GENERALIZATION IN HIGHER-ORDER NEURAL NETWORKS
AN EVALUATION OF FORECASTING PERFORMANCE WHEN RECREATING NONLINEAR SURFACES

James E. Mitchell and Hao Sun
Geohydrology Section
Kansas Geological Survey
The University of Kansas
Lawrence, KS 66047
913-864-3965
JIM_MITCHELL@MSMAIL.KGS.UKANS.EDU
FAX: 913-864-5317

KANSAS GEOLOGICAL SURVEY
OPEN FILE REPORT 92-15

Keyword I:
- Performance Analysis

Keywords II:
- 3-D Surface Estimation and Approximation
- 3-D Free-Form Surface Estimation and Approximation
- Higher-Order Neural Networks
- Performance Analysis, Measure and Evaluation
- Experimental Design
- Nonlinear Function Approximation
- Parameter Estimation
- Forecasting
- Learning Accuracy
- Non-Smooth Interpolation
- Noise Sensitivity
- 3-D Vision
- Back-Propagation Learning Algorithm
- Curve Fitting

ABSTRACT

Representing three-dimensional surfaces is a problem that is frequently encountered in spatial analysis and computer graphics. It is particularly difficult in applications where the only source of data is nonexperimental samples. In traditional geoscience applications, the researcher generalizes information from randomly spaced points, incorporating his knowledge of local geology, topology, and other factors, to recreate the character of a free-form surface.

This paper describes a higher order neural network (HONN) approach to this problem. Regular (grid) and irregular (random) samples are used to train the HONN. Experimental surfaces are created using polynomials with specified parameters and three levels of random error. Generalization of a HONN is evaluated by comparing the networks' estimation and prediction mean square error (MSE) with those from polynomial regressions. Weight vectors are visualized and compared for networks which are presented data from the same surfaces using regular and irregular samples.

That HONN successfully identifies the underlying character of the polynomial surfaces. The MSE of prediction for HONN compares favorably with polynomial regressions on the same data.

Free-form surface prediction is difficult because location and direction provide little information on the form of the response surface beyond the local area. The results suggest a HONN performs well in this application because, (1) it is able to identify the important components in a complex relationship and (2) it is not constrained by a parametric mathematical representation of the surface.

INTRODUCTION

Surface representation has been an important issue in many research areas. In computer vision research, the surface of an object is reconstructed from sparse range data which are grid sampled. In geoscience, surfaces are typically predicted from a set of irregularly spaced samples. The nonexperimental data collection process makes these problems more difficult [Mi92] and [Da86].

Conventional reconstruction algorithms rely on smoothing constraints and stabilizing functionals. In this manner, the discontinuities are pre-detected to segment a surface into piecewise smooth patches. This segmentation has proved to be a difficult task. Another method attempts to extract global surface parameters. To attain reliable parameter estimates, it is necessary that all the selected points lie on the same surface patch. Therefore, a preprocessing grouping of raw data is necessary for these algorithms. This is a complicated exercise [Bo91].

Free-form surface representation has not received much attention in the research literature, recently. An important property of a free-form surface is that it is not constrained to any mathematical model, such as piecewise-superquadrati, but the surface normal is continuous everywhere. A good example is land surface elevation. Algorithms to represent elevation are very limited in the sense of efficiency and accuracy [Be90].

Neural network approaches have been intensively studied for nonlinear function approximation [Jo90]. With efficient hidden layers and units, multilayer networks are able to compete with statistical estimators. They have been referred to as universal estimators [Ho89]. However, using neural networks to approximate 3-D functions is seldom mentioned and less addressed in detail in the research literature.

Chen and Jain [Ch91] propose a multilayer feedforward network extended with a robust learning algorithm. Their network has shown the ability to approximate an underlying 3-D function even perturbed with noise. However, this work is limited to a piecewise smooth surface. Free-form reconstruction is considered to be much more difficult because of its high degree of nonlinearity. Mitchell, et al., discuss how first-order-multilayer network performance can drop when nonlinearity increases in 3-dimensions [Mi92].

In this paper, quadratic multilayer neural networks are employed to reconstruct surfaces from either grid or random samples from surfaces which are generated with differing error characteristics. Generalization of networks trained using different sampling regimes is visualized using star charts of the trained networks' weights. Accuracy of the networks are analyzed by comparing the estimation and prediction mean squared error with polynomial regressions on the same data. Finally, a HONN is used to represent two free-form surfaces derived from land surface elevations.

Visualization of the converged weights indicates that identical networks which are trained by different sampling regimes appear to "focus" on similar features. Forecasting performance appears to be stable when small relative noise is present. However, as noise increases, grid sampling outperforms random sampling of data.

METHODS

Polynomial Surfaces

The first phase of this study generates several polynomial surfaces with differing error characteristics. Using interactive graphics, care was taken to choose functions which displayed a wide range of z-dimension variation and balanced this with the amount of random error added to the surface. The purpose of this error was to simulate "real" data by providing a measure of difficulty in determining the true underlying function. Previous work on similar problems has indicated that higher order neural networks are quite adept at identifying and predicting three dimensional surface functions [Mi92]. As a result, the following four equations were selected:

$$z = x^2 + y^2 + xy \quad (1)$$

$$z = x^2 - y^2 + xy \quad (2)$$

$$z = x^3 + y^3 \quad (3)$$

$$z = x^3 + x^2 y \quad (4)$$

Using these relationships, surfaces were generated as:

$$z = f(x,y) + \sigma^2 \quad (5)$$

Where, σ^2 represents identically normally distributed random error defined as:

$$\sigma^2 \sim N(0, 0.31^2) \quad (6)$$

Two approaches were used to sample from these surfaces. A regularly spaced 11 by 11 grid was the first sampling regime. The second generated 121 random (x,y) pairs, uniformly, along the x and y domains and used them for network training and statistical estimation. The mean squared error (MSE) of the eight polynomial regressions and the MSE of the eight converged neural networks were then compared. All statistical operations were performed using SAS [SAS89].

Neural networks in this phase were comprised of a quadratic processing unit, plus one hidden layer with three nodes, and a single output. The spatial location (x, y) and the corresponding surface value z are input-output pairs. All layers are fully connected to the subsequent layer. Figure 1 depicts this system.

Network Visualization

An analysis was made of the weights assigned to network connections, using the converged networks described above. This was accomplished through the use of "star charts." This is a graphical method which displays the weights assigned to the connections to each neuron. Each is depicted as a spine or ray emanating from a central point. The three nodes from the hidden layer in each network can then be easily compared. This is done for each pair of samples (grid and random), taken from each generated surface.

