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

OFFICE NOTE 266

Optimum Interpolation: Practical Aspects of Operational Application

Ronald D. McPherson
Development Division

SEPTEMBER 1982

1. Introduction

Operational experience with optimum interpolation at NMC and elsewhere has stimulated close examination of the characteristics of the method. Some of the questions have arisen out of the basic formulation of the method; for example, how sensitive is the resulting analysis to variations in the theoretical forecast error covariance model? Others result from compromises necessary to implement optimum interpolation on existing computers for execution within operational deadlines. For example, computer limitations require restricting the amount of data influencing the analysis at any point; one can then enquire what the implications of this are, and what is the proper way to select observations to be used.

At NMC, examination of these aspects of optimum interpolation came about as a result of questions concerning the method's ability to resolve relatively small-scale features in atmospheric flow patterns. In the 6-hour analysis/forecast cycle which constitutes the NMC Global Data Assimilation System, occasions have been frequently noted in which rapid cyclogenesis is not adequately represented. Usually, this is manifest by the 6h predictions of the cyclogenesis being too slow and of insufficient intensity. The corrections which the optimum interpolation analysis makes in such cases tend to be localized, of relatively small scale, but sometimes of considerable amplitude. Maps of these correction fields display "bullseyes", with horizontal dimensions ranging from about 10 degrees latitude in diameter to about 40 degrees with amplitudes (in the geopotential field) of more than 100m in mid-troposphere, in extreme cases. Figure 1 displays one such example, from a case examined by Kistler and Parrish (1982).

Early in the operational life of the optimum interpolation system at NMC, it was noted that such features on the small end of the length scale range were not analyzed as accurately as those with larger dimensions; that is, the analyses in such cases did not reflect the data as faithfully as might be desired. In particular, small-scale features were not analyzed as small enough in horizontal dimensions, or with enough intensity. To explore the factors influencing this apparent scale limitation, a series of one-dimensional analysis simulation experiments was performed.

The next section discusses the design of the experiments, followed by a discussion of the main results. A summary concludes the note.

2. Experimental Design.

The characteristics of the analysis model are outlined as follows:

- Analysis grid: 1-dimensional, along latitude 45, with analysis points at 2^o intervals

- Data: True convection field specified analytically as a function of longitude:

$$h(\lambda) = A\cos\frac{2\pi\lambda}{L}$$

$$v(\lambda) = \frac{g}{fr\cos\phi}\frac{\partial h}{\partial\lambda}$$

Observations composed of true field plus random error with zero mean and standard deviation E(h), E(V).

32 equally spaced observation points at 2.5^o intervals
64 observations total

- Statistics: First guess error covariance model

$$\overline{h_i h_j} = \sigma_h^2 \; e^{-ks^2}, \; s = \text{separation distance between i and j}$$

$$\overline{h_i v_j} = \frac{g}{fr\cos\phi}\frac{\partial(\overline{h_i h_j})}{\partial\lambda}$$

$$\overline{v_i v_j} = \frac{g^2}{f^2 r^2\cos^2\phi}\frac{\partial^2}{\partial\lambda^2}(\overline{h_i h_j})$$

First guess error standard deviation

$$\sigma_h = 40 \text{ m}$$

$$\sigma_v = \left(\frac{g}{f}\sqrt{2k}\right)\sigma_h$$

Observational error standard deviation

$$E(h) = 20 \text{ m (in most experiments)}$$

$$E(v) = \left(\frac{g}{f}\sqrt{2k}\right)E(h)$$

To illustrate the problem of scale limitation, the simulated true correction field was given horizontal dimensions varying from 40° latitude to 10° latitude with $k=1x10^{-6}km^{-2}$. Figure 2a shows the response of the analysis to a feature with dimensions of 40° longitude from zero value to zero value. The solid lines depict the "truth" - that is, the analytic field - and the dotted lines represent the analysis. Simulated observations are shown by crosses. In this experiment, all 32 geopotential observations and 32 wind observations were used in the analysis at every point. It will be noted that the analyzed geopotential is a close approximation of the truth, and that the random noise in the data has been eliminated: the minimum value in the true field, -100m, is analyzed as -91m, and the analysis is agreeably smooth.

Figures 2b-2d show the same depiction but with the dimensions of the true correction field progressively reduced to 25°, 15°, and 10° longitude. For the 10° case, the minimum analyzed value is only -53m, and the analyzed dimensions are larger - the wave length is about 22° latitude.

The effort to investigate this behavior included simulation experiments on:

- o multivariate vs. univariate analysis;
- o variations in the forecast error covariance model ($k=1,2,4 \times 10^{-6}km^{-2}$);
- o variations in the number of observations used (5, 8, 10, 20, 64);
- o variations in the observational error standard deviations (E = 20, 10, 5, 2m);
- o variations in data selection method (closest vs. most highly correlated).

3. Results

Univariate vs. Multivariate Analysis.

The first set of experiments was suggested by the work of Lorenc (1981), who showed an example of the improvement of the mass analysis when wind observations are used to augment the mass observations. Figure 3 displays the results of the present experiment, for the case of the shortest-length disturbance (10° longitude). In the univariate analysis, only the 32 mass observations were used in the mass analysis, and only the 32 wind observations were used in the wind analysis. All 64 reports were used in both mass and wind analyses in the multivariate case. The experiments were otherwise identical.

The univariately-analyzed mass field (dashed line) exhibits somewhat greater distortion than does its multivariate counterpart. The minimum value analyzed is about -40m in the univariate case, compared to -53m in the multivariate case, and -100m in the analytic "truth". The analyzed wavelength is also larger in the univariate analysis than in the multivariate. Virtually no difference was noted in the wind field, so only the dashed line is included. The entry in Table 1 for these experiments shows a 3.3m reduction in height analysis error, and a 0.1m/s reduction in wind error in the multivariate compared to the univariate analysis.

These results suggest that the mass analysis does indeed benefit from using wind as well as mass reports, and although it has not been shown here, the greatest improvement occurred with the smallest scale perturbation. It is not clear from these experiments that mass data improved the wind analysis to any appreciable degree.

Variations in the Forecast Error Covariance Model

The shape of the forecast error covariance model is given by the Gaussian function $\exp(-ks^2)$ where s is the separation distance and k is a constant which governs how rapidly the correlation curve falls off with distance. Three values of k were used, and the resulting functions are displayed in Figure 4: height-height correlations in the upper part and height-wind in the lower part.

For larger values of k, the correlation curve becomes sharper, and it would be anticipated that the narrower the function the more easily small-scale features might be represented. This is confirmed in Figure 5, showing the results obtained for the three values used, in the case of the 10° perturbation. It is clear that the narrower correlation function produced a sharper, more accurate response in the analysis as expected. The extreme in the mass field decreases from -53m to -78m and the RMS error (Table 1) declines from 15.4 m to 7.1 m. Similar improvements are noted in the wind analysis.

This experiment clearly demonstrates the important role of the forecast error covariance model in determining the scale response of the analysis methods.

Variations in the Number of Observations

As noted in the Introduction, the current computational capabilities available to NMC require restricting the number of reports used at each analysis point to no more than 20. The experiment reported in this section was motivated by two considerations: first, if the n observations selected (n less than the total available) are those closest to the analysis point, then the correlation functions displayed in Figure 4 are effectively truncated, thus becoming narrower and therefore capable of greater response to small-scale features. On the other hand, a recent

theorem by Phillips (1982) concerning the completeness of multivariate analysis for the "slow" modes requires as a necessary condition the use of all available data in the analysis at each analysis point.

The experiment progressively reduced the number of reports used at each point from the maximum of 64 to 20, 10, and then 5. An intermediate value of the shape factor k was used: $k=2\times10^{-6}km^{-2}$. In each of the reduced-data cases, observations were taken in pairs of height and wind, the closest pair to the analysis point being selected first. Thus, in the case of 20 reports, the procedure at each analysis point used the 10 closest geopotential reports and the 10 closest wind reports. In the last experiment (5 reports) both the mass and wind analyses used three heights and two winds at each analysis point.

