

THE IMPORTANCE OF UNDERSTANDING ERROR IN LIDAR DIGITAL ELEVATION MODELS

Smith, S.L.^a, Holland, D.A.^a, Longley, P.A.^b

^a Research & Innovation, Ordnance Survey, Romsey Road, Southampton, SO16 4GU
e-mail sarah.smith@ordnancesurvey.co.uk; david.holland@ordnancesurvey.co.uk

^b Centre for Advance Spatial Analysis, University College London, 1-19 Torrington Place, Gower Street London, WC1E 6BT e-mail plongley@geog.ucl.ac.uk

KEY WORDS: DEM/DTM, Laser Scanning, Error, Urban, Accuracy

ABSTRACT:

The paper presents the results of an accuracy analysis of a number of gridded Digital Surface Models (DSMs) which have been created from first return laser scanning (LiDAR) data. DSMs are created using a variety of methods, and differences between the models and the raw LiDAR data are quantified and the spatial patterns of the errors explored. The results presented in this paper demonstrate the importance of understanding not only the global accuracy of elevation models, but also the location and magnitude of individual errors which are created during DSM creation. Results to date indicate that errors of up to 1m can be created by using different techniques, and that these errors will affect the results of any subsequent analysis. This investigation highlights the importance for both data suppliers and data users to better understand errors which are created within digital elevation models and incorporated within any subsequent analysis.

1. INTRODUCTION

1.1 Applications of LiDAR DEMs

LiDAR derived elevation models are used in a multitude of applications, including town planning, urban micro-climate analysis, and telecommunications planning. Despite the importance of the analysis which is based on these models, the accuracy of the models is often overlooked, and there is currently a paucity of information regarding the characteristics of the errors created using different modelling methodologies. A large proportion of the total error is introduced during the processing of the raw data, and the subsequent modelling with and analysis from it. Each stage of the modelling process potentially introduces error into the DEMs (Digital Elevation Models). This error is in addition to the error incorporated during the capture process (Latypov, 2002). Modelling stages include data formatting, filtering, which is the removal of above ground features to create bare earth Digital Terrain Models (DTMs), segmentation, which is the differentiation of above ground objects, and object reconstruction (Haithcoat et al (2001)). This paper investigates the effect of the first stage of modelling: data formatting.

Many software packages require that the irregularly spaced raw points are interpolated onto a regular grid for analysis and visualisation of a height model. Despite the fact that laser scanner points are sampled at very small separation distances, the interpolation from points onto a grid can introduce a degree of uncertainty into the model. The level of this uncertainty can vary greatly with different interpolation methods and grid sizes. Therefore, the choice of gridding methodology is potentially a very significant decision in determining the accuracy of the final LiDAR model. There are two principal processes which affect the success of the gridding process: the choice of interpolation algorithm, and the choice of grid size for resampling. In this investigation, four deterministic interpolation techniques (bilinear, bicubic, nearest neighbour, and biharmonic splining) are created at 3 different grid resolutions (1m, 2m, and 4m). Each of the surfaces created are compared to the values of the raw

LiDAR data in an attempt to quantify the errors introduced by different gridding methods.

1.2 Previous Research

Modelling from interpolated, or gridded, data is generally considered to be less accurate than modelling from the raw data, and most recent academic research has shown a preference for modelling from the raw data (Rottensteiner and Briese, 2002; Vosselman, 2000; Maas and Vosselman, 1999; Gruen and Wang, 1998) for this reason. Yet, the gridded approach remains the one favoured in many commercial and business environments, mainly as a result of the speed of processing and accessibility of software which requires regularly spaced data. Quantifying the amount of error incorporated during the gridding process is therefore important if users of gridded data are to assess the accuracy of the DEMs they create.