Analysis of Forecasting Performance

Grid and random samples were compared for their ability to train higher order networks with the objective of forecasting (predicting) surface elevations at points which have not been used to train the network. A "Mexican Hat" function was used for this purpose [SAS90]. This is a useful function for this purpose because it is NOT considered a piecewise smooth surface [Ch91]. Random error was added to the observations in order to analyze their effect on surface identification and forecast performance.

The Mexican hat function can be seen in Figure 2 and is generated by:

$$f(x,y) = \mathrm{SIN}[(x^2 + y^2)^{1/2}] \qquad (7)$$

Where x and y are defined over the range of -5.0 to 5.0. Three datasets were used in this phase of the study. The first was a perfect, no-error surface of the function. the second and third were error contaminated using the same approach as in equations (5) and (6), except the error variance was manipulated to provide standard errors of 0.08 and 0.2, respectively. This resulted in sample surfaces which were "rough," but retained the basic shape of the function.

Each dataset was comprised of a 26 X 26 (=676) grid of points. For the grid sampling experiments all 676 were used to train the network. Preliminary work indicated that this number of grid points provides sufficient resolution to recreate this particular surface. The random sample experiments were performed by sampling 121 points from the population of 676. Both the x and y domain were sampled with a uniform distribution. These random points were then used to train the network.

Networks employed in this experiment were comprised of a quadratic processing unit (as shown in figure 1), two hidden layers, the first with five neurons, the second with three, and a single output. All layers are fully connected to the subsequent layer.

Forecasting performance was tested using a second grid of dimension 26 X 26. These points were chosen to avoid locations used in the original training set. The same points were used to evaluate each network. In this manner an appropriate validation of the networks' performances could be made [Ke92]. Data were analyzed, using [SAS86] employing a repeated measures design.

Free-form Surface Representation

Digital Elevation Models (DEM) were used as samples from a free-form surface. Data from Douglas County, Kansas were used for this purpose. The sample contained a mixture of sharp, local changes in slope and areas of little to moderate elevation change [USGS87].

An area of approximately 2.80 X 10^7 meters^2 was represented with a 90 meter^2 grid sampling. From this a grid of dimension 31 X 31 was selected with a grid spacing of approximately 180 meters. An additional subset of these data were taken, using a 41 X 26 (x, y) grid to define an area of approximately 7.7 X 10^6 meters^2, with a resolution of 90 meters.

These surfaces were analyzed using networks comprised of a quadratic input layer and two hidden layers, with 5, 14, and 7 neurons, respectively. Results of these networks are graphically displayed as three-dimensional surfaces.

RESULTS

Polynomial Surfaces

The comparison of polynomial regression estimation and network training MSE is portrayed in Table 1. In both the grid and random samples, the neural networks consistently outperform the polynomial regressions on MSE. Under grid sampling, this difference is well above one order of magnitude for most of the functions. Overall the random sampling appears to have inconsistent effects on the performance of the neural networks. Some appear to improve slightly under random sampling, others do not. At the same time, the polynomial regressions do substantially better under randomly sampled data.

Network Visualization

The star charts in Figures 3 - 6 are presented as a means of comparing converged network weights. Each figure surface, using different sampling approaches. Differences in the weights (and therefore the shapes of the charts) are assumed to be attributable to the difference between grid and random sampling.

Figure 3 shows the networks representing equation (1). The first two nodes appear to have similar shape with different size, comparing the two sampling schemes. The third node is very different between the two. This is not the case for equation (2) in Figure 4. Each network appears to focus on different aspects of the input layer. Figure 5, pertaining to equation (3), and Figure 6, pertaining to equation (4), appear to show shape similarity and size difference for all three nodes.

It is interesting to note that the forms of equations (1) and (2) are both quadratic and equations (3) and (4) are both cubic. The network structure includes only a quadratic processing unit. This suggests that the third node in the hidden layer responds to some characteristic which is present in the higher order surface and not accounted for in the processing of the second order information in the polynomial.

Forecasting performance

Table 2 compares the forecasting performances of a HONN as a function of grid versus random sampling and as a function of the error inherent to the generated surface (the Mexican hat). Figure 7 shows the distributions of prediction errors for each combination of sampling scheme and standard error. As indicated in the table, all differences between the experiments are statistically significant.

This suggests that the interaction between the sampling scheme and the amount of error inherent in the surface combine to create a significant interaction. This indicates that with little or no error, the random sampling scheme is sufficient to represent the surface. However, as error increases, it appears that a grid sample is required to capture the nature of the surface. The increase in the generated standard error from 0.08 to 0.20, which elicits this effect increases the squared error loss by a factor of three. The other statistics reported in the table demonstrate similar behavior.

Testing the main effects of sampling scheme indicates no significant differences between the two approaches, taken in combination. However, as the standard error of the underlying surface increases, prediction errors become significantly larger ($p < .0001$).

Free-form Surface Representation

Two surfaces were reconstructed from the DEM data. These are depicted in Figures 8 and 9. Figure 8 shows a larger area with a coarser sampling grid. In this dataset, each point represents approximately 29,487 meters^2. In the smaller area this figure drops to 7,243 meters^2. Both networks converged to a standard error of approximately 0.2 meters.

CONCLUSIONS

The networks outperform polynomial regression approaches for reconstructing rough, error contaminated surfaces. This can be attributed to the fact that the polynomial regression approach is constrained to predicting the expected surface elevation, according to a specific (parametric) polynomial equation. On the other hand, the HONN is able to more readily adept to local anomalies. The HONN prediction is therefore not strictly $z = E[f(x,y)]$. It is able to accommodate more information about the local surface.

The sampling scheme appears to have a much larger effect on the estimation MSE of polynomial regressions than on the network MSE. The neural network appears to hove the property of being more robust to the different sampling designs.

Visualizing weights (Figures 3-6) leads to two conclusions: (1) Star charts from identical networks presented with data sampled by different methods show similar shape, but vary in size. Thus the information from each node in the input layer is weighted the same at each node, in a relative sense. The output weights are similarly proportional. (2) When the order of data presentation changes, the influence of particular neurons can change greatly. This can be seen in the third neuron in the star charts, where the quadratic surfaces create larger output weights than do the cubic surfaces.

Prediction of surface elevation by the networks can be greatly influenced by the interaction of sampling method and the amount of local variation on the surface. When the amount of error on the surface is low, random samples appear to do relatively better than grid samples measured by prediction MSE. As variation on the surface increases, the

data suggest that grid sampling outperforms random sampling. Clearly, random samples from a surface with a relatively large amount of error are not as efficient as grid samples at capturing the character of the surface, in this study. Methods which can identify and utilize information about relative local and global variation may be able to improve forecasting performance.

