



WESTERN REGION TECHNICAL ATTACHMENT

NO. 98-14
APRIL 7, 1998

AN EVALUATION OF NCEP GUIDANCE FOR CALIFORNIA JANUARY 31, 1997 TO FEBRUARY 8, 1998

Jon Mittelstadt - WRH, SSD, Salt Lake City, UT

[Editor's Note: This Technical Attachment was written as one component of a Western Region Storm Survey. Several of the figures referenced here are only available via the Western Region Web page, <http://nimbo.wrh.noaa.gov/wrhq/TA.html>]

Introduction

An El-Nino regime with an energetic jet stream crossing the Pacific Ocean brought a series of strong storms and widespread precipitation to the west coast over the period January 31 to February 8. Past storm surveys have adequately documented the performance of coastal WSR-88Ds. No substantial changes were made to the WSR-88Ds to modify their performance.

Many users need long lead times to prepare for events. Various local and federal officials require up to several days to take corrective actions and pre-deploy personnel and other resources. Since NCEP models are the primary source of forecast guidance beyond 6 hours, this section of the storm survey focuses on model performance.

Methodology

Forecasts from the Navy NOGAPS and NCEP MRF, AVN, NGM, Eta-48 and Eta-29 models, and MRF ensemble and Hydrometeorological Prediction Center (HPC) Quantitative Precipitation Forecast (QPF) charts and discussions were examined. Several web-based resources greatly reduced the time spent gathering data. They are listed below. In order to quickly, yet carefully, evaluate the various models and HPC guidance, specific parameters and forecast problems were selected for the short, medium and extended ranges. The forecast problems were chosen primarily on the basis of their relevance to a public forecast. The model forecasts were evaluated at 0000Z and 1200Z.

Short-Range Guidance

1. 24-hour model and HPC QPF

Note: Except for the Eta-29, the QPFs are for a 24-hour period from 12 to 36 hours after the 0000 UTC model initialization time. The Eta-29 QPF is for a 24-hour period from 9 to 33 hours after its 0003 UTC initialization time. HPC forecasts are compared with the model guidance that was available to HPC prior to issuance time.

Objective QPF Skill Scores

Figure 1(a) shows model biases for February 1998 calculated over a domain covering California, eastern Oregon, and eastern Washington (Baldwin 1998). The Eta-29 is the only model without a strong bias of under-forecasting amounts greater than 1 inch (25 mm). The NGM bias is large and starts for smaller amounts, i.e., amounts greater than about .5 inches (12.5 mm). Equitable Threat skill scores for the same domain are displayed in Fig. 1(b). The Eta-29 had the highest skill, followed by the AVN and then the Eta-48. The skill of the NGM is significantly lower than the other models. These scores are fairly typical of NCEP model skill scores in general.

Forecast Problems

Q1. Will the storm produce a large area of excessive precipitation tomorrow (i.e., 12 to 36 hour forecast)?

Model and HPC QPFs were ranked from highest to lowest based on areal coverage and precipitation amount over California and compared to the ranking of observations. The forecast and observed rankings compare well, indicating forecasters received adequate guidance from NCEP models with respect to this forecast problem. The one exception was February 6, a day when HPC and the models over-forecast the amount and areal coverage of precipitation.

Q2. Were QPF amounts as large as observed amounts?

No. It is important to remember that QPF forecasts are for areas and not specific points. Thus, QPF maxima will generally be lower than observed maxima. The Eta-29 is the only model that did not consistently under-forecast the larger precipitation events. A typical example can be seen in Fig. 2 where the east-west mountains north of Los Angeles (Santa Ynez, San Gabriel and San Bernardino) recorded many observations greater than 2 inches and two observations greater than 4 inches (Fig. 2a). The highest corresponding model QPF was from the Eta-29 (2 inches) and Eta-48 (1 inch). For most heavy precipitation events like this one, HPC was successful in forecasting amounts higher than the models, 3 inches for this example (Fig. 2g).

Q3. Do QPFs capture detailed orographic effects?

