Computing Uncertainty of Natural Neighbour Interpolation

T. R. Etherington*[1]

[1] Manaaki Whenua – Landcare Research, PO Box 69040, Lincoln, New Zealand
*Email: etheringtont@landcareresearch.co.nz

Abstract

Interpolation techniques provide a method to convert point sampling data into a continuous estimate of a field phenomenon and have become a fundamental geocomputational technique of spatial and geographical analysts. Natural neighbour interpolation is one method of interpolation that has several useful properties: it is an exact interpolator, it creates a smooth surface free of any discontinuities, it is a local method, and is spatially adaptive. However, as with any interpolation method, there will be uncertainty in how well the interpolated field values reflect actual phenomenon values. Using a method based on distance error rates calculated for data points via leave-one-out cross-validation, reasonable estimates of interpolation error can be made, at least within the convex hull of the data points. While this method does not replace the need for analysts to use sound judgement in their interpolations, it does provide a valuable tool to aid in assessing the uncertainty associated with those interpolations.

Keywords: interpolation, uncertainty, natural neighbour, convex hull, error

1. Introduction

Natural neighbour interpolation is an interpolation technique that was first presented by Sibson (1981). Previous authors (Sambridge et al., 1995; Watson, 1999) have noted several useful properties of natural neighbour interpolation: (i) the method is an exact interpolator, in that the original data values are retained at the reference data points; (ii) the method creates a smooth surface free of any discontinuities; (iii) the method is entirely local, as it is based on a minimal subset of data locations that excludes locations that while close are more distant than another location in a similar direction; and (iv) the method is spatially adaptive, automatically adapting to local variation in data density or spatial arrangement. These properties make natural neighbour interpolation particularly well suited for the interpolation of continuous phenomena from reference data points that have a highly irregular spatial distribution. However, there is a great deal of uncertainty associated with any interpolation, and therefore being able to associate predictions from natural neighbour interpolation with some form of uncertainty would be highly desirable. I present an approach to estimate the likely error as a function of distance from natural neighbours and the known error rates at data points.

2. Computing Uncertainty

2.1. Method

Given a smooth continuous field that is sampled at a series of locations (Figure 1a), natural neighbour interpolation can then estimate that field (Figure 1b). The difference between the actual field and the estimated field becomes the known error (Figure 1c) that we are interested in estimating. The proposed method uses leave-one-out cross-validation to calculate the absolute error at each data point when the field value at that data point is estimated via natural neighbour interpolation. This absolute error is then converted to a distance error rate by dividing the error by the distance to the natural neighbours. Using localised error rates is highly advantageous as it allows for error estimates to reflect local changes in the field, with lower error rates in smoother areas and higher error rates in rougher areas. This contrasts with other spatial interpolation uncertainty methods such as kriging which estimate uncertainty of interpolation using a global variogram model fitted to all the data simultaneously. An estimate of the absolute error for the interpolation is then made by using natural neighbour interpolation for a second time to interpolate the distance error rate for each data point, and then using map algebra to multiply the interpolated distance error rate field by a distance to data points field (Figure 1d). The difference between the known and estimated absolute error represents how well the proposed method performs (Figure 1e).

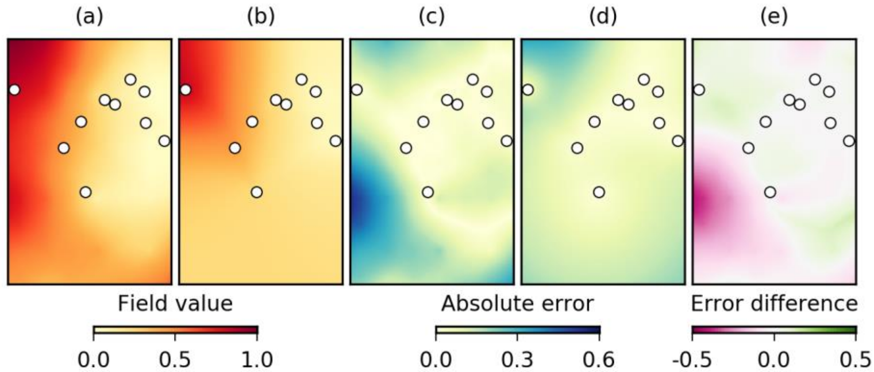

Figure 1: The natural neighbour interpolation uncertainty computational process. (a) A randomly generated field with random sampling points, and (b) the resulting natural neighbour interpolation from the sampling points. The (c) known absolute error between the random field and the interpolated field, and (d) the estimated absolute error. (e) The error difference between the known and estimated absolute errors, showing areas of error underprediction and overprediction.

2.2. Experiments

To examine the performance of the proposed method a series of 15 computational experiments were run using a Python computational framework (Pérez et al., 2011). Simulated random field phenomena for grids of 100 × 150 cells were created using the mid-point displacement method (Fournier et al., 1982) implemented in the NLMpy package (Etherington et al., 2015). The smoothness of the mid-point displacement method can be controlled by varying the *h* parameter which was set at two for a smooth field. Simulated data points were created by extracting the underlying value from ten points randomly located across the simulated spatial phenomena grids. The proposed computational method (Figure 1) was applied to each of the 15 experiments to produce 15 error difference fields (Figure 2).