Figures 8 and 9 clearly indicate that, at the scales presented, the HONN generally captures the nature of the underlying surface. More work on the effect of random and grid sampling on free-form surfaces must be performed.

TABLE 1
A COMPARISON OF ESTIMATION AND TRAINING ERROR
FOR POLYNOMIAL REGRESSIONS AND NEURAL NETWORKS

Grid Sampling of Data
(Mean Squared Error)

Function	HONN	Polynomial Reg.
(1)	0.0036	0.4021
(2)	0.0104	0.4011
(3)	0.0090	0.3958
(4)	0.0081	0.3993

Random Sampling of Data
(Mean Squared Error)

Function	HONN	Polynomial Reg.
(1)	0.0042	0.0069
(2)	0.0036	0.0067
(3)	0.0042	0.0064
(4)	0.0081	0.0066

These figures represent the mean squared error for the datapoints used to train the network or fit the indicated polynomial function. In each of the eight experiments, 121 points were used, either from an 11 X 11 grid or a random (x,y) sample.

TABLE 2
PREDICTION LOSS FOR THE EXPECTED VALUE
OF THE SURFACE OF THE MEXICAN HAT FUNCTION

Sample Scheme	Std. Error of Data	Expected Loss	Median Loss	Std. Error of Mean Loss	Interquartil Range of Loss	Maximum Loss
Random	0.00	0.0889	0.0542	0.0040	0.1316	0.5025
			$p < 0.0123$			
Grid	0.00	0.0955	0.0411	0.0048	0.1507	0.4251
			$p < 0.0001$			
Random	0.08	0.1580	0.1033	0.0066	0.2094	0.8555
			$p < 0.0093$			
Grid	0.08	0.1723	0.1270	0.0058	0.1753	0.6591
			$p < 0.0001$			
Grid	0.20	0.4719	0.3931	0.0151	0.5644	2.0094
			$p < 0.0025$			
Random	0.20	0.4924	0.3700	0.0181	0.6586	2.2848

Values in this table represent squared error loss functions for 676 points sampled on a 26 X 26 grid. Statistical tests were performed using a repeated measures design and MANOVA. All differences between experiments were significant at the p-value indicated.

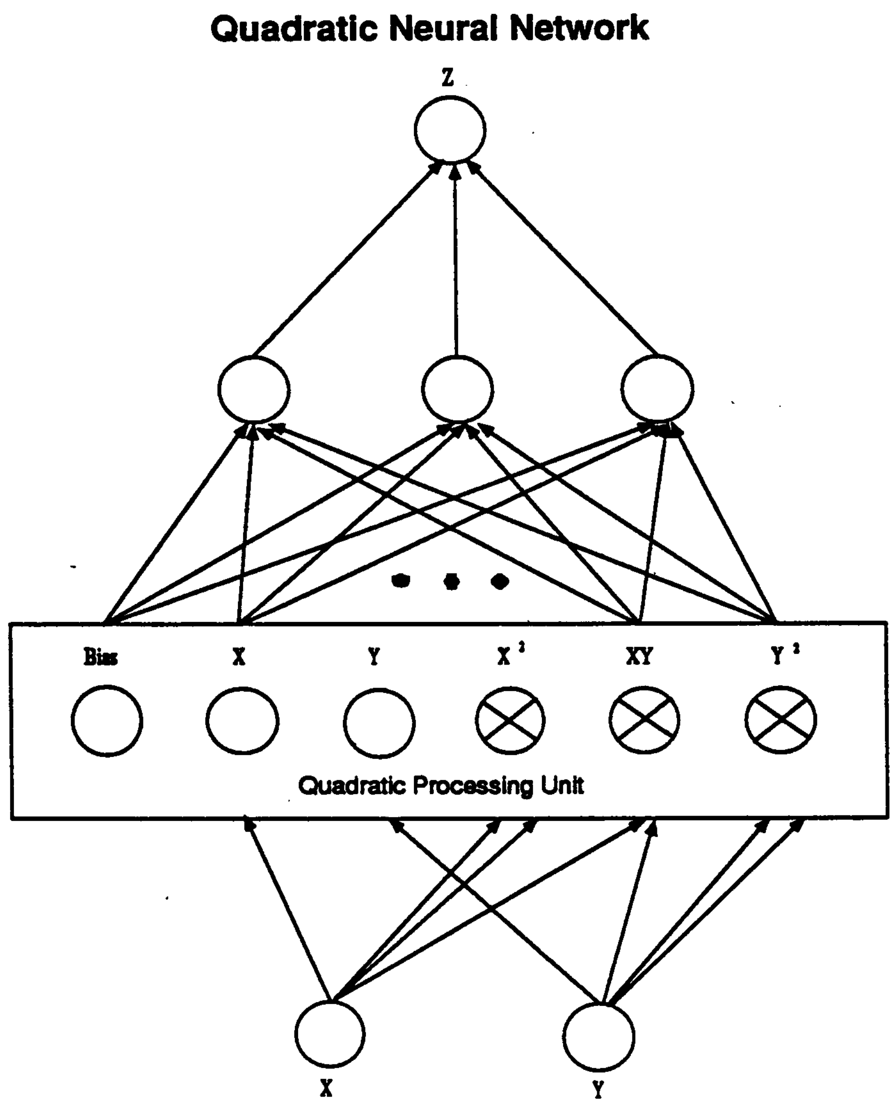


Figure 1 Organization of a quadratic neural network with one hidden layer.

