SPE 38264

A Comparison of Mapping Schemes for Reservoir Characterization

Eric D. Zwahlen, LBNL, and Tad W. Patzek, SPE, University of California, Berkeley

This paper was prepared for presentation at the 1997 SPE Western Regional Meeting held in Long Beach, California, 25–27 June 1997.

Abstract

Often one is interested in the relationships between various types of geological data, such as depth, formation thickness, position, permeability, porosity, etc. However, geological data are difficult to analyze due to their spatial complexity, and because obvious analytical maps of one type of data onto another seldom exist. Typically, multivariate regression techniques are used to correlate geological data. This approach requires often stringent assumptions about the functional form of the data. Here, we compare a method that does not require such assumptions: Alternating Conditional Expectation (ACE), with two other approaches that do: Geostatistics and MATLAB®. The ACE method, developed by Breiman and Friedman (1985), is an iterative optimal transformation that maximizes the correlation between the transformed dependent variable and the sum of the transformed independent variables. Even with multiple independent variables, the specific functional dependence of each variable can be found. A thorough geostatistical analysis of the full data set is treated as the ground truth, against which the two other methods are compared. The field example from the South Belridge diatomite, Kern County, CA, involves correlating the true vertical depths of diatomite layer boundaries with their *x* (EW) and *y* (NS) coordinates. The geological data from approximately 70 wells in an area about 1000 feet by 1000 feet have been used. We compare how well the three methods can smooth and interpolate the data to produce layer mappings. In the end, the most laborious of the three methods, geostatistics with Surfer® as a plotting tool, wins hands down. The other two methods are far more limited.

Introduction

Geologic data may be a function of many variables. The data also may contain scatter due to inherent heterogeneity and data collection errors. One must perform some type of regression to correlate the data with each of the variables. The functional form of a specific variable may give physical insight into the data, such as faulting or the location of a different type of rock.

In most regression schemes, a functional form is assumed, and then the equation is fit to the data to find the optimal coefficients. This introduces a bias into the regression. In this paper, we examine a nonparametric regression scheme for fitting data and generating surface plots of the data.

The purpose of this paper is to compare various methods for visualizing reservoir layer elevation ASL, *z*, as a function of *x* (easting) and *y* (northing). We use layer depth data from 70 wells in the Shell Phase III Steam Drive Pilot in the South Belridge Diatomite near Bakersfield, California (**Fig. 1)**. We investigate geostatistics, the Alternating Conditional Expectation[1] (ACE) method, and MATLAB®[2].

Geostatistics. The usual geostatistical procedure[3-7] has been applied to the field data. The top surfaces of three cycles (G, K and M) are used here as an illustration.

An experimental variogram[5] of the G-Cycle top is shown in **Fig. 2**. The white color at the corners denotes lack of data. The maximum data variation occurs near the corners. From this variogram, it is obvious that a SE-NW trend exists in the data, approximately at 45-55 degrees counterclockwise from the EW axis. The experimental variograms along three different directions (55, 145 and 100 degrees), **Fig. 3**, are then fit with a nested model. A linear and a spherical semi-variogram have been invoked to match the data in the direction of 55^0:

$$\gamma(h) = 1.31h, \quad \text{Anisotropy ratio} = 3.4, \qquad \ldots\ldots\ldots\ldots\ldots\ldots (1)$$

and

$$\gamma(h) = 210\left[\frac{3}{2}\frac{h}{910} - \frac{1}{2}\left(\frac{h}{910}\right)^3\right], \text{ if } h \le 910 \ldots (2)$$

and $\gamma(h) = 210$ otherwise, Anisotropy ratio = 1. The fit is quite good and matches the field data in several other directions (e.g., 100 and 145^0).

The model variogram, shown in **Fig. 4**, explains 1800 ft, out of 2200 ft of the total data variation. As stated above, the maximum data variability occurs near the corners where the wells are missing, and the model does a good job explaining the spatial correlation of data elsewhere.

The two variograms above are then input into Surfer®[8], and the resulting kriged surface is shown in **Fig. 5**. The surface has multiple SE-NW ridges and valleys, and a depression near the NW corner caused by an outlier.

An experimental variogram of the K-Cycle top is shown in **Fig. 6**. The maximum data variation occurs near the corners. As before, a SE-NW trend exists in the data, approximately at 45-55 degrees counterclockwise from the EW axis. The experimental variograms along three different directions (55, 145 and 100 degrees), **Fig. 7**, are fit with a nested model:

$$\gamma(h) = 2.4545h, \text{ Direction of } 55^0, \text{ Anisotropy ratio = } 4.5 \quad (3)$$

and

$$\gamma(h) = 302\left[\frac{3}{2}\frac{h}{780} - \frac{1}{2}\left(\frac{h}{780}\right)^3\right], \text{ if h } \le 780, \quad (4)$$

and $\gamma(h) = 302$ otherwise, Direction of 145^0, Anisotropy ratio = 4.9. The fit is good and matches the field data in several other directions.

The model variogram, shown in **Fig. 8,** explains 4500 ft out of 5800 ft of the total variation of data. Again, the maximum data variability occurs near the corners where the wells are missing and the model does a good job explaining the spatial correlation of data elsewhere.

The two variograms above are input into Surfer®, and the resulting kriged surface is shown in **Fig. 9**. The resulting surface is visibly smoother than that in Fig. 5.

An experimental variogram of the M-Cycle top is shown in **Fig. 10**. As before, the maximum data variation occurs near the corners. The experimental variograms along three different directions (55, 145 and 100 degrees), **Fig. 11**, are fit with a nested model in the direction of 55^0:

$$\gamma(h) = 2.66h, \text{ Anisotropy ratio = } 4.7 \quad (5)$$

and

$$\gamma(h) = 487\left[\frac{3}{2}\frac{h}{633} - \frac{1}{2}\left(\frac{h}{633}\right)^3\right], \text{ if h } \le 633 \quad (6)$$

and $\gamma(h) = 487$ otherwise, Anisotropy ratio = 0.68. Again, the fit matches well the field data in several other directions.

The model variogram is shown in **Fig. 12**. The model explains 4100 ft out of 4700 ft of the total variation of data. The maximum data variability occurs near the corners, where the wells are missing, and at a localized depression near one of the wells. The model does a good job explaining the spatial correlation of data elsewhere.

The two variograms are input into Surfer® and the resulting kriged surface is shown in **Fig. 13a**. **Fig. 13b** shows the back side of the same surface to illustrate the influence of a single outlier.