Figure 6 displays the results for the 10° disturbance. The curves for 64 and 20 reports are virtually indistinguishable from each other; 10 reports produces a slightly more accurate analysis for this case both for the extreme values and in the RMS sense (Table 1), but there is clearly substantially more noise apparent, especially in the flat areas outside the disturbance itself. A further reduction to 5 reports is definitely disadvantageous.

Thus, this set of experiments suggests that attempts to improve the short-length scale response by greatly reducing the amount of data used in each analysis point will be attended by unpleasant side effects. This should not be surprising, given that a reduction in data amounts to a truncation of the correlation function, with predictable results in the response. On the other hand, the reduction to 20 reports from 64 produced no detectable deterioration in the analysis. This suggests that although Phillips' theorem formally requires all data to be used at each analysis

point, practically it is sufficient to use only those closest to each analysis point, provided that enough are used to suppress the effects of random errors. The number required would then depend in part on the magnitude of the observational errors.

Variations in Observational Error Variances

Optimum interpolation will exactly reflect error-free data located at analysis points. Therefore, it was thought that the inability to represent the 10° disturbance as in figure 2d perhaps might be alleviated, at least in principle, if the analysis model were given error free data, and told of that fact by setting the observational error variances to zero. To examine this possibility, the $(20m)^2$ geopotential error variance was reduced in successive runs to $(10m)^2$, $(5m)^2$, and $(2m)^2$ respectively. An experiment with error-free data was not possible. The iterative method used to solve the analysis equations at each point produced erratic solutions, due to weak diagonal dominance present in the coefficient matrix when the diagonal terms are not augmented by the observational error variance. It should also be noted that the data were not relocated to analysis points so interpolation continued to be required.

Figure 7 displays the results. Clearly, the analysis model attempts to more closely reflect the data as the error variance is reduced. But it appears to be approaching a limit asymptotically such that continued halving of observational would produce a progressively smaller change in the solution. It would appear that, even with perfect data, the solutions would still reflect the underestimate of amplitude and overestimate of length scale evident in Figure 2d, due to the interpolation with a fairly broad correlation function. No doubt this response could be determined analytically through information theory, but no attempt has been made to do so for this note.

Variations in the Method of Data Selection

Given the necessity of using less than the total available data base at each analysis point, a selection procedure must be developed to identify those observations which contain the most information about the analysis. At the inception of the NMC optimum interpolation analysis system, the decision was made to select those observations which exhibit the highest correlation with the analysis point, on the grounds that this minimizes the analysis error variance for each variable. Because the shapes of the height, u-, and v-correlation functions are quite different from each other (see Bergman, 1979), observations were selected separately for the mass, u-, and v-analyses. In general, this means that a different set of observations may be used in the analysis of each parameter at any point. In particular, where data of only one type is available - for example, satellite-derived geopotentials over oceans - an analysis of geopotential will select those observations which are closest to the analysis point. Those reports have the highest correlation with the analysis point, as may be seen in the upper part of Figure (the analysis point being assumed at the central point) 4. For the wind analysis, however, the most highly correlated height reports, are those located several degrees of latitude away, as may be noted from the ZV correlation curves in the lower part of Figure 4. If the number of data permitted to influence the analysis are small, then the data sets for the mass and motion analyses may be completely different. The potential for mass-motion imbalance is evident.

Upon examination of the changes made to analyzed mass and motion fields, by the normal mode, initialization step in the vicinity of large, small-scale corrections such as in Figure 1, it became apparent that not only was the mass correction underestimated by the optimum interpolation

procedure, the initialization step made the situation even worse. This suggested that the selection procedure, even though minimizing the analysis error variance, led to wind and mass corrections which were unbalanced; the imbalance was then resolved by the initialization according to geostrophic adjustment principles.

The final experiment reported in this note was intended to examine this problem. Only height observations were used, and only eight reports at any analysis point. For the mass analysis, the eight most highly correlated were selected at each point. These are marked with circumscribing circles for the analysis point at the center of the grid. It should be noted that the data spacing was decreased to 1.25° for this experiment in order to amplify the results for easy illustration.

For the wind analysis, the experiment marked "best" in Figure 8 selected the eight most highly correlated height reports with the analyzed wind at the analysis points. Again, for the center point, these are marked by circles in the lower part of Figure 8. It will be noted that these are not the same set selected for the mass analysis.

The dashed curve in the lower part of Figure 8 is the resulting wind analysis. It is severely distorted, with very small magnitude in general, and displaying relative maxima where the analytic solution is a minimum and vice versa. It clearly differs from the geostrophic wind that might be calculated from the analyzed height correction curve in the upper part of Figure 8. The analyzed height and wind correction curves thus are poorly balanced.

Alternatively, if the selection procedure is based on the eight closest height reports in the height analysis, and the same set is used in the wind analysis, the dotted curve (marked "closest") in the lower part of

Figure 8 results. It is obviously an underestimate of the analytic wind correction, reflecting the inability to represent the intensity of small features and therefore a weaker gradient. But at least the principal extrema are in approximately correct positions, so that the mass and motion corrections are much better balanced and less prone to confuse the initialization procedure.

As a note of passing interest, the upper row of numbers between the height and wind parts of Figure 8 represents the estimated analysis error standard deviations at each analysis point for the "best" experiment, while the lower represents the same parameters for the "closest" experiment. The former are indeed smaller than the latter, but the actual solutions of the latter are clearly superior.

IV. Summary

The one-dimensional simulation experiments reported in this note suggest the following conclusions:

- The scale limitation in optimum interpolation is controlled mostly by the shape of the covariance model, and to some extent by the quality of the data.
- For the convariance model in use at NMC (Gaussian, $k=10^{-6}km^{-2}$), features in the correction field of scale less than about 20° latitude will not be represented faithfully.
- For data with random errors, it is important to use enough data to reduce the effect of errors, but it does not appear necessary to use all of the data.
- Data selection should result in the same set of observations being used in both mass and motion analyses.

References

Bergman, K., 1979: A multivariate optimum interpolation analysis system for temperature and wind fields. Mon. Wea. Rev., 107, 1423-1444.

Kistler, R., and D. Parrish, 1982: Evolution of the NMC Data Assimilation System: September 1978-January 1982. Mon. Wea. Rev., 110, 1335-1346.

Lorenc, A., 1981: A global three-dimensional multivariate statistical interpolation scheme. Mon. Wea. Rev., 109, 423-442.

Phillips, N., 1982: On the completeness of multivariate optimum interpolation. Mon. Wea. Rev., 110, 1329-1334.

Table 1. Summary of one-dimensional analysis simulations, in terms of RMS errors (exp. minus anlaytic truth).

WAVELENGTH EXPERIMENTS
Multivariate, all data, $k = 1 \times 10^{-6}$ km^{-2}, $e_h = 20$ m

Exp	RMS h_e (m)	RMS V_e (m/s)
$\lambda = 40^0$	3.3	1.3
$\lambda = 15^0$	11.3	6.5
$\lambda = 10^0$	15.4	10.8

UNIVARIATE VERSUS MULTIVARIATE
10^0 wavelength, all data, $k = 1 \times 10^{-6}$ km^{-2}, $e_h = 20$ m

Exp	RMS h_e (m)	RMS V_e (m/s)
Multivariate	15.4	10.8
Univariate	18.7	10.9

VARIATIONS IN THE COVARIANCE MODEL
10^0 wavelength, multivariate, all data, $e_h = 20$ m

Exp	RMS h_e (m)	RMS V_e (m/s)
$k = 1 \times 10^{-6}$ km^{-2}	15.4	10.8
$k = 2 \times 10^{-6}$ km^{-2}	11.4	8.7
$k = 4 \times 10^{-6}$ km^{-2}	7.1	5.6

NUMBER OF OBSERVATIONS
10^0 wavelength, multivariate, $k = 2 \times 10^{-6}$ km^{-2}, $e_h = 20$ m

Exp	RMS h_e (m)	RMS V_e (m/s)
NOBS = 64	11.4	8.7
NOBS = 20	11.0	8.4
NOBS = 10	7.1	5.8
NOBS = 5	11.5	5.4

Table 1, continued.