The Eta-29 is the only model with adequate resolution to capture some of the orographic effects of California and place precipitation maxima in the correct locations. For example, the Eta-29 QPF (Fig. 2c) correctly highlights Mt. Shasta, the Sierra Nevada, and the east-west mountains north of Los Angeles. The other models, due to inadequate terrain resolution, tend to place maxima in unrealistic locations such as the central valley. However, forecasters adjust model QPF based on their knowledge of both actual and model terrain (Martin 1996). HPC QPF forecasts also reflected these orographic effects (e.g., Fig. 2g). The next question addresses the forecaster's ability to adjust model QPF to actual terrain.

Q4. Since forecasters attempt to adjust model QPF to actual terrain, one can ask, "Do the QPFs place precipitation maxima in *approximately* the correct locations"?

Twenty-four hour rainfall charts created at NCEP and the University of Utah (Fig. 2a and 2b) were evaluated to identify areas of maximum observed rainfall. Figure 3 is a table summarizing model performance relative to these charts. The models had enough skill to forecast some but not all of the important locations of precipitation. For example, Fig. 2 shows that none of the models nor HPC forecast the precipitation along the coast south of Monterey. False alarms also limit the usefulness of model QPFs -- the Eta models seemed to over-forecast the southern Sierras and the AVN and MRF seemed to over-emphasize the Los Angeles Basin. The poor performance of the NGM was obvious. The NGM QPF rarely provided useful guidance with respect to location. HPC sometimes improved on the models with respect to location, for example, for February 5, HPC placed a QPF contour covering San Diego, an area of precipitation not forecast by the models. However, HPC seemed to use the Eta-29 as a "first-guess" for most forecasts, and as mentioned above, the Eta-29 did not always provide the best guidance. The fact that HPC forecasts were no better than the best model in 5 of 10 cases (Fig. 3) shows the difficulty of choosing the "model of the day" with respect to this forecast problem.

QPF Summary

Overall, the subjective evaluation was similar to the objective scores, i.e., the Eta-29 had the most skill, followed by the AVN and Eta-48, and the NGM had the lowest skill. The AVN model is coarser than the Eta models, so its skill could be due to its global domain. Mesinger (1998) for example, demonstrated a slight loss of skill in QPFs longer than 24 hours caused by the limited domain of the Eta-29 model.

Twenty-four hour QPFs were successful in forecasting heavy vs. light statewide precipitation. Useful but limited skill was also evident in forecasting the locations of heavy precipitation. The Eta-29 clearly produced the most realistic forecast in the complex terrain.

A Comparison of NCEP and University of Utah Observed Rainfall Charts

Two types of 24-hour total rainfall charts were evaluated: (1) from NCEP, and (2) from the University of Utah (Figs. 2 a and b). The NCEP charts are derived from an algorithm that uses radar estimates to fill in values where observations are sparse. (Detailed information is available via the multi-sensor web page listed below.) The NCEP charts show values that represent areal coverages, unlike the University of Utah charts that display the raw NWS RFC rainfall observations. Radar precipitation estimates during the cold season are often too low in the West, primarily because most radars are located at mountain locations where they tend to overshoot winter storms. Since rainfall observations are also sparse in the West, the NCEP algorithm will tend to under-represent precipitation. A comparison of the two charts shows that the NCEP charts indeed do tend to eliminate too many of the large observed amounts in Western Region, for example over the east-west mountains north of Los Angeles in Fig. 2b.

Why are these differences important? Since the values shown on the NCEP charts have been used for model validation, these errors could be detrimental in efforts to improve model skill over the West. For example, the Eta-29 QPF (Fig. 2c) shows more skill for the Santa Ynez mountains when compared with the University of Utah chart (Fig. 2a) than when compared with the NCEP chart (Fig. 2b). Further, NCEP plans to use the multi-sensor data during model assimilation to improve QPF accuracy (Lin et al., 1998). The NCEP charts are experimental but are posted to the web page listed below. NCEP scientists are aware of the weaknesses of the current multi-sensor approach and are experimenting with new approaches, including the use of a complex cloud-model and 4DVAR assimilation (Mike Baldwin, personal communication).

2. 36-hour (33-hour for the Eta-29) model forecasts of Mean-Sea-Level Pressure (MSLP) compared to surface observations

Forecasting surface fronts is an important problem because of the role they play in the timing and intensity of precipitation and strong onshore/offshore winds.

