U. S. DEPARTMENT OF COMMERCE
NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION
NATIONAL WEATHER SERVICE
NATIONAL METEOROLOGICAL CENTER

OFFICE NOTE 106

Note on Insertion Procedures for Meteorological Data Assimilation

Ronald D. McPherson
Robert E. Kistler
Development Division

JANUARY 1975

Copies of Office Note 106 were distributed to:

Dr. Brown
Dr. Shuman
Dr. Phillips
Dr. Stackpole
Dr. Gerrity
Dr. Bonner
Mr. Kistler
Mr. McPherson
Dr. Miyakoda, GFDL
Dr. Schlatter, NCAR
Dr. Halem, GISS
Dr. Bengtsson, European Centre, England
Dr. Seaman, CMRC, Australia
Dr. Rutherford, Canada

NOTE ON INSERTION PROCEDURES FOR METEOROLOGICAL DATA ASSIMILATION

I. Introduction

The original stimulation for research in four-dimensional assimilation was the anticipation of global observations of atmospheric temperature by satellite-borne radiometer sensors. By contrast with conventional observation systems, these data are not only distributed in time and space, but also yield information on only one meteorological variable. Studies with simulated data (e.g., Charney, Halem and Jastrow, 1969) established that incomplete information can be used to induce a complete representation of the meteorological variables through the use of a prediction model, provided a sufficiently long sequence of historical data is available. However, the evolution toward such a representation was found to be exceedingly slow and therefore impractical for operational usage. Morel, Lefevre, and Rabreau (1971) pointed out the basic difficulty: inserted observations of the mass field, without an accompanying adjustment of the motion field, tend to be "rejected" by the model. Hayden (1973) proposed a geostrophic correction to the motion field based on the observed gradient of the mass field. Kistler and McPherson (1975) found that this technique substantially improves the model's memory of the observations. However, the geostrophic correction method may not be applied in low latitudes, and its use elsewhere may not be entirely beneficial.

Another method, similar to one suggested by Miyakoda (1973), seeks to reinforce the model's memory by inserting the mass field as corrected by the observations, and then restoring that field at the end of each of several succeeding time steps. This idea traces its ancestry to the initialization method of Nitta and Hovermale (1969).

In this note, we report on assimilation experiments in which the restoration method and the geostrophic correction method are compared to simple insertion of asynoptic data.

II. Experimental Procedures

a. Model

The prediction model which served as the basic element in these experiments is a primitive equation barotropic model previously described by Kistler and McPherson (1974). A lattice of 27 x 29 points superimposed on a polar stereographic projection of the Northern Hemisphere is the computational grid. The mesh length is 762 km at 60N. Spatial finite differencing has been described by McPherson (1971). Time integration is by the centered-difference, or "leapfrog" method, with a time step of 10 minutes. A time filter (Asselin, 1972; coefficient of 0.5) is included to suppress gravitational oscillations resulting from the insertion of alien data.

b. Initial data and initialization

Operational 500-mb analyses of height and wind components for 9 April 1973, 0000 GMT, produced by the National Meteorological Center

(NMC), were the initial data for these experiments. The analyses were initialized by integrating forward and backward between 0600 GMT and 1200 GMT 10.5 times, using the Euler-backward method (Kurihara, 1965) to damp the gravity waves produced by the initial imbalance. The final fields of height and winds for 0000 GMT were used as the initial state for the assimilation experiments. Figure 1 shows the features of the initialized height field.

c. Asynoptic data

Observations of 500-mb height, as calculated from operational Vertical Temperature Profile Radiometer (VTPR) soundings, during the 15-hour period from 2230 GMT 8 April 1973 to 1330 GMT 9 April 1973 were taken as the data to be inserted. These observations were stratified into 3-hour blocks centered on 0000 GMT, 0300 GMT, 0600 GMT, 0900 GMT, and 1200 GMT. There were originally 140 observations. Elimination of reports south of 20N, as well as several apparently unrealistic observations, reduced the total to 103. The number of observations in each data set is given in Table 1, and each set is plotted in Figure 1.

Table 1. Number of 500-mb height observations, as calculated from VTPR soundings, in each of the five 3-hour time blocks.

Time (GMT)	Number
0000	8
0300	42
0600	8
0900	14
1200	31

d. Interpolation

Each data set was interpolated to grid points in the vicinity of the observations by means of a successive-correction algorithm similar to that used operationally at NMC for many years (Cressman, 1959). The current model state was used as a first guess, and four scans through each data set were made, with influence radii of 2.375, 1.8, 1.1, and 0.9 grid increments, respectively. A filter designed by Shuman (1957) was applied to the resulting corrections prior to insertion.

