#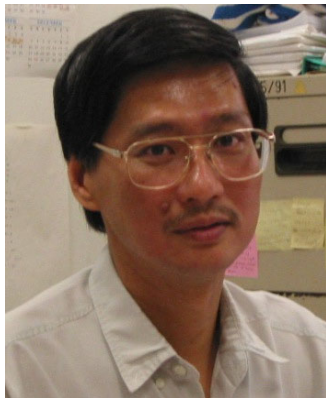 The Optimum Exponent In Inverse-Distance Weighted Method Of Contour Generation – Case Of Environmental Gamma-Ray Dose Rate Distribution

K. K. Lai [1] and S. Minato [2]

1. Physics Department, Universiti Brunei Darussalam, Bandar Seri Begawan BE1410, Brunei Darussalam.
     EMAIL: kklai@fos.ubd.edu.bn

2. Seto Site, Chubu Center, National Institute of Advanced Industrial Science and Technology, 110 Nishiibara-machi, Seto-shi, Aichi prefecture, 489-0884, Japan.

ABSTRACT

The inverse-distance weighted interpolation routine is a very widely used method of contour generation. Despite this fact, hitherto there has been no firm basis that one can refer to when deciding on the optimal exponent value of the weight. We report here our success with a series of comprehensive computer simulations and verification analyses that have produced useful guidelines for the optimal choice of exponent values. The test data were simulated isotropic spatial gamma ray distributions characterised via one-dimensional power spectrum analysis. Since the analysis is independent of the true identity of the data, the results may be extended to other applications, particularly scattered sets of isotropic environmental data. Our results show that the optimum exponent value is not the prevailing favourite of 2, but varies from 3 to 7; lower values for greatly fluctuating or noisy types of data field and higher values for smooth types of data field.

Introduction

In many diverse fields of scientific research, particularly those in the atmospheric, environmental, and geostatistical sciences, it is often necessary to make predictions about the wider spatial distribution of various attributes using information provided by a limited number of sample measurements. The predictions are usually based on some form of interpolations between measurements that have been acquired at sampling sites either randomly located or neatly arranged in a grid pattern. The primary goal is to infer the real distribution patterns, and the predictions are most often presented as contour maps (Franke, 1982). Examples are plentiful in the literature and our own interest is in the generation of contour maps of terrestrial gamma radiation dose rates (Lai *et al.*, 1996).

One of the interpolation techniques is the inverse-distance weighted method, first proposed in 1968 (Shepard, 1968). In its simplest form the value at an evaluation site is calculated by summing the weighted contributions from all of the other sampling sites. The weight is *inversely* proportional to an integer p^{th} power of the distance between the evaluation and the sampling sites. If an evaluation site coincides with one of the sampling sites, then the summation is replaced with the measured value at that site. In some modified forms of this method, the weighted contributions may be taken from only those sampling sites within a specified zone-of-influence around the evaluation site (Franke, 1982).

Because it is simple, easy to calculate and the results are generally acceptable, the inverse-distance weighted method became widely adopted, even though it has an inherent shortcoming of ignoring any underlying trends that may exist in some types of environmental data. In this paper, we address only those cases where the data is assumed to be isotropic in the two-dimensional plane.

For a given set of samples, the value of the exponent p is crucial in determining the outcome of the interpolation. The original paper (Shepard, 1968) had suggested an exponent of 2. In the commercial data analysis software *Sigma Plot* that we have used to perform the contour calculations, the default value was 3, with no choice of zone-of-influence. There have also been some reports on the values of exponents that were found to produce satisfactory results for the specific attributes tested (Gotway *et al.*, 1996; Weber and Englung, 1992 and 1994; Laslett *et al.*, 1987; Phillips *et al.*, 1992; Van Kuilenburg *et al.*, 1982). However, there has never been a comprehensive study that examines systematically the optimum values of p to use for arbitrary data sets of different characteristics.

The question that prompted this paper was this: On what basis do we decide on the best value of exponent for an arbitrary set of data? Since a theoretical analysis is not available, we have opted to perform numerical experiments to obtain the answer. Our strategy was as follows: Find a number of different testing grounds, characterise them fully, collect a sufficient amount of sample data from each, generate multiple series of interpolations with different values of the relevant parameters, and finally, compare the interpolations with the testing grounds to decipher the influence of the various parameters.