1.2.1 Interpolation Methods

A number of previous studies have looked at the effect of interpolation as it is a routine process in many disciplines including digital image processing, cartography, soil science, electronic imaging, and mathematics. The effects and quality of different interpolation algorithms have been heavily investigated in the past in these disciplines (Desmet, 1997) including studies of how interpolation algorithms perform at different resolutions (Turner et al, 1996). However these studies have not investigated the interpolation onto a regular grid, nor have they examined how the interpolation effects change for different grid resolutions and very few have looked at this problem where the grid heights represent the urban surface. The conclusions from these studies may not be directly transferable to models of urban surface form. Much real world urban spatial variation is not smooth and continuous, but is jagged and irregular (Longley et al, 2001), this is certainly the case for models of elevation in urban surfaces. According to fractal geometry theory, jagged irregularity is a property which can be observed across a

range of scales for many types of surfaces. Studies of fractals in elevation surfaces have shown how they can be used to model bare earth terrain (Batty and Longley, 1994) at different resolutions on the basis of their self-repeating properties. However, this same argument is not directly transferable to the modelling of large scale urban surface form (say at the level of modelling buildings) due to the inherent complexity and lack of self-repeating structures at this level in urban areas. If urban surfaces cannot be modelled with fractals at these large scales, then it follows that we do not understand how the characteristics of these surfaces may alter at higher resolutions. In effect, this means that we do not understand how the pattern of interpolation errors may change for different grid resolutions.

However, several **global** characteristics of the various surfaces created by different interpolation methods have been noted previously in the literature. Zinger et al (2002) commented that linear interpolation will tend to overly smooth and deform building edges. However, such general characteristics reveal little about the exact spatial pattern of error within a surface model. Lloyd and Atkinson (2002) further investigated the quantification of error within interpolated surfaces. The authors focused on a comparison of Inverse Distance Weighting (IDW) and kriging interpolation, and quantified the inaccuracy in each surface. Such measures are useful general indicators of error within surface models, again however they do not reveal anything about the spatial pattern of errors across the surface. One of the most relevant comparisons of interpolation algorithms was that of Rees (2000) who investigated the interpolation of gridded DEMs to higher resolutions - whilst this study did not look at the interpolation of irregularly spaced data onto a regular grid (the subject of this paper) many of Rees' (2000) conclusions are nevertheless relevant. Rees (2000) concluded that simple bilinear and bicubic interpolations are adequate for most elevation model requirements in non-urban areas. Rees (2000) conclusions are tested within this study for urban areas.

1.2.2 Grid Spacing

Whilst the effect of different interpolation methods on the form of the surface has been investigated in the past (eg. Zinger et al, 2002; Morgan and Habib, 2002; Lloyd and Atkinson, 2002; Smith et al, 2003a) there has been little research into the effect of changing grid size in the interpolation stage save for that of Behan (2000). Behan (2000) quantified error within models produced from different interpolation algorithms. It was found that the most accurate surfaces were created using grids which had a similar spacing to the original points. Behan's (2000) study looked at global or average error differences between two interpolation methods.

1.3 Aim of the Investigation

The investigation presented here quantifies the amount of model error introduced during the interpolation process, and specifically examines the pattern of errors created when modelling at different spatial resolutions. From previous literature it is known that the interpolation method and scale will influence the derived urban DSM, however we do not know to what extent.

2. METHODOLOGY

2.1 Creating the Surfaces

2.1.1 The Data and the Study Area

DSMs were created from a subset of a first return laser scanning dataset, supplied by the Environment Agency. The data were captured from an airborne sensor, at a point density of ~2m. The area used for modelling is shown in Figure 1, which shows that this sample region comprises a complex roof structure (church), some bare earth, a flat roof and a variety of vegetation. Despite being a small area (1315 points over a 80m by 50m region), the surface was considered to be representative of the typical types of structure found in the wider region. In addition, the investigation has been conducted over 2 more study areas to ensure the reliability of the results.



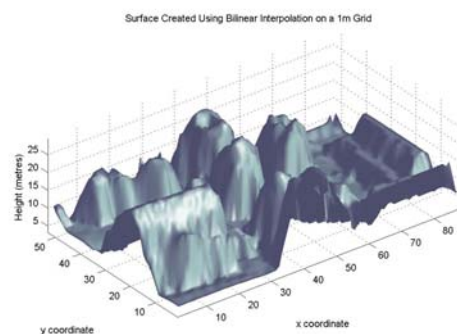
Figure 1: Orthorectified photograph of the corresponding area. Aerial photography reproduced with permission of Ordnance Survey © Cc Ordnance Survey. All rights reserved.

