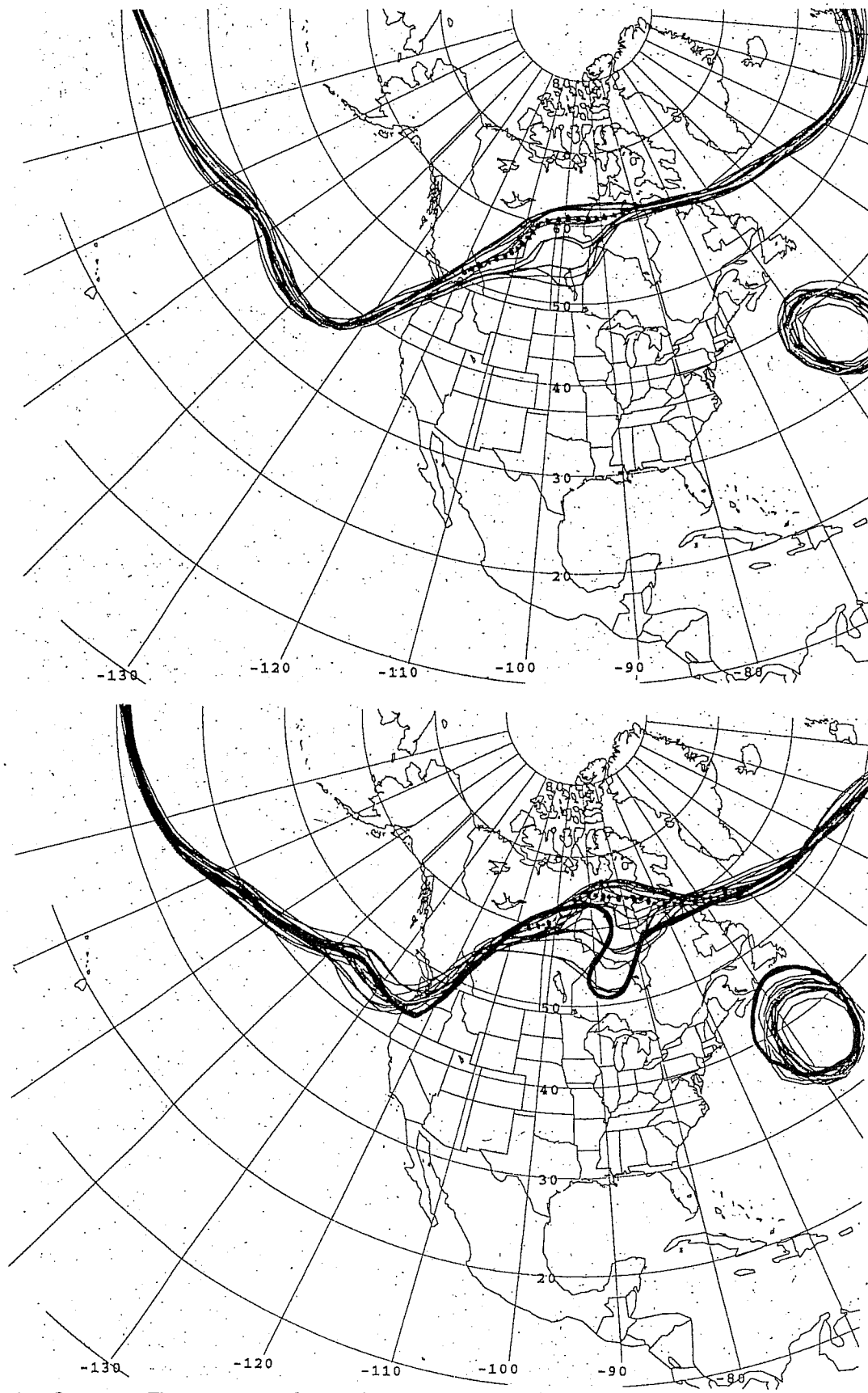

Fig. 18a,b: Same as Fig. 12 except for 12–hour (top, verification is not shown) and 36–hour (bottom) forecasts initiated at 95/03/14/00Z.