The Testing Grounds

The choice of the testing ground is an important issue. Before the advent of computers, one would have chosen a plot of terrain with a regular grid marked out on it, and used the soil parameters as attributes for testing. However, the use of such a testing ground would suffer from several shortcomings: The collection of both the sample and reference data would cause permanent damages to the testing ground. The attributes could change while data were being collected. The data could only be as accurate as the grid pattern and measurement accuracy could provide. Most significantly, the biggest shortcoming was that we could not have complete knowledge of the underlying properties of the terrain, and could not alter it at will.

To overcome the above difficulties, we opted to use computer-generated terrains, which we could create with different properties, and of which we would have complete and detailed knowledge. Our testing grounds were in the form of simulated plots of terrain on two-dimensional square grids. The height from "sea-level" would represent an arbitrary environmental attribute, such as the terrestrial radioactivity dose rate.

The underlying property that we incorporated into the terrain was an isotropic one-dimensional power spectral density (1-D PSD) of the form $a\,k^{-n}$, where a is a constant, k the wave number and n a positive real number. The power index n was used as the characterising parameter for the testing ground and the sample data.

An isotropic 1-D PSD means that the PSD of a height profile along a straight line in the testing ground does not depend on the orientation or location of the line. The spatial distributions of environmental attributes frequently exhibit 1-D PSD of the power law form (Lai *et al.,* 1996; Burrough, 1981; Minato, 1996; Gilbert, 1989). The procedure for calculating the 1-D PSD and extracting the power index from a scattered data set has been published in details elsewhere (Lai *et al.,* 1996).

The power index n can be converted to the fractal dimension D by comparison with the power spectrum of the Weierstrass-Mandelbrot fractal function (Mandelbrot and Van Ness, 1968):

$$D = (5 - n) / 2 \tag{1}$$

Fractal behaviour is observed in the range $1 \leq D < 2$, or equivalently, $1 < n \leq 3$.

The Monte-Carlo simulation used for creating the terrain on a two-dimensional square grid of N intervals and total length L on each side is as follows:

$$z^{*}(x, y) = \sum_{q=1}^{Q} \sqrt{k^{-n}} \cos\left\{ 2\pi(k_{x} x + k_{y} y + R_{1,q}) \right\}$$

$$k = \frac{N}{2L} + R_{2,q}\left(\frac{1}{L} - \frac{N}{2L} \right) \tag{2}$$

$$k_{x} = k \cos(2\pi R_{3,q}) \qquad k_{y} = k \sin(2\pi R_{3,q})$$

$R_{1,q}$, $R_{2,q}$ and $R_{3,q}$ are uniformly distributed independent random numbers between 0 and 1, q is the run number and (x, y) denotes a grid point. Q is the total run number and is set to 10000. The final terrain is constructed as

$$z_{n} (x, y) = m + \frac{\sigma}{\sigma^{*}} z^{*}(x, y) \tag{3}$$

where m is a user-specified mean for offset adjustment, σ a user-specified value for scaling purpose, σ^{*} the standard deviation of the z^{*} values across all the (x, y) points, and the subscript n denotes the power index of the 1-D PSD.

We produced a total of four plots of terrain with the following values of n: 0.5, 1.5, 2.5 and 3.5 respectively. A given terrain was generated on a 120 x 120 grid with 1-unit intervals, but only the portion spanning the central 100 x 100 grid was used as the testing ground. The shoulders on the four sides provided the extra data needed during subsequent smoothing analyses. Without the shoulders, data points on the edges of the ground would get only partial smoothing.

The testing grounds are shown in Figures 1 (a) to (d), and their characteristics are listed in Table 1. As can be seen from the figures, lower values of n correspond to greater extent of roughness. As mentioned earlier in relation to Equation (1), the intermediate cases of $n = 1.5$ and $n = 2.5$ are expected to exhibit fractal behaviour.

Contour Generation and The Blurring Test

For the random sampling and subsequent contour analyses, a non-recurrent random function was used to pick a set of N_s number of sampling points $\{\, z_{ns}(x_i,\, y_i)\,,\quad i=1,\, 2,\dots,\, N_s\,\}$. The contour landscape, which is the 3-D form of a contour map, was generated from the samples on a 20 x 20 square grid of 5-unit intervals. This gave an overall plot size that was the same as the testing ground, but at a lower resolution. We denote the contour landscape as $z_{nc}\,(x_c,\, y_c)$ where $x_c,\, y_c = 10, 15, 20, \dots, 110$.