OBSERVATIONAL ERROR
10^{o} wavelength, multivariate, $k = 1 \times 10^{-6}$ km^{-2}

Exp	RMS h_e (m)	RMS V_e (m/s)
e_h = 20 m	15.4	10.8
e_h = 10 m	13.8	10.4
e_h = 5 m	12.7	9.9
e_h = 2 m	12.4	9.8

DATA SELECTION
10^{o} wavelength, multivariate, 8 height obs only, $k = 1 \times 10^{-6}$ km^{-2}

Exp	RMS h_e (m)	RMS V_e (m/s)
BEST	17.8	12.1
CLOSEST	17.2	11.0

Table 2. Experimental variations and corresponding entries in the legends of the diagrams.

EXPERIMENTAL VARIATIONS		
	Parameter	
Variables		
Univariate or multivariate	CROSS = 1	M/V
	CROSS = 0	U/V
Covariance Function		
$k = 1 \times 10^{-6}$ km^{-2}	CH = 40	
$k = 2 \times 10^{-6}$ km^{-2}	CH = 80	
$k = 4 \times 10^{-6}$ km^{-2}	CH = 160	
Number of Observations	NOBS	
Observational Error	E	
Data Selection		
Heights only	CROSS = -1	
Most highly correlated	BEST	

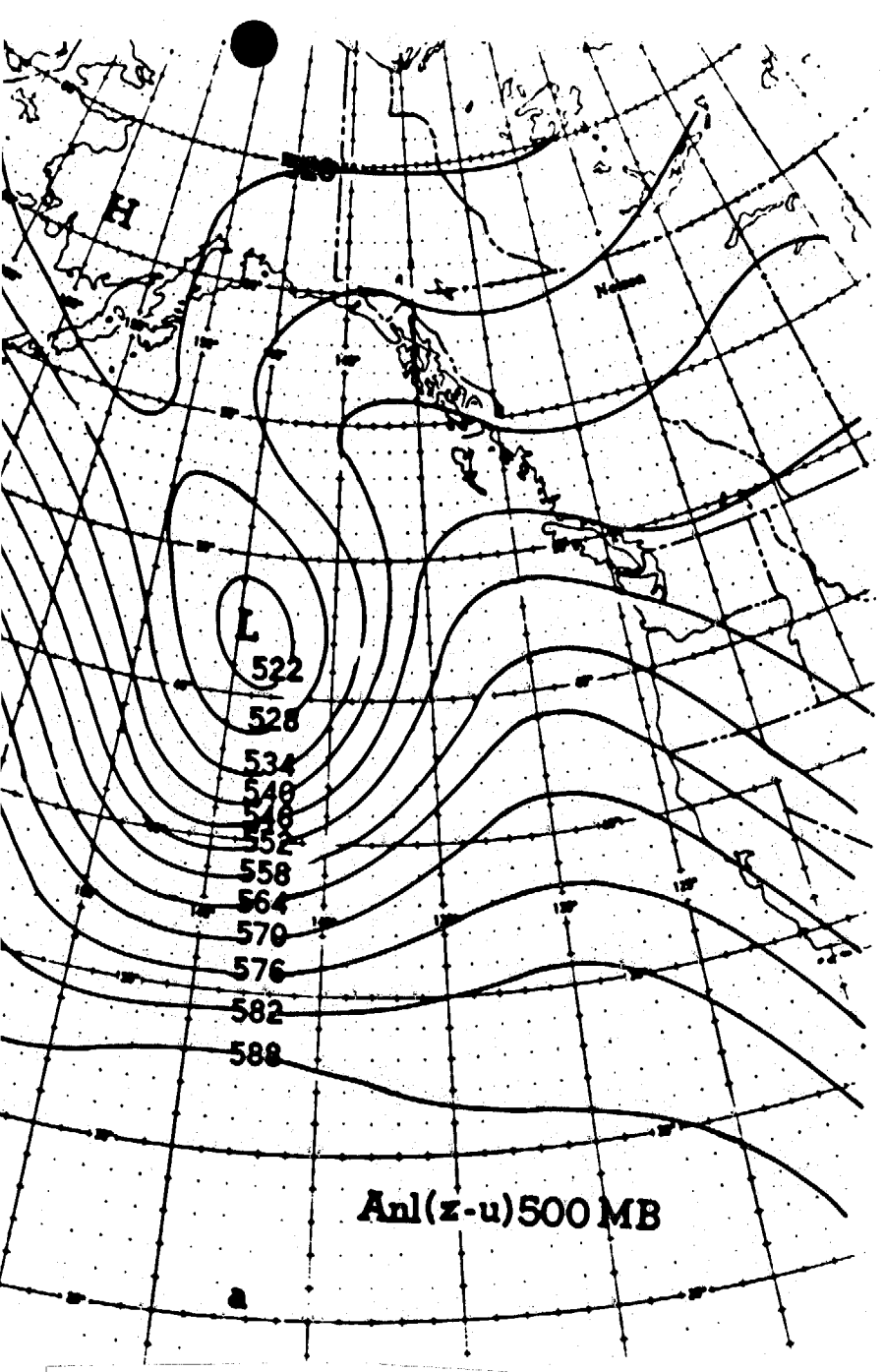

Figure 1a. 500-mb geopotential analysis for 12Z 21 October 1979 produced by the NMC Data Assimilation Cycle. Units are decameters.

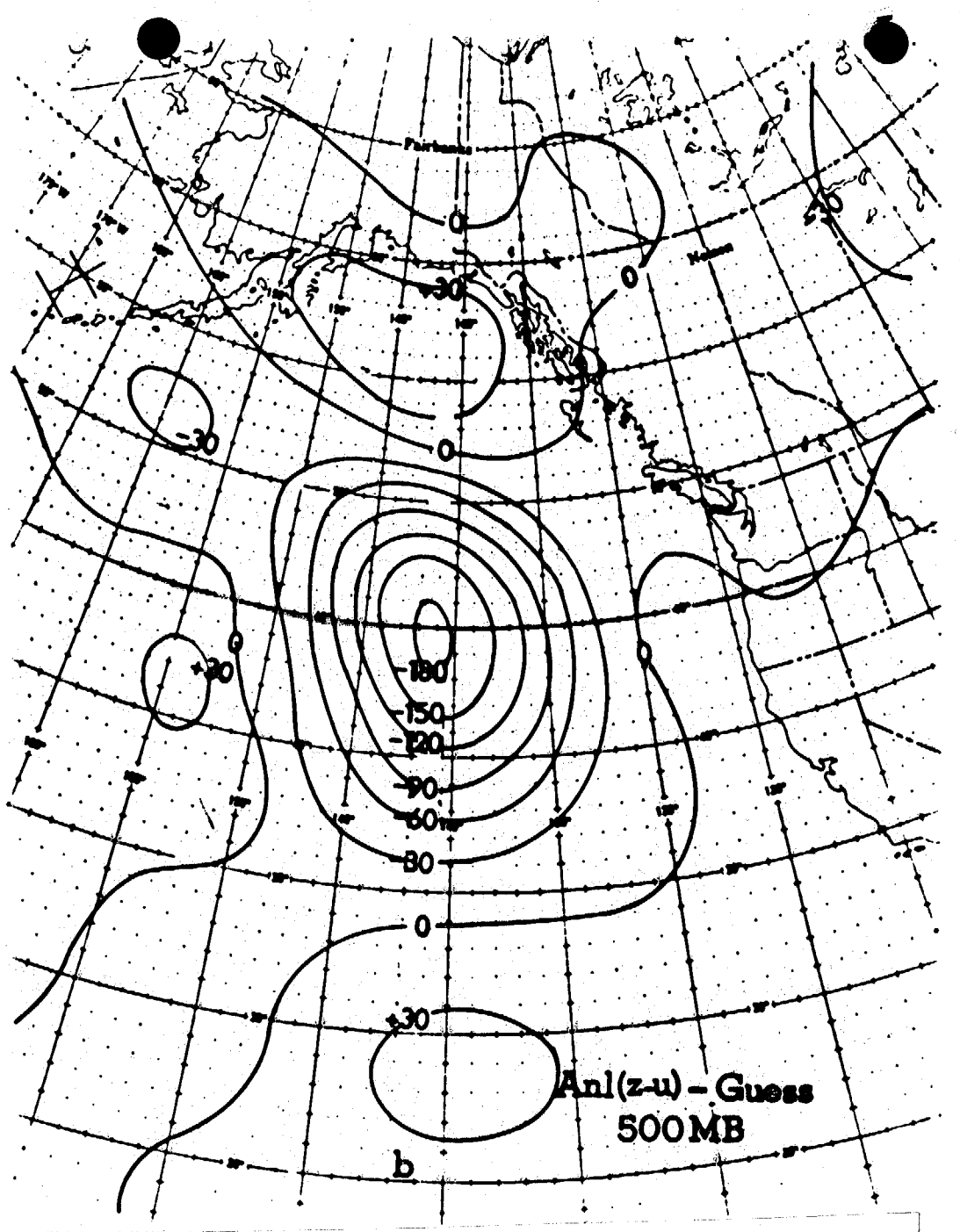

Figure 1b. 500-mb geopotential correction (difference between first guess and analysis) corresponding to (a) Units are meters. (After Kistler, and Parrish, 1982)