For the first data set, the initialized field served as a first guess. The data were introduced through the interpolation procedure, and a 3-hour prediction was made. This predicted field served as a first guess for the interpolation of the 0300 GMT data set. From the corrected fields, another 3-hour forecast was made, and so on until all data sets were introduced. Each interpolated data set therefore had the benefit of the data sets previously inserted, except for the first one.

e. Experiments

Four variations of the restoration method were tested.

(1) Complete restoration of the corrected height field for nine time steps (1.5 hours). At each insertion time, the corrected height field was obtained through the successive-correction method described in the previous section, and inserted into the model. Following each

of the succeeding nine time steps, this height field was reinstated, but the winds were allowed to respond freely.

(2) Partial restoration of the corrected height field for nine time steps. At each insertion time, and for nine time steps thereafter, the corrected height field was blended with the predicted height field according to a linear weighting scheme,

$$h_b^{\tau} = \left(\frac{\tau-\tau_o}{9}\right) h_f^{\tau} + \left(1 - \frac{\tau-\tau_o}{9}\right) h_a^{\tau_o}; \quad (\tau-\tau_o) \leq 9, \qquad (1)$$

where h_b^{τ} is the blended height value at time step τ, h_f^{τ} is the predicted height value at the same time, and $h_a^{\tau_o}$ is the corrected height value at the insertion time τ_o. At $\tau = \tau_o$, full weight is given the height field as corrected by the observations; at $\tau = \tau_o + 9$, full weight is given the predicted height field.

(3) Complete restoration of the corrected height field for 18 time steps (3 hours), centered on the insertion time. In this experiment, the model state existing nine time steps prior to insertion time was stored. Following the successive-correction procedure, the model was returned to the stored state and the integration restarted. At the end of each of the subsequent 18 time steps, the corrected height field was completely restored.

(4) Partial restoration of the corrected height field for 18 time steps. This followed the procedure of (3), but with the blending outlined in (2).

For comparison, two additional experiments were performed:

(5) Simple insertion of the corrected height field only at each insertion time τ_0.

(6) Insertion of the corrected height field at each insertion time, with geostrophic correction of the wind field as described in Kistler and McPherson.

f. Evaluation

In order to systematically evaluate each experiment, it is necessary to establish objective criteria for successful assimilation. Observations may be inserted into a model, but they cannot be said to be assimilated unless their influence is retained in the subsequent integration. The first criterion of successful assimilation is therefore that the model must "remember" the inserted data. Suppose error-free observations were inserted into a reversible, perfect model, and the integration were resumed. If, after a few time steps, the model were reversed and integrated backward to the identical state existing immediately after insertion, then the observations have been assimilated perfectly. However, neither observations nor model are perfect, and so the second criterion of successful assimilation must be that the model representation reflect a mutual adjustment between the observations to a level near the expected error level of the data.

To formulate these criteria objectively, at each insertion time the root-mean-square (RMS) difference ε_f between the observations and the first guess interpolated to the observation location was computed:

$$\varepsilon_f = \frac{1}{N} \left[\sum_{i=1}^{N} (h_i^o - h_i^f)^2 \right]^{1/2} \tag{2}$$

where h_i^o is the (i)th observation, and h_i^f is the first guess interpolated to the (i)th observation point. N represents the number of observations in the data set. Similarly, immediately after the successive-correction interpolation procedure, the RMS difference between the observations and the corrected field ε_a was calculated,

$$\varepsilon_a = \frac{1}{N} \left[\sum_{i=1}^{N} (h_i^o - h_i^a)^2 \right]^{1/2} \tag{3}$$

where h_i^a is the corrected value interpolated back to the (i)th observation point. The quantity ε_a is used as an objective measure of the second criterion.

To evaluate the degree to which the model retains the influence of the data, at the end of each experiment the model was reversed and integrated back to the initial time with the time filter disabled. At each insertion time, Eq. (2) was recalculated and identified as ε_b. The observations were not reinserted. If the inserted observations were completely retained by the model, $\varepsilon_b = \varepsilon_a$; if completely forgotten, $\varepsilon_b = \varepsilon_f$. A memory index M combines these measures; it is identified as

$$M \equiv 1 - \frac{\varepsilon_b - \varepsilon_a}{\varepsilon_f - \varepsilon_a} \tag{4}$$

When $\varepsilon_b = \varepsilon_a$ (perfect memory), M = 1; when $\varepsilon_b = \varepsilon_f$, M = 0.

Finally, it is convenient to express these objective measures in terms of the entire set of observations in addition to the stratification into 3-hour blocks. For this purpose, "pooled" measures may be formed by extending the calculations in (2) and (3) over all of 103 observations. These "pooled" measures are indicated by an overbar, i.e., $\bar{\varepsilon}_f$, $\bar{\varepsilon}_a$, $\bar{\varepsilon}_b$, $\bar{M}$.

III. Results

The results of these six experiments are summarized in Table 2.

