Variable Resolution Spatial Interpolation Using the Simple Recursive Point Voronoi Diagram

B. Boots¹ and R.D. Feick²

¹Department of Geography and Environmental Studies, Wilfrid Laurier University,
Waterloo, Ontario N2L 3C5, Canada.
Telephone: 1 519 884 0710 (Ext. 2604)
FAX: 1 519 725 1342
Email: bboots@wlu.ca

²Department of Geography, School of Planning, Faculty of Environmental Studies,
University of Waterloo,
Waterloo, Ontario N2L 3G1 Canada
Telephone: 1 519 888 4567 (Ext. 5493)
FAX: 1 519 725 2827
Email: rdfeick@fes.uwaterloo.ca

Abstract

This paper introduces a procedure for progressively increasing the density of an initial point set that can be used as a basis for interpolating surfaces of variable resolution from sparse samples of data sites. The procedure uses the Simple Recursive Point Voronoi Diagram in which Voronoi concepts are used to tessellate space with respect to a given set of generator points. The construction is repeated every time with a new generator set, which comprises members selected from the previous generator set plus features of the current tessellation. We show how this procedure can be implemented in Arc/Info and present an illustration of its application using a known surface. Initial results suggest that the procedure has considerable potential and we discuss further methods for evaluating and extending it.

1. Introduction

Spatial data consists of measurements (*data values*) of an attribute taken at specific locations (*data sites*) in a geographic space (*study region*). In much data collected in the environmental sciences, the attribute is assumed to be spatially continuous (possibly piecewise) so that the data values can be considered as a sampling of the attribute at the data sites. Suppose that our primary aim is to use the data to interpolate the values of the attribute at locations other than the data sites thus enabling us to create a visual representation of the underlying surface (Watson, 1992). Here we focus on such activity in situations where the data sites are sparse and there is no possibility of supplementing their number by increasing the size of the sample. Typically, this applies to most historical data. Further, notwithstanding advances in remote sensing and GPS data collection methods, in some instances it can still be impractical or too costly to collect appropriately detailed data. In addition, with historical data, we may also lack knowledge

of the spatial nature of the sampling design (e.g., random, systematic, etc.).

In this paper we propose a procedure for progressively increasing the density of an initial point set that provides a basis for interpolating surfaces of variable resolution from sparse samples of data sites. The procedure makes use of the Simple Recursive Point Voronoi Diagram (SRVPD).

Our proposal is motivated by both practical and conceptual considerations. Practically, the procedure can be fully automated and implemented within existing COTS software. Conceptually, it is supported by recognition of the dual relationships between Voronoi and Delaunay constructions. The latter have a long history of successful use in terrain modelling and visualization with the Delaunay triangulation being the most popular method of constructing a triangulated irregular network (TIN) (Hutchinson and Gallant, 1999). In part, this is because it is the only triangulation that satisfies both local and global max-min angle criteria (Sibson, 1978; Okabe et al., 2000, p.93). Wang et al. (2001) also demonstrate that the basic Delaunay triangulation even outperforms data dependent triangulations in modelling terrain surfaces. Further, several recursive forms of Delaunay triangulations have already been used in top-down, multi-resolution TIN models (see De Floriani et al., 1996 for a review). These include strictly hierarchical TINs (HTINs), which involve the recursive subdivision of the initial triangle(s) into a set of nested triangles (De Floriani et al., 1992; De Floriani and Puppo, 1995), and pyramidal TINs (PTINs) in which a new structure is computed every time new points are inserted (De Floriani, 1989; De Floriani et al., 1985; Voightmann et al., 1994). In essence, the SRVPD is a dual form of PTIN.

We begin our presentation by describing recursive Voronoi diagrams in general terms and providing a specific definition of the SRVPD. We then show how these constructions can be used to create surfaces of varying spatial resolution. In Section 3 we describe how these procedures can be implemented in Arc/Info. This is followed in Section 4 by an application of the procedure to a sample set of points from a known topographic surface. We conclude by discussing outstanding issues and directions for future work.

2. Recursive Voronoi Diagrams

The basic Voronoi concept involves tessellating an m -dimensional space with respect to a finite set of objects by assigning all locations in the space to the closest member of the object set. This concept can be also be applied recursively by tessellating the space with respect to a given set of generators and then repeating the construction every time with a new generator set consisting of objects selected from the previous generator set plus features of the current tessellation. More formally, consider a finite set of n distinct generators, G . First, construct the Voronoi diagram $\mathcal{V}(G)$ of G . Next, extract a set of features Q from $\mathcal{V}(G)$ and create a new set of generators G' which comprises of Q plus selected members of G . Then construct the Voronoi diagram $\mathcal{V}(G')$ of G' . This step is then repeated a number of times. At each step, the number of generators that are retained may range from none to all. Boots and Shiode (2003) show that such recursive

constructions provide an integrative conceptual framework for a number of disparate procedures in spatial analysis and modelling.