ACE[1]. (Alternating Conditional Expectation), has been used successfully to fit permeability vs. porosity data.[9] Of the methods used in Ref.[9], ACE was the best for an example with a single independent variable. ACE can also be used with multiple independent variables[9-11], which is what we will show here.

ACE is a nonparametric method developed at Berkeley by Breiman and Friedman for finding optimal transforms. The source code can be found at http://playfair.stanford.edu. Nonparametric means that no assumption is made about the functional form of the regression, such as linear, power, exponential, etc. The only assumption is that the optimal transform of the dependent variable is the sum of the optimal transforms of the independent variables

$$\theta(Z) = \sum_{i=1}^{p} \phi_i(X) \quad (7)$$

where θ and ϕ are the transform functions and p is the number of independent variables. Also *Z* and *X* are vectors of the data.

Both X and Z can have random error. The method will find the optimal θ and ϕ, from which a smoothed *Z* can be calculated from *X*.

ACE can exactly find the functional dependence of each variable and eliminate noise if the data are separable, that is Z really is a function of the sum of independent functions of the independent variables

$$Z = f_1(x_1) + f_2(x_2) + \ldots + f_p(x_p) \quad (8)$$

However, if the data are not separable, the smoothing resulting from the optimal transforms may *not* be satisfactory. ACE tends to smooth quite severely and outliers are basically assumed to be error.

MATLAB® Surface Plotting. MATLAB[2] is a powerful mathematical programming environment, particularly useful for matrices. It has routines for generating surfaces from a set of (x,y,z) data, where the x's, and y's form a rectangular grid. Each point inside this rectangular grid can be thought of as connected to the four nearest neighbors. The MATLAB mesh algorithm then induces four-sided patches on the surface. This defines a mesh of quadrilaterals, or *quad-mesh*. The resulting surface obeys each mesh vertex point exactly. This may lead to anomalies near outliers, with the surface going through the point but also showing a large kink in the other direction near the data point. Obviously, deleting the outliers eliminates the anomalies. With only two independent variables, we can easily visualize the data with MATLAB, but it tells us nothing about the structure of the data.

Problem Statement

We want to visualize seven layers ("cycles") of the South Belridge Diatomite, Kern County, CA, in the Shell Phase III Steam Drive Pilot. The depth data come from the directional surveys in approximately 50 wells within the 1000 foot by 1000 foot pilot area, with an additional 20 wells in the 200 feet surrounding the pilot, Fig. 1. Thus each layer G through M has approximately 70 points for each layer. We will visualize each layer using the three methods and compare the surfaces generated.

We further subdivide the 70 points for layers G, K, and M into training and testing data sets. For example, we have selected 50 random points from the 70 points for each layer as training points and the remaining 20 points as testing points. We have made 3 such sets for each layer. We compare the ACE optimal transforms predicted by the training sets for each layer with those predicted by the entire data set. This shows the sensitivity of ACE to the data.

Results

First, we show the comparison between the three methods for layers G, K, and M. On all of these surface plots, the vertical elevation *z* is shown on the vertical axis, *x* (easting) is shown on the right axis, and *y* (northing) is shown on the left axis. All lengths are measured in feet. Note that some of the mesh lines have missing segments that are due to errors in cutting-and-pasting between different programs.

Fig. 14 shows the results for layer G. ACE predicts that the layer top surface is made up of two sections with different slopes and a distinct intersection near $y = 0$. Other than that, the two regions appear quite flat. ACE smoothes out waves that may be caused by a single data point.

The MATLAB® surface shows quite different behavior. It appears wavy, but except for the front corner near (-500,-500), it is relatively flat with only a uniform slope. It doesn't look anything like the ACE figure. The peak and valley in the front corner is an anomaly. The peak is real, but valley to the left of the peak near y = 0 is not real. Obviously, the outlier has seriously distorted the surface.

The geostatistical model shows a wavy surface with two regions having differing slopes. The intersection appears along a definite SE-NW line. The peak in the front corner that gave MATLAB® such trouble is shown only as small hump, and no anomalous valley appears.

Layer K, shown in **Fig. 15,** is relatively simple, with no outliers or strange behavior. All of the methods capture the downward sloping trend from back to front.

As in layer G shown in Fig. 14, ACE predicts quite a sharp bend in the surface near y = 0. Also, ACE predicts a smooth not wavy surface.

MATLAB® again shows some gentle waves. But more important, it shows totally different behavior from ACE along the *x*-axis. Whereas ACE predicts the lowest part of the surface along the *x*-axis, MATLAB predicts even lower values for the front corner, but quite high values near x = 500.

The geostatistical model gives a surface nearly identical to the surface from MATLAB, except in the front corner at (-500,-500). Because there are no data in this region, the geostatistical model gives a flat, horizontal surface. MATLAB extrapolates linearly the downward trend as though it were real.

Layer M is shown in **Fig. 16**. We note here that the deep valley shown clearly in the MATLAB surface is actually an erroneous data point, which from reviewing other data, should be around -800 instead of -950.

As before, ACE predicts a non-wavy surface that slopes downward from the back left of the plot. This seems quite consistent with the other layers. We note that there is no effect of the erroneous data point.

MATLAB also predicts a similar shape as the other layers. Apart from the aforementioned valley, it shows a gentle downward slope from the back to the front. We plotted layer M with the erroneous point removed, and the resulting surface appeared very smooth.

The geostatistical model shows a valley at the erroneous point, Fig 13b, but not nearly as large as that from MATLAB. It respects outliers, but their influence is more localized than in the MATLAB® mesh algorithm.

We now show the effect that data has on the ACE prediction. **Fig. 17** shows the surface generated for layer G, previously shown in Fig. 14, along with three other visualizations using subsets of the data. The transforms used to generate the smoothed data are shown in **Fig. 18**.

All of the surfaces in Fig. 17 show the same general trend, a downward slope from the left back and a kink near y = 0. The figure G2 appears nearly identical to G, but G1 and G3 have other kinks as well. G1 show an upward trend near the y-axis.

We can also compare the transforms shown in Fig. 18. Notice that the transforms for G and G2 are quite similar and their surfaces in Fig. 17 were also similar. G1 and G3 are quite different.

Figs. 19 and **20** show the surfaces and transforms for layer K. All of them are quite similar in both trend and magnitude except that K2 in Fig. 19 has a serious kink near y = 0. Also, K3 appears very smooth. Obviously the transforms of K2 in Fig. 20 are quite different than the others.