Table 2. RMS differences between observations and (a) first guess (ε_f); (b) height field corrected by observations (ε_a), and (c) model state during backward evaluation (ε_b), over the 12-hour interval containing the observations. The memory index $\bar{M}$ is defined in the text.

Experiment	$\bar{\varepsilon}_f$ (m)	$\bar{\varepsilon}_a$ (m)	$\bar{\varepsilon}_b$ (m)	$\bar{M}$
1. complete restoration, 9 time steps	42.5	23.7	33.6	0.47
2. partial restoration, 9 time steps	43.0	23.7	34.5	0.44
3. complete restoration, 18 time steps	41.8	23.9	36.9	0.27
4. partial restoration, 18 time steps	43.0	24.0	33.4	0.51
5. simple insertion	44.0	23.9	27.7	0.81
6. geostrophic correction	41.4	23.0	23.4	0.98

The table indicates that in all six experiments the differences between the observations and the height field corrected by the observations, $\bar{\varepsilon}_a$, are very similar. But significant differences appear in the degree to which the corrected height field is remembered. With the aid of the geostrophic correction, the model's memory is nearly perfect. Insertion of the corrected height field only at insertion times results in a reduction of the memory index to 0.81. By contrast, only one of the restoration experiments exhibits a memory index exceeding 0.5.

The restoration methods therefore have a detrimental effect, by comparison to simple insertion. It has been noted, by Williamson and Kasahara (1971), and Morel and Talagrand (1974), among others, that the insertion frequency in a data assimilation system should be such as to allow sufficient time between observations to damp the gravity wave generated by the insertion. Since the restoration methods effectively insert at several successive time steps, allowing no time for damping, the amount of gravity wave noise should be greater than in the simple insertion experiment. This is confirmed in Figure 2, which presents the RMS height tendency, a quantity very sensitive to noise, as a function of time during the backward evaluation integration. No data are inserted, nor is the time filter operative during this part of the experiment, so that there is neither generation nor damping of gravity waves. The noise levels of the simple insertion and nine-time-step partial restoration experiments are very similar, approximately 2.5 m/time step, but the remaining restoration experiments exhibit higher noise levels.

The experiments reported here were conducted within the framework of a relatively simple model. While it is true that indications of usefulness with respect to any of the insertion techniques are not necessarily transferable to a sophisticated baroclinic model, experience has shown that a technique which fails in a simple model cannot be resurrected by adding complexity. Thus, these experiments may be viewed as part of a process of elimination which will hopefully provide useful guidance in the design of an operational four-dimensional data assimilation system.

REFERENCES

1. Asselin, R., 1972: "Frequency filter for time integrations," Monthly Weather Review, vol. 100, no. 6.

2. Charney, J., M. Halem, and R. Jastrow, 1969: "Use of incomplete historical data to infer the present state of the atmosphere," Journal of the Atmospheric Sciences, vol. 26, no. 5.

3. Cressman, G., 1959: "An operational weather analysis system," Monthly Weather Review, vol. 87, no. 10.

4. Hayden, C., 1973: "Experiments in the four-dimensional assimilation of Nimbus 4 SIRS data," Journal of Applied Meteorology, vol. 12, no. 3.

5. Kistler, R., and R. McPherson, 1974: "On the use of local balancing in four-dimensional data assimilation," submitted for publication in Monthly Weather Review.

6. Kurihara, Y., 1965: "On the use of implicit and iterative methods for the time integration of the wave equation," Monthly Weather Review, vol. 93, no. 7.

7. McPherson, R., 1971: "Note on the semi-implicit integration of a fine mesh limited area prediction model on an offset grid," Monthly Weather Review, vol. 99, no. 3.

8. Miyakoda, K., 1973: "The four-dimensional analysis," paper presented at the Second Conference on Numerical Prediction of the American Meteorological Society, 1-4 October 1973, Monterey, California.

9. Morel, P., G. Lefevre, and G. Rabreau, 1971: "On initialization and non-synoptic data assimilation," Tellus, vol. 23, no. 3.

10. Morel, P., and O. Talagrand, 1974: "The dynamic approach to meteorological data assimilation," submitted for publication to Tellus.

11. Nitta, T., and J. Hovermale, 1969: "A technique for objective analysis and initialization for the primitive forecast equations," Monthly Weather Review, vol. 97, no. 9.

12. Shuman, F., 1957: "Numerical methods in weather prediction: II. Smoothing and Filtering," Monthly Weather Review, vol. 85, no. 1.

13. Williamson, D., and A. Kasahara, 1971: "Adaptation of meteorological variables forced by updating," Journal of the Atmospheric Sciences, vol. 28, no. 8.