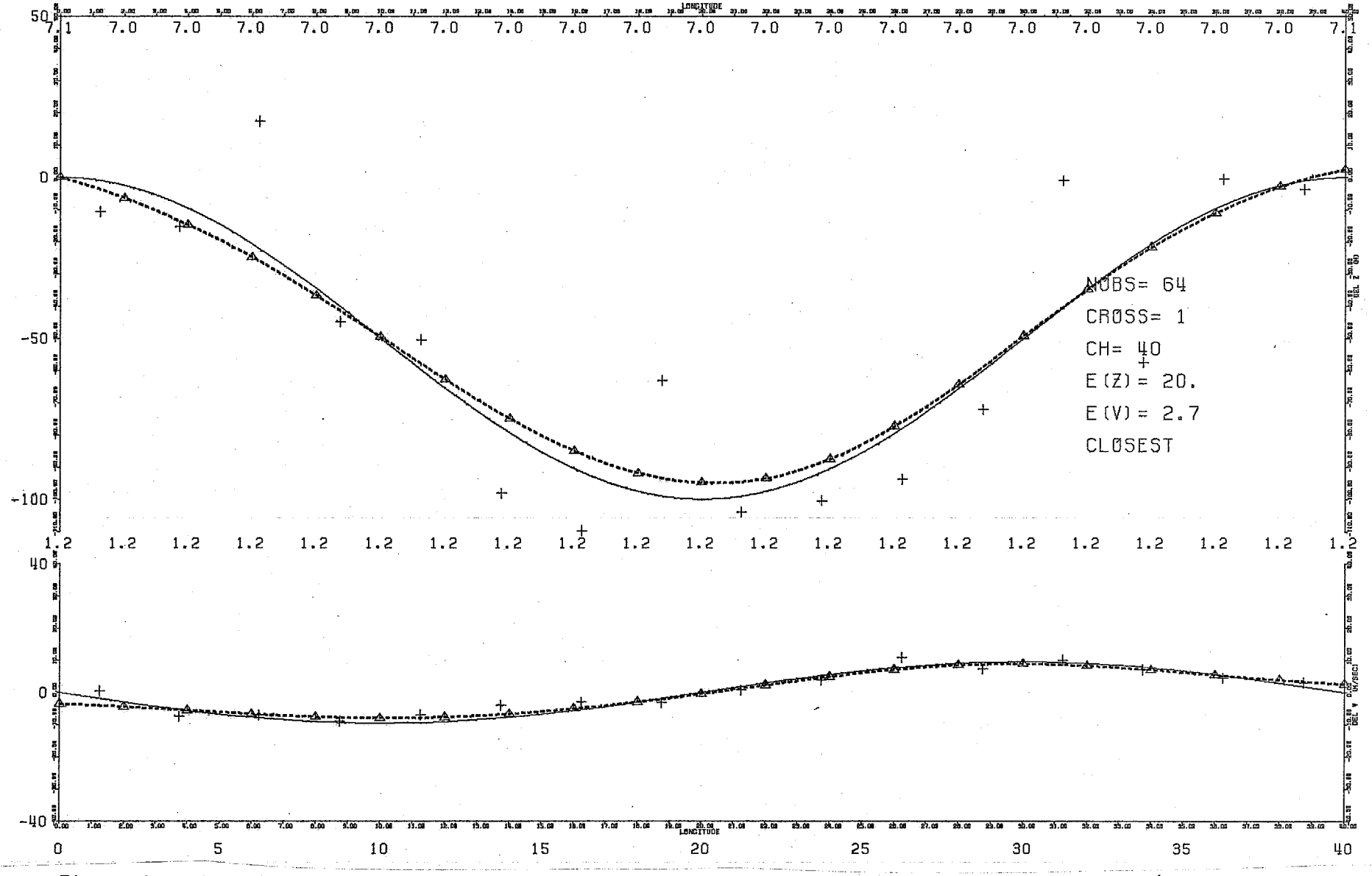

Figure 2a. One-dimensional analysis simulation, with the length of the true correction field (forecast minus observed) specified as 40 degrees latitude. Solid lines represent the given analytic true field, and the dashed lines depict the analysis. Crosses indicate locations and values of simulated observations, with random errors corresponding to RMS values of E(z) and E(v). The upper part of the diagram shows the geopotential, and the lower part shows the wind component normal to the diagram. See text and Table 2 for further explanation.

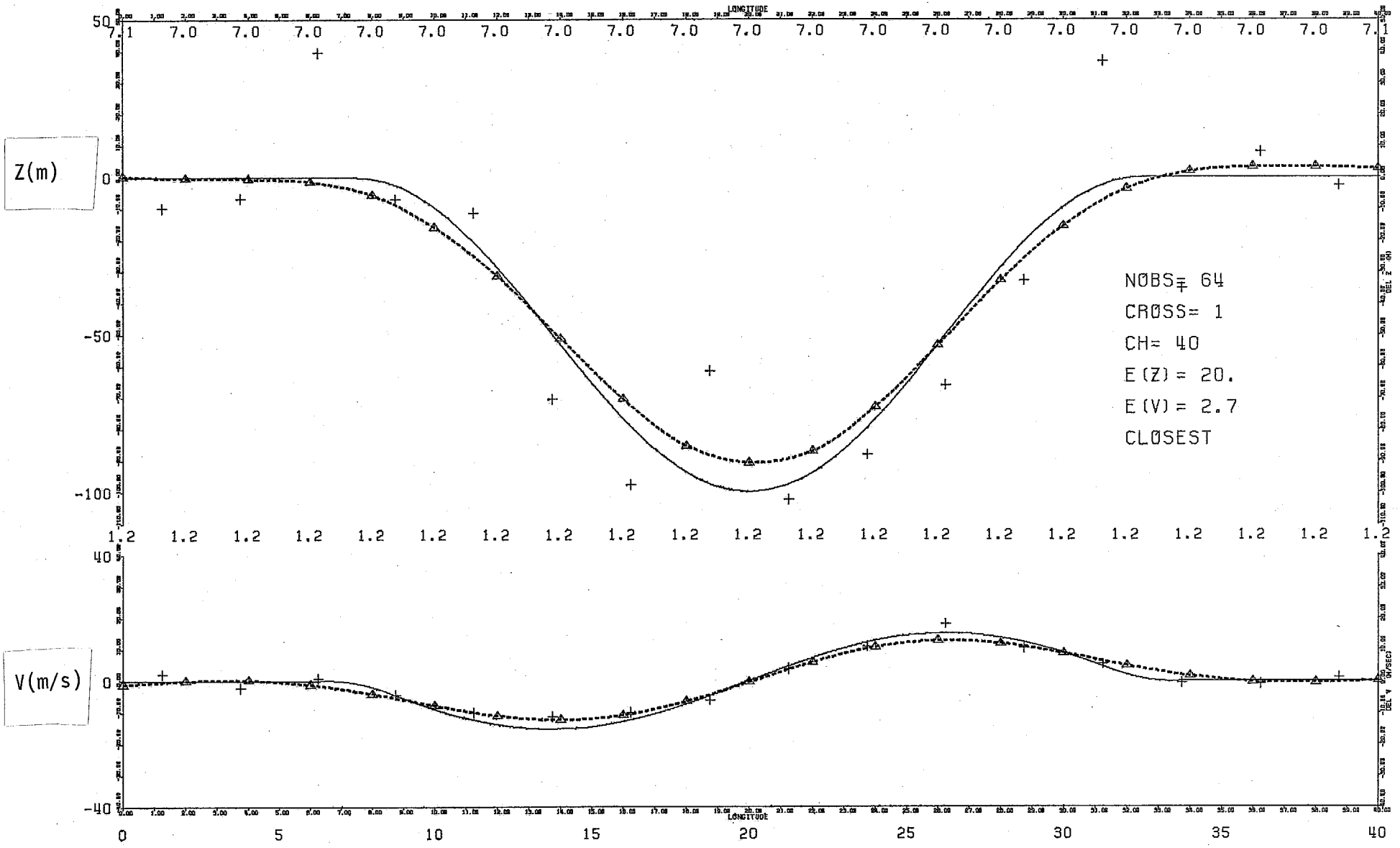

Figure 2b. Same as Figure 2a, except wavelength is reduced to 25 degrees latitude.