2.1.2 The Choice of Interpolation Methods

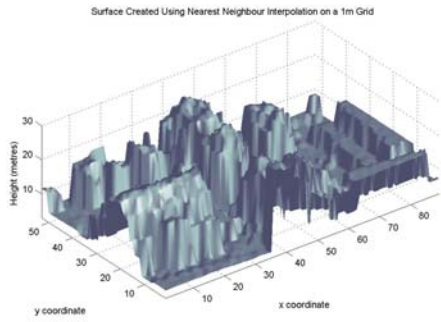
There are many routines for spatial interpolation available and these have been widely documented in the past (Watson, 1992). However, not all of the methods are suitable for elevation modelling from LiDAR data. In particular, for the urban surface environment where there are frequent discontinuities, local interpolation rather than global or fitted function methods are preferable in order that the local complexity in the surface be retained as much as possible. For this reason, this study compared only local deterministic interpolation techniques.

2.1.3 Performing the Interpolation

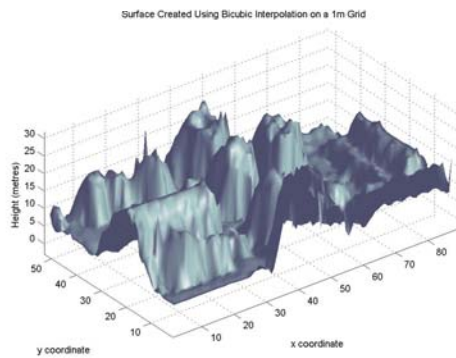
The raw points were first resampled onto a regular 1m, 2m and 4m grid, using four interpolation methods: bilinear, bicubic, biharmonic splining, and nearest neighbour, and the resultant surface forms produced are shown below in Figure 2.



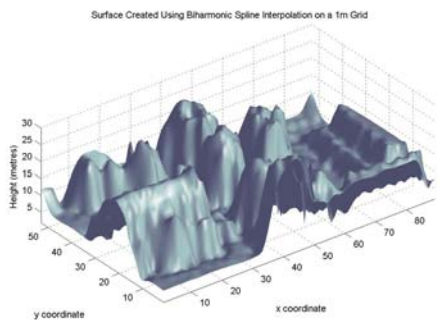
(a)



(b)



(c)



(d)

Figure 2: Showing the (a) bilinear, (b) nearest neighbour, (c) bicubic, and (d) biharmonic spline surfaces created using the interpolation methods at 1m grid spacing. For information regarding the specifics of the four algorithms see Sandwell (1987), Watson (1992), and Smith et al (2003b).

The differences in surface form between the DSMs shown in Figure 2 can clearly be seen. The nearest neighbour surface is blocky and ‘stepped’ in appearance, whilst the spline surface is much smoother, with many of the building and vegetation edges appearing curved. The quantified differences between the surface height predictions and the raw data are presented below.

2.2 Comparing the Surfaces

For the purposes of this investigation, the error (\mathcal{E}) at each investigated point within the surfaces was considered to be

the difference between the raw data point, $Z(x)$, and the interpolated value, $Z_i(x)$, for that location (see eq.1 below).

$$\mathcal{E}(x) = Z(x) - Z_i(x) \quad (1)$$

where $\mathcal{E}(x)$ = measured error at location (x), Z = height value, Z_i = interpolated height value.

This calculation was repeated across the surface to assess the success of the four interpolation algorithms at the investigated grid spacings. Model suitability was assessed in relation to how much error was introduced to the surface by each of the techniques - the most suitable model being the one which introduced the least error (\mathcal{E}).

In order to assess model suitability, some of the raw data points had to be omitted from the surfacing process. If all of the data points were used to interpolate a surface, then the goodness-of-fit of the surface could not be assessed with these same data points as this would yield an overly optimistic (low) prediction error. For this reason a standard validation procedure was employed which involved omitting some raw data from the interpolation procedure and assessing the success of the procedure to model in the absence of these known values. The method chosen is called the split-sample validation routine which is advocated by Declercq (1996). In this procedure part of the sample values are omitted, the interpolation is performed, and the difference between the predicted and the raw data values at these locations are calculated. This difference is then used as a measure of the success of the algorithm. The usefulness of this technique is increased when it is used iteratively, where the number of omitted points is progressively increased and the differences calculated. This can return useful information regarding the stability of the algorithm, and its ability to cope with differences in input point density. The split-sample methodology was used in this way in Lloyd and Atkinson (2002). The authors used sample sizes of 95%, 50% and 25% of original points to measure the effects of interpolation for rural DEMs. The methodology outlined in Lloyd and Atkinson (2002) is adhered to in this study, and the same proportions of omitted data are used here for the comparison of interpolation methods. For the grid spacing investigation a random selection of 5% of the raw points were omitted and the surfaces produced at different resolutions. The success of each resolution was then assessed by calculating the difference between the omitted data and the surface predictions at these locations. The amount of omitted data points in each sample was not varied for this part of the investigation.