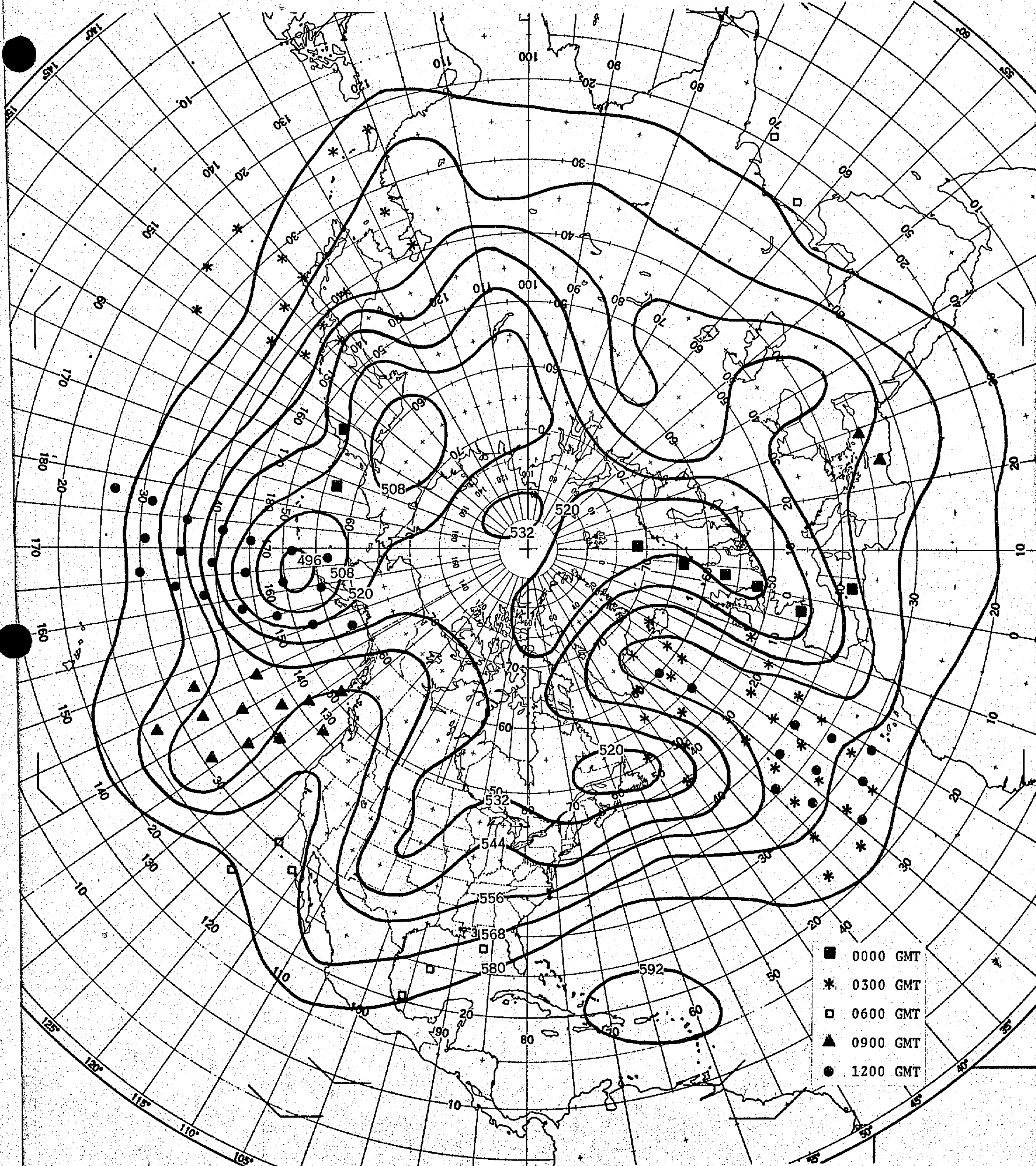

Figure 1. Initialized 500 mb height field, 0000 GMT 9 April 1973, with VTPR observations between 2230 GMT 8 April 1973 and 1330 GMT 9 April 1973.

