PERFORMANCE EVALUATION OF MULTIVARIATE INTERPOLATION METHODS FOR SCATTERED DATA IN GEOSCIENCE APPLICATIONS

Matthew P. Foster and Adrian N. Evans

Department of Electronic & Electrical Engineering, University of Bath, UK

ABSTRACT

Interpolation methods are widely used in geoscience applications to reconstruct multivariate data from irregular samples. This paper describes a quantitative methodology for assessing the performance of various state-of-the-art interpolation methods. The methodology consists of simulation-validation and cross-validation using simulated and real data respectively, and has recently been applied to study the reconstruction of total electron content maps of the ionosphere. These two approaches are described and a study of the various artefacts associated with different interpolation methods also presented, including their origins and typical locations. Finally, the use of the statistical moments of error histograms as a method of evaluating techniques for biases and skew is described, as well as providing confidence bounds on error values. The methodology and artefact analysis should be of use to anyone who uses multivariate interpolation methods.

Index Terms— Image reconstruction, interpolation, remote sensing, geophysics

1. INTRODUCTION

Scattered data sets are common in geosciences, arising wherever irregular sampling patterns are employed or point measurements made. Application areas where such data are found are very diverse and include, for example, salinity data returned from the over 3100 freely moving Argo floats, core samples of mineral deposits, aquifer head height measurements, global precipitation measures (see, e.g. [1]) and line of sight measurements of the total electron content (TEC) of the ionosphere.

As scattered data-sets rarely include samples in the desired configuration, multivariate interpolation methods (MIM) are commonly applied for reconstruction, or calculating values at desired positions. Some example geoscience applications that employ MIM include visualisation, gridding, contouring, slicing, mapping, modelling, reconstruction and analysis. In fact, a great many data-products which are in widespread, daily use are the result of various degrees of interpolation.

Although many geoscientists may use MIM on a regular basis, the relative performances and nuances of the different MIM that are available and the affect that data sparsity can have on the quality of reconstructed data are not always widely appreciated. As these issues and effects can influence the quality of the scientific work it is important that practitioners understand both how the various MIM operate as well as the problems that are associated with specific techniques. This paper attempts to address these issues by posing and answering the following four questions:

M. P. Foster is funded by a UK EPSRC doctoral training award.

1. How can the relative performance of MIM be assessed under varying input conditions?
2. How can the performance of MIM on real data be verified?
3. What artefacts and behaviour do different MIM exhibit?
4. How can MIM be checked for biases and inconsistencies?

The first two questions were the subject of a recent study on the reconstruction of ionospheric TEC maps, which proposed a methodology for the quantitative assessment of the performance of MIM [2]. This methodology employed both simulated and real data and is described in section 2. The interpolation methods that were considered included several triangulation-based MIM including linear, cubic [3] and natural-neighbour (NN) interpolation [4]; radial basis function interpolation [5], biharmonic spline interpolation[6], ordinary kriging [7, 8] as well as adaptive normalised convolution (ANC) [9], a technique based on convolution with rotated filters that is new to geoscience applications. Although a quantitative evaluation is the method of choice for determining the performance of MIM, it is unable to provide specific information on the types of artefacts produced by various MIM and the situations in which they are likely to occur; these are discussed in section 3. The error distributions of the MIM are also considered in section 4 and conclusions drawn in section 5.

2. QUANTITATIVE EVALUATION METHODOLOGY

Quantitative evaluation methodologies can conveniently be divided into two main classes: those which require a full and correct output to be known *a priori*, and those which can operate on real data for which the correct outputs are unknown. These are known as simulation-validation (SV) and cross-validation (CV), respectively. To fully characterise the performance of MIM, the methodology in [2] proposed a combination of the two approaches, first using SV in conjunction with simulated data to quantify the performance of the MIM and then secondly – as simulated data are only an approximation of the data found in real studies – examining their performance on real data using CV.

2.1. Simulation Validation

SV facilitates validation using a ground truth, as the entire input field is generated and sampled as appropriate. This allows for a complete characterisation of output errors, as error values are available for every output point. Fields of data, in which the statistical properties can be tightly controlled, are generated and then randomly sampled to produce data with varying degrees of sparsity. The sampled fields are then reconstructed using various different interpolation methods and the output fields compared to the original inputs. By reconstructing from many different sets of samples and over a range of sparsities, the performance of the various MIM can be characterised.