For layer M, shown in **Figs. 21** and **22**, the same conclusions apply. All surfaces are relatively smooth and similar except M1, which has a large kink and very different transforms.

In each of these layers, changing the data set had a large effect on the surface that ACE generated. It appears that no simple structure was found and the data were *not* really separable.

Conclusions

1. All three methods seem to capture the major trends in the layers.

2. ACE drastically smoothes the data and eliminates

outliers. It does show a dramatic change in layer slope at about $y = 0$. This trend is also shown by the other methods, but the change in slope occurs more along a diagonal line across the field.

3. ACE is very sensitive to the data. Different subsets of the data can generate quite different views of the total surface.

4. If the data are not separable, i.e., the z is not the sum of a function of x and a function of y, the resulting ACE transforms and subsequent smoothed prediction of z may *not* be satisfactory.

5. MATLAB® generates uneven surfaces that obey the data points exactly. It distorts the area near an outliers, and creates artificial valleys and peaks. However, by showing the outliers, erroneous data points can be detected. MATLAB should be used only for fast screening of the data and simplified visualization.

6. The full geostatistical model shows the trends in the data, along with some peaks and valleys. Outliers are retained, but their influence is very localized. It does not generate anomalous peaks or valleys near outliers.

7. In summary, the fastest and easiest visualization method (MATLAB®) has serious limitations. ACE smoothes excessively and is very sensitive to input data. Geostatistics is cumbersome, but it yields by far the best results. Surfer® is a superb software package to generate and display kriged surfaces.

Acknowledgments

This work was supported by two members of the U.C. Oil® Consortium, Chevron Petroleum Technology Company, and CalResources, LLC. Partial support was also provided by the Assistant Secretary for Fossil Energy, Office of Gas and Petroleum Technology, under contract No. DE-ACO3-76FS00098 to the Lawrence Berkeley National Laboratory of the University of California. The field data set was provided by CalResources, LLC.

Nomenclature

γ = a positive definite variogram function, length
h = spacing or lag between data points, length
ϕ = optimal transforms of independent variables
p = number of independent variables
θ = optimal transform of dependent variable
X = vector of independent variables
Z = vector of dependent variable

References

1 L. Breiman, and Friedman, J. H., Journal of the American Statistical Association**,** 580 (1985).

2 (The Math Works, Natick, 1995).

3 E. H. Isaaks, and Srivastave, R. M., *Applied Geostatistics* (Oxford University Press, New York, 1989).

4 C. V. Deutsch, and Journal, A. G., *GSLIB, Geostatistical Software Library and User's Guide* (Oxford University Press, New York, 1992).

5 Y. Pannatier, *VARIOWIN, Software for Spatial Data Analysis in 2D* (Springer Verlag, New York, 1996).

6 J. C. Davis, *Statistics and Data Analysis in Geology*, 2 ed. (John Wiley & Sons, Inc., New York, 1986).

7 N. A. C. Cressie, *Statistics for Spatial Data*, Revised ed. (John Wiley & Sons, Inc., New York, 1993).

8 D. Smith, et al., , 6.12 ed. (Golden Software, Inc., Golden, CO, 1996).

9 G. Xue, Datta-Gupta, A., Valko, P., and Blasingame, T. A., in *Paper SPE 35412, Optimal Transformations for Multiple Regression: Application to Permeability Estimation from Well Logs*, Tulsa, OK, 1996 (SPE).

10 G. Xue, and Datta-Gupta, A., in *Paper SPE 36500, A New Approach to Seismic Data Integration During Reservoir Characterization Using Optimal Non-parametric Transformations*, Denver, CO, 1996.

11 A. Datta-Gupta, Xue, G., and Lee, S., *Non-parametric Transformations for Data Correlation and Integration: From Theory to Practice*, in Proceedings of the Fourth International Reservoir Characterization Conference, Houston, TX, 1997.

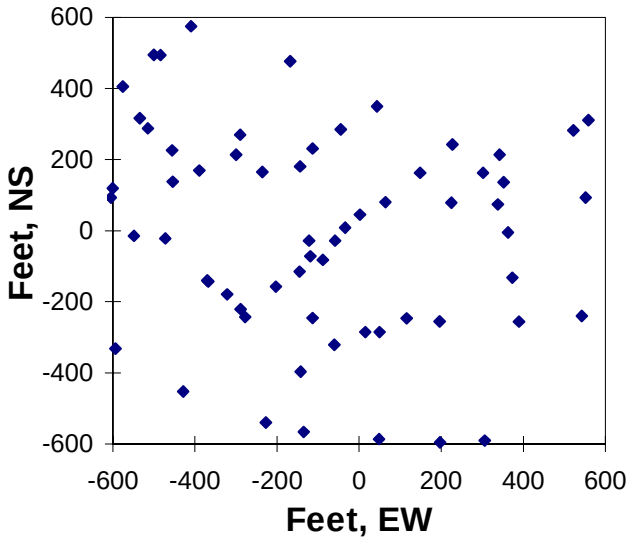

Fig. 1—A map of the well locations in and around the Shell Phase III Pilot in the South Belridge Diatomite, Kern County, CA.

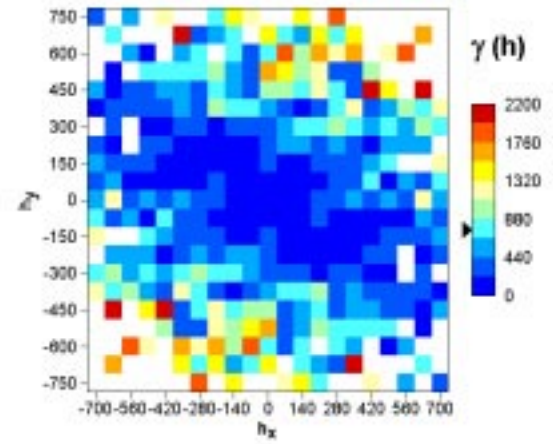

Fig. 2. - Experimental variogram of the G-Cycle top.