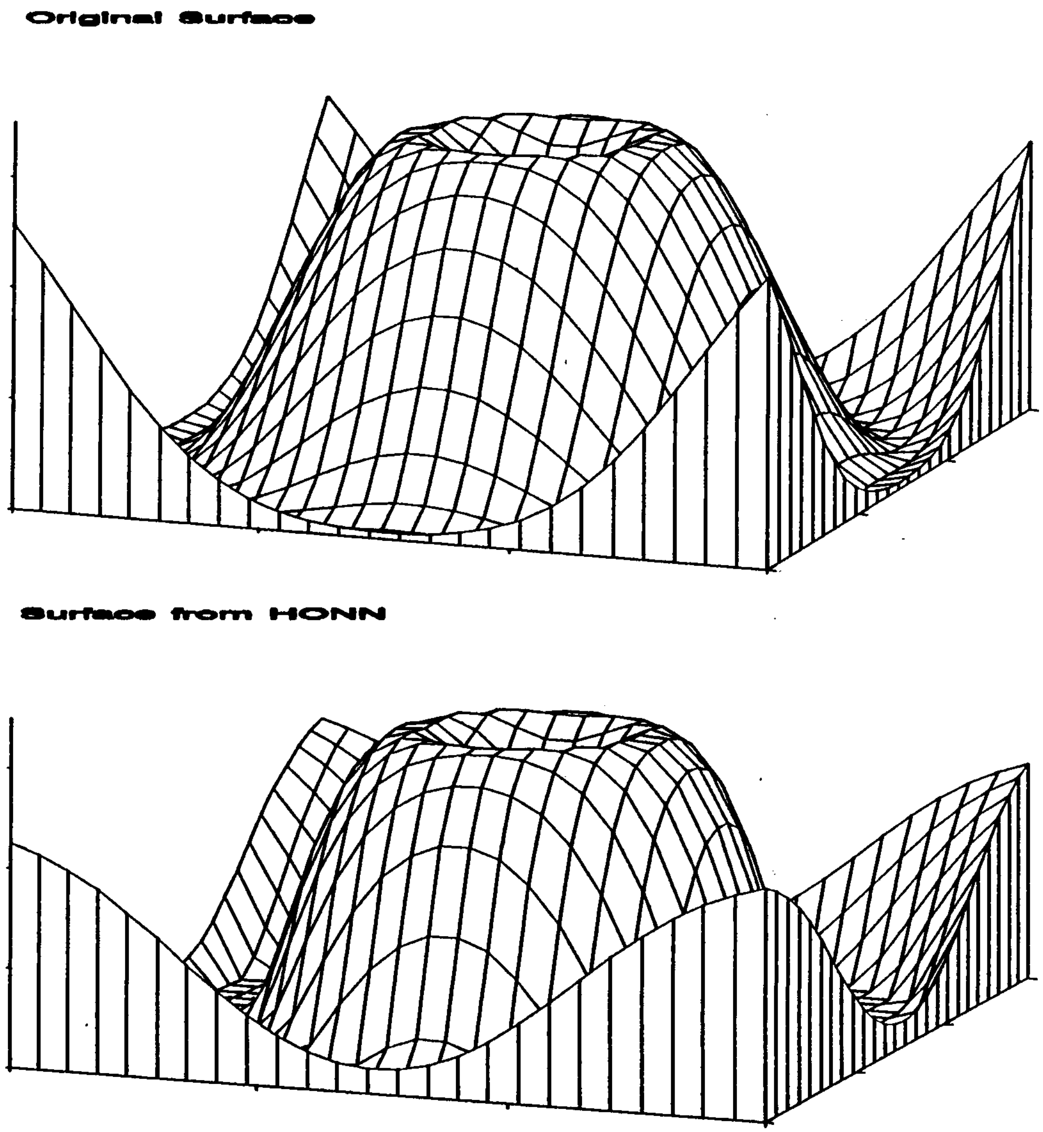


Figure 2 Example of Mexican hat surface and result of HONN reconstruction from 26x26 grid samples.

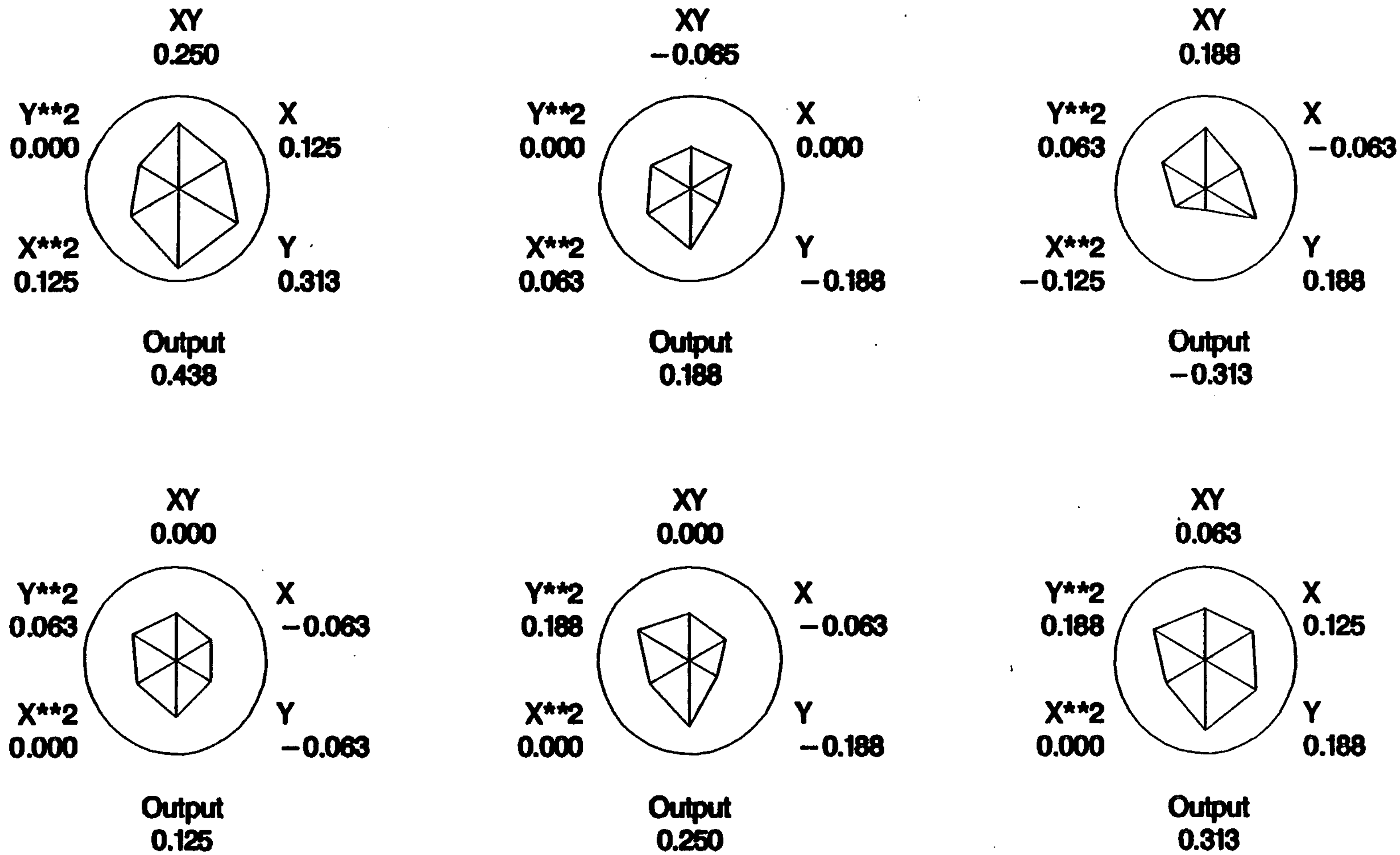


Figure 3. Weights in hidden layers trainned by grid (row 1) and random (row 2) samples from equation (1). The six spines represent one output weight and five input wieghts, as noted.