Forecast Problems

Q1. How accurate were 36-hour forecasts of low-pressure centers (position and depth) and of warm and cold fronts?

Q2. How accurate were 36-hour forecasts of MSLP gradients along the coast of northern and southern California?

Results

- In general, the models provided useful guidance for MSLP and gradients.
- The Eta-29 model showed the most skill followed by the NOGAPS (e.g., Fig. 4). This is a surprising result because the NOGAPS model (80km) has coarser resolution than the Eta-48 (48km). In general, these two models placed the surface features within 100 miles of their correct locations. This magnitude of error would certainly be noticeable to an emergency manager using NWS forecasts. However, forecasters use observations and satellite imagery to reduce the magnitude of these errors as forecast lead-time shortens. Eta-48 forecasts were frequently quite different from Eta-29 forecasts (e.g., Fig. 4).
- The Eta-29 was best at forecasting the position and depth of low pressure centers in 8 of 12 forecasts studied and it placed second in the other 4 cases.
- The NOGAPS was the best model at handling lows as they traversed the mountains and either decayed or reformed over the Great Basin. For example, in Fig. 4, the NOGAPS was the only model to forecast a deep enough low over the Great Basin and along the California/Arizona border.
- In roughly half the cases, a MSLP ridge over southern California created a much weaker gradient there than over northern California. Fig. 4 shows this distinction but not as well as other cases. The Eta-29 and NOGAPS again demonstrated superior 36-hour skill in forecasting this distinction.

Medium Range

1. NOGAPS, AVN (48 and 72-hour) and Eta (48-hour) forecasts of 500mb vorticity compared to model analyses.

Q. Did the models correctly forecast the timing of shortwaves traveling through the mean flow?

During the evaluation period, several major rainfall episodes came from dynamics associated with shortwaves rotating through synoptic-scale troughs. Eta, AVN and NOGAPS analyses were examined to determine the major shortwaves, i.e., those captured by all three models.

Results

- In most cases, the models forecast synoptic-scale trough positions fairly well. However, no model gave reliable guidance for 48 and 72-hour forecasts of shortwaves. Fig. 5 shows poor model forecasts of an energetic shortwave west of the California/Arizona

border. Splitting systems and the unusually strong southern-branch jet stream across the Pacific Ocean probably led to the poor skill of the models (beyond 24-hr) with respect to the timing of shortwaves.

- The differences between the initial analyses of vorticity were also large. This reflects, in part, the “noisy” nature of the vorticity field, but also highlights the difficulty of initializing models over the data-sparse Pacific Ocean.

Extended Range

1. MRF and NOGAPS 120-hour Forecasts of 500-mb Heights Compared to Model Analyses

Q1. Did the models place the large-scale trough and ridge axes in the correct locations?

The models showed useful skill with respect to this problem February 1 to February 5 including during the largest event as a negative-tilt trough came onshore February 3 (Fig. 6). Both models did poorly the last three days of the evaluation on February 6-8. An ongoing NCEP/WR evaluation of these models has shown: (1) this variability in skill is typical at this forecast range; and (2) the MRF has, for winter 97/98, been better at forecasting changes in the large-scale pattern at the 120-hour range. Figure 7 indicates the MRF and UK models scored better than NOGAPS but worse than the ECMWF model over the period December 1, 1997 through February 28, 1998.

Q2. Did the models forecast the amplitude of the troughs as they reached California?

No. At 120 hours both models tended to under-forecast amplitude.

2. MRF ensemble 132-hour 500mb height spaghetti charts

Note: The MRF ensemble consists of 17 forecasts. Each forecast begins with slightly different initial conditions. The spread of the 17 forecasts can be used to forecast uncertainty, i.e., if the “spread” is large then forecast certainty is considered small.

Q1. Did the verification fall within the spread of the 17 ensemble forecasts?

Yes, in 8 of 10 cases. Statistically, one should expect the verification to fall outside the spread of a 17-member ensemble in about 11% of the cases. However, due to model errors, the verification will fall outside the spread more often than the statistical estimate. For the MRF ensemble, this occurs about 25% of the time (Mittelstadt, 1995). This is close to the ratio (2 of 10) that was observed here.