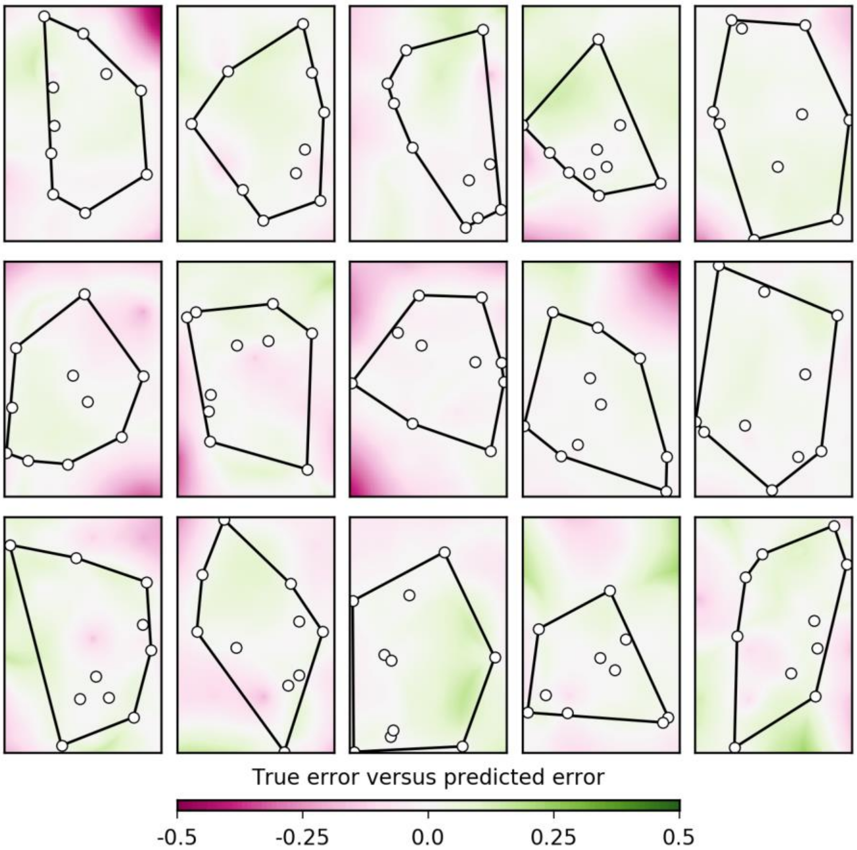

Figure 2: Error difference between the known and estimated absolute errors for 15 natural neighbour interpolation experimental replicates.

3. Results and Discussion

The results from the computational experiments (Figure 2) would indicate that at least for a smooth field sampled with 10 data points, a reasonable estimate of interpolation error can be made. Within the convex hull encapsulating the data points, the estimated error was very close to or slightly above the known error. Overestimates of error while not ideal still envelop the actual error and therefore are not considered as significant an issue as underestimates of error. Large differences between the estimated and known error, and underestimates of error, were largely limited to areas outside the convex hull. A reduction in performance of natural neighbour interpolation outside the convex hull is not unexpected as it has been noted before (Watson 1999), but this is also likely to be true of all spatial interpolation techniques as beyond the convex hull interpolation becomes extrapolation. However, we do not suggest that interpolation should be restricted to within the convex hull as there may be occasions where the area of interest may occur slightly outside the convex hull. For example, when interpolating rainfall data from weather stations that are usually sited in settlements, there are likely to be areas of coastline along peninsulas and headlands that will not fall within a convex hull around the weather stations. Therefore, it is logistically useful for an interpolation technique to be able to estimate values beyond the convex hull of the available data points. Therefore, while this technique for estimating interpolation error shows promise, the responsibility of appropriate use of natural neighbour interpolation still belongs with the spatial analyst who must make decisions about whether interpolation is useful based on their knowledge of: the smoothness of the phenomenon being interpolated, the number and distribution of data points, and the location of the areas for which interpolations are required.

4. Acknowledgements

We acknowledge funding from the Strategic Science Investment Funding for Crown Research Institutes from the New Zealand Ministry of Business, Innovation and Employment's Science and Innovation Group.

5. References

Etherington TR, Holland EP, O'Sullivan D (2015) NLMpy: a Python software package for the creation of neutral landscape models within a general numerical framework. Methods in Ecology and Evolution 6:164-168

Fournier A, Fussell D, Carpenter L (1982) Computer rendering of stochastic models. Communications of the ACM 25:371-384

Pérez F, Granger BE, Hunter JD (2011) Python: an ecosystem for scientific computing. Computing in Science & Engineering 13:13-21

Sambridge M, Braun J, McQueen H (1995) Geophysical parametrization and interpolation of irregular data using natural neighbours. Geophysical Journal International 122:837-857

Sibson R (1981) A brief description of natural neighbour interpolation. In: Barnett V (ed) Interpreting Multivariate Data. Wiley, Chichester, pp 21-36

Watson D (1999) The natural neighbor series manuals and source codes. Computers & Geoscience 25:463-466