WESTERN REGION TECHNICAL ATTACHMENT
NO. 96-30
NOVEMBER 26, 1996

THE INITIALIZATION PROCEDURE IN THE MESO ETA MODEL

Mike Staudenmaier, Jr. - WRH-SSD/NWSFO SLC

Introduction

The Meso Eta model has proven to be a valuable model in the forecasting process due to its high resolution and model physics (Burks and Staudenmaier 1996; Janish and Weiss 1996; Schneider et al., 1996). However, as with all models, the solution generated by the model equations is highly dependent on the data which goes into the model. Thus, the initialization procedure should be of great importance to the forecaster in determining the potential use of the model output. This Technical Attachment will investigate the current initialization procedure used in the Meso Eta model and discuss future improvements to the process.

Basic Procedure

The initialization procedure for the Meso Eta begins three hours prior to the actual start of the forecast model run. At t-3 hours, or at 0000 UTC and 1200 UTC, a first guess is provided by the Global Assimilation System (GDAS) using all available data. This first guess is applied to the Meso Eta coordinate system. The original analysis is converted from spectral space (off of the Aviation (AVN) model grid) to the Eta model native grid and interpolated vertically to Eta coordinate surfaces. This adjusted "first guess" is then interpolated to each observation location and the observed increments (observed - first guess) are computed. A multi variate Optimum Interpolation (OI) analysis of the observed increments is performed on the Meso Eta model grid and is used to modify and update the "first guess". The only variables in the model which are updated during the OI analysis are temperature (T), the u- and v-components of the wind field (u,v), the specific humidity (q), and the pressure at the model terrain level (p*). The model then integrates for three hours, at which time another OI analysis is performed on all observed data which has been received over the past 1.5 hours that is in the form just mentioned (T,u,v,q,p*). The grids are modified based on these observed increments from the OI analysis and these new grids provide the initial analyses for the 0300 UTC and 1500 UTC runs of the model.

These new analyses benefit from a later cutoff time for the 0000 UTC and 1200 UTC data, and also utilize the new data from 0130 UTC to 0300 UTC and from 1330 UTC to 1500 UTC. This additional data during the three-hour integration is comprised of numerous aircraft reports, surface observations, profilers, and limited satellite observations. Once all of this data has been assimilated into the initial analysis, the 33-hour forecast is run. Boundary conditions for both the assimilation and model forecast are obtained directly from the AVN run of the NMC Global Spectral Model, thus one could consider the Meso Eta to be a nest inside the AVN model run.

Discussion of Present Problems

Due to resolution differences between the model and reality, there are still locations that will likely not see the full potential of this assimilation process. Many places in the Western Region fall into this category. Even at 29 km resolution, many of the sharp gradients in topography, especially narrow valleys or mountain ranges, are not represented in the model topography. Numerous cities in the Western Region are located in these types of locations. Thus, there are many observing stations in the Western Region which actually lie below the model surface. *Any data which in reality is greater than 25 mb below the model surface is not used in the assimilation process.* Additionally, many second and third order surface observation points (103 of them in the Western Region) do not report station pressure at the station elevation. These sites must be thrown out since the assimilation scheme does not know at what level to place the data. Because of these facts, only 14% (20/144) of the NWS surface observation network in the Western Region will generally make it into the initialization of the Meso Eta. In the Western Region, the locations most likely to have their surface data used in the Meso Eta model are much of Washington, portions of southern and central California, northeastern Arizona, western Idaho, and portions of central Montana (Fig 1). Currently, no surface information from Nevada or Utah makes it into the Meso Eta model.