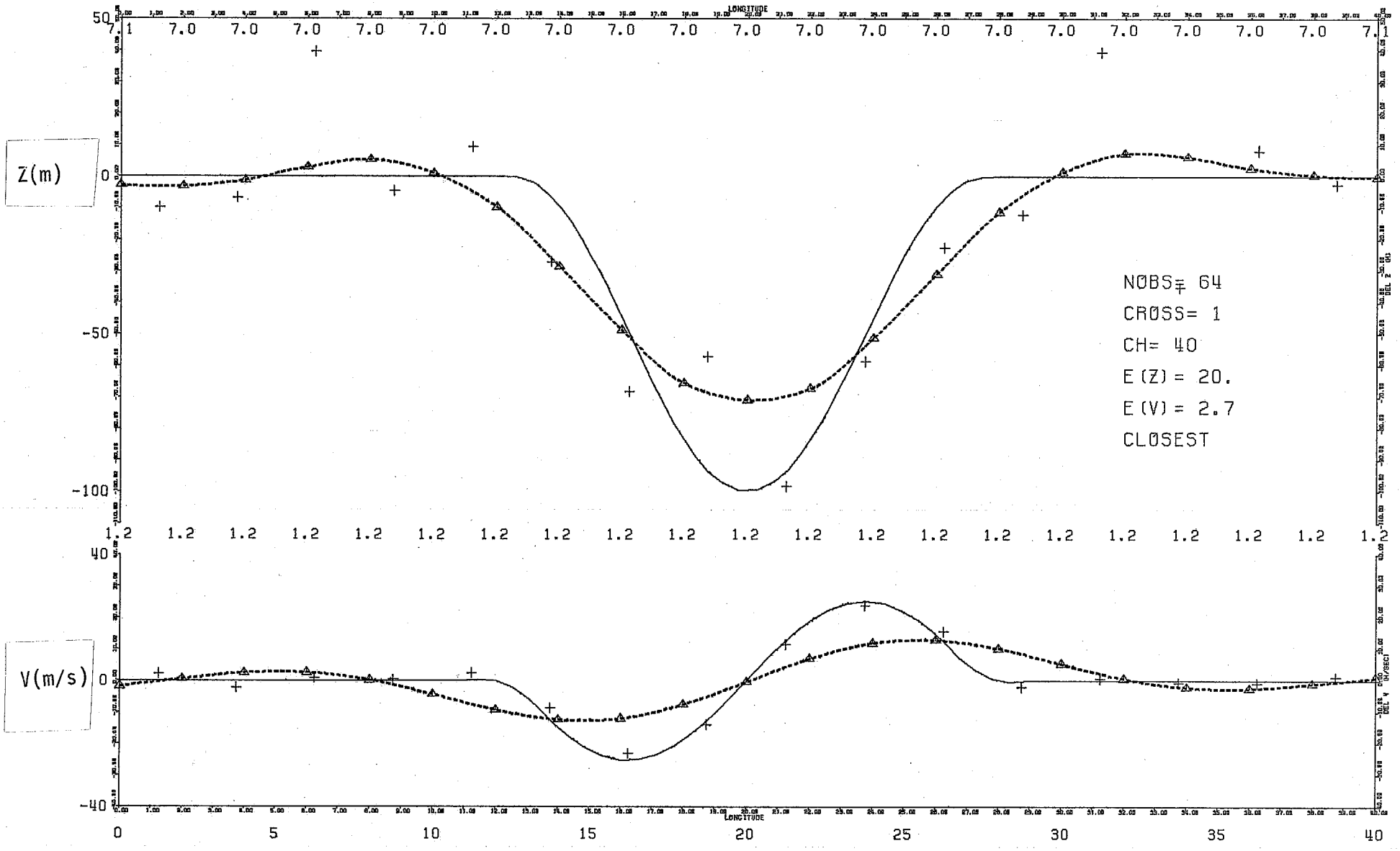

Figure 2c. Same as Figure 2a, except wavelength is reduced to 15 degrees latitude.

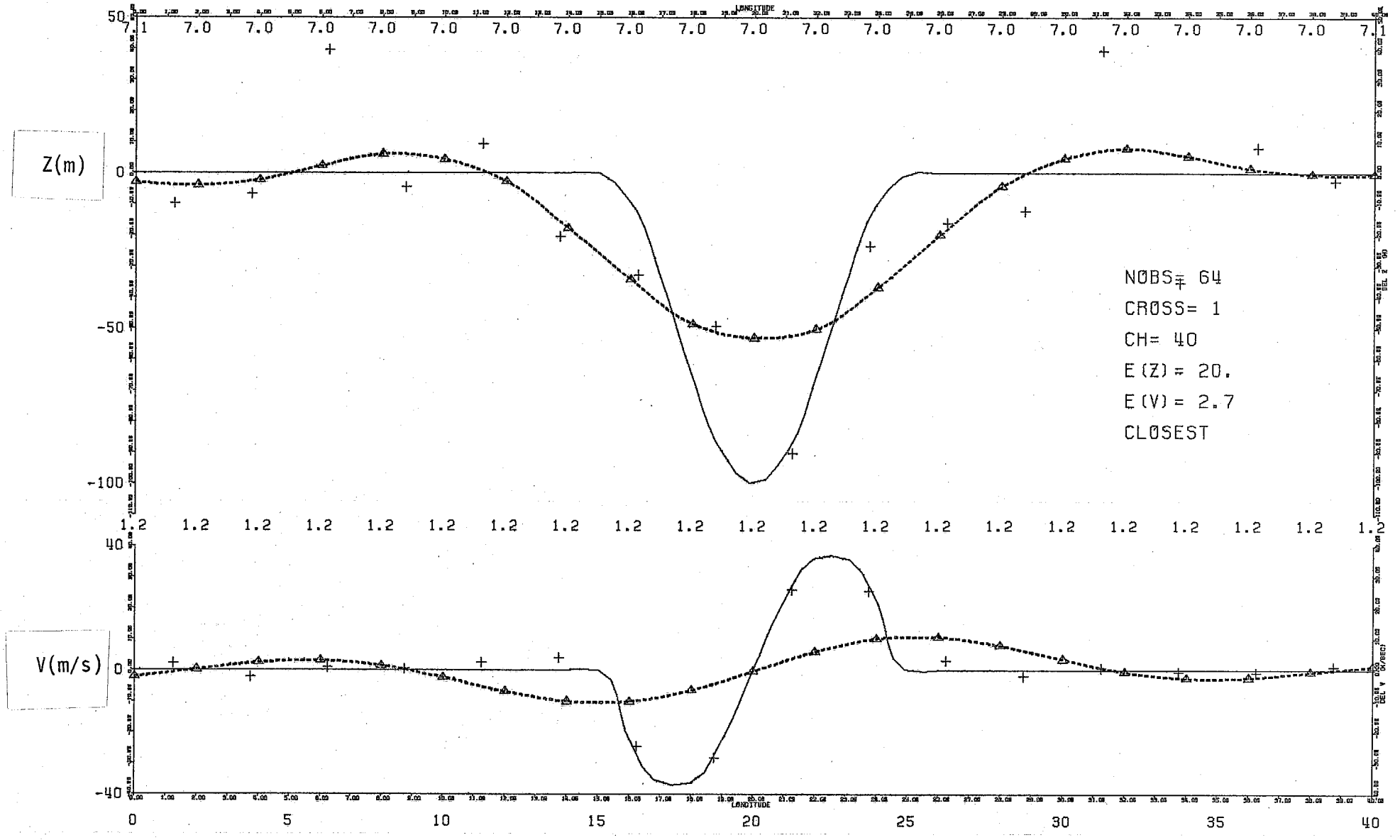

Figure 2d. Same as Figure 2a, except wavelength is reduced to 10 degrees latitude.