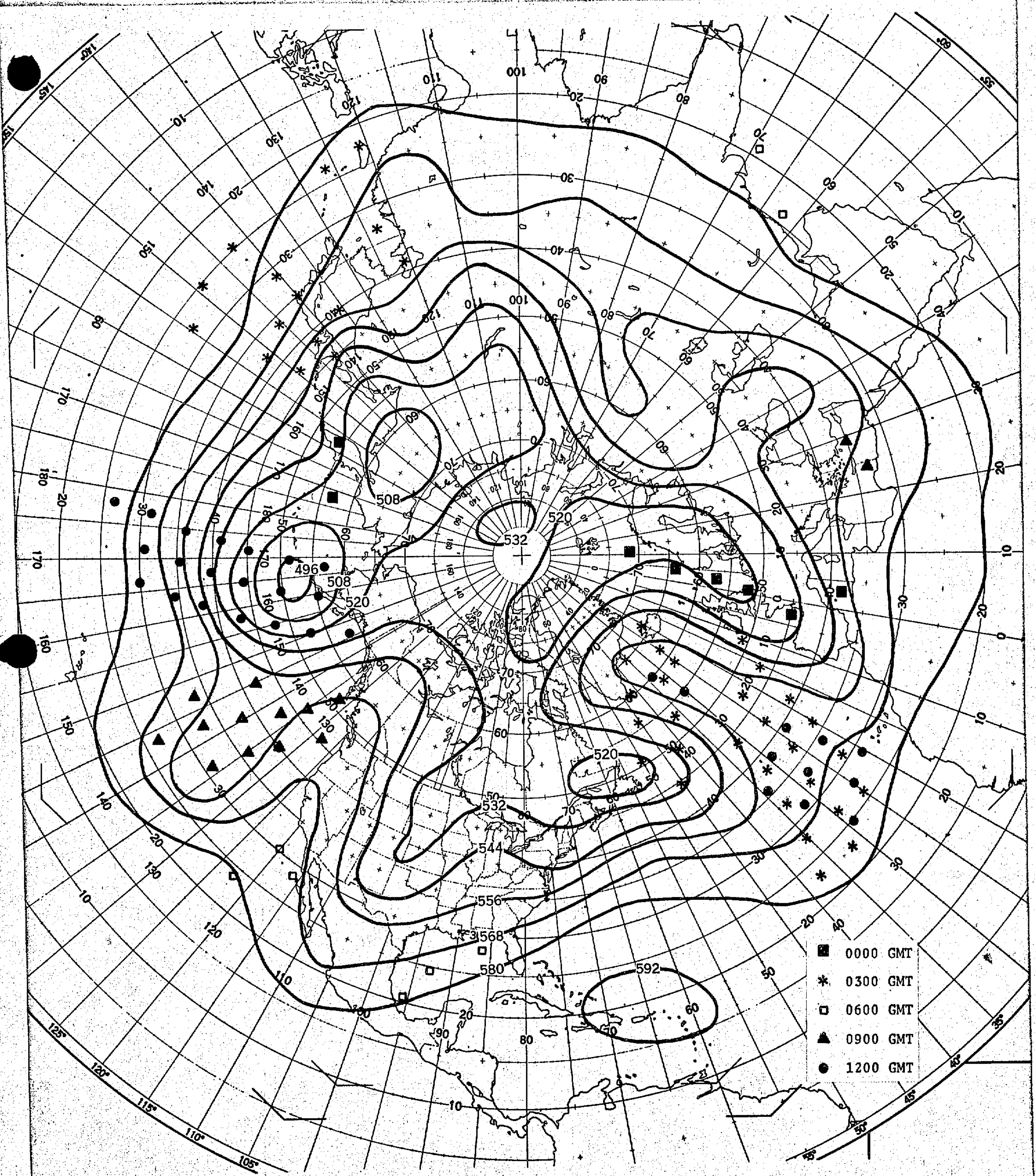

Figure 1. Initialized 500 mb height field, 0000 GMT 9 April 1973, with VTPR observations between 2230 GMT 8 April 1973 and 1330 GMT 9 April 1973.