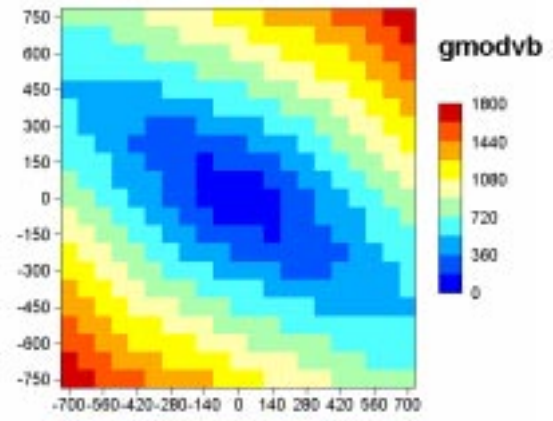

Fig. 4. - Model variogram of the G-Cycle top.

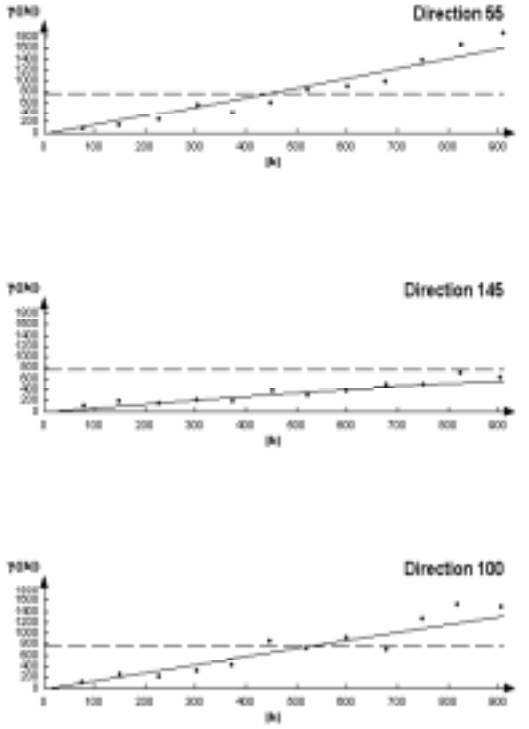

Fig. 3. - Optimal model semi-variograms of the G-Cycle top surface. The directions are in degrees counterclockwise from the EW (x)-axis.

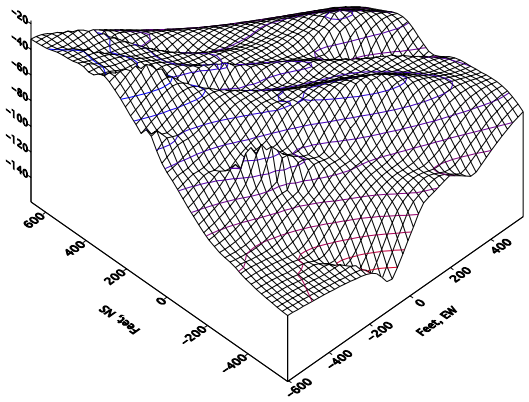

Fig. 5. -The kriged G-Cycle top surface.

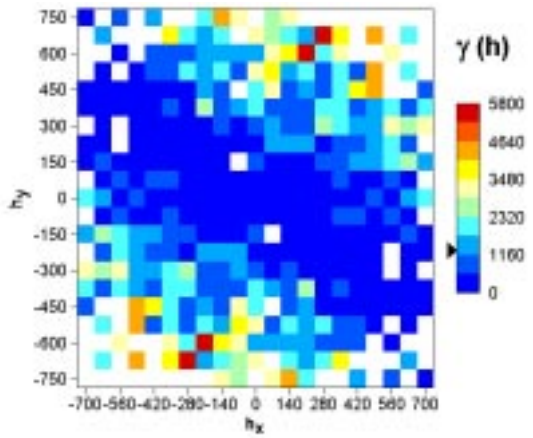

Fig. 6. - Experimental variogram of the K-Cycle top.

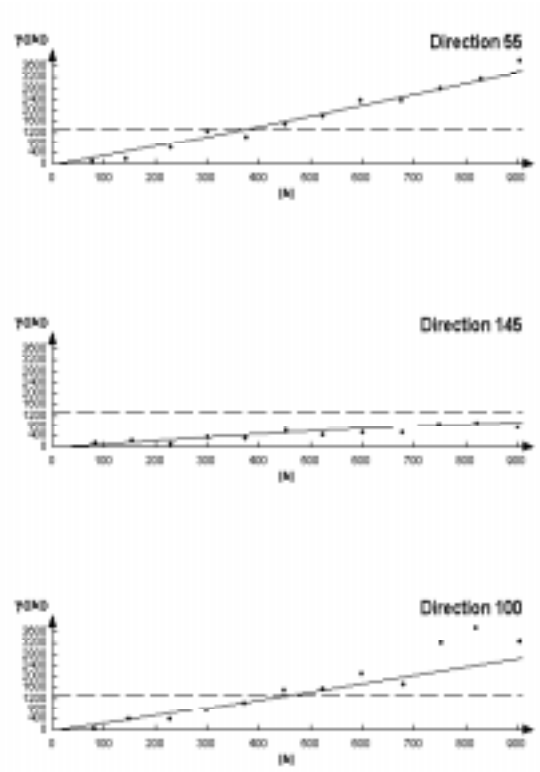

Fig. 7. - Optimal model semi-variograms of the K-cycle top surface. The directions are in degrees counterclockwise from the EW (x)-axis.

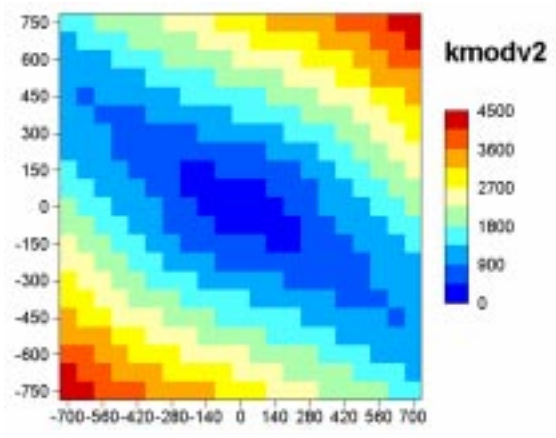

Fig. 8. - Model variogram of the K-Cycle top.