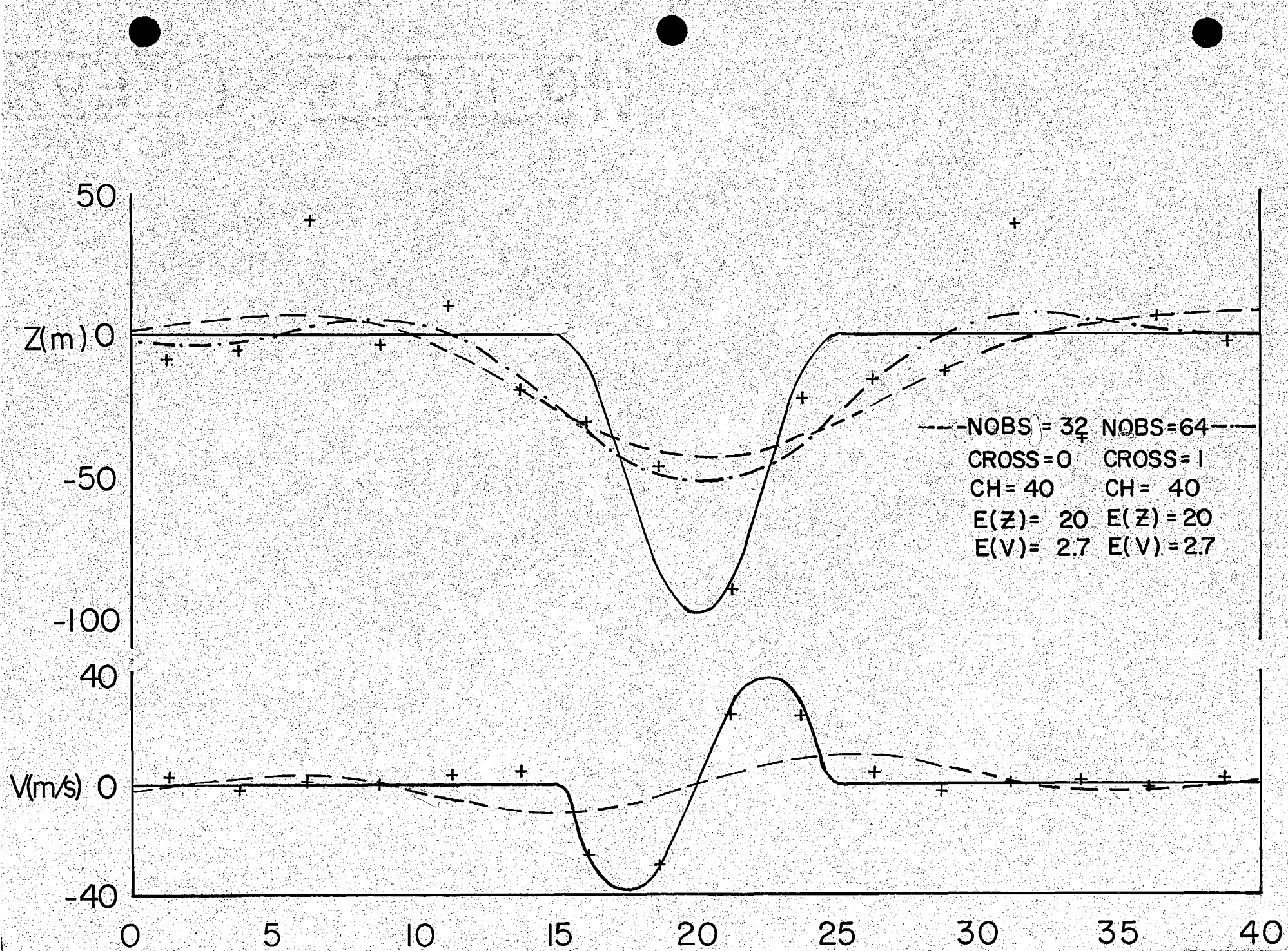

Figure 3. One-dimensional analysis simulation comparing univariate analysis (dashed lines) with multivariate analysis (dash-dot lines). See text and Table 2.

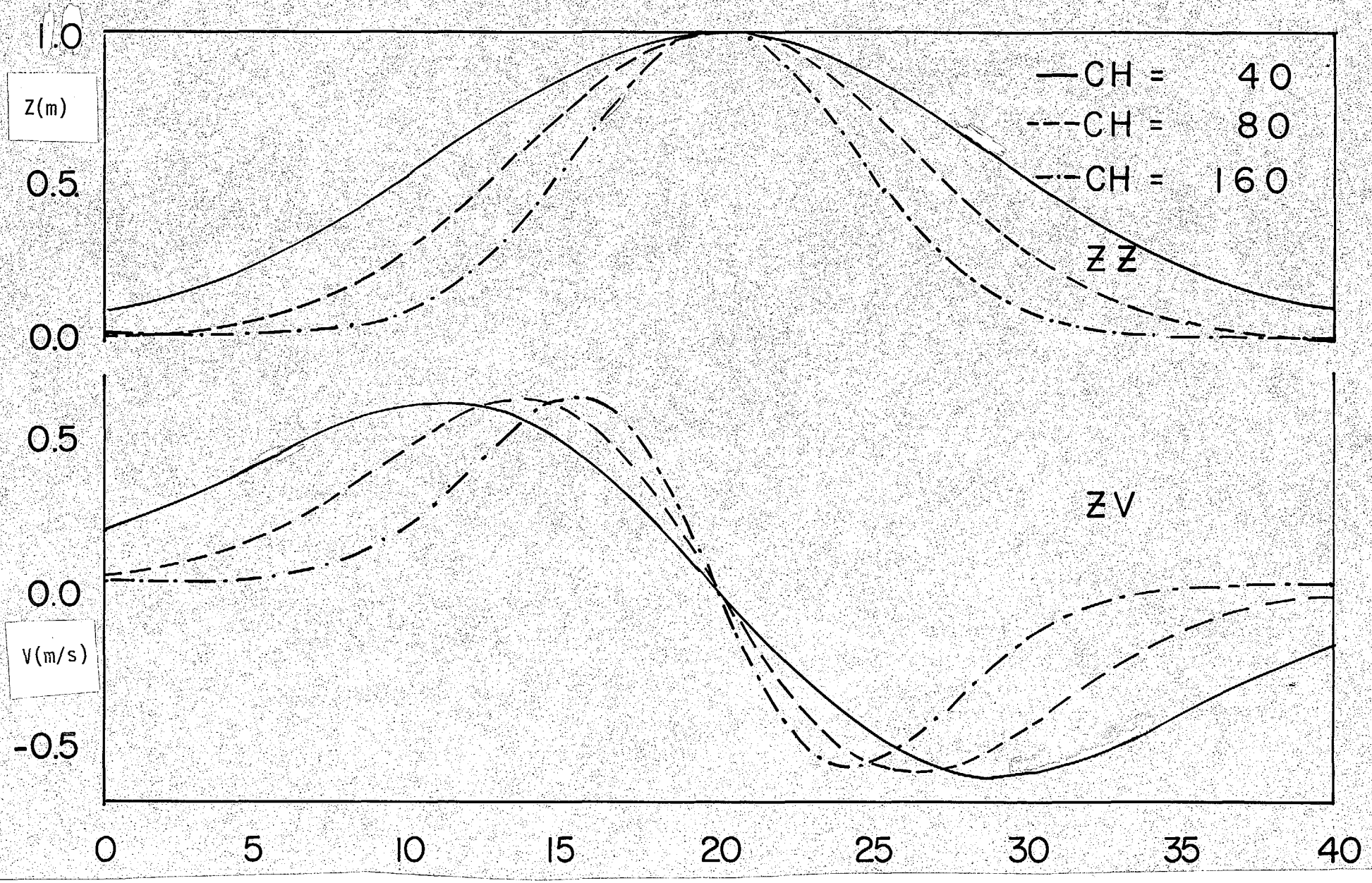

Figure 4. Shapes of the forecast error covariance functions used in the analysis simulation. See text and Table 2.