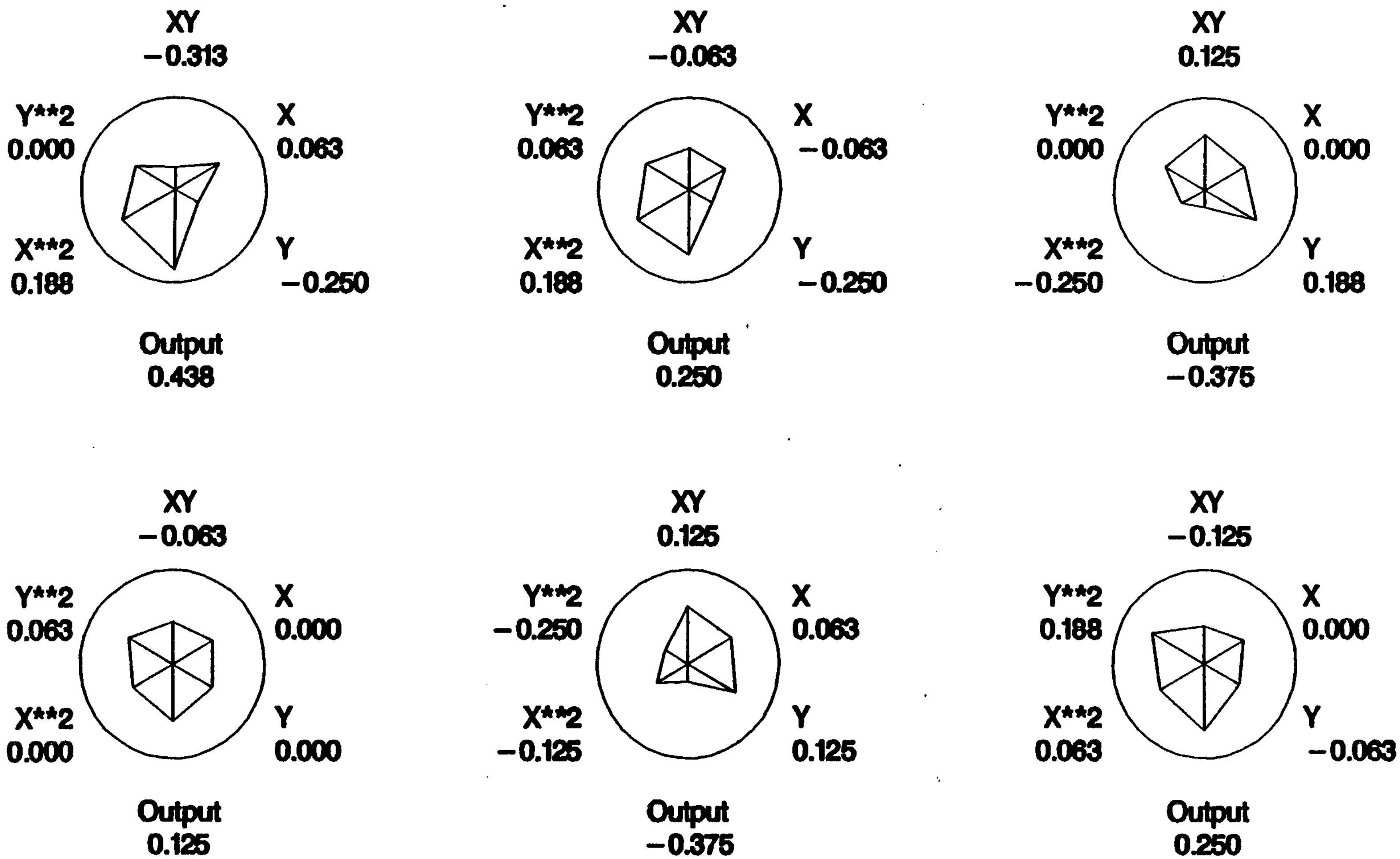


Figure 4. Weights in hidden layers trainned by grid (row 1) and random (row 2) samples from equation (2). The six spines represent one output weight and five input wieghts, as noted.