A brief explanation of how surface data is processed during the assimilation procedure is needed at this point. As mentioned above, only T, u, v, q, and p* are used in the assimilation process. Additionally, because of how the models are set up for integration between levels, these variables exist at the mid-point of the layers, not at the layer interfaces (at the bottom or top of the layer). *Thus, in reality, there is no model surface.* Nor is there a level at 2 meters, or at 10 meters. These values are derived from the values at the mid-point of the lowest layer in the model domain above the model terrain. Because of this, the first place in the model where information is needed to correct the first-guess towards reality is actually at the mid-point of the first layer above the model's terrain, which over elevated terrain can be a substantial distance above the model terrain and therefore even further away from reality where the actual observation was taken. In order to generate the corrections needed to the first guess when the observation is below the lowest data level in the model requires extrapolation 'underground' through the model surface and upward into the model's atmospheric portion of the profile. This is a very risky proposition for which no real effective, or realistic, procedure exists. Thus, to make the assimilation procedure easier, this extrapolation is limited to no more than 25 mb. So if a piece of observed data requires extrapolation of more than 25 mb to reach the mid-point of the first model layer, it doesn't get used. This is the current method used in both the NGM and the Eta models since the development of the NGM model in 1983. The global model does not allow any extrapolation of data below the first sigma layer mid-point, so even more surface data in the Western Region is thrown out.

Additionally, data cutoff occurs in radiosonde data as well because of this effect. In the model assimilation process, the model will use any data from a radiosonde which is above the mid-point of the lowest model layer and throw out any data which is located below this level, even if the model surface is located well above the real surface. Since the free atmosphere is being used as the new 'surface', low-level surface moisture gradients may be weakened or destroyed, and many times this will create a drier and cooler surface than what actually exists in reality. Any low-level temperature gradients may also be weakened by this process as well as destroying low-level radiative inversions and frontal structure. The most susceptible locations for this loss of near-surface data appears to be in a belt from Medford, OR to Grand Junction, CO, covering much of the Great Basin region. Figure 2 is a map showing the most likely areas for data loss in radiosondes. Figure 3 shows typical depths of data loss for many of the radiosondes in the western portion of the United States. As can be seen, the Meso Eta improves over the Eta model on the data loss problem in most places, however, at some sites it has become slightly worse. Some sites, like Salt Lake City (SLC) are still losing the lowest 65 mb of the raob during the assimilation process. Clearly though, the benefits outweigh the drawbacks.

Why is the Great Basin so prone to data loss during the initialization process? A likely reason may have to do with the way model topography is derived. In determining the surface elevation for a particular area, each 29 km horizontal grid box is first divided into 16 subboxes. Mean elevations for each of these 16 subboxes are calculated from official United States Geological Survey (USGS) topographical data. Using these values, the maximum mean value from each of the four rows and four columns are taken to yield an intermediate value for the grid box surface elevation. The mean of these eight values are taken to yield an intermediate value for the model grid box. The final grid elevation is found by moving this mean value up or down to match the closest vertical layer in the model domain. Due to the particular nature of topography in the Great Basin, with numerous narrow north-south oriented mountain ranges with broad valleys, the model topography would be biased toward the higher elevation of the narrow mountains. Almost all of the surface data in the Great Basin, however, comes from locations on the valley floors. This likely leads to most surface observation points being located more than 25 mb below the model topography and thus, being thrown out during the assimilation process.

Currently, no mesoscale information is implemented in the initialization process. Even if mesoscale data sources become available in the Western Region for model ingestion, it is still questionable if much of this data would even pass the assimilation process, or if most of it would be thrown out, especially when referring to mesonets of surface observations occurring in areas previously mentioned. Additionally, only temperature, specific humidity, and station pressure are being ingested from those surface observations which make it into the assimilation process. *No observed surface wind data is currently being used in the model, nor in any of the NCEP suite.*

At this time, the cloud model (Zhao et al., 1996) in the Meso Eta model is not initialized with cloud water or ice. Thus, the model must 'spin-up' or create cloud water/ice during the first few hours. This creates a slight lag in the development of precipitation as the cloud model must reach saturation before any precipitation can occur. Additionally, once the cloud model saturates, model clouds may develop in different locations than reality, as model clouds must develop due to model physics, without any basis on their actual location in reality.