In each investigation the omitted data points were chosen in a random selection process, and the tests run multiple times to ensure different points were selected each time and similar results obtained. This ensured the reliability of the investigation.

3. RESULTS

The model errors were calculated in accordance with Equation (1), and are recorded in Table 1 below.

Table 1: The Results from Split-Sample Routine Designed to Assess the Suitability of Each Interpolation Technique

INTERPOLATION METHOD	Error (ε) Descriptive Statistics				
	Max	Min	Mean	Std Dev	Rmse
95% Split sample					
Bilinear	5.3886	-6.8551	-0.1444	1.9168	1.8999
Bicubic	5.1996	-3.0855	-0.0085	1.6887	1.6499
Nearest Neighbour	13.34	-11.63	0.1398	3.5087	3.4848
Biharmonic Spline	7.1825	-12.7806	-0.3186	2.6709	2.6697
%50 Split sample					
Bilinear	10.1808	-14.8968	0.1296	2.9364	2.9367
Bicubic	12.3605	-15.709	0.1721	3.0472	3.049
Nearest Neighbour	15.37	-15.4	0.2402	3.4617	3.4674
Biharmonic Spline	15.1701	-15.7385	0.216	3.0555	3.0608
25% Split Sample					
Bilinear	13.3931	-15.1774	0.1575	2.9994	3.0018
Bicubic	11.3781	-16.7194	0.1026	2.9863	2.9861
Nearest Neighbour	15.99	-16.14	0.1597	3.6569	3.6585
Biharmonic Spline	11.7705	-18.2399	0.1919	3.0508	3.0553

Table 1 above shows the results from the interpolation method comparison on a 1m grid. It can clearly be seen that there is a difference in the statistics of the calculated errors. A discussion of these results follows in section 3.1.

Table 2: The Results from the Split-Sample Routine Designed to Quantify the Differences in Errors Introduced by the Methods at a Variety of Grid Resolutions.

INTERPOLATION METHOD	Grid Resolution (m)	Error (ε) Descriptive Statistics				
		Max error (m)	Min error(m)	Mean (m)	Std Dev (m)	RMSE (m)
Bilinear	1m	5.56	-7.14	-0.02	1.98	1.96
	2m	5.24	-7.26	-0.16	1.93	1.92
	4m	4.29	-7.44	-0.73	2.12	2.22
Bicubic	1m	6.13	-6.30	0.02	2.64	2.01
	2m	5.83	-6.57	-0.15	1.96	1.94
	4m	2.94	-7.50	-0.49	2.02	2.05
Nearest Neighbour	1m	8.40	-7.14	0.29	2.58	2.58
	2m	8.40	-6.94	0.36	2.41	2.42
	4m	8.33	-14.72	-0.06	2.67	2.65
Biharmonic Spline	1m	7.30	-6.27	0.13	2.18	2.17
	2m	9.28	-6.15	0.16	2.40	2.39
	4m	13.09	-22.73	-0.28	3.93	3.91

3.1 Discussion

The results presented in Table 1 show that bilinear and bicubic algorithms were found to produce the lowest RMSE of all the interpolators. It was surprising that the biharmonic spline method did not produce lower errors. This was thought to be caused by the tendency for this algorithm to produce strong artificial oscillations in the surface reconstruction in unconstrained regions as indicated by the large maximum and minimum errors. The results also showed that, in general, the nearest neighbour interpolator produced the highest RMSE value. Despite preserving discontinuities across the surface, the nearest neighbour algorithm was found to introduce a large amounts of error. This was probably due to its inability to model oblique surfaces - as there are no slopes the changes between groups of values are very steep and create discontinuities which are not necessarily present in the raw data.