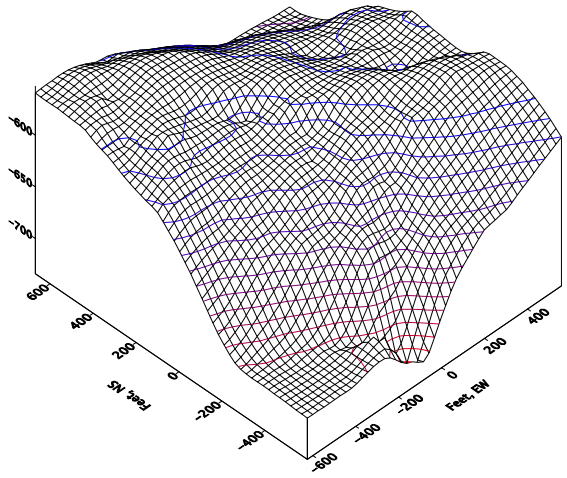

Fig. 9. -The kriged K-Cycle top surface.

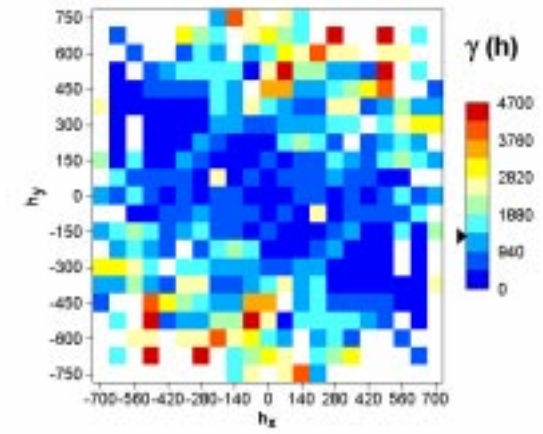

Fig. 10. - Experimental variogram of the M-Cycle top.

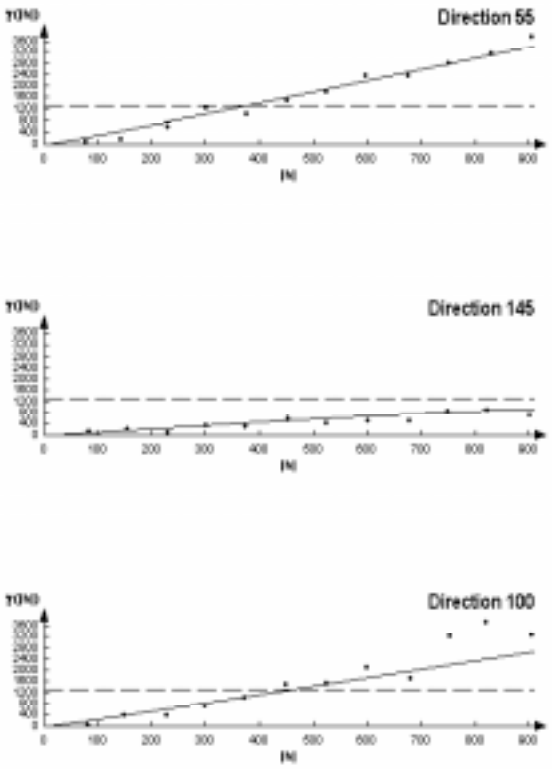

Fig. 11. - Optimal model semi-variograms of the M-cycle top surface. The directions are in degrees counterclockwise from the EW (x)-axis.

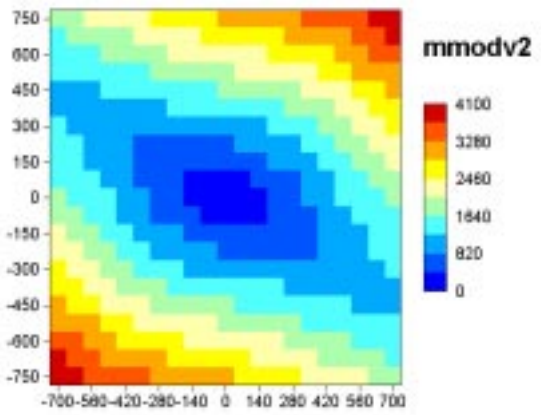

Fig. 12. - Model variogram of the M-Cycle top.

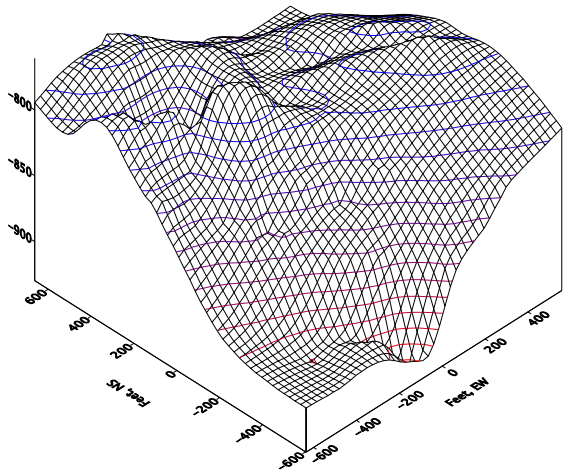

Fig. 13a. -The kriged M-Cycle top surface.

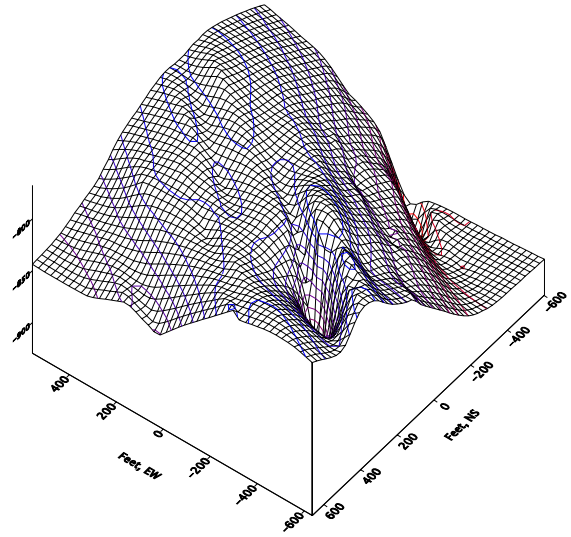

Fig. 13b. -The back side of the kriged M-Cycle top surface. The gaping hole present in the corresponding MATLAB grid (Fig. 16) is localized here.