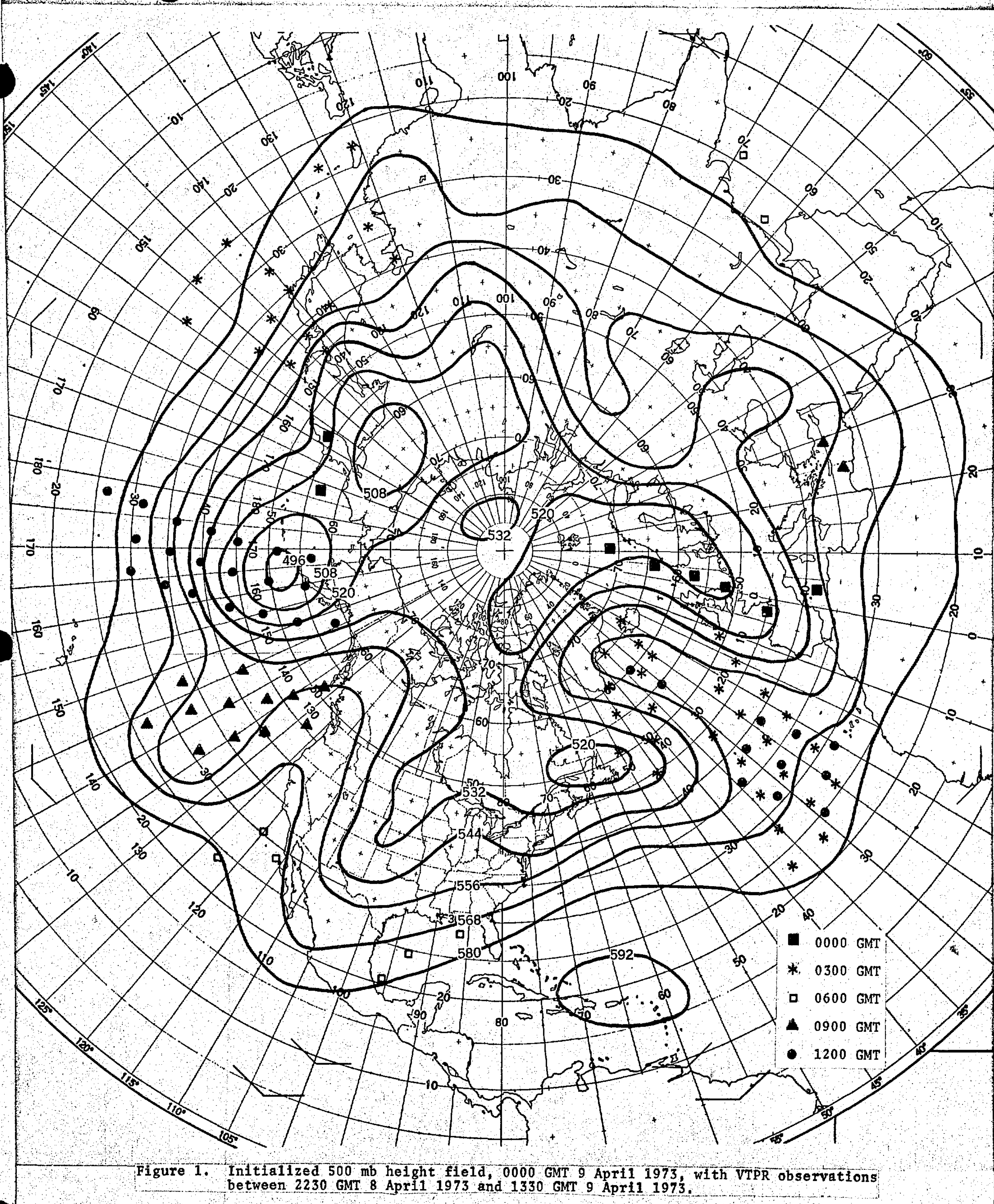

Figure 1. Initialized 500 mb height field, 0000 GMT 9 April 1973, with VTPR observations between 2230 GMT 8 April 1973 and 1330 GMT 9 April 1973.