The inverse-distance interpolation algorithm used in our contour generation has the following form:

$$z_{nc}(x_c, y_c) \;=\; \begin{cases} \dfrac{\displaystyle\sum_{i=1}^{N_s} w_{ci}\, z_{ns}(x_i, y_i)}{\displaystyle\sum_{i=1}^{N_s} w_{ci}} & \text{if} \quad x_c \neq x_i \quad \text{or} \quad y_c \neq y_i \\[4ex] z_{ns}(x_i, y_i) & \text{if} \quad x_c = x_i \quad \text{and} \quad y_c = y_i \end{cases} \tag{4}$$

and,

$$w_{ci} \;=\; \left[(x_c - x_i)^2 + (y_c - y_i)^2\right]^{-\frac{p}{2}} \;=\; \left(\text{distance}_{ci}\right)^{-p} \tag{5}$$

where p is a positive integer. Contribution is taken from all the sampling points, i.e. we did not set any zone-of-influence. Since our testing grounds are characterised by underlying fractal behaviour, we should expect some small influence even from very far away points. If an evaluation point coincides with one of the sampling points, then the weighted summation is discarded and replaced with the sampled value at that point. We took care to reduce the number of such coincidences by using a lower resolution for the contour grid versus a higher resolution for the testing ground grid.

To assess how closely the contour landscape reproduces the testing ground we used the root-mean-squared error analysis, RMSE. For every grid point on the contour landscape, (x_c, y_c), $c = 1, \dots N_c$, we locate the identical grid point on the testing ground and calculate the square of the difference between the pair of values. The square root of the average difference over all grid points is the RMSE:

$$\text{RMSE} = \sqrt{\dfrac{\displaystyle\sum_{c=1}^{N_c} \left[z_{nc}(x_c, y_c) - z_n(x_c, y_c)\right]^2}{N_c}} \tag{6}$$

The RMSE calculates the absolute errors and places equal weight on all points. The errors were not weighted by z_n, as doing so would have enhanced mismatches in the valleys and decreased the impact of mismatches in the peaks.

If the RMSE value is zero it implies a perfect fit at all the grid points. However, strictly speaking, the same cannot be inferred for areas in between the grid points. Therefore, although a lower RMSE value can be interpreted as indicating probably an overall closer match than one with a higher RMSE value, it may not necessarily indicate a better choice. The interpretation must be made with caution and in relation to other observations.

We thought that the RMSE by itself is not an absolute best-fit indicator, but its derivative might be. Generally, a contour landscape is a simplified image of the testing ground put together using the information provided by a collection of samples. At best it may be expected to resemble a "blurred" image of the testing ground. Thus, we proposed to perform RMSE analyses between the contour landscape and a series of increasingly "blurred" versions of the testing ground. A plot of the RMSE versus the blurring parameter should then show a minimum corresponding to the best-match condition. We called this the "blurring test".

To produce the blurring effect, we used a moving-average routine that calculates the blurred height at the evaluation grid point G as the average of all the original un-blurred heights on a set of grid points surrounding G. The set is defined as consisting of all the grid points on or inside the boundary of a square centred on G, and includes G itself. The term "moving-average" refers to the fact that as the evaluation point moves through the grid, the "square-of-influence" also moves along with it. The side of the square is of length $2N$ intervals, where the integer $N \geq 0$. This routine is sometimes referred to as a $(2N+1)^2$-point moving average routine, in reference to the total number of grid points in the set. The blurring parameter corresponds to N. Thus, $N = 0$ means no smoothing, and $N = 2$ means 25-point smoothing. The grid resolution is not altered by the smoothing routine. The all-round borders of 10-intervals wide discussed earlier allowed the smoothing to be carried out right to the edge of our testing ground provided $N \leq 10$.

Collections of samples were done using the following procedures. From each testing ground we extracted measurements from 400 randomly located sampling points to form one set of data. This was then repeated 40 times to yield 40 sets of data for each testing ground. From each set of data, we generated 8 contour landscapes in a series with p varying progressively from 1 to 8. RMSE calculations were then performed on each contour landscape by comparing it systematically with a series of increasingly smoothed versions of the testing ground, i.e. with N varied progressively from 0 to 10. Thus, for each testing ground we generated a total of 320 contour landscapes and performed a total of 3520 RMSE calculations. To ensure that valid comparisons could be made between the different testing grounds, we took care to use the same collection of 41 sequences of random numbers: one for generating the testing ground and 40 for extracting the random samples.

Analysis and Discussions of the RMSE results

The RMSE results have been condensed into Figure 2. Every point on the graph is the average value obtained from the 40 sets of data with the standard deviations indicated by the error bars. Points with the same value of p are joined in a line.