In terms of differences in errors created at different resolutions, the bilinear, bicubic, and biharmonic splining interpolators produced relatively stable range and mean errors, and there appeared to be only a minimal difference between the surfaces produced at different resolutions. It was, however, noted that the resolution which produced the

lowest error was that which was as close as possible to the original point spacing, which in this instance was ~2m. The nearest neighbour interpolator produced higher errors at larger grid spacings than any of the other methods. In all cases interpolation onto the 4m grid resulted in higher errors, due to the loss of information in this approach. The increase in error was in the order of 50-80cm. In such cases it remains the decision of the end-user as to whether the decreases in accuracy are outweighed by faster computation and smaller file sizes.

3.2 The Spatial Pattern of Error

The pattern of individual errors was examined by plotting the locations of the interpolated points and assigning them a size in proportion to the error calculated for that point (Figure 3). It was observed that there was a strong spatial dependence of the highest magnitude errors in the biharmonic splined surface, and that many of the highest magnitude errors occurred at the edges of the dataset. It was considered that these edge errors were skewing the statistical analysis. As such it is suggested that the biharmonic spline should be used with an edge buffer to eliminate some of the largest errors (see Smith et al, 2003b). It was also noted that there was a general patterns of larger errors which coincided with the occurrence of breaklines in the dataset (in this instance breaklines included building and vegetation edges). However, there was a clear difference in the amounts of errors over different breaklines, with the errors caused by building breaklines tending to be smaller than those caused by vegetation.

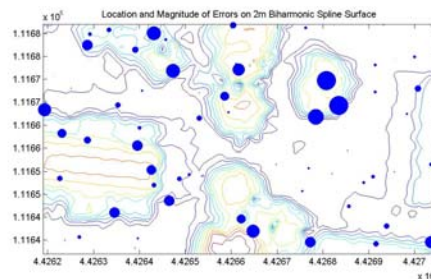


Figure 3 Errors created in the interpolation of DSM using Biharmonic Spline method on a 2m grid. Size of the disks indicates amount of error at that location. Contours underlain for context. Note the occurrence of some large errors at the edges of the dataset. This edge effect for the biharmonic splining algorithm is explored further in Smith et al (2002b).

There was also a difference noted in the pattern of errors produced by different methods at different resolutions. Figure 4 shows the differences in the distribution of errors for both the bilinear and the nearest neighbour methods at the 3 different scales. It can be seen that the errors produced by the nearest neighbour method are much more dense and of higher magnitude over vegetated areas than the bilinear errors. The increase in errors over breaklines at higher grid resolutions can also be observed. Refer to Figure 1 for context.

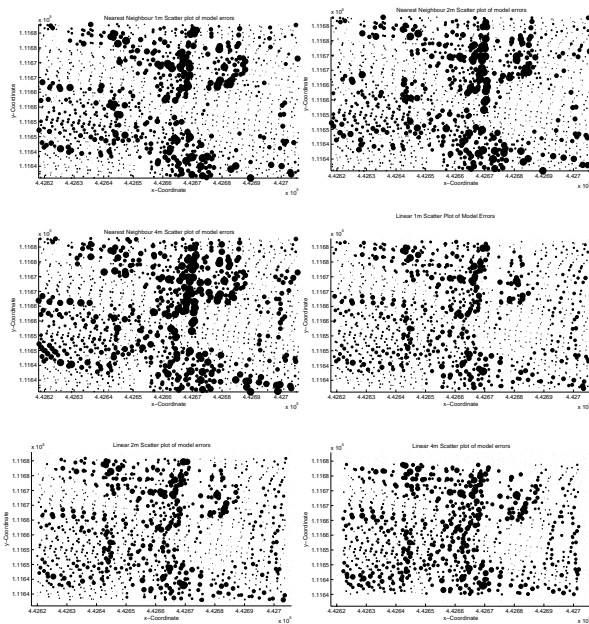


Figure 4 The pattern of errors in the bilinear and nearest neighbour surfaces at 1m, 2m and 4m grid resolutions.

4. SUMMARY OF TECHNICAL CONCLUSIONS

This investigation has shown that there is significant variation between the forms of DSMs created using different interpolation algorithms and different grid sizes. It was found that the most error was introduced by the nearest neighbour algorithm, and the least error was introduced by the bilinear and bicubic methods, with some evidence that the biharmonic spline may produce lower errors where used with a buffer zone to reduce the strongest oscillations in the unconstrained regions. Despite being shown to produce the least overall error, bilinear and bicubic interpolation were observed to over smooth building edges and may therefore be unsuitable for some applications. Ultimately the choice of optimal algorithm for a particular application must be decided by the user – understanding spatial variations in accuracy can therefore promote better informed decision making.