Q2. Was the ensemble spread a useful indicator of the skill of the operational MRF forecast?

Yes. In general, skill was high and spread small prior to February 6 (e.g., Fig. 8a) and skill was low and spread large on and after February 6. Fig. 8(b) shows the ensemble spaghetti chart verifying on February 6, the MRF operational forecast (thin black line) is distant from the verifying contour (thick black line), the spread is large, and almost all the ensemble members are closer to the verification than the operational run.

Internet Resources

The following Internet web pages were used as a means to quickly and easily collect data for this model evaluation.

sgi62.wwb.noaa.gov:8080/scores (NCEP model QPF skill scores)
sgi62.wwb.noaa.gov:8080/verf/pcpgifs.html (NCEP model 24-hour QPFs and 24-hour total observed)
nic.fb4.noaa.gov:8000/research/gcp/hdpprec.html (Multi-Sensor products and information)
sgi62.wwb.noaa.gov:8080/ens/enshome.html (NCEP Ensemble Homepage)
sgi62.wwb.noaa.gov:8080/reanl2/wd22sl (AVN Model First Guess Difference Charts)
precip.fsl.noaa.gov/hourly_precip.html (RAWS hourly precip meteograms)

Summary

In general, as expected, NCEP model guidance was useful and reliable for one and two day forecasts. A notable exception was poor two-day forecasts of shortwaves. Model skill varied from good to very bad for medium and extended ranges. NCEP MRF ensemble output was a good indicator of uncertainty for extended range forecasts.

The higher resolution of the Eta-29 clearly led to improved QPF and MSLP forecasts. However, this model evaluation (and NCEP statistics) suggest that a global domain can have a positive impact over a limited domain for forecasts longer than 24 hours.

During the period studied the strengths of the models were:

- All the models provided good 36-hour MSLP forecasts. The Eta-29 was most accurate.
- NOGAPS 36-hour MSLP forecasts were nearly as good as Eta-29 forecasts and were better at forecasting lows as they traversed the Sierra Nevada Mountains.
- The MRF Ensemble was a reliable indicator of uncertainty for 132-hour forecasts of 500mb height.
- Model and HPC QPFs were good indicators of the intensity of precipitation on a statewide scale.
- HPC improved the model QPF by increasing the forecast amounts in the correct locations.

The weaknesses during the period were:

- In half the cases, HPC did not improve on the best "model of the day" in terms of forecasting the location of precipitation maxima (a very difficult thing to do!).
- The NGM 24-hour QPF charts were almost always very poor.
- The AVN, Eta and NOGAPS models struggled with 48-hour forecasts of energetic 500mb vorticity centers.

Acknowledgments

Thanks are due to Brett McDonald (University of Utah) for providing observed 24-hour rainfall total charts, Mike Eckert (HPC) for HPC charts and discussions, Pete Caplan (EMC) for medium-range model statistics, and Dean Hazen (NWSFO Pocatello) for archived gridded model data.

References

- Lin, Y., K. Mitchell, E. Rogers, M. Baldwin, 1998: Assimilation of Real-Time Multi-Sensor Hourly Precipitation Observations into the NCEP Eta Model. Preprints, 12th AMS Conference on Numerical Weather Prediction.
- Martin, G., 1996: A Dramatic Example of the Importance of Detailed Model Terrain in Producing Accurate Quantitative Precipitation Forecasts for Southern California. WR-Technical Attachment 96-07.
- Mesinger, F., 1998: Comparison of Quantitative Precipitation Forecasts by the 48- and 29-km Eta Model: An Update and Possible Implications. Joint Preprints, 12th AMS Conference on Numerical Weather Prediction or 16th Conference on Weather Analysis and Prediction.
- Mittelstadt, J.C., 1995: Introduction to Ensemble Forecasting. WR-Technical Attachment 95-29.