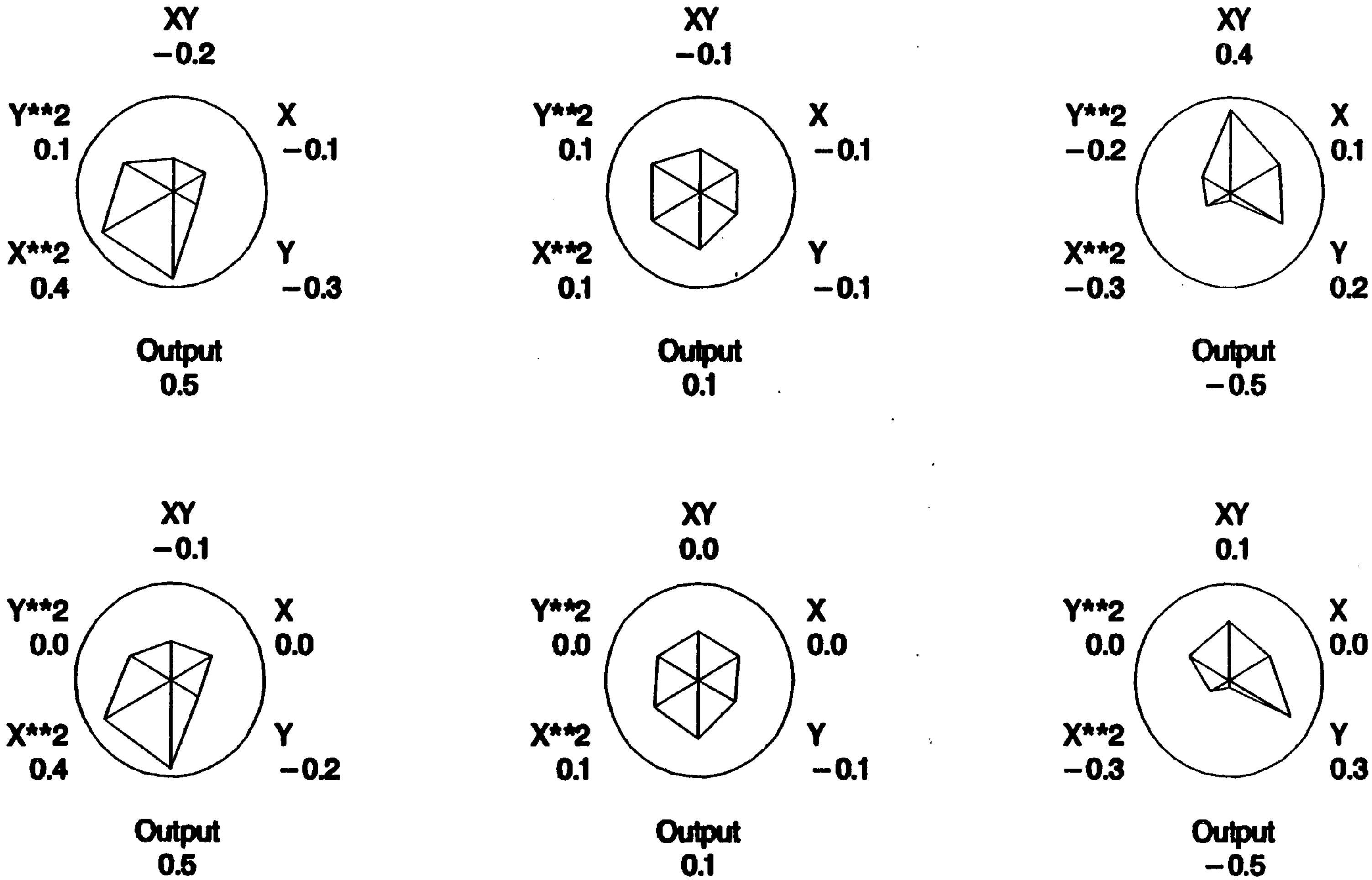


Figure 5. Weights in hidden layers trainned by grid (row 1) and random (row 2) samples from equation (3). The six spines represent one output weight and five input wieghts, as noted.

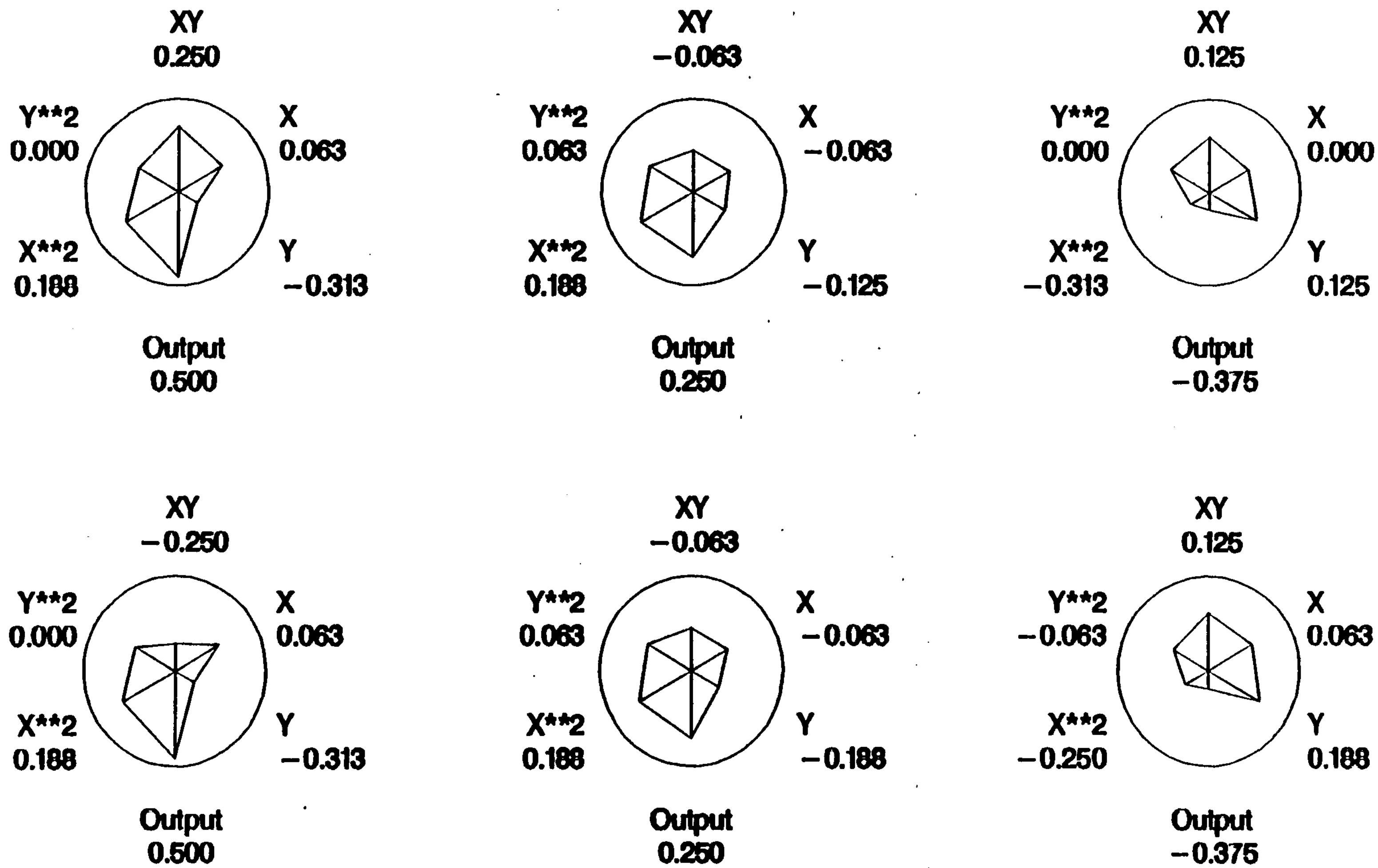


Figure 6. Weights in hidden layers trainned by grid (row 1) and random (row 2) samples from equation (4). The six spines represent one output weight and five input wieghts, as noted.