The soil model in the Meso Eta is currently not initialized with real-time moisture or temperature values, but rather with climatological values. Currently there is no real-time accessible nationwide network of root-zone soil moisture observations available for assimilation. This poor initialization of soil moisture and temperatures can cause problems in the development of precipitation in the model. One reason is that soil moisture and temperature gradients can occasionally act as focusing mechanisms for the initiation and sustenance of convection. Additionally, evaporation of soil moisture into the free atmosphere can enhance low-level moisture in the atmosphere, leading to heavier rainfall potential and possible convection. Vegetation type and soil type in the model are based on 1 degree by 1 degree fixed climatological values. The green vegetative fraction, however, is based on 0.15 degree by 0.15 degree monthly fields based on a 5 year climatology. These monthly fields of green vegetative fraction are interpolated to actual days of the year, so the values can change slightly from day to day. Additional resolution of these fields is necessary at 29 km resolution, and will be even more necessary at 10 km resolution.

The radiation scheme in the model uses climatological values of carbon dioxide and ozone concentrations, which are not allowed to evolve during the forecast integration. Surface albedo is derived from 1 degree by 1 degree quarterly climatological fields which are then interpolated to actual days of the year. This first guess of albedo is then allowed to evolve based on snow cover and snow depth, green vegetation fraction, sea ice, and model clouds, among other things. In a mesoscale model with a short forecast period, it might not be that important to have actual values of the chemical composition of the atmosphere. However, having values of actual surface albedo for initialization might improve the radiation scheme slightly. Satellite data is not used in the initialization of the radiation model, and since there is no cloud water/ice in the cloud model, the radiation package does not reflect reality in terms of cloud cover.

Satellite data is currently used only to nudge the current analysis towards reality. With the OI analysis, satellite observations must be in the form of temperature, wind, specific humidity, or pressure at the model surface to be used. Almost all satellite data is not in this form, and thus cannot be used currently. In the current analysis scheme, the usefulness of satellite data only includes GOES cloud drift winds, TOVS deep-layer thicknesses over the oceans, and SSMI total column precipitable water. The cloud drift winds are used to augment upper air data and to add some detail to the upper atmospheric structure. Over the ocean, moisture profiles are subjectively determined based on certain cloud signatures and then added to the OI scheme. This data is low resolution (around 200 km resolution with 6 levels of relative humidity data) and can only be determined when certain cloud signatures exist over the ocean. SSMI precipitable water products are also being used in the assimilation process. This data has good horizontal resolution, but has NO information on how that moisture is distributed through the vertical column. Thus, this moisture can only be used to do an adjustment of the model precipitable water guess field. As stated above, no information from the satellite data is used in conjunction with the cloud model or the radiation package. Currently, the retrieved single layer temperatures and satellite-derived vertical profiles of temperature and moisture are not accurate enough to use in the OI analysis.

Hope for the Future

Even with all these problems in the way data is assimilated into the model, the Meso Eta still manages to capture much of the details of the day to day weather pattern. Due to its 29 km resolution, the Meso Eta likely manages to assimilate more data than any of the other models in the NCEP suite. Modelers are now at a crossroads in terms of the philosophy of data assimilation. Up to this point, real surface data had little importance in the model, since at coarse resolution, both in the vertical and the horizontal, any pertinent information from real surface data would be quickly washed out during the model integration. Now, however, with finer resolution, and a much more resolved planetary boundary layer, the need for actual surface information in the model is becoming more and more of a necessity. Forecasters were not presented with surface temperature, relative humidities, and surface winds until the 29 km Meso Eta appeared. This is the first time that forecasters have a model with a fine enough resolution to be able to start using the boundary layer information as a forecasting tool. However, much of the information near the surface is not very accurate, and modelers will now have to look for ways to improve the mesoscale boundary layer forecasts, just as they have worked in the past to improve the skill of the AVN, NGM and Eta models in forecasting synoptic scale features.