This investigation has also shown that changes in grid sizes have very different effects on the magnitude of error introduced by different interpolation algorithms. Where accuracy is the most important factor, optimal grid spacing for any interpolation method should be as close as possible to (or slightly less than) the original point spacing. This supports Behan's (2000) conclusions. The pattern of highest magnitude error appeared to occur in the areas of greater surface roughness. There is some potential for this pattern to be used in subsequent image segmentation as an indicator of surface roughness. This is currently being investigated by the authors.

5. SIGNIFICANCE OF THE FINDINGS AND REAL WORLD APPLICATIONS

One common application of LiDAR DEMs is in flood modelling and flood inundation prediction. Small differences in values in the DEM used as an input to the flood modelling program can have a large effect on the predictions. The effect of these differences are shown in figure 5, in which the four surfaces created are 'flooded' at 4m and then at 5m. Figure 5

shows the marked difference between the flood predictions, particularly the limited flooding in the nearest neighbour surface. This is caused by the 'stepped', or blocky, nature of this surface which was noted in section 2.1. In the nearest neighbour flood prediction it is clear that the level of the 'step' must be exceeded in order for the flood to propagate. In this instance the 'step' characteristic of this surface clearly acts as a flood break, altering the results, and producing a potentially erroneous prediction. In contrast, the biharmonic splined surface, which was noted to produce a very smooth surface, permitted the largest flood prediction. Despite these noted differences it is, of course, impossible to comment on which of the predictions is closest to reality. Clearly, however, if these models were to be used in the calculation and mapping of flood risk areas, the use of different gridding techniques could substantially alter the results. The same argument holds for the use of DEMs in similarly sensitive applications, such as mobile phone wave propagation modelling, and noise pollution modelling. Thus differences in the height values of DEMs can have significant implications where the models are used in a predictive or analytical capacity.

6. SUMMARY

The introduction of errors in the gridding of data remains only one source of error in the process of modelling from LiDAR data. It has been the purpose of this paper to demonstrate how these gridding errors may be introduced, and how the magnitude and the spatial structure of these can change with different methodologies. Whilst differences have been highlighted between different interpolation methods, it should be noted that there is no optimal creation methodology, as the final decision regarding interpolation algorithm, grid spacing, filtering method, and segmentation procedure must be driven by the requirements of the application for which the DEM is intended. However, what has been shown is that in order that informed decisions can be made regarding the specific modelling processes, users must be provided with error information which may come in the form of a map of the spatial structure of error – such as those presented in this paper. The provision of this error information further requires that there are software processes which are able to cope with the communication of the error, and with the incorporation of this information in subsequent analysis. This in turn requires that the data users understand the error information, and can use this intelligently in order to reduce the introduction of error in their LiDAR processing.

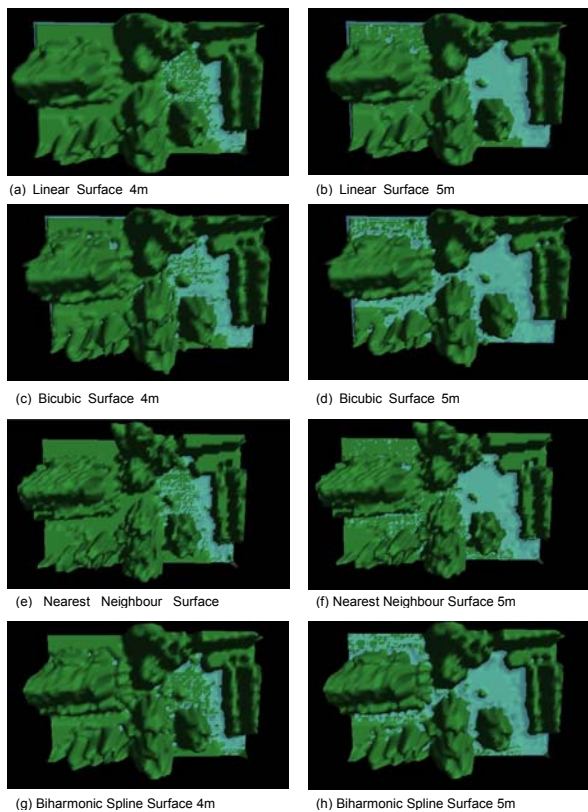


Figure 5 Showing overhead view of study area flooded at different levels (4m on left, and 5m on right) for the bilinear and the nearest neighbour surfaces.