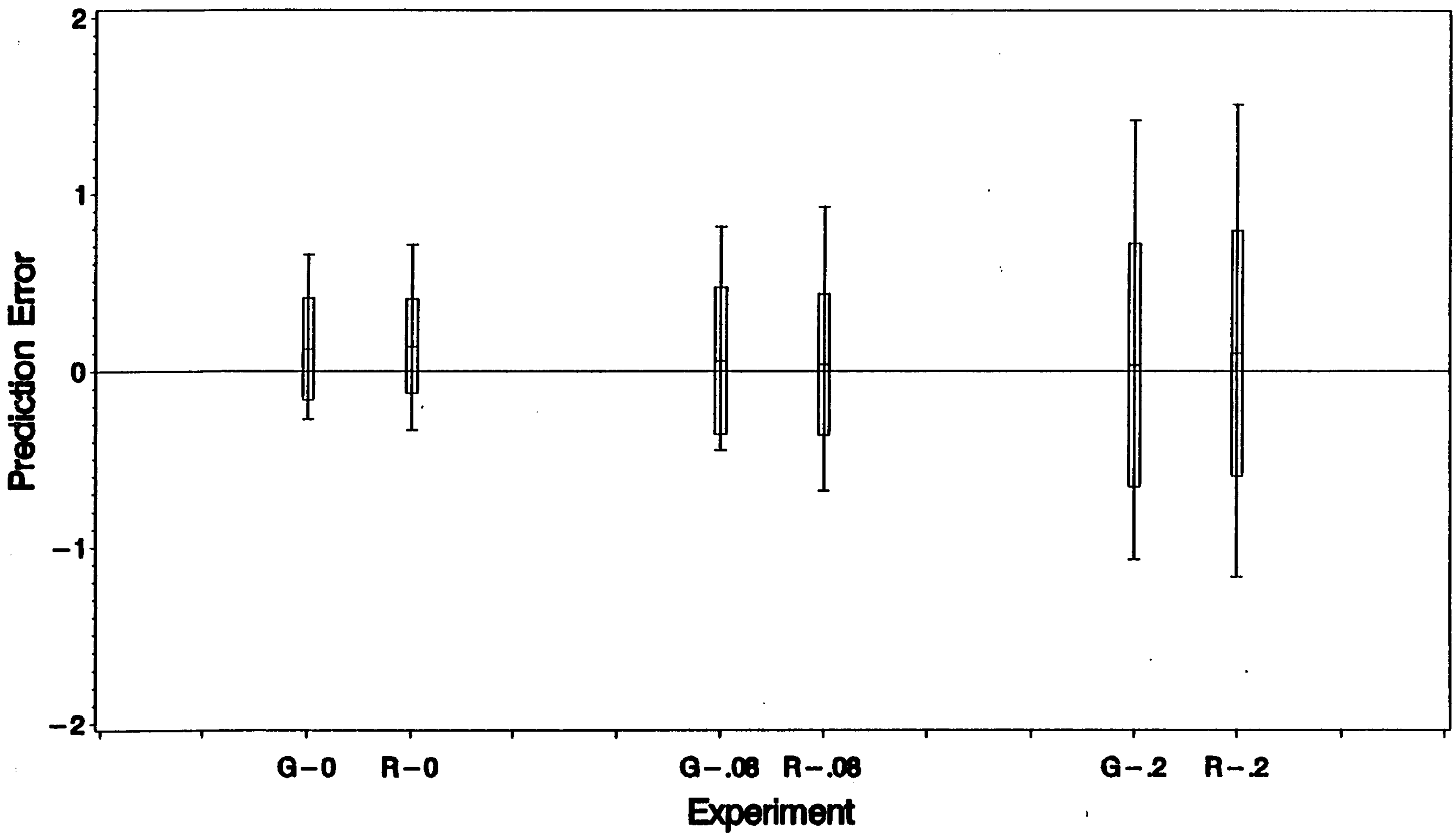
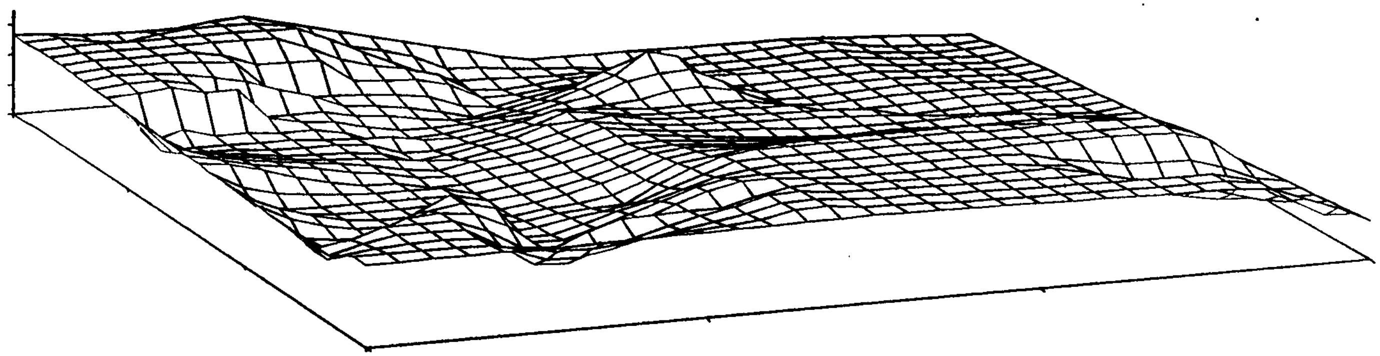


Figure 7. Prediction error for the Mexican hat function. Horizontal axis values indicate the sampling scheme, (G)rid or (R)andom and the standard error for the generated surface. Bars mark 1 standard deviation above and below the mean.

Original Elevation Surface



Surface from HONN

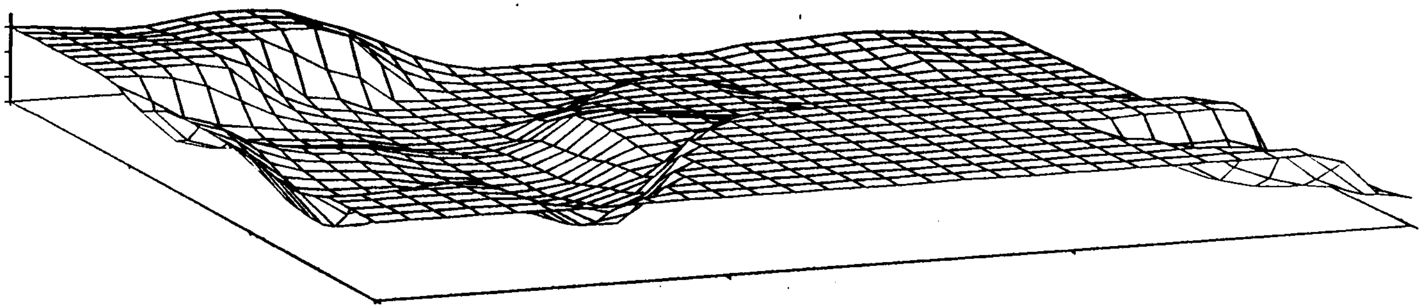


Figure 6 A free—form surface derived from Digital Elevation Data representing land surface elevation in Douglas County, Kansas and result of HONN reconstruction from 31x31 grid samples.