One interesting feature in Figure 2 is that for a given p, the line typically starts with a negative gradient then bends upwards forming a dip. The significance of this is centred on our use of RMSE to measure the extent of mismatch between a contour landscape and smoothed copies of the testing ground. The presence of a dip at $N > 0$ indicates that statistically, the contour landscape matches a smoothed copy of the testing ground better than the original.

As p increases, the dip tends to a lower limit in N, marked as N_0 in the figures. The existence of a lower limit indicates clearly that the inverse-distance routine is not capable of generating a contour landscape with any more details than a $(2N_0+1)^2$-point smoothed copy of the testing ground. In other words, it is reasonable to regard N_0 as an indicator of the limit of correspondence. This important observation vindicates our use of the "blurring tests", defying the traditional wisdom of judging a contour landscape against the original testing ground.

N_0 appears to have the same value for all the cases in Figure 2. Since all these cases

have data sets of the same size, could changing the size of the data set affect the N_0 limit? In Figure 3, we varied the size of the data set while maintaining the same grid resolution. As the number of sampling points increases, N_0 can be clearly seen to decrease, indicating an improved limit of correspondence. This agrees with the general notion that a larger data set should yield a more accurate contour landscape.

Since N_0 can be regarded as an indicator of the limit of correspondence, then in our search for the optimal p value, the choice should not be based on the RMSE scores at N values that are less than N_0. On the other hand, neither should it be based on the overall lowest RMSE score, since this may occur at a large N value. The simplest choice might be the p value that gives the lowest RMSE score at N_0. However, this choice may turn out be one whose RMSE dip occurs far away from N_0, as is the case in Figures 2a and 2b, implying a closer resemblance to a more highly smoothed landscape. Since our desire is to produce a contour landscape that overall resembles the original as closely as possible, we should look for a RMSE dip at a low N value, limited only by N_0. Therefore, we propose a more prudent selection rule, which is to choose the p value that gives rise to a RMSE dip that is the lowest among those at or very near to N_0.

If the above selection rules were applied to Figure 2b where $N_0 = 3$, the resultant optimal choices would be: i) $p = 3$ using the first rule, ii) $p = 5$ or 6 using the second rule. The two cases of 5 & 6 have overlapping standard deviations, but since the dip of $p = 6$ is closest to N_0, the preference is on 6. Although $p = 4$ has a lower RMSE dip, the position of the dip is mid-way between $N = 3$ and $N = 4$. This indicates resemblance to a landscape of greater extent of smoothing than the 49-point smoothing of $N = 3$. Therefore $p = 4$ is not considered the optimal choice.

Could we have obtained a different set of optimal p values if we had changed the size of the data sets? A detailed examination of Figure 3 shows that although the limit N_0 shifts with the number of sampling points, this change does not appear to have any influence on the ordering of the curves, nor the proximity of the RMSE dips to N_0. Therefore, the same optimal values would have been obtained had the data sets been of a different size.

Whatever the values, it is interesting that none of the above cases gives the commonly assumed value of $p = 2$. We summarise in Table 2 the optimal values of p for the various testing grounds obtained using the first and second criteria.

To gain more insights, we plot in Figure 4 the mean contour landscapes for $p = 2$, 3 and 6. Each of these is the mean of the 40 contour landscapes that were used in the calculation of Figure 2b. The smoothed testing grounds with $N = 3$, 4 and 10 have also been shown in the bottom row such that the vertical pairings relate to the dips of the respective p curves in Figure 2b: ($p = 2$, $N = 10$), ($p = 3$, $N = 4$), ($p = 6$, $N = 3$).

A glance at Figure 4 would confirm that visually, the resemblance between any vertical pair is much better than any diagonal pair. For example, the $N = 3$ landscape at the right of the bottom row shows how the original testing ground would look like after a 49-point smoothing routine is performed on it. A casual glance at the top row would suggest that both the $p = 3$ and $p = 6$ mean contour landscapes appear similar to it. However, closer examination shows that the valleys along X = 40 and the blue peaks above the green bands are better reproduced in $p = 6$ than in $p = 3$. In other words, the right-most vertical pair has the best overall resemblance between the partners. This observation lends supports to the second criterion that gives $p = 6$ as the optimal choice based on the proximity of its RMSE dip to N_0, located at $N = 3$.

It is interesting to note that in fact, 6 out of the 9 possible pairs in Figure 4 have lower RMSE values than the optimum pair. As discussed earlier, the trend and dip of the RMSE curves are of more importance than the absolute RMSE values. The reason is that an isolated

RMSE calculation can over-emphasise a small number of localised medium or large differences between the contour landscape and the testing ground. Therefore we need to track the derivative of the RMSE calculations, as we have just demonstrated, in order to discover the true picture.