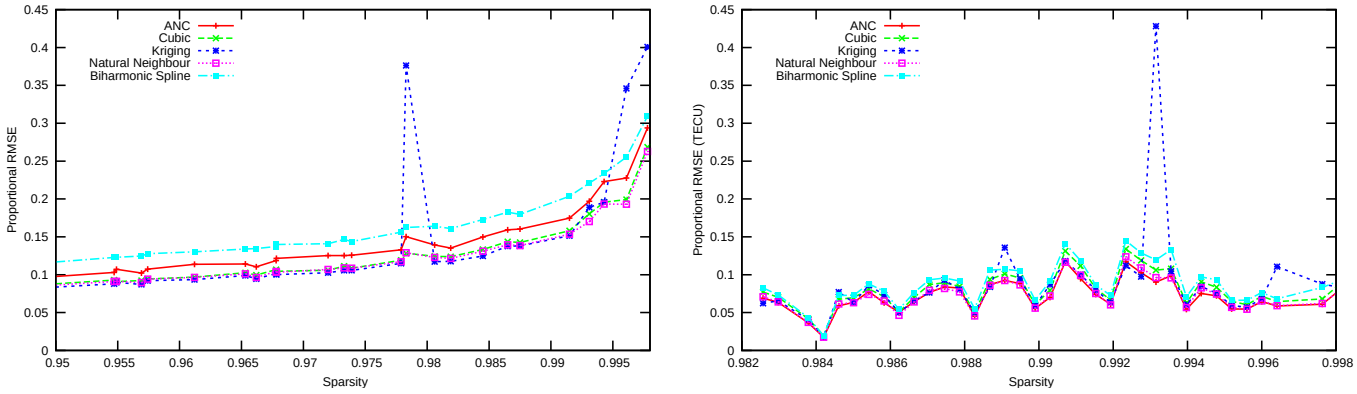


Fig. 1. Normalised RMSE for simulated data (left), and TEC data (right)

Figure 1 presents the proportional root mean square error (RMSE) for different MIM where, for clarity, only five techniques are shown. At each sparsity, the result is the average of multiple reconstructions from different samples, thus ensuring that any sampling effects were removed from the output. The results show the expected increase in error with increasing sparsity. While the relative performance of the different MIM remains reasonably constant over the range of sparsities, there are some anomalous results, particularly when using kriging.

SV using ground truth data provides enough results to allow the spatial and statistical distribution of errors to be examined. Of particular importance is the examination of the statistical error distribution to check the bias of a given MIM. This is discussed in more detail in section 4.

2.2. Cross-validation

CV enables performance analysis using only input data and works by partitioning the data into two sets, an input set and a validation set. The output field is then formed by reconstructing using the input set and the output errors are calculated by subtracting the points in the validation set from the output. By altering the relative sizes of the input and validation sets it is possible to change the sparsity of the data used for the reconstruction. An example study was conducted by applying a number of MIM to ionospheric TEC data, extracted from global positioning system (GPS) signals. Here, k-fold CV was used, where the data are first randomised and then partitioned into several blocks. The MIM are applied to the first block and the remaining blocks used to calculate the errors at the corresponding positions. This process is repeated for various different combinations of block configuration, in order to generate enough data for the reliable calculation of error statistics.

Results from [2] generated using CV are shown in Figure 1 (right) and differ from the simulated results in several respects. Firstly, there is much less of a clear trend of increasing error with sparsity. Secondly, the results are much noisier, because of the nature of the real TEC data. Finally, the relative performance of the MIM is different, with ANC showing the lowest errors, thus confirming the need for evaluation on real data in addition to simulation. However, some similarities are also present, such as the RMSE spikes in the kriging method.

3. RECONSTRUCTION ARTEFACTS

Examining the outputs of different MIM spatially, as well as in relation to one another, can yield useful and important information on specific problems and artefacts occasionally manifested in some schemes. Knowledge of the kind of problems which present themselves, and with which MIM could prove vitally important. Some examples of problems occurring in different schemes include: peaks around data-points in the output of NN interpolation; overshoot in cubic interpolation methods and highly-inaccurate output fields due to ill- or non-fitting models in kriging.

To illustrate how surfaces reconstructed using various interpolation methods differ relative to the input, and one-another, Fig. 2 shows example Shuttle Radar Topography Mission (SRTM) data, sampled to a sparsity of approximately 99% and then reconstructed using five different MIM. Whilst there are no *major* differences between the surfaces produced by the interpolation methods, some subtle differences are evident.