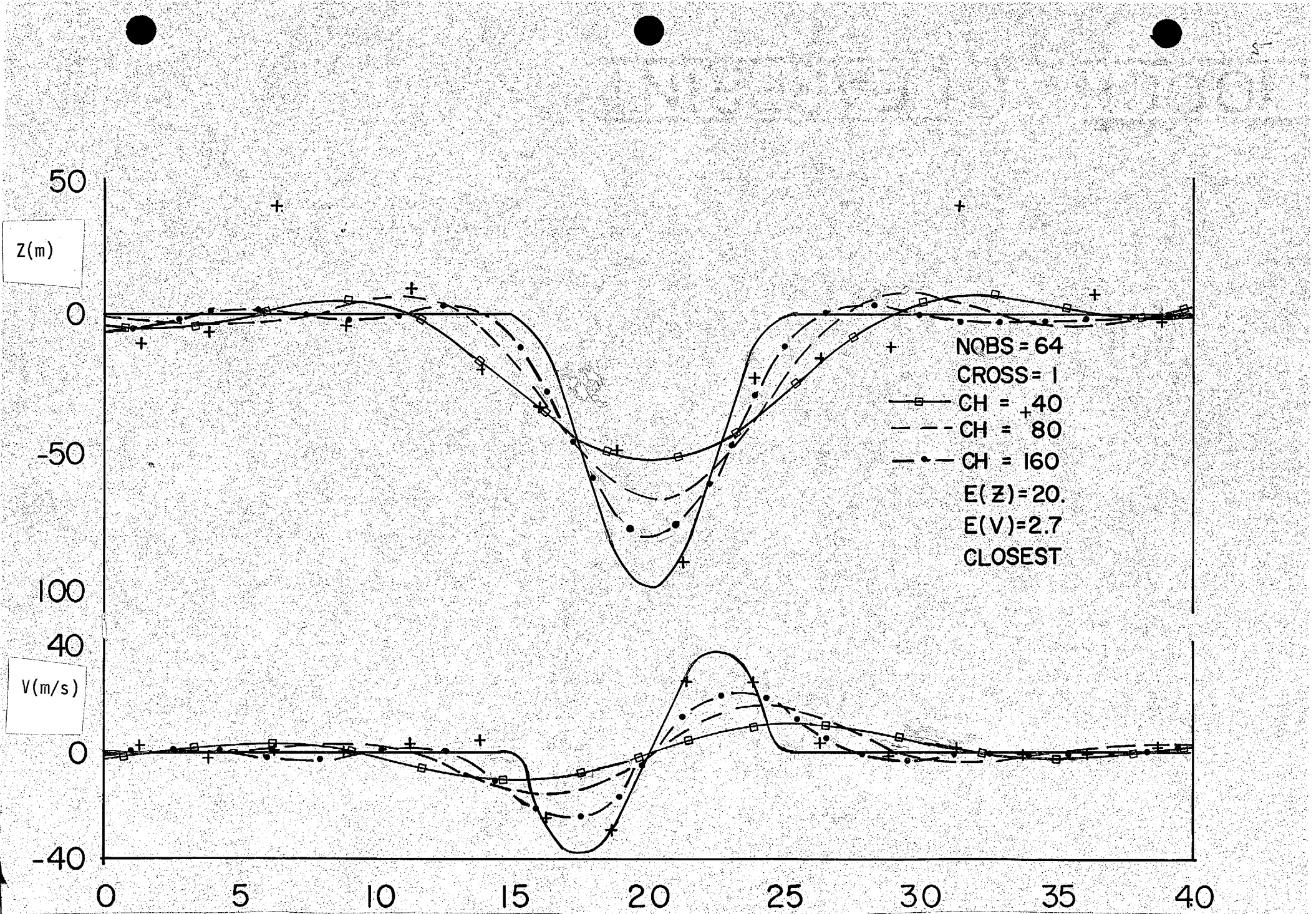

Figure 5. Analysis simulation comparing variations in the forecast error covariance model. See text and Table 2.

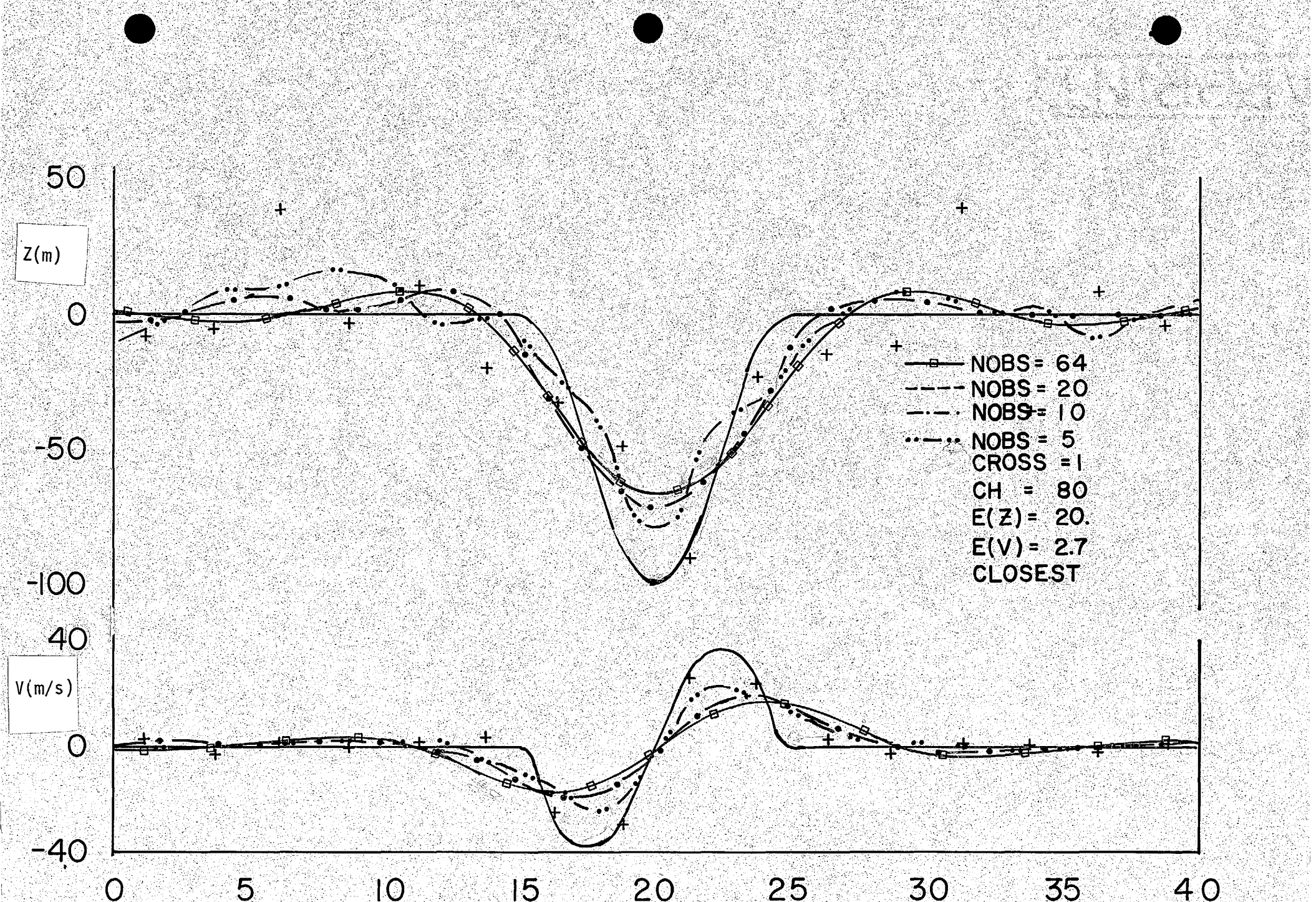

Figure 6. Analysis simulation comparing variations in the number of observations permitted to affect each analysis point. See text and Table 2.