Theoretically, it is difficult to explain the occurrence of high exponent values. However, if we bear in mind that the weight in Equation 4 is a merely a means of apportioning influence from the randomly distributed sample points that surrounds the evaluation site, then it can be appreciated that high exponent value would result in the filtering out of far away influences better than low exponent value. In smoothly varying terrains, this would be desirable as the local trend is more crucial than long-distance similarities. The net effect of a high exponent value would have been equivalent to restricting contributions to come only from a small zone-of-influence around the evaluation site. This might explain why the use of a zone-of-influence was found to be useful by some authors (Franke, 1982). With rough terrains like those in Figure 1a and 1b, the situation would be opposite: long-distance similarities are more prominent than local trend. In such cases, low exponent values would be more suitable.

Although our test data originally arose from simulated isotropic spatial gamma ray distributions, the simulation process itself was characterised by only one parameter: the 1-D PSD power index, k. Thus, the analysis is actually independent of the true identity of the data. Therefore, we believe that our results may be extended to other applications, particularly scattered sets of isotropic environmental data.

Conclusion

In conclusion, we have shown through a series of numerical experiments that the most suitable exponent for use in the usual form of the inverse-distance weighted method of contour generation may not be the commonly assumed value of 2. On the contrary, the optimum power may be from 3 to 7 depending on the characteristic of the data field, measured via the power index of an isotropic one-dimensional power spectral density. The results indicated that a higher value should be used for a smooth type of data field and a lower value for a greatly fluctuating or noisy data field.

REFERENCES

Burrough, P. A., 1981. Fractal dimensions of landscapes and other environmental data, Nature 294, 240-242.

Franke, R., 1982. Scattered Data Interpolation: Test of some Methods, Mathematics of Computations 38, 181-200.

Gilbert, L. E., 1989. Are topographic data sets fractal? PAGEOPH 131, 241-254.

Gotway, C. A., Ferguson, R. B., Hergert, G. W. and Peterson, T. A., 1996. Comparison of Kriging and Inverse-Distance Methods for Mapping Soil Parameters, Soil Sci. Soc. Am. J. 60, 1237-1247.

Lai, K. K., Hu, S. J., Minato, S., Kodaira, K. and Tan, K. S., 1999. Terrestrial gamma ray dose rates of Brunei Darussalam, Appl. Radiat. Isotopes 50, 599-608.

Laslett, G. M., McBratney, A. B., Pahl, P. J. and Hutchinson, M. F., 1987. Comparison of several spatial prediction methods for soil pH. Journal of Soil Science 38, 325-341.

Mandelbrot, B. B. and Van Ness, J. G., 1968. Fractal Brownian motions, fractal noises and applications. SIAM Review 10, 422-437.

Minato, S., 1996. The structure of spatial fluctuation in terrestrial gamma-ray dose rate levels observed through one-dimensional surveys. Radioisotopes 45, 619-625.

Phillips, D. L., Dolph, J. and Marks, D., 1992. A comparison of geostatistical procedures for spatial analysis of precipitation in mountainous terrain. Agricultural and Forest Meterology 58, 119-141.

Shepard, D., 1968. A two-dimensional interpolation function for irregularly-spaced data, Proc. 23rd Nat. Conf. ACM, pp. 517-524.

Weber, D. D. & Englung, E. J., 1992. Evaluation and comparison of spatial interpolators, Math. Geol. 24, 381-391.

Weber, D. D. & Englung, E. J., 1994. Evaluation and comparison of spatial interpolators II, Math. Geol. 26, 589-602.

Van Kuilenburg, J., De Gruijter, J. J., Marsman, B. A. & Bouma, J., 1982. Accuracy of spatial interpolation between point data on soil moisture supply capacity, compared with estimates from mapping units. Geoderma 27, 311-325.

Table 1: Characteristics of the four testing grounds, where n is the power index of the 1-D PSD, $a\,k^{-n}$.

Value of n specified	Value of n from PSD analysis of 10 sets of 400 random samples each	Reasons for the choice of n
0.5	0.65 ± 0.29	non-fractal, noisy terrain
1.5	1.73 ± 0.14	fractal behaviour, value of n is typical of the distribution of terrestrial gamma radiation
2.5	2.69 ± 0.23	fractal behaviour, value of n is typical of a wide range of environmental attributes
3.5	3.44 ± 0.37	non-fractal, smooth terrain

Table 2: The optimal values of p obtained using the first and second criteria (see text) respectively.