The modelers at NCEP are looking at improving the skill of the Meso Eta through a variety of ways over then next two years. Most importantly, the assimilation process of the Meso Eta, which currently is only a three-hour assimilation using the first guess from the GDAS, will be changed so that it cycles on itself, and the assimilation will last six hours rather than three. Thus, the information used to initialize the model will come from the 48 km Eta as opposed to the coarser resolution GDAS. A first guess generated from the same model is preferable to one based on a global spectral model-based first guess because of the matching physics, the better resolution, and less of a 'spin up' problem. This will allow many of the small scale models, like the soil model, the cloud model, and the radiation model, to have initial values from the previous run, rather than having to initialize with zero information and 'spin up' to reality.

Another change to the assimilation process will be with the method used in creating the initialization for the model. NCEP plans to go from the current OI analysis technique to a newer and potential better method, the 3-dimensional variational analysis technique (3DVAR). OI and 3DVAR are actually two methods for solving the same problem, which is finding the best fit of both observational data and the model first guess forecast. The formal mathematical solution to this problem involves the solution of a matrix problem of dimension equal to the number of observations, which is impractically large. The OI procedure for solving this problem has been to make a local approximation, by using a small number (30-100) of nearby observations to compute smaller matrices at each grid point. Because of the way the OI procedure works, it has only been possible to use observations directly related to model variables, as mentioned above.

The 3DVAR method for solving this matrix problem makes no local approximation. The entire matrix problem is solved. Two procedures make this possible. First, the matrix being solved for is not computed directly, but instead is represented by a sequence of simple operations. Second, the problem is solved by iteration, using a technique known as the conjugate gradient method. One of the main advantages of using the 3DVAR technique is that observations no longer need to be the same as model variables. It is only necessary to have a procedure which can compute a simulated observation from standard model variables. This is called the forward model. For example, a radial wind measurement from a doppler radar is only a partial measurement of the wind, along the direction of the radar beam. The forward model in this case is simply the projection of the model wind along the beam direction at the observed location. Thus, the 3DVAR system will be much more flexible than the OI technique and will make it possible to use many new data sources and will better utilize existing data.

The cloud model will also be initialized with data in the future, through the use of real-time, high-resolution nephanalyses (cloud maps), and hourly raingage data over the U.S. These will be used to initialize the cloud water/ice, moisture and latent heating fields to be internally consistent with observed precipitation rates. Also, with the cycling EDAS system during the assimilation process, and a longer assimilation period, the model will have cloud water/ice values to initialize with, and more time to 'spin up' to reality. This should have a positive impact on precipitation scores, along with more realistic cloud fields. This process will also be used to moisten the soil model, thus improving evaporation and boundary layer moisture fields. Because of the 3DVAR system, it will also be possible to use cloud observations inferred from the GOES satellite to help initialize the cloud model.

Additional sources of satellite data are expected to be used more often in the future as well. Current GOES 8/9 moisture profiles are now accurate enough that the resulting integrated layer precipitable water values are good enough to be of use in the model. Black-body radiance temperature information from the satellite is also expected to be used to initialize the model once the 3DVAR technique is implemented. Satellite soundings still do not have enough resolution to be of use in the assimilation process, although in the future, it is hoped they will improve to the point of being useful for initializing the model in data sparse regions, especially over the oceans.

NCEP has also developed a smaller-scale model with a resolution of 10 km, hereafter called the Eta-10. This model, with much more resolved topography in the Western Region, should allow for some more surface data to be used in the assimilation process, although how much more will make it into the model is not yet known. To produce useful forecasts at such small grid spacing, it will be more critical than ever to initialize the model with mesoscale information, both to allow the model to be more accurate, and to allow the boundary layer to evolve realistically. At high resolution parameters such as vegetation type and soil type play a crucial role in the soil moisture pattern, which is reflected in boundary layer processes. Thus, much higher resolution data will need to be found to initialize the model with, rather than the current 1 degree by 1 degree data. The Eta-10 is expected to be run over the Western Region domain in the near future.