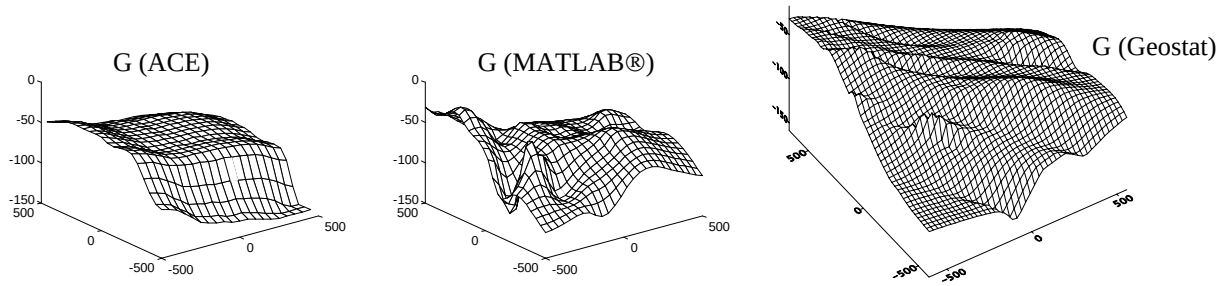

Fig. 14—Visualization of Layer G using the three methods.

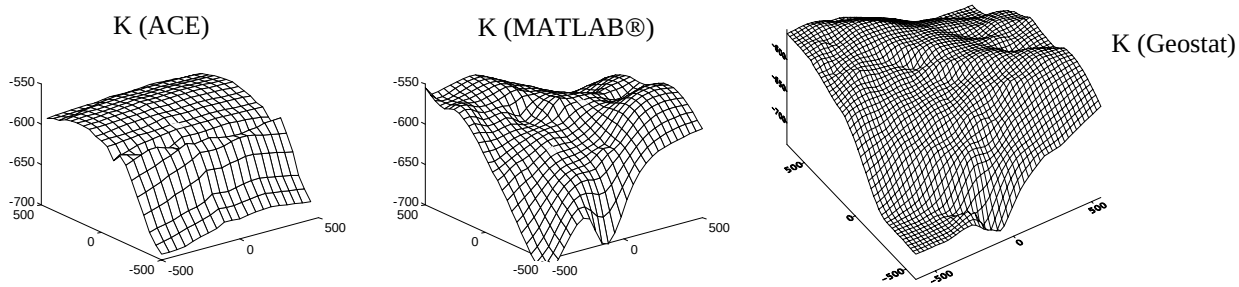

Fig. 15—Visualization of Layer K using the three methods.

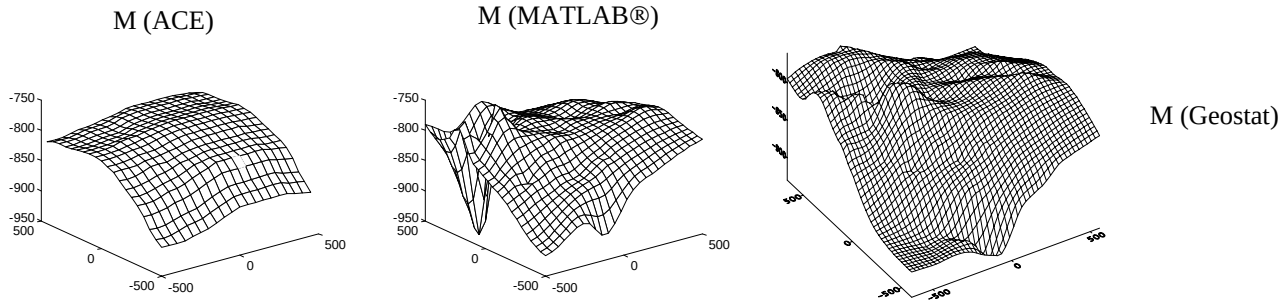

Fig. 16—Visualization of Layer M using the three methods.

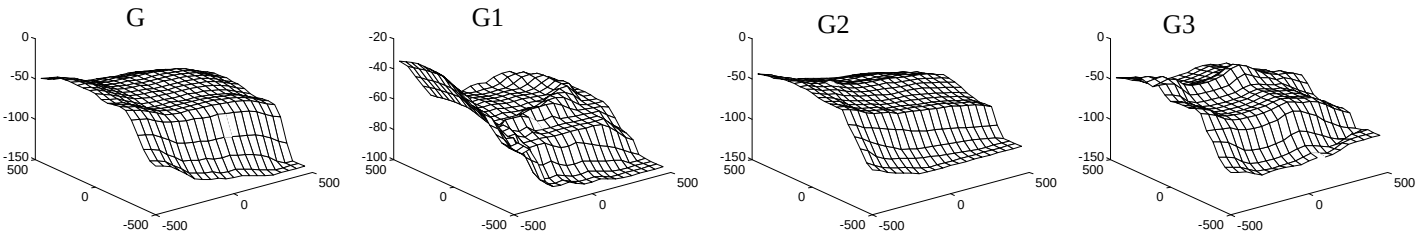

Fig. 17—An ACE visualization of Layer G using all of the data and also three training sets.