Fig. 2b was interpolated using linear interpolation based on triangulation. Linear interpolation is analogous to fixing triangular plates together over a frame whose vertices are the data points, and with edges following those of the triangles. This gives rise to the characteristic faceted surfaces that are evident in the figure. Reconstruction using radial basis function (RBF) interpolation, Fig. 2c, produces a similar output to the triangulation-based output but does not appear faceted, since the interpolation is not based on triangulation. The main disadvantage of linear RBF interpolation is the fact that being a global method, it is computationally expensive.

Fig. 2d was interpolated using TPS RBF. This technique produces results which are almost identical to biharmonic spline interpolation (BSI), and are characterised by smooth, isotropic surfaces. TPS uses a cubic spline basis function and so has a tendency to overshoot when it tries to maintain continuous derivatives at edge points. A good physical analogy to spline-based interpolation is imagining the image having been interpolated using metal sheets, which are able to bend a certain amount, and whose joins with other plates must be continuous to at least the first derivative.

NN interpolation, see Fig. 2e, again uses the Delaunay triangulation, but uses the ratio of overlapping areas to determine the weighting of data points. This results in surfaces which vary more smoothly than linear interpolation. The most obvious artefacts caused by NN interpolation are sharp peaks around the input datum. The surfaces produced are similar in appearance to a rubber sheet, stretched over

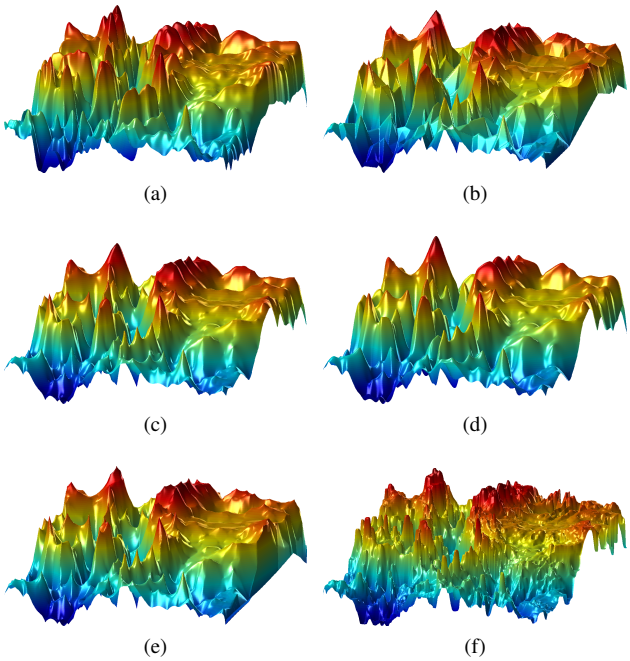


Fig. 2. False colour surfaces for SRTM elevation data. (a) Original input data. Reconstructed outputs from (b) linear triangulation based interpolation; (c) linear RBF interpolation; (d) TPS RBF interpolation; (e) NN interpolation; (f) ANC interpolation.

and attached to the input datum. This is because NN interpolation creates output values based on the area of overlap between the Voronoi cells before and after the input point is added to the image.

Fig. 2f was interpolated using ANC [2, 9] which uses convolution with rotated and scaled Gaussian kernels to perform interpolation. It produces outputs which include anisotropy where necessary, and in this case its output is somewhere between those of NN and TPS interpolation. Artefacts produced by ANC generally result from poorly chosen filter size limits, and usually manifest as narrow peaks around most datum. In extreme cases, where filters are forced to be far too large, the data are over-smoothed. This results in outputs whose values can be significantly below those of the input data.

Examination of the reconstructed surfaces in Fig. 2 shows that artefacts tend to occur in regions characterised by a high rate of change, or at local extrema. To further illustrate these effects, Fig. 3 shows some examples of peaks and edges which have been reconstructed using various interpolation methods. Fig. 3a shows the effect that using different methods has on the smoothness of interpolated peaks. This particularly illustrates the effect that NN interpolation has around input datum, where it tends to produce steep peaks. Generally the areas around these peaks will be underestimates of the input values. Fig. 3b shows the possible overshoot effect that occurs when interpolating steep edges using cubic (and higher order polynomial) interpolation.