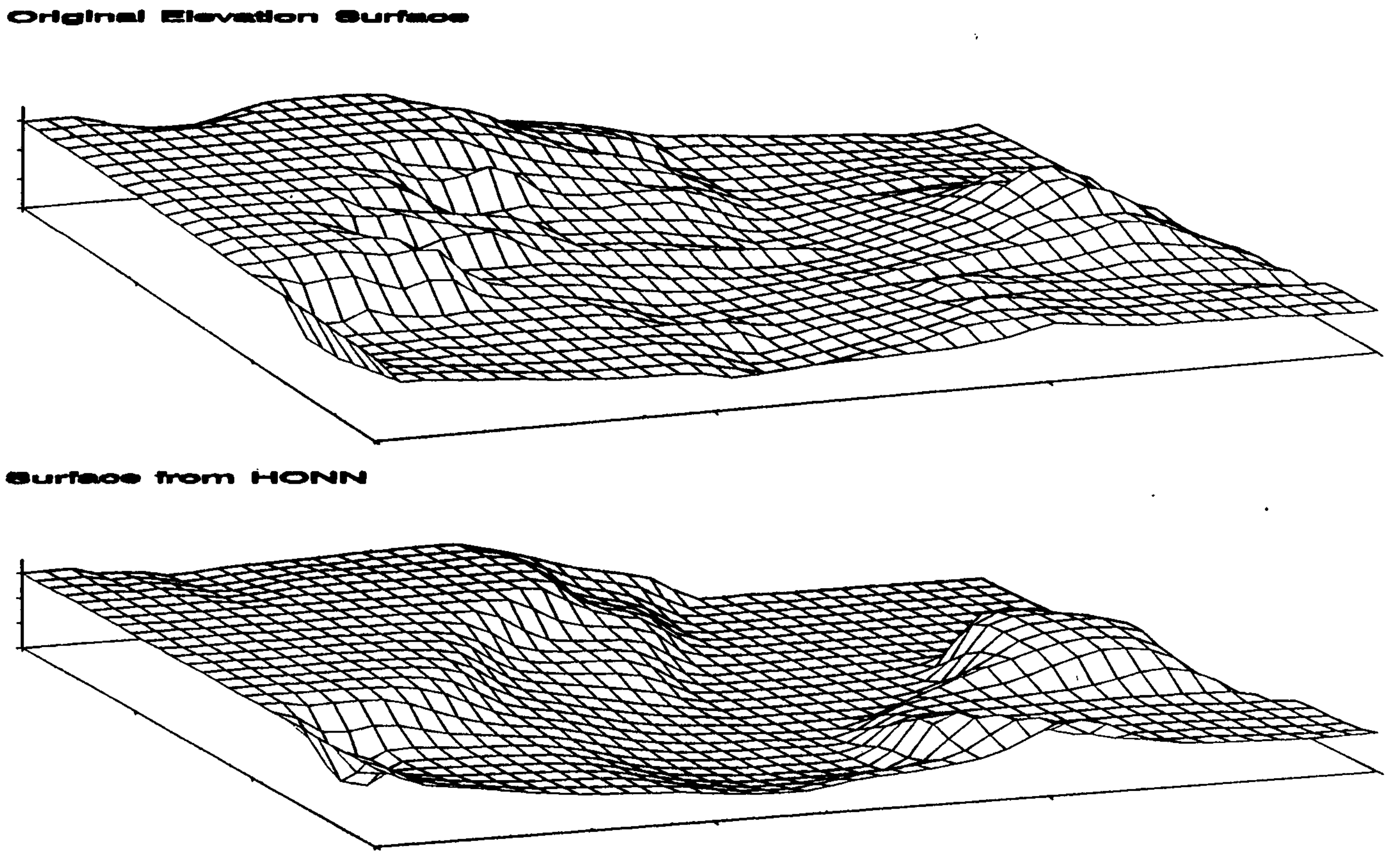


Figure 9 A free—form surface derived from Digital Elevation Data representing land surface elevation in Douglas County, Kansas and result of HONN reconstruction from 41x26 grid samples.

REFERENCES

[Bo91] Ruud M. Bolle and Baba C. Vemuri, "On Three-Dimensional Surface Reconstruction Methods," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 13, No. 1, pp. 1-13, Jan. 1991.

[Be90] Paul J. Besl, "The Free-Form Surface Matching Problem," Machine Vision for Three-Dimensional Scenes. Academic Press, Inc. San Diego, CA, 1990.

[Ch91] David S. Chen, Ramesh C. Jain, and Brian G. Schunck, "Discontinuity-Preserving and Outlier-Removing Surface Reconstruction Using Multilayer Neural Networks," Proceedings of Artificial Neural Networks in Engineering 1991, pp. 461-466, 1991.

[Mi92] James E. Mitchell, Hao Sun and Swapan Chakrabarti, "Predicting the Form of a Surface from Random and Nonexperimental Samples Using Neural Networks," Proceedings of First Midwest Electro-Technology Conference, pp.78-81, 1992.

[Ho89] K. Hornic, M. Stinchcombe and H. White, "Multilayer Feedforward Networks Are Universal Approximators," Neural Networks, Vol. 2, pp. 359-366, 1989.

[Mu90] B. Muller and J. Reinhardt, Neural Network: An Introduction. Springer-Verlag, Berlin, Heidelberg, 1990.

[He91] John Hertz, Anders Krogh, and Richard G. Palmer, Introduction to the Theory of Neural Computation. Addison-Wesley, Redwood City, California, 1991.

[Da86] John C. Davis, Statistics and Data Analysis in Geology. John Wiley & Sons, 1986.

[Vo90] Dennis J. Volper and Steven E. Hampson, "Quadratic Function Nodes: Use, Structure and Training," Neural Networks, Vol. 3, pp. 93-107, 1990.

[SAS90] SAS Software Reference V. 6 First Edition. SAS Institute Inc., Cary, NC, 1990.

[SAS90] SAS Software Reference 1986 Edition. SAS Institute Inc., Cary, NC 1986.

[Sp90] Lilly Spirkovska and Max B. Reid, "Connectivity Strategies for Higher-order Neural Networks Applied to Pattern Recognition," Proceedings of IEEE International Joint Conference on Neural Networks, Vol. 1, pp. I-21 - I-26, 1990.

[Jo90] R. D. Jones and Y. C. Lee, "Function Approximation and Time Series Prediction with Neural Networks," Proceedings of IEEE International Joint Conference on Neural Networks, Vol. 1 pp. 649-959, 1990.

[USGS87] US Geological Survey, "Digital Elevation Models," US Geological Survey National Mapping Program Technical Instructions, Users Guide 5, Reston, VA, 1987.

[Ke92] Peter Kennedy, A Guide to Econometrics Third Edition. The MIT Press, Cambridge, Massachusetts, 1992.