Testing Ground, n	First Criterion	Second Criterion
0.5	2	3
1.5	3	6
2.5	4	6
3.5	4	7
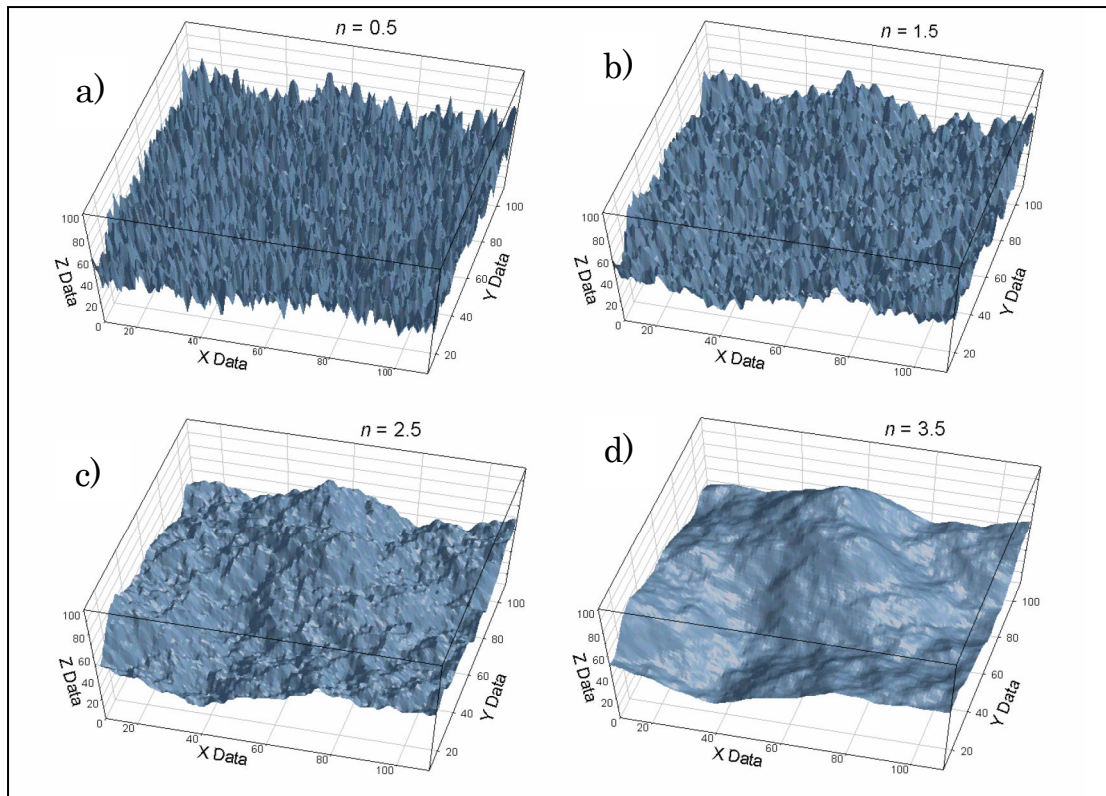