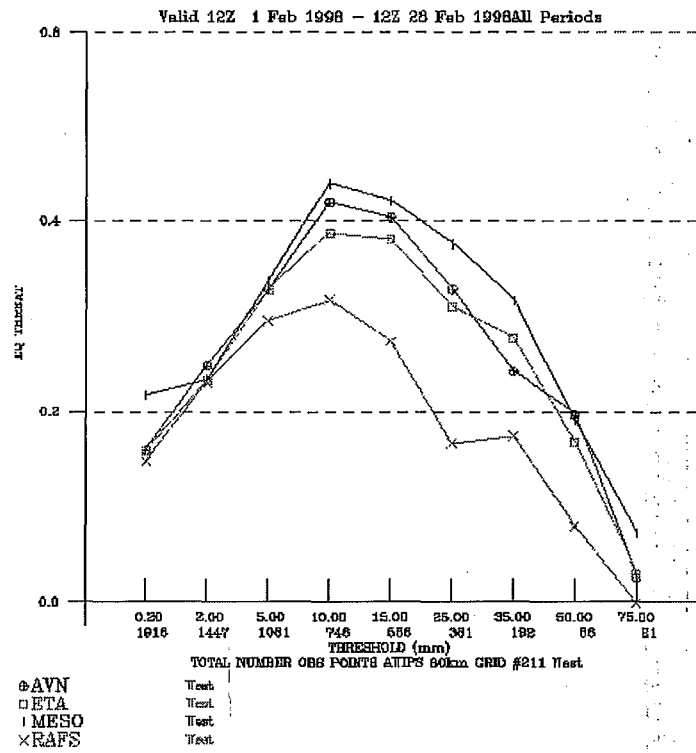
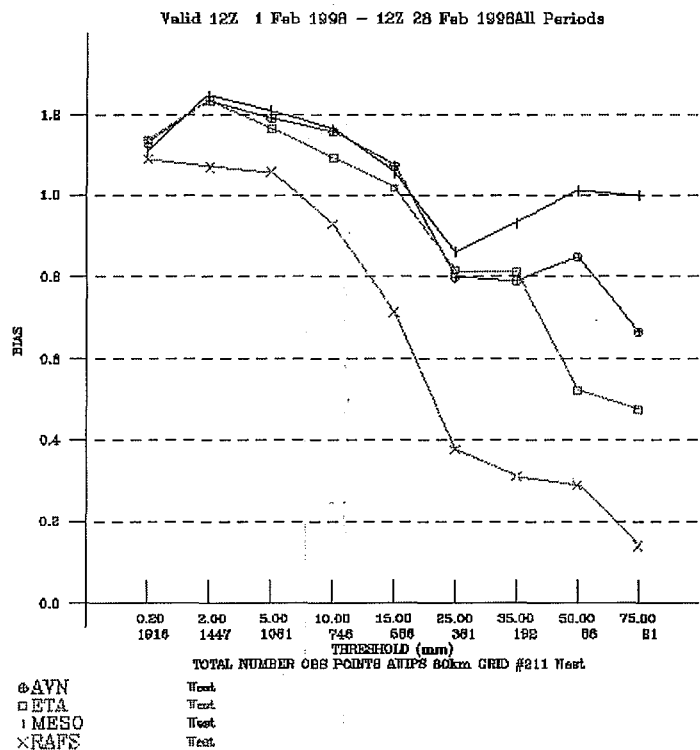
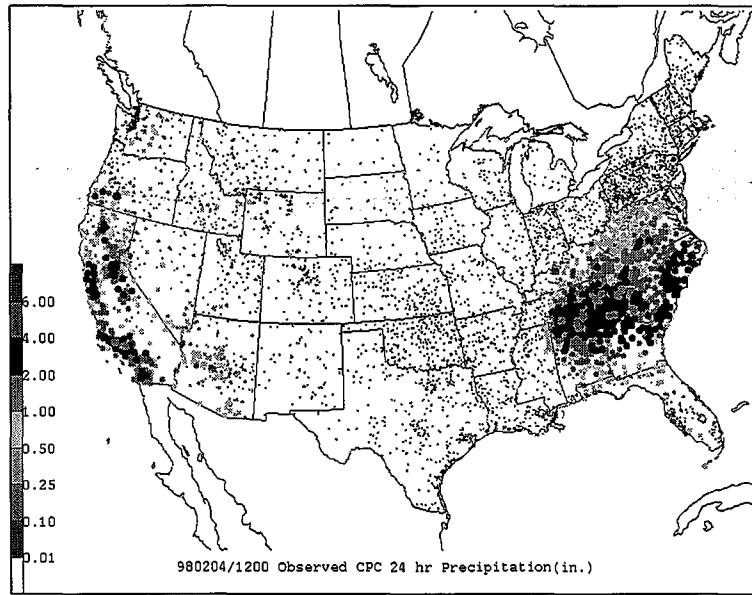