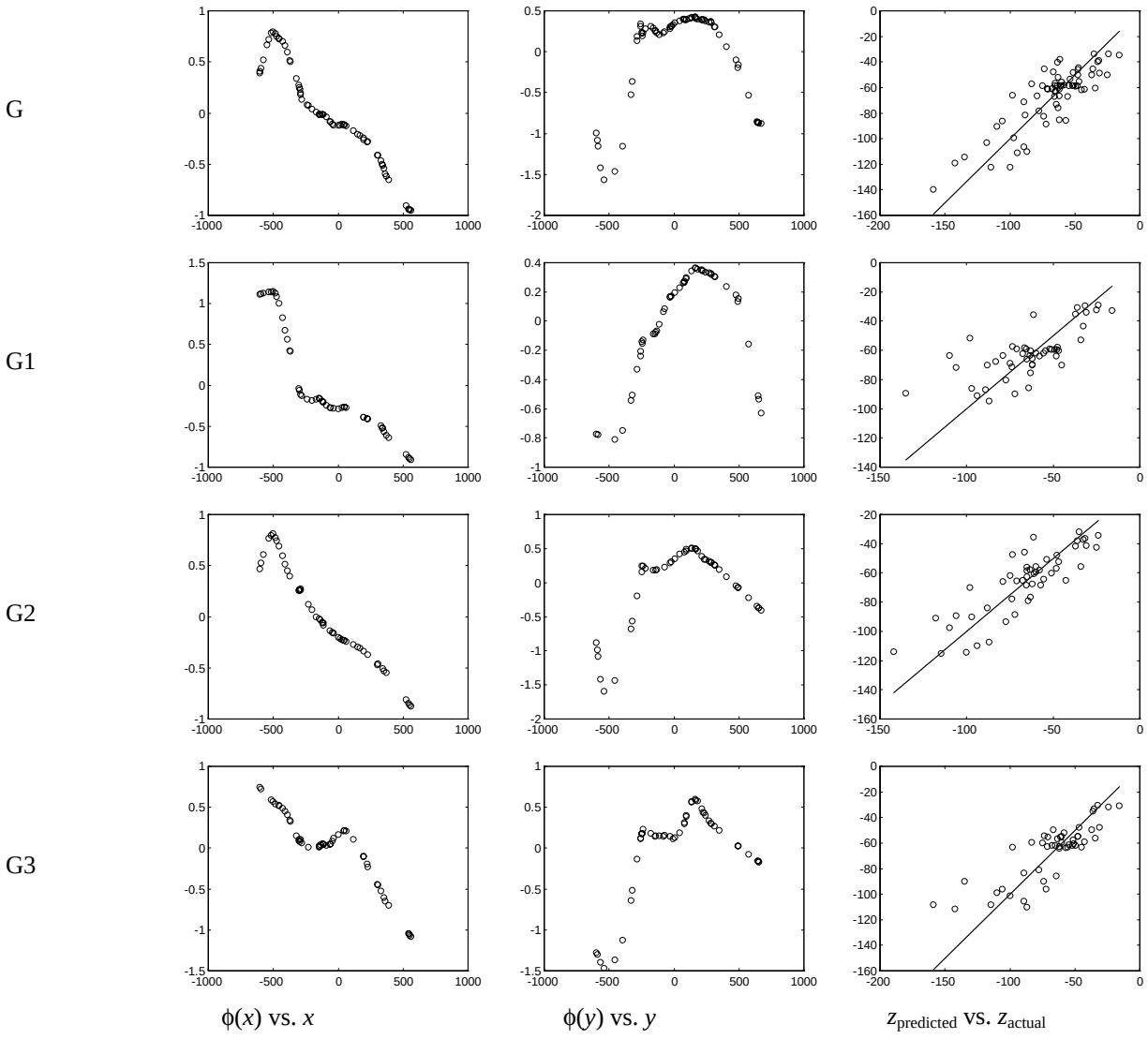

Fig. 18—Comparison of optimal transforms between complete data set and random subsets for layer G. Also shown is a comparison of predicted vs. actual layer depth.

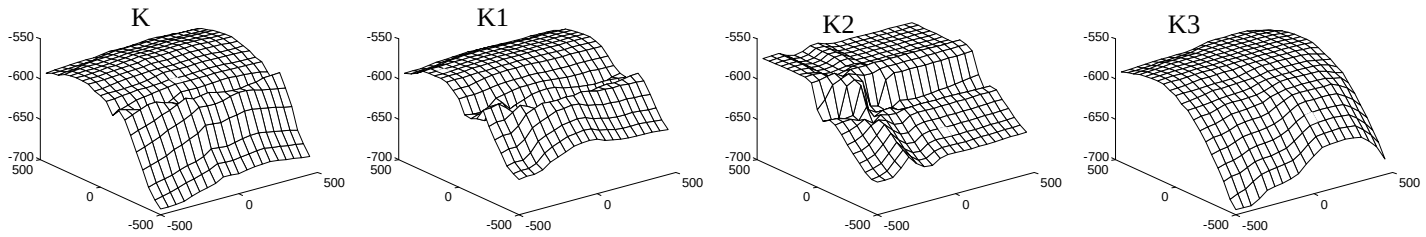

Fig. 19—An ACE visualization of Layer K using all of the data and also three training sets.