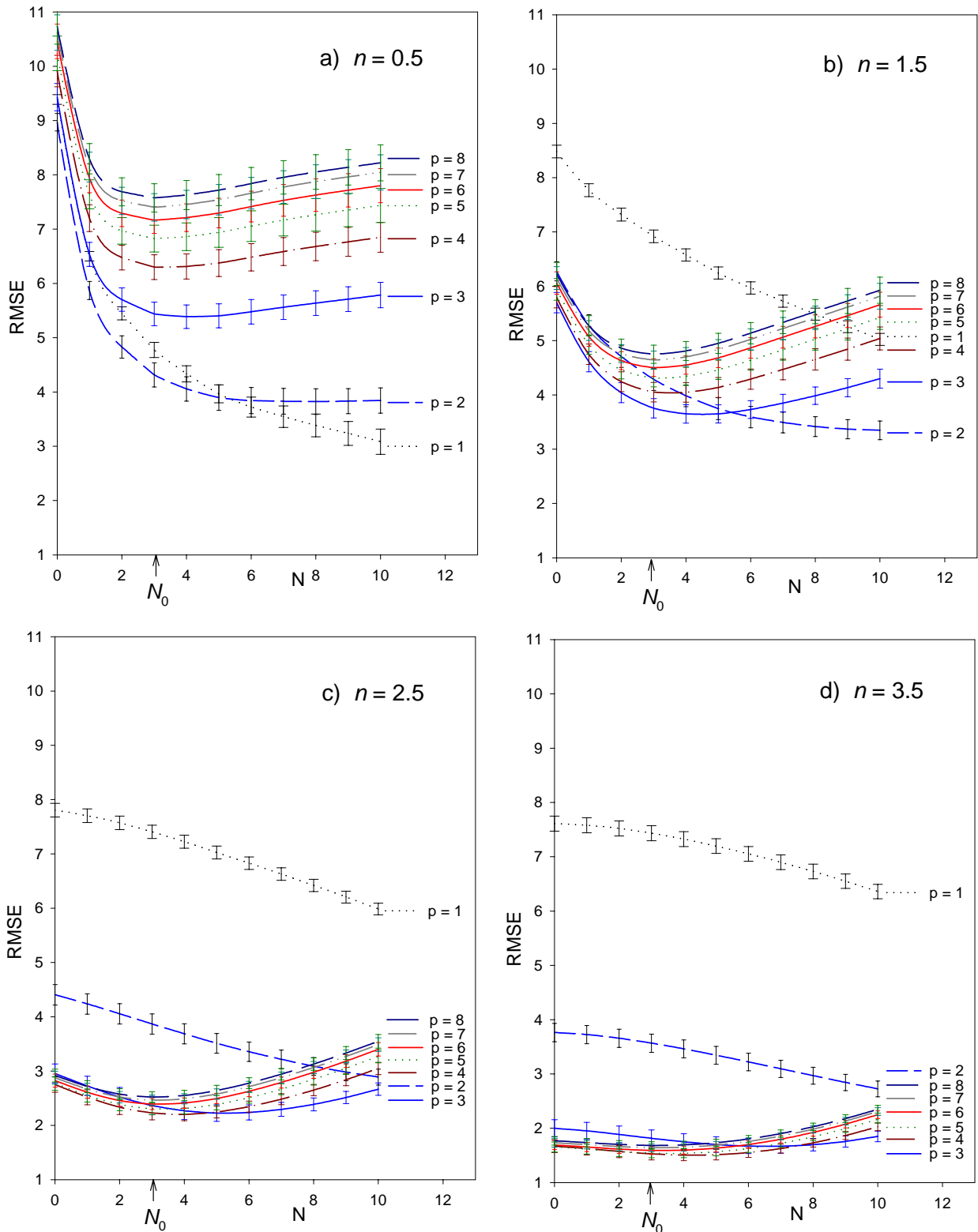**Figure 1.** The four testing grounds with their respective power indices.