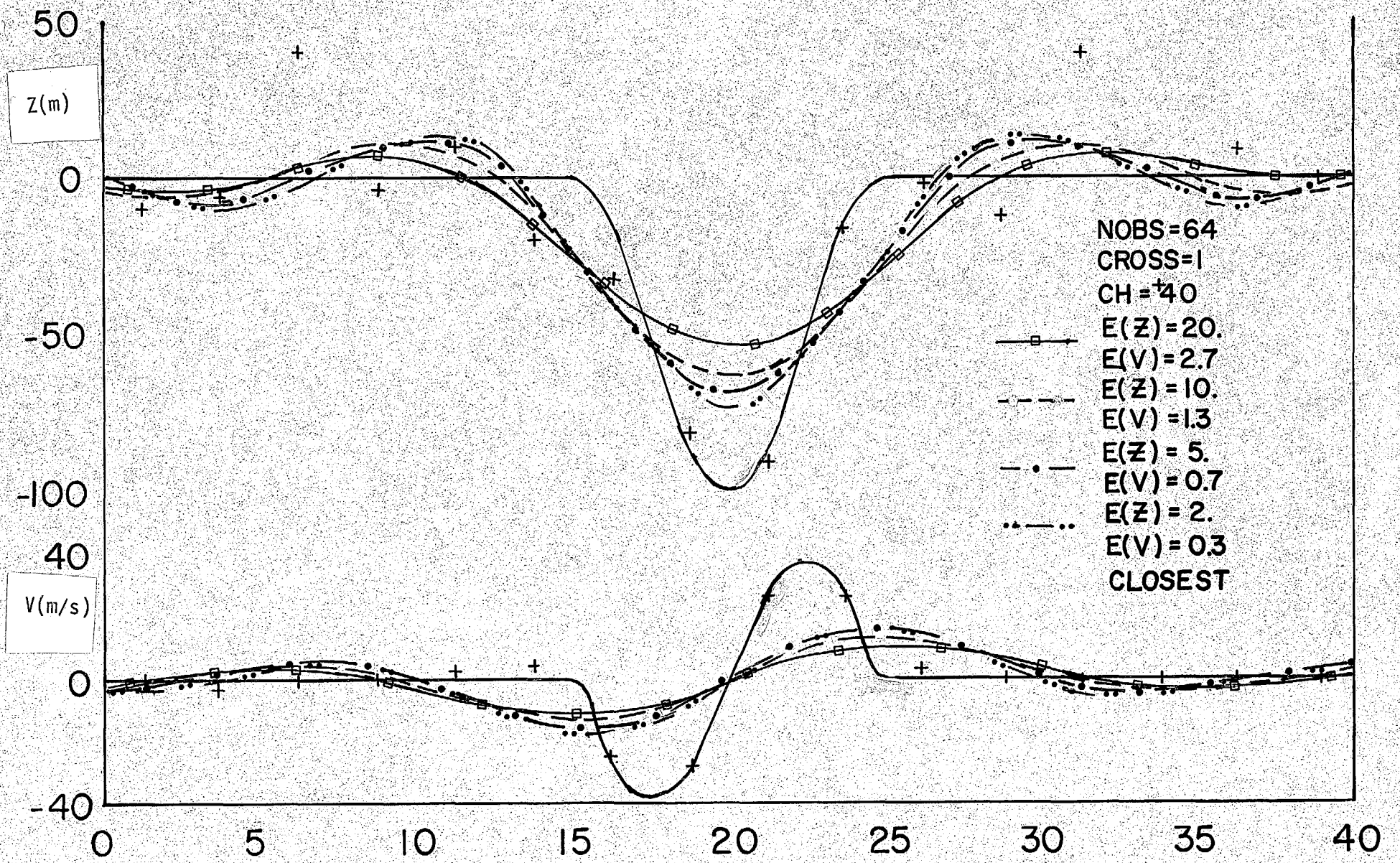

Figure 7. Analysis simulation of the effect of observational error variance. See text and Table 2.

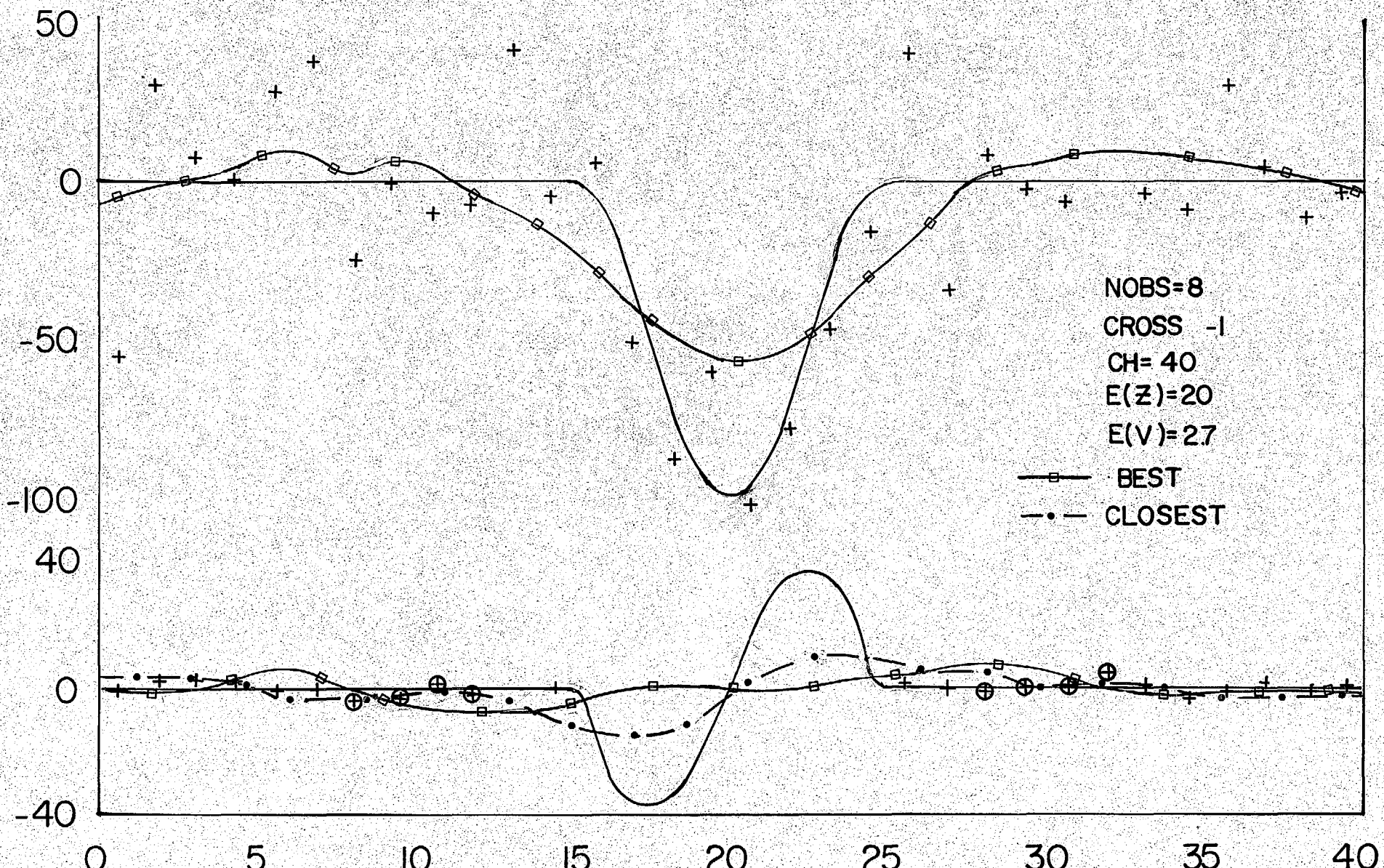

Figure 8. Analysis simulation of the effect of data selection procedures. See text and Table 2.