Knowing the artefacts that different interpolation methods can cause, as well as when, where and how they occur can be very useful in situations where analysis techniques are sensitive to specific effects. In these cases, the ability to choose the interpolation method which best suits the application is very important. In situations where it is important that no overshoots are introduced, and no extraneous spikes are added, the most stable technique to use is

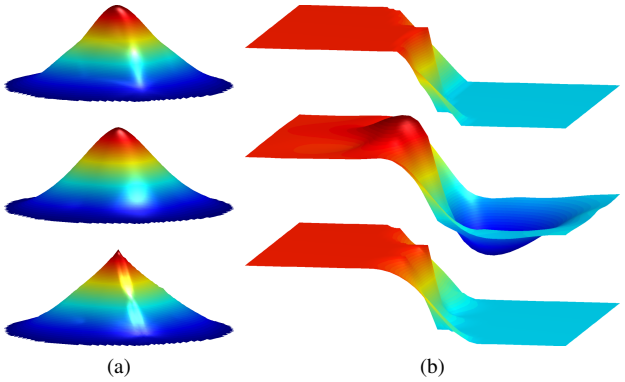


Fig. 3. Images demonstrating (a) peak artefacts and (b) overshoot problems occurring at edges. Images were interpolated using (bottom) NN interpolation, (middle) TPS RBF, and (top) cubic in (a) linear in (b).

probably linear interpolation. Use of linear RBF interpolation will have the added advantage that the output will not contain the faceted appearance that typifies triangulation-based linear interpolation.

4. INTERPOLATION ERROR DISTRIBUTIONS

One very effective way of examining interpolation methods for possible problems is to create a histogram of the errors between an interpolated output, and a simulated full-field input. The histogram should describe an approximate Gaussian distribution centred on zero. Information about any possible interpolation errors can be obtained by examining the distribution using the first few standardised moments. The first of these moments, the mean, defines the centre point of the distribution, and should be approximately zero. If this is not close to zero, then the interpolation method is biased. In this case, the result should be discarded, and the technique's implementation checked for errors.

The variance is useful for characterising the spread of error values, and can be used to derive confidence limits. Confidence limits are often more useful when calculated using the absolute error distribution. The third moment, the "skewness", describes the asymmetry of a distribution. A non-zero skewness is indicative of a tendency for the interpolation method to under- or over-estimate output values which do not lie on input datum. This is the most common problem

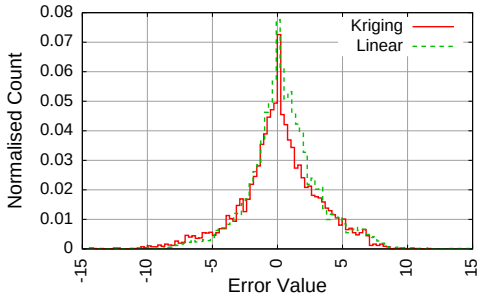
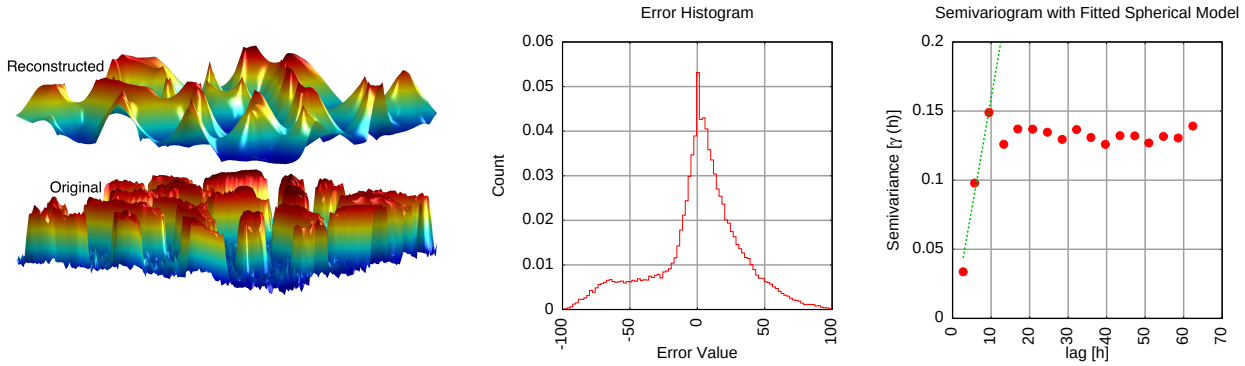


Fig. 4. Histograms showing error distributions from reconstructing Fig. 2a, using kriging and linear triangulation-based interpolation.