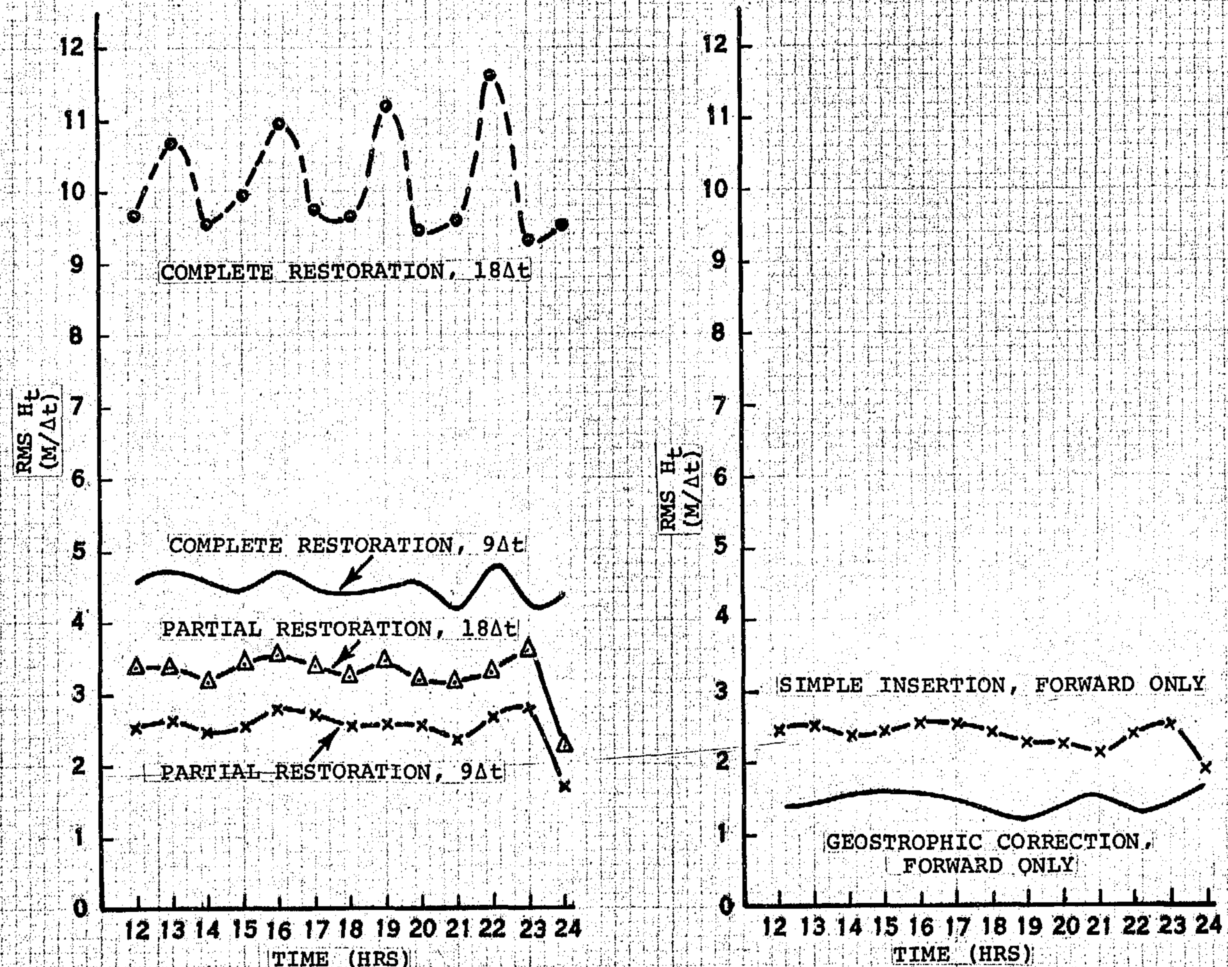

Figure 2. RMS H_t AS A FUNCTION OF TIME DURING BACKWARD EVALUATION

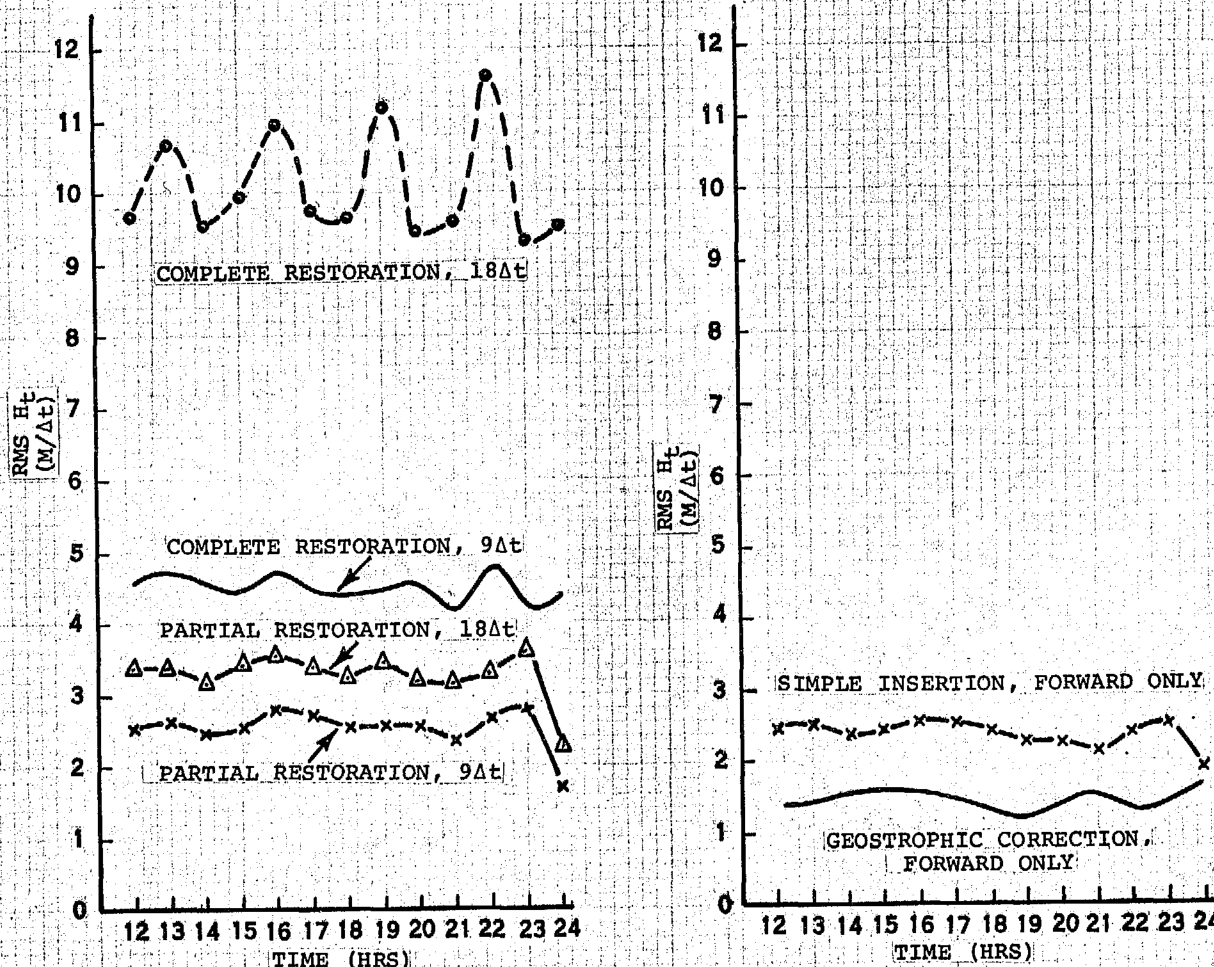