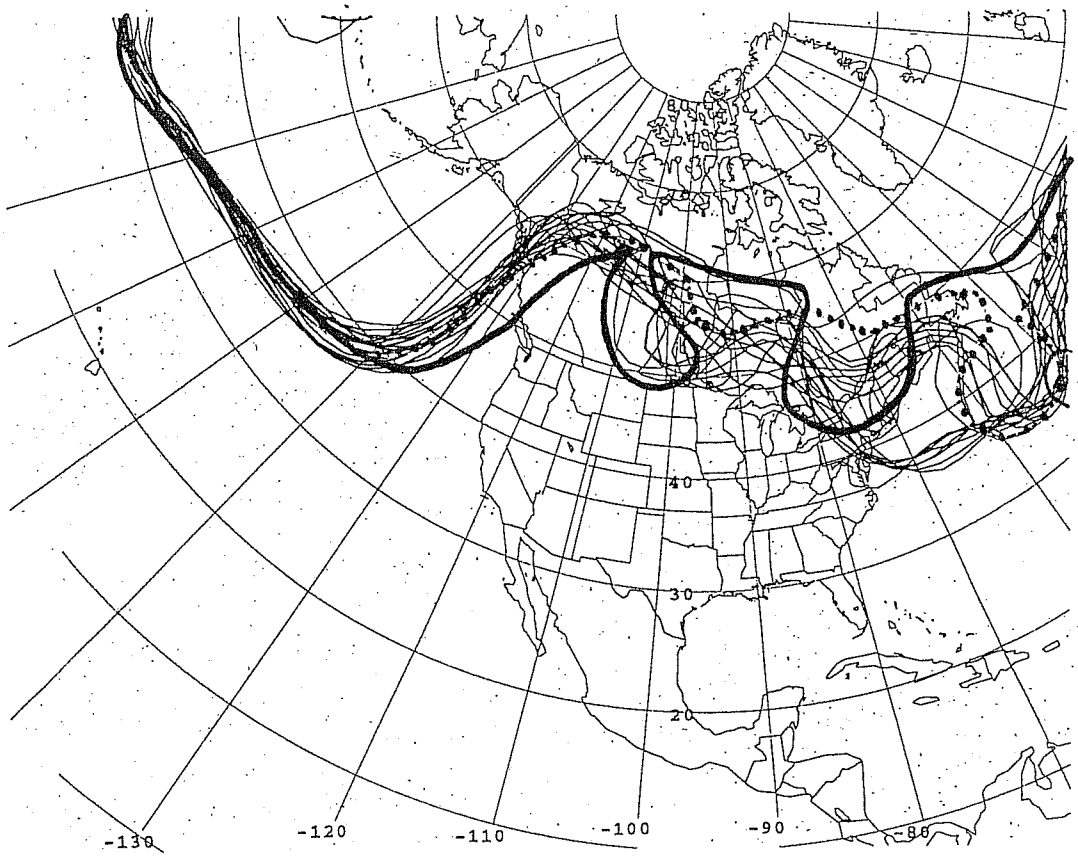

Fig. 18c: Same as Fig. 18a,b except for 84–hour forecasts initiated at 95/03/14/00Z.

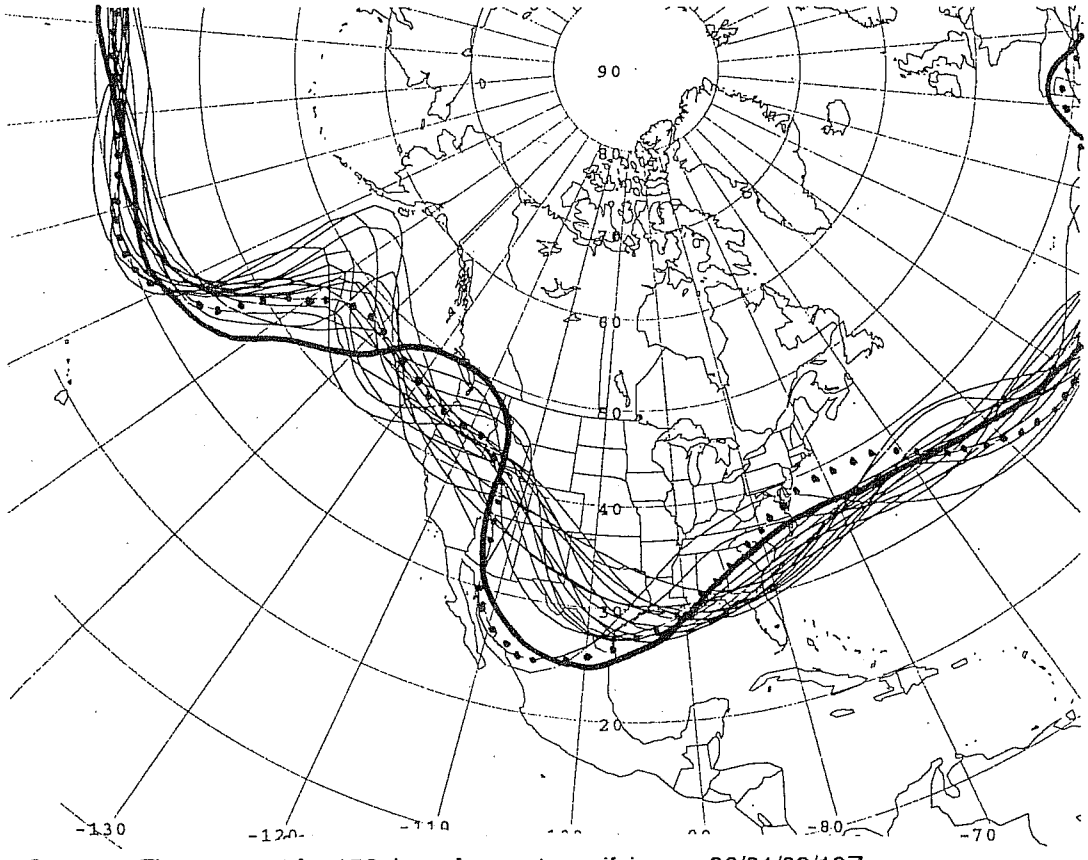

Fig. 16: Same as Fig. 7 except for 156–hour forecasts verifying on 96/01/02/12Z.