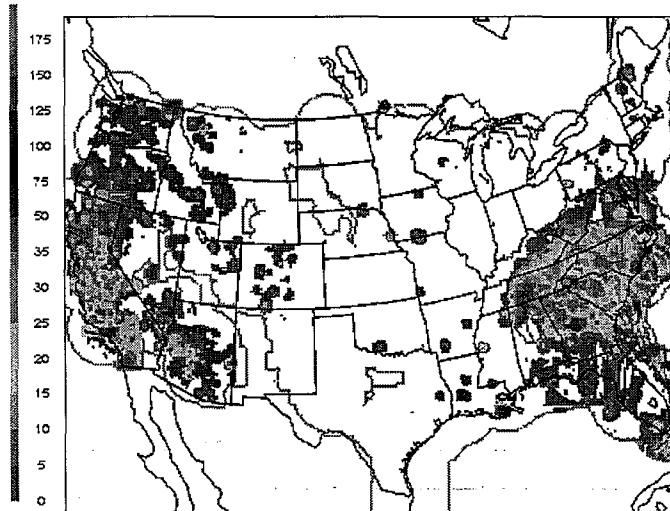
Figure 1.
 NCEP model QPF (a) bias and (b) equitable threat scores calculated over a domain covering California, western Oregon and western Washington for the month of February, 1998.



(a)

PRECIP (mm)
24h accum
VALID 12Z 04 FEB 98

MULTI-SENSOR
14.3 KM POL STR GRD

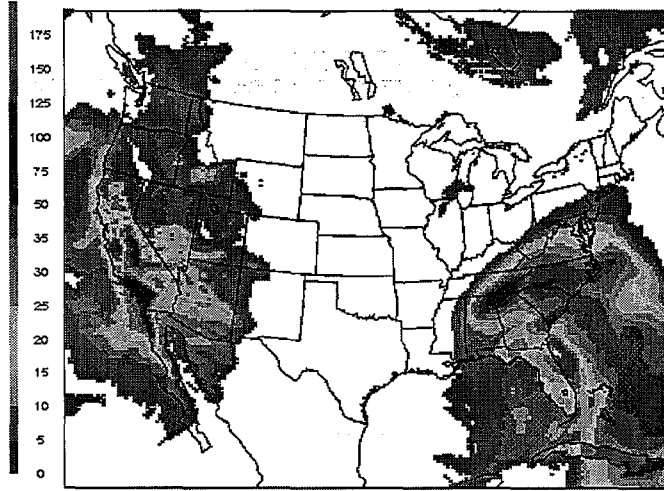


(b)

Figure 2. Twenty-four hour precipitation totals from observations (a) plotted at the University of Utah and (b) plotted by NCEP (see text); and QPF from (c) Eta-29, (b) Eta-48, (c) AVN, (f) NGM models and (g) HPC, all valid on February 4, 1998 1200 UTC.

PRECIP (mm)
24h accum
VALID 12Z 04 FEB 98

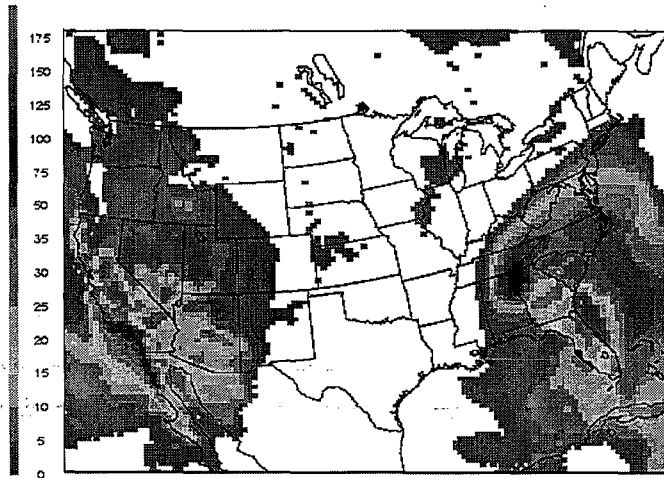
MESO ETA
33-H ETA FCST
29KM/50LYRS



(c)