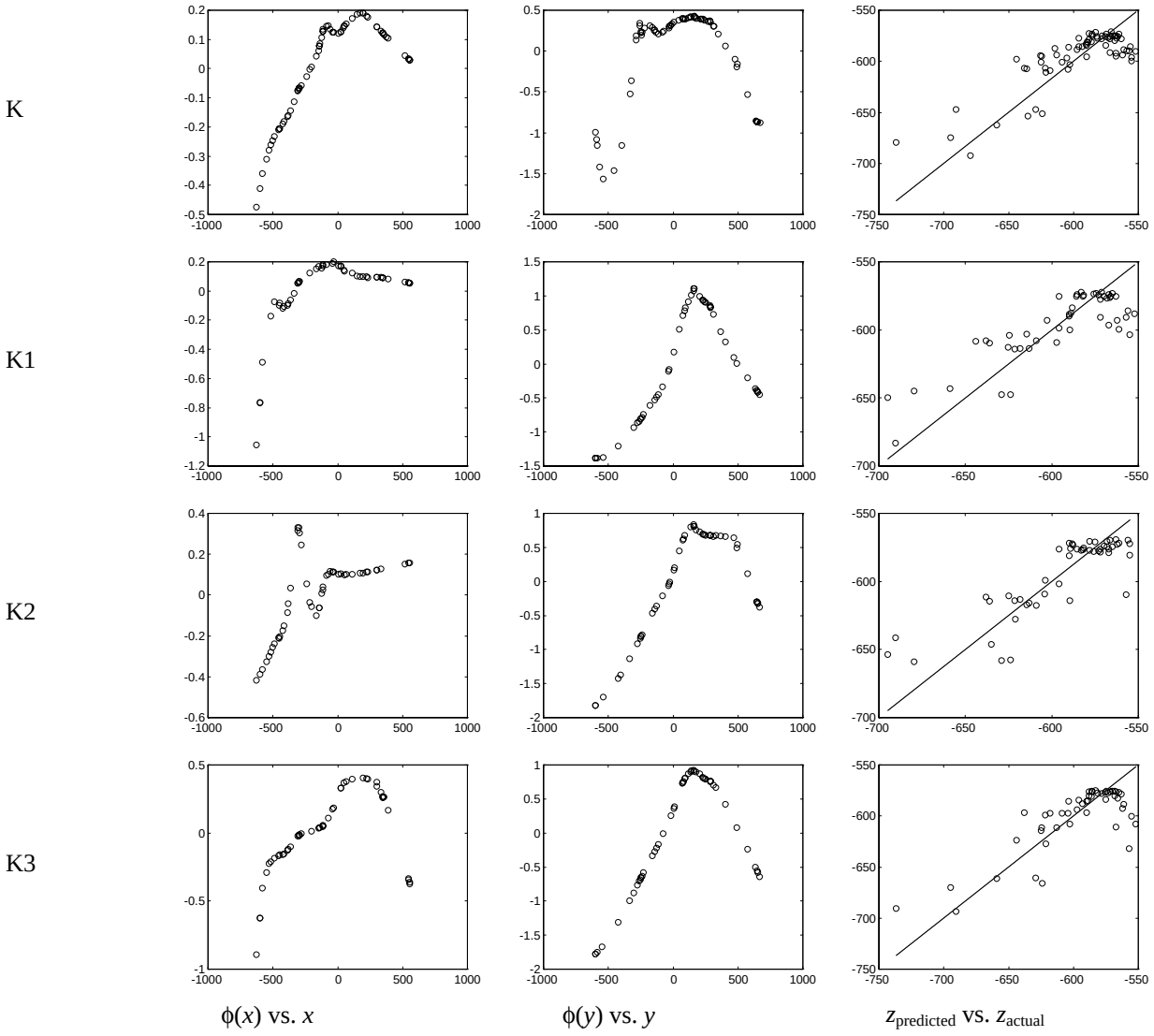

Fig. 20—Comparison of optimal transforms between complete data set and random subsets for layer K. Also shown is a comparison of predicted vs. actual layer depth.

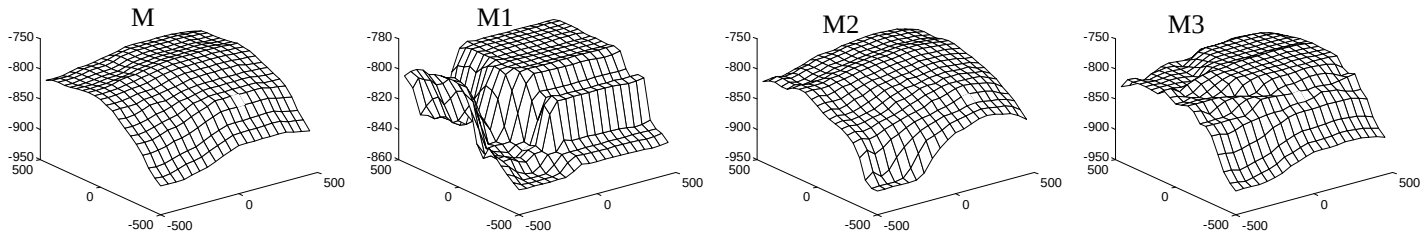

Fig. 21—An ACE visualization of Layer M using all of the data and also three training sets.

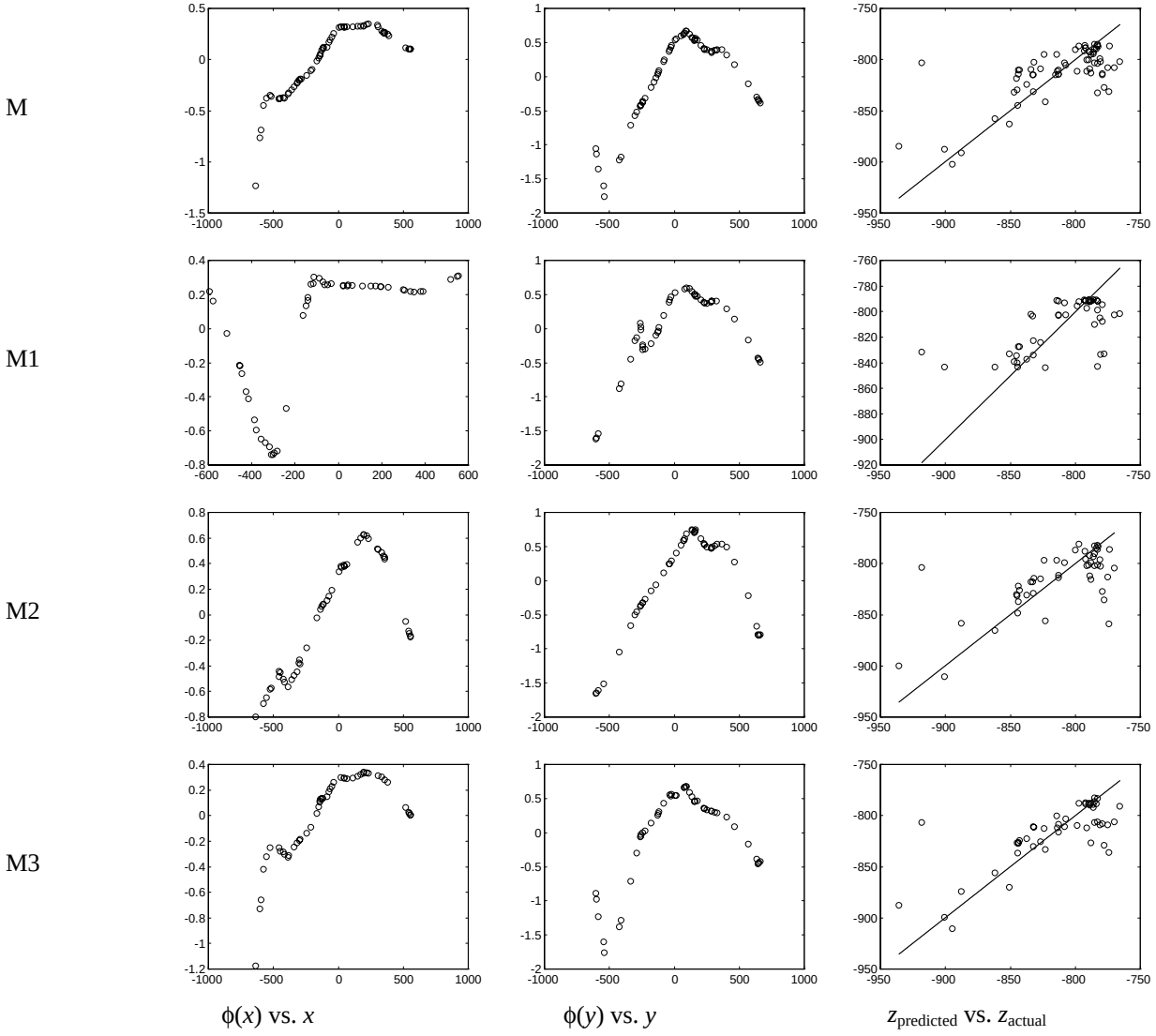

Fig. 22—Comparison of optimal transforms between complete data set and random subsets for layer M. Also shown is a comparison of predicted vs. actual layer depth.