Figure 2. RMS H_t AS A FUNCTION OF TIME DURING BACKWARD EVALUATION

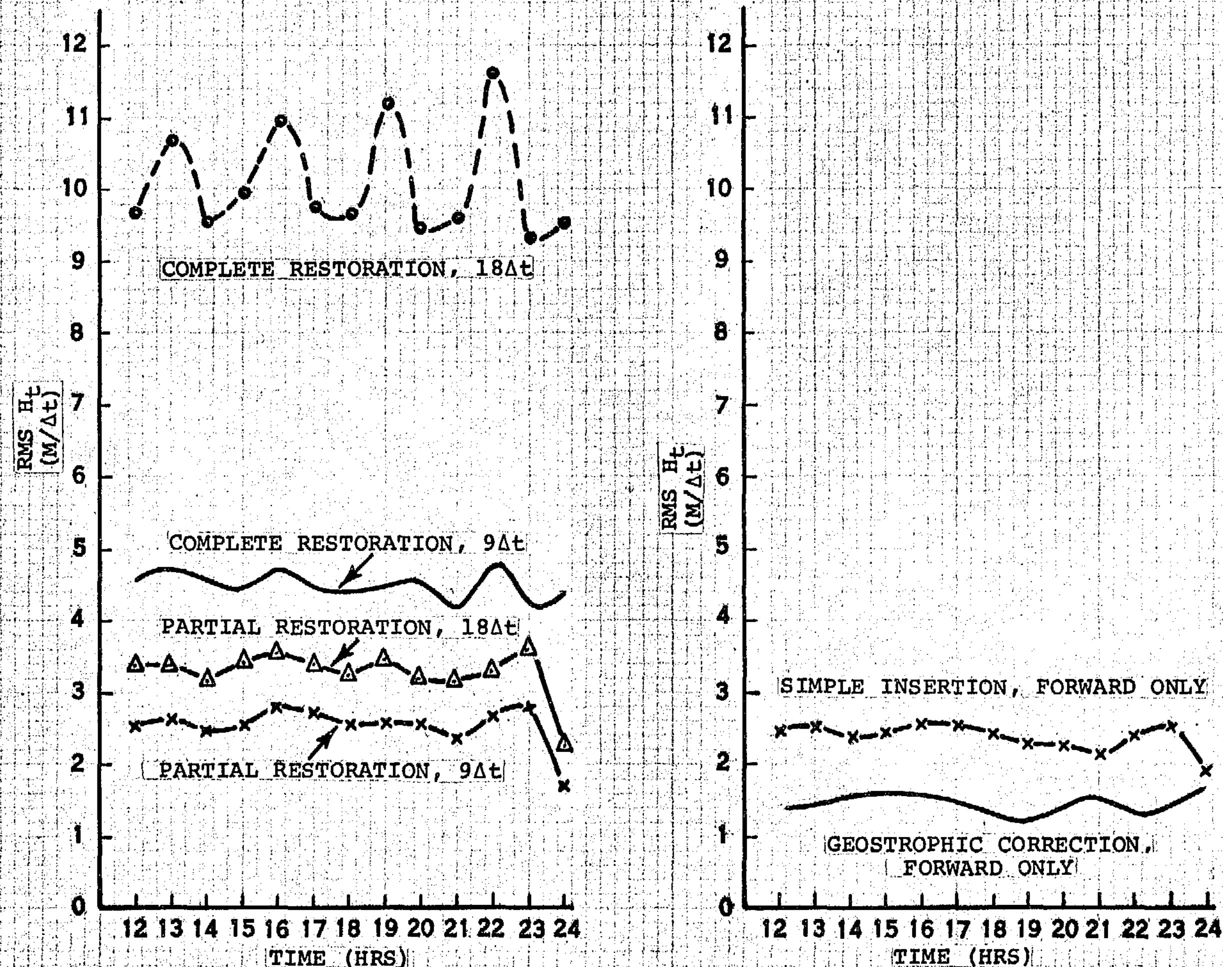

Figure 2. RMS H_t AS A FUNCTION OF TIME DURING BACKWARD EVALUATION