References:

- Batty, M. and Longley, P.A (1994) *Fractal Cities*. London: Academic Press Limited.
- Behan, A (2000) On the Matching Accuracy of Rasterised Scanning Laser Altimeter Data. *IAPRS, Vol. XXXIII, Amsterdam*
- Declercq, F.A.N (1996) Interpolation methods for scattered sample data: Accuracy, spatial patterns, processing time. *Cartography and Geographical Information Systems* 23(3), 128-144
- Desmet, P.J.J (1997) Effects of interpolation errors on the analysis of dems. *Earth Surface Processes and Landforms* 22, 128-144
- Gruen, A. and Wang, X. (1998) CC-Modeler: a topology generator for 3-D city models. *ISPRS Journal of Photogrammetry and Remote Sensing* 53 286-295
- Haithcoat, T.L, Song, W. and Hipple, J.D (2001) Building Footprint Extraction and 3-D Reconstruction from LiDAR Data. *IEEE/ISPRS Joint Workshop on Remote Sensing and Data Fusion over Urban Areas*.
- Latypov, D. (2002) Estimating relative lidar accuracy information from overlapping flight lines. *ISPRS Journal of Photogrammetry and Remote Sensing* 1211 xxx-xxx
- Lloyd, C.D and Atkinson, P.M (2002) Deriving DSMs from LiDAR data with kriging. *International*

Journal of Remote Sensing (23), 12. Pg. 2519 – 2524

Longley, P.A, Goodchild, M.F., Maguire D.J, and Rhind, D.W (2001) *Geographic Information Systems and Science*. Chichester. John Wiley and Sons.

Maas, H-G and Vosselman, G (1999) Two Algorithms for Extracting Building Models from Raw Laser Altimetry Data. *ISPRS Journal of Photogrammetry and Remote Sensing* 54 (1999) 153-163

Morgan, M and Habib, A (2002) Interpolation of LiDAR Data and Automative Building Extraction. *ACSM-ASPRS 2002 Annual Conference Proceedings*

Rees, W.G. (2000) The accuracy of Digital Elevation Models interpolated to higher resolutions. *International Journal of Remote Sensing*, vol. 21, no 1, 7-20

Rottensteiner, F and Briese (2002) A New Method for Building Extraction in Urban Areas from High-Resolution LiDAR Data. *ISPRS Commission III, Vol.34,Part 3A, Photogrammetric Computer Vision*, Graz, 2002

Turner, D.P., Dodson, R. and Marks D. (1996) Comparison of alternative spatial resolutions in the application of a spatially distributed biogeochemical model over complex terrain. *Ecological Modelling* 90, 53-67

Sandwell, David T., "Biharmonic Spline Interpolation of GEOS-3 and SEASAT Altimeter Data", *Geophysical Research Letters*, 2, 139-142,1987

Smith, S.L, Holland, D.A, and Longley, P.A (2003a) Interpreting Interpolation: The Pattern of Interpolation Errors in Digital Surface Models Derived from Laser Scanning Data. *Proceedings of GISRUK 2003*.

Smith, S.L, Holland, D.A, and Longley, P.A (2003b) The Effect of Changing Grid Size in the Creation of Laser Scanner Digital Surface Models, *Proceedings of Geocomputation 2003*

Vosselman, G (2000) Slope Based Filtering of Laser Altimetry Data. *IAPRS, Vol.XXXIII, Amsterdam*

Watson, David E., (1992) *Contouring: A Guide to the Analysis and Display of Spatial Data*, Tarrytown, NY: Pergamon (Elsevier Science, Inc.):

Zinger, S, Nikolova, M., Roux, M, and Maître, H (2002). 3D Resampling for Airborne Laser Data of Urban Areas. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XXXIV, 3B, Commission III

Acknowledgements

The authors would like to thank the Environment Agency for supplying the raw LiDAR data, and the Ordnance Survey for sponsoring the research and supplying orthorectified photography of the study region. An earlier version of this paper was presented at the ISPRS WG III/3 meeting on 3D Reconstruction from airborne laser scanning and InSAR data.