Doppler winds from the NEXRAD network are also being examined for as sources of mesoscale initialization in the Eta-10. This data set can only be used once the 3DVAR analysis technique is implemented, as mentioned above. This is because the wind measurements are incomplete, with only the component along the radar beam being measured. Conventional methods, like OI, require overlapping beams from two radars to get the full wind vector. In the NEXRAD network, there are not many locations with such a set-up. With the 3DVAR method, the along-beam component of the wind can be used, rather than the complete wind vector. More information regarding the use of Doppler winds in the initialization of the Eta-10 can be found in Parrish et al., 1996. Local mesoscale networks of surface observations will also be used in the future, assuming that they are not thrown out during the assimilation process.

Conclusions

The current method of assimilating data during the initialization procedure in the Meso Eta model has been discussed, with numerous problems being addressed. Even with all of the problems mentioned, the Meso Eta is clearly the most accurate model in the NCEP suite, with the best resolution, the benefits of a later data cutoff, and the best physical package. This accuracy can be seen best in the precipitation threat scores of the NCEP suite over the past year (Staudenmaier, 1996). However, much of the data in the model boundary layer, is still lacking in the detail needed by forecasters in the field. Until more mesoscale information, including more surface information, can be initialized into the model, this trend will likely continue. With the development of the Eta-10 model, the cycling EDAS system, along with the 3DVAR analysis technique, improvements are expected, with much of the improvement expected in the model boundary layer. The future looks bright for mesoscale forecasting in the NWS.

Acknowledgments

The author wishes to thank all the good folks at NCEP for their help during the research phase of this paper. Special thanks go to Geoff DiMego, who was extremely helpful in tracking down data files and answering questions. Without their help, this paper would not have been possible.

References

Burks, J.E., and M.J. Staudenmaier, 1996: A comparison of the Eta and the Meso Eta models during the 11-12 December 1995 storm of the decade. *WR-Technical Attachment* **96-21**.

Janish, P.R., and S.J. Weiss, 1996: Evaluation of various mesoscale phenomena associated with severe convection during VORTEX-95 using the Meso Eta model. *Preprints, 15th Conf. on Weather Analysis and Forecasting,* AMS, Norfolk, VA. Aug, 1996.

Parrish, D., J. Purser, E. Rogers, and Y. Lin, 1996: The regional 3D-variational analysis for the Eta model. *Preprints, 11th AMS Conf. on Numerical Weather Prediction,* Norfolk, VA. Aug, 1996.

Schneider, R.S., N.W. Junker, M.T. Eckert, and T.M. Considine, 1996: The performance of the 29 km Meso Eta model in support of forecasting at the Hydrometeorological Prediction Center. *Preprints, 15th Conf. on Weather Analysis and Forecasting,* AMS, Norfolk, VA. Aug, 1996.

Staudenmaier, M.J., 1996: Precipitation verification statistics from the NCEP operational model suite. *WR-Technical Attachment* **96-28**.

Zhao, Q., T.L. Black, and M.E. Baldwin, 1996: Implementation of the cloud prediction scheme in the Eta model at NCEP. Submitted to *Wea. Forecasting.*

Figure 1: Map of the western Uñited States showing those areas most likely to have surface observations included in the Meso Eta data assimilation process (shaded areàs).