PRECIP (mm)
24h accum
VALID 12Z 04 FEB 98

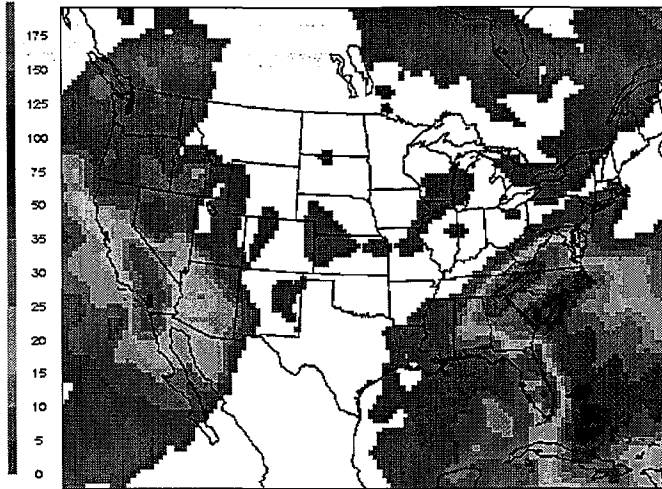
EARLY ETA
36-H ETA FCST
48KM/36LYRS



(d)

PRECIP (mm)
24h accum
VALID 12Z 04 FEB 98

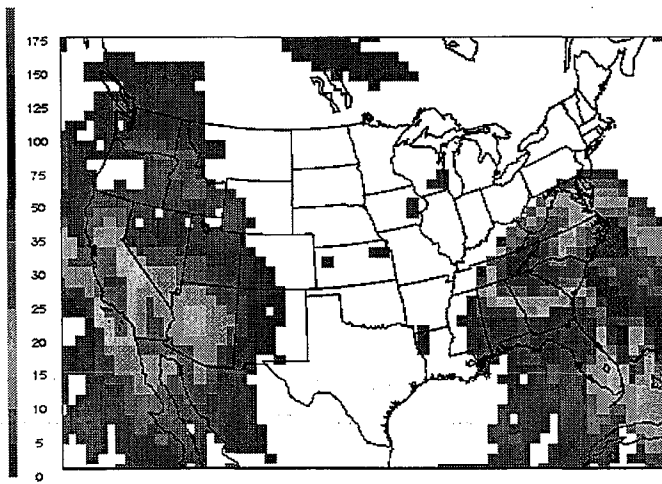
AVN
36-H FCST
40.6 KM LMB CON GRD



(e)

PRECIP (mm)
24h accum
VALID 12Z 04 FEB 98

NGM
36-H FCST
90.8 KM POL STR GRD



(f)

.25 .25
.50

1
2
3
SNOW
HIER
ELEVS

.50
SNOW
HIER
ELEVS

.50129
OPRNL DAY 1-
24HR TBF 603
ENDING 12Z
WED 04 FEB 1998
FCSTR: KORTY

ISOHYETS DEPI
RAIN/SNOW LIN
PCPN NORTH/WE

(g)

Date	Best Model	Notes	Did HPC improve on best model?
01/30/98	Eta-29 or AVN		no, similar to Eta-29
01/31	all okay	fairly dry day	no, similar to Eta-29
02/01	Eta-29	MRF; AVN, NGM incorrectly emphasize LA basin	no, but not worse
02/02	AVN or Eta-29	AVN only model to emphasize South CA	no, forecast is similar to Eta-29
02/03	Eta-29		yes, increased amounts
02/04	Eta-29	all models miss MTR area	yes, increased amount
02/05	none	all models has serious weaknesses	yes, excellent forecast
02/06	AVN or Eta-29		no, over-forecast South
02/07	all okay except NGM		yes, increased amounts
02/08	all okay except NGM		no, but not worse.

Fig. 3.
Table showing subjective evaluation of model and HPC QPF performance.