Figure 2. Results of the RMSE analyses between contour landscapes and smoothed testing grounds. The testing grounds are identified by n, the contour landscapes are characterised by p, and N is the blurring parameter. 40 sets of random samples were extracted from each testing ground and used to generate 40 contour landscapes for each p value. The data point on the curves represents the average result obtained from the RMSE analyses performed on those 40 contour landscapes, and the error bars indicate the standard deviations.
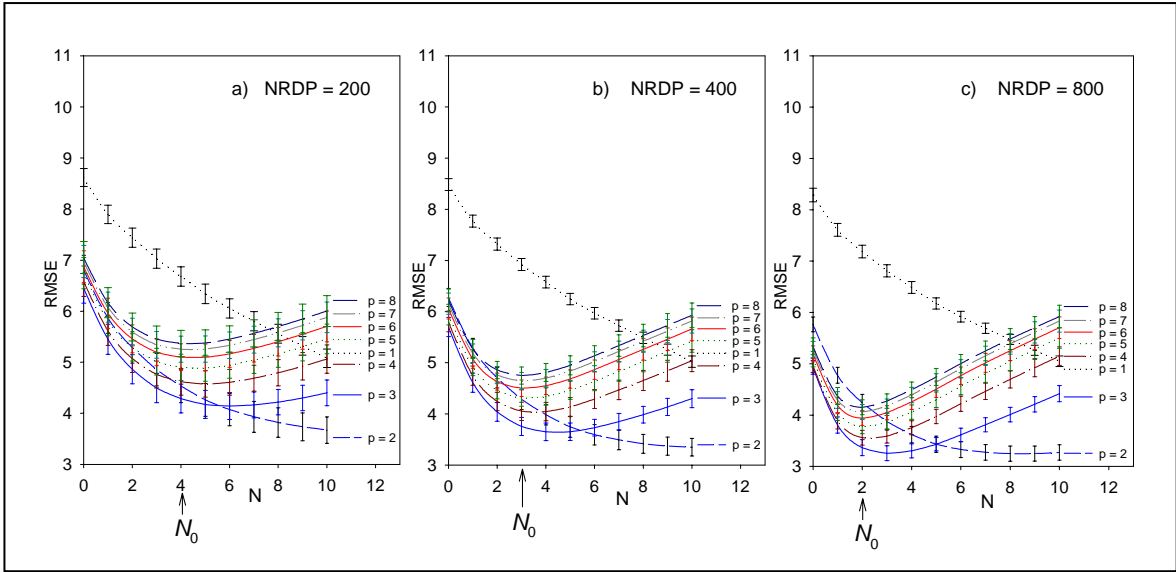
Figure 3. The effect of changing the number of random data points (NRDP) on the shape of the RMSE curves and the value of N_0. The testing ground is $n = 1.5$. All symbols have the same meaning as in Figure 2.
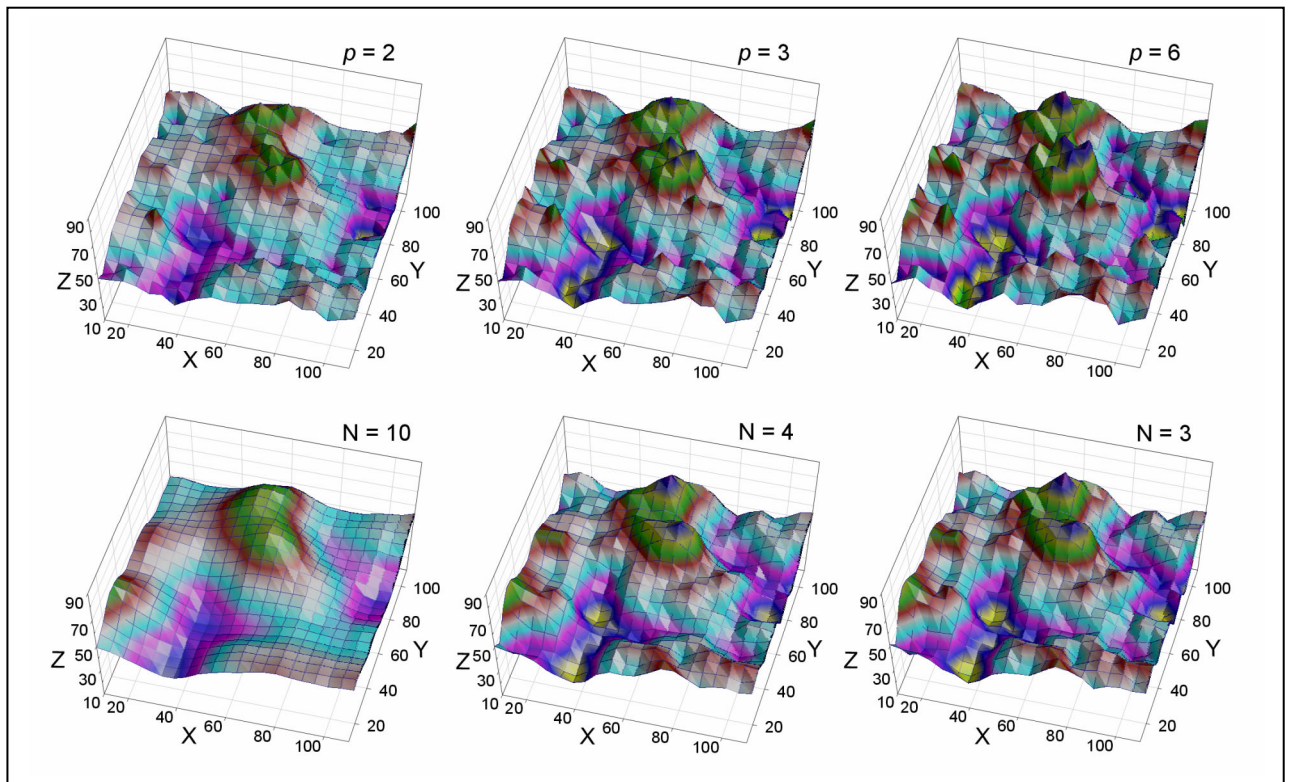

Figure 4. The top row shows 40-sets-averaged contour landscapes generated using the indicated p values for the testing ground of $n = 1.5$. When each of these averaged contour landscapes is evaluated against a series of smoothed testing grounds, the lowest RMSE is obtained when the evaluation is against the smoothed testing ground shown directly below it.