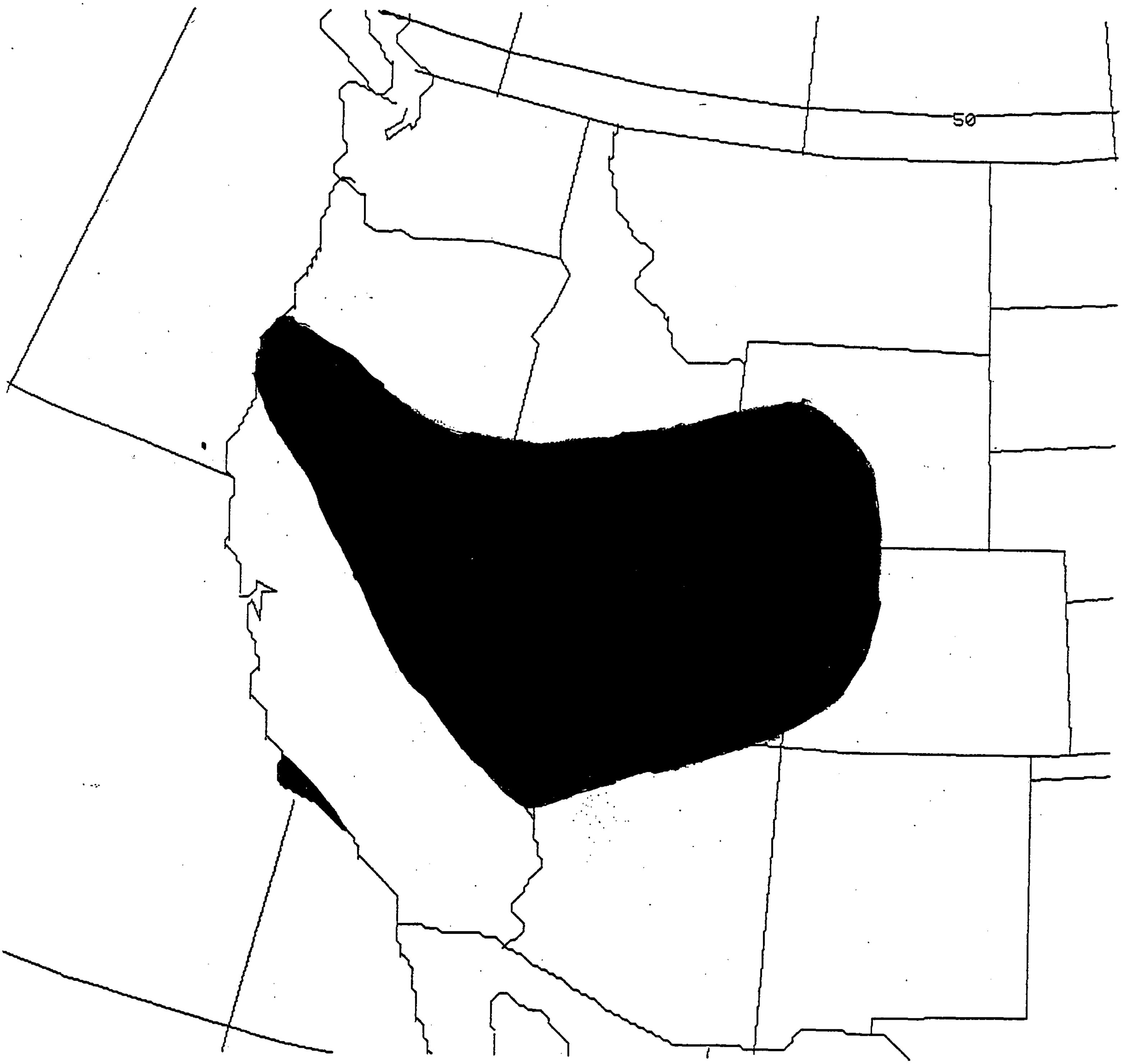

Figure 2: Map of the western United States showing those areas most likely to have data loss in the lowest portion of radiosonde data of greater than 20 mb in the Meso Eta assimilation process (shaded areas).

STATION IDENTIFICATION	ETA (MB)	MSO (MB)
72582 LKN	29	21
72376 FGZ	0	18
72274 TUS	60	13
72572 SLC	59	68
72597 MFR	109	48
72493 OAK	18	15
72489 REV	36	46
72293 SAN	12	5
72387 DRA	42	29
72476 GJT	98	105
72576 RIW	104	27
72786 GEG	4	11
72776 GTF	26	4
72797 UIL	14	9
72694 SLE	50	15
72681 BOI	55	21
72768 GGW	29	13
72393 VBG	5	32
72381 EDW	12	4

Figure 3: Table of western United States radiosonde sites showing the average depth of data loss of the lowest portion of the raob in both the 48 km Eta model and the 29 km Meso Eta model for the period of November 12-15 1996. Values are likely similar throughout the year.