(a) (b) (c)

Fig. 5. Example reconstruction errors using kriging interpolation. (a) Reconstructed and original images, (b) normalised histogram showing the distribution of error values between the two images in (a) and (c) semivariogram showing incorrectly fitted spherical model.

seen in interpolation outputs, and typically occurs when convex or concave surfaces are fitted to approximately linearly varying data. Finally, the fourth moment, the "kurtosis" describes the distribution of outliers. A kurtosis value of greater than three indicates the proportion of outliers is higher than for a standard Gaussian distribution.

Fig. 4 shows two example normalised histograms of reconstructed data from the SRTM. Whilst they are quite similar in extent, they also show some differences, particularly on the right-hand side of the linear interpolation error distribution. This corresponds to a variance of 7.78, a skewness of -0.01, and a kurtosis of 5.22. The kriging distribution has a variance of 9.07, a skewness of -0.3, and a kurtosis of 4.21. This indicates that the kriging distribution has a lower spread than the linear interpolation distribution, outside of the body of the distribution, which is slightly wider in the case of kriging.

Fig. 5a shows an example reconstruction using ordinary kriging interpolation. Visual comparison with the original image reveals that there are significant errors in the reconstructed data. Examination of the error histogram (see Fig. 5b), shows that it is skewed to the left. This is confirmed by a skewness value of -0.41 which indicates that, in this case, the kriging method consistently under-estimated the true output values. Kriging works in several stages, the first of which is the estimation of an *experimental semivariogram*, which describes the spatial autocorrelation of the data to be interpolated. A model is then fitted to the semivariogram and used as a basis function for a global interpolation. Fig. 5c shows that the semivariogram model failed to correctly fit the data. This failure caused the interpolation basis function to be inappropriately chosen, leading to the erroneous output. This type of error distribution can occur whenever the semivariogram model does not fit the experimental semivariogram, or when the semivariogram sampling fails to correctly capture the spatial-autocorrelation of the data. A failure at the model fitting stage can also lead to a complete failure of the interpolation process. Problems such as this are by no means specific to kriging, although its complexity makes it more prone to the propagation of errors through its multiple stages.

5. CONCLUSIONS

This study has demonstrated some useful ways in which MIM can be characterised, by using SV where full correct fields are available and CV where only scattered input data are available. An examination of artefacts inherent in interpolation methods has also been presented, along with specific illustrative examples demonstrating how, and where these artefacts may appear. Finally, a discussion of interpolation error histograms, including how they can be analysed, and the useful information that the first four standardised moments can provide has been presented. These evaluation methodologies should prove useful to geoscientists and engineers who use multivariate interpolation methods in their work.

6. REFERENCES

[1] M. Chen, W. Shi, P. Xie, V. Silva, V. E. Kousky, R. W. Higgins, and J. E. Janowiak, "Assessing objective techniques for gauge-based analyses of global daily precipitation," *J. Geophys. Res*, vol. 113, 2008.

[2] M. P. Foster and A. N. Evans, "An evaluation of interpolation techniques for reconstructing ionospheric TEC maps," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 46, no. 7, pp. 2153–2164, 2008.

[3] K. Sugihara, A. Okabe, and B. Boots, "Spatial tessellations: Concepts and applications of voronoi diagrams," *Probability and Statistics*, 2000.

[4] R. Sibson, "A brief description of natural neighbour interpolation," *Interpreting Multivariate Data*, pp. 21–36, 1981.

[5] J. C. Carr, W. R. Fright, and R. K. Beatson, "Surface interpolation with radial basis functions for medical imaging," *IEEE Trans. Med. Imag.*, vol. 16, pp. 96–107, 1997.

[6] D. T. Sandwell, "Biharmonic spline interpolation of GEOS-3 and SEASAT altimeter data," *Geophysical Research Letters*, vol. 14, no. 2, pp. 139–142, 1987.

[7] N. A. C. Cressie, *Statistics for Spatial Data*, John Wiley & Sons, Inc., 1991.

[8] M. H. Trauth, *MATLAB Recipes for Earth Sciences*, Springer, 2006.

[9] T. Q. Pham and L. J. van Vliet, "Normalized averaging using adaptive applicability functions with applications in image reconstruction from sparsely and randomly sampled data," *Image Analysis, Proc.*, vol. 2749, pp. 485–492, 2003.