2.1 Simple Recursive Voronoi Point Diagram (SRVPD)

In this paper we consider a specific form of recursive Voronoi diagram, the simple recursive point Voronoi diagram (SRPVD), in which the initial set of generators consists of n distinct points $G_{(0)} = \{g_{(0)1}, g_{(0)2}, \dots, g_{(0)n}\}$. The construction of this diagram involves the following steps:

0. Define the initial set of generators $G_{(0)}$;
1. Generate the ordinary Voronoi diagram $V(G_{(0)})$ of $G_{(0)}$;
2. Extract all $m(0)$ Voronoi vertices $Q_{(0)} = \{q_{(0)1}, q_{(0)2}, \dots, q_{(0)m(0)}\}$ of $V(G_{(0)})$;
3. Create a new set of generator points $G_{(1)} = G_{(0)} + Q_{(0)}$;
4. Repeat steps 1 through 3.

We call the result of the k^{th} construction, $V(G_{(k)})$, the k^{th} generation of the simple recursive point Voronoi diagram. Similarly, we call its generator set $G_{(k)}$ and its vertices $Q_{(k)}$, the k^{th} generation of the point set and the k^{th} generation of the vertices, respectively. The results of applying these steps for five recursions to an initial generator set consisting of five points are shown in Figure 1. Note that the Voronoi polygons at generation k are either entirely contained within a Voronoi polygon at generation $(k-1)$ (i.e., $V(G_{(k) i}) \subset V(G_{(k-1) i})$) or are composed of pieces of three polygons from generation $(k-1)$, i.e., $V(G_{(k) i}) = (V(G_{(k) i}) \cap V(G_{(k-1) l})) \cup (V(G_{(k) i}) \cap V(G_{(k-1) m})) \cup (V(G_{(k) i}) \cap V(G_{(k-1) n}))$ (see Figure 2).

Figure 1. The first five recursions of a Simple Recursive Point Voronoi Diagram

As noted above, a recursive Voronoi structure will have a dual recursive Delaunay structure. The recursive Delaunay construction that is equivalent to the SRVPD is defined as follows.

0. Define the initial set of generators $G_{(0)}$;
1. Generate the Delaunay triangulation $D_{(0)}$ of $G_{(0)}$;
2. Identify the circumcentres $C_{(0)} = \{c_{(0)1}, \dots, c_{(0)m}\}$ of all the triangles of $D_{(0)}$;
3. Create a new set of generator points $G_{(1)} = G_{(0)} + C_{(0)}$;
4. Repeat steps 1 and 2.

2.2 Using the SRVPD to Construct Variable Resolution Surfaces

Using the SRVPD, the following procedure can be used to produce topographic surfaces that have increasing spatial resolution.

Let $G(0)$ be an initial set of generators each with an associated data value and label the set of data values $D(0)$.

Generate $V(G(0))$.

Extract the set $Q(0)$ consisting of the $m(0)$ vertices of $V(G(0))$.

Extrapolate data values for $Q(0)$ using Sibson's natural neighbour interpolation (Sibson, 1981) and $D(0)$.

Label extrapolated values $D(1)$.

If desired, create surface representation from $D(0) + D(1)$.

Let $G(1) = G(0) + Q(0)$.

Generate $V(G(1))$.

Extract the set $Q(1)$ consisting of the $m(1)$ vertices of $V(G(1))$.

Extrapolate data values for $Q(1)$ using Sibson's natural neighbour interpolation and $D(0) + D(1)$.

Label extrapolated values $D(2)$.

If desired, create surface representation from $D_{(0)} + D_{(1)} + D_{(2)}$.

·
·
·

Let $G(k) = G(k-1) + Q(k-1)$.

Generate $V(G(k))$.

Extract the set $Q(k)$ consisting of the $m(k)$ vertices of $V(G(k))$.

Extrapolate data values for $Q(k)$ using Sibson's natural neighbour interpolation and $D(0) + D(1) + \dots + D(k)$.

Label extrapolated values $D(k+1)$.

If desired, create surface representation from $D_{(0)} + D_{(1)} + \dots + D_{(k+1)}$.

For generators in general quadratic position, each interpolated value in $D_{(k+1)}$ will be interpolated from three values in $D_{(0)} + D_{(1)} + \dots + D_{(k)}$. If $D_{(k+1)i}$ is the interpolated value at $Q_{(k)i}$ and $V(Q_{(k)i})$ is the Voronoi polygon of $Q_{(k)i}$, and $*V(Q_{(k)i})*$ is the area of $V(Q_{(k)i})$, then

$$D_{(k+1)i} = \sum_{p=1}^3 \frac{|V(Q_{(k)i}) \cap V(Q_{(k-1)p})|}{|V(Q_{(k)i})|} D_{(k)p} \quad [1]$$

Note that:

$$|V(Q_{(k)i})| = \sum_{p=1}^3 |V(Q_{(k)i}) \cap V(Q_{(k-1)p})|.$$

By selecting vertices exclusively, an entire set of interpolated vales can be generated in one pass. Further, vertices may be considered as locally optimal sites for new data interpolation locations since they maximize the distance from triples of existing interpolation locations.

3. Operationalising SRVPD Construction in Arc/Info

The SRPVD constructions described in this paper were built with the Arc/Info 8.1 GIS platform. The input consists of a set of generator seed points, the $G_{(0)}$ generator set described in Section 2, and an associated set of attribute values. These data, which we refer to as Gen_*i*_Pts, were used as the basis for each of five Voronoi recursions.

The general procedure for generating the Simple RPVD construction is as follows with specific Arc/Info commands identified for reference in `courier` font:

1. Create a Voronoi polygon coverage for Generation *i* (Gen_*i*_VD) Gen_*i*_Pts using Thiessen.
2. Build node and line topology for Gen_*i*_VD.
3. Add a “Gen” field to the node attribute table of the Gen_*i*_VD coverage to track the “age” of each point. The value of the Gen field is set to *i* for all nodes.
4. Convert all nodes in Gen_*i*_VD to a point coverage (VD_*i*_Pts) using NodePoint.
5. Create Gen_*i*+1_Pts coverage by Appending Gen_*i*_Pts and the VD_*i*_Pts.
6. Repeat steps 1 to 5 based on the updated Gen_*i*+1_Pts coverage for *k* generations.

Figure 2 illustrates this procedure for generations 0 to 2. In practice, although the Voronoi structure is not bounded in a conceptual sense, a spatial limit needs to be established to minimize the impact of edge effects on subsequent analyses. This is discussed further in Section 4.

Figure 2. SRVPD generations 1 to 3

Next, attribute values for the new data points created using the process described above are interpolated on a generation-by-generation basis. The interpolation procedure is summarised below and is also illustrated in Figure 3:

1. Union Gen_i_VD and Gen_{i+1}_VD Voronoi polygon coverages to create a $Union_i_{i+1}$ polygon coverage
2. Establish *relates* between $Union_i_{i+1}$, Gen_i_VD and Gen_{i+1}_VD . Calculate the ratio of each $Union_i_{i+1}$ polygon's area relative to the area of the corresponding "parent" VD_i polygon. Store the value in a *PercentArea* field.
3. Calculate the attribute value being interpolated for each $Union_i_{i+1}$ polygon as $PercentArea * \text{the attribute value recorded for the parent } Gen_i_VD \text{ polygon}$. Store the value in a *CalcAttribute* field.
4. Use *statistics* on the $Union_i_{i+1}$ coverage to create a summary table (Gen_i_stats) that lists the sum of the *CalcAttribute* field for each Gen_{i+1}_VD polygon.
5. *Reselect* the Gen_{i+1}_VD and Gen_{i+1}_Pts features for generation $i+1$ and transfer the corresponding interpolated values from the Gen_i_stats table using *relates*.

Figure 3. Interpolation procedure example

4. Illustration

To demonstrate our procedure, we applied it to a sample of data values drawn from a known surface. The surface is the fifth-order trend surface shown in Figure 7 which has an elevation range of 80 metres (352 – 432 metres). The coefficients for generating this surface are given in Table 1. The 225 points in the initial generating set are located on a randomly jittered square grid (see Figure 7). We used the procedure described in Section 2.2 to generate additional data sites and data values up to the 5th generation of the SRPVD. Column 2 of Table 2 gives the numbers of the data sites by their generation.

Table 1. Coefficients of trend surface

Coefficients			
Constant	350.318		
		X4	2.503E-19
X	-.0001063	X3Y	1.021E-20
Y	.0003144	X2Y2	-9.690E-20
		XY3	-1.484E-19
X2	9.863E-09	Y4	8.927E-20
XY	-2.513E-09		
Y2	4.007E-10	X5	-3.229E-25
		X4Y	3.612E-26
X3	-7.548E-14	X3Y2	2.751E-26
X2Y	-3.321E-16	X2Y3	1.548E-25
XY2	3.264E-14	XY4	2.712E-25
Y3	-1.224E-14	Y5	-2.011E-25

As indicators of the performance of our procedure, we undertook a series of error analyses. However, recognizing that our procedure, like all interpolation procedures, is subject to edge effects, we did not consider errors associated with data points located within a 20,000 metre buffer of the boundary of the convex hull of the sample points. Column 3 in Table 2 lists the number of data points that remain in the analysis.

Table 2. RVP generated data points and errors by generation

Generation	All points	Points examined	# extreme errors ¹	% extreme error points
0	225	169	-	-
1	452	338	0	0
2	1356	1012	0	0
3	4068	3040	6	0.20
4	12204	9116	145	1.59
5	36612	27342	65	0.24
Totals	54917	41017	216	0.53

¹ An RVP data point is classed as an extreme error if its interpolated value deviates (+ or -) by 4 metres from the corresponding trend surface value.

To examine overall performance, we calculated both the mean absolute error (MAE) and the root-mean-squared (RMSE) error by generation (see Table 3). MAE is less sensitive to large errors than RMSE and so gives a better picture of the overall performance of the interpolation. As Table 3 shows, the average MAE over all generations is small, approximating 0.6 metres (0.75 per cent of the elevation range). There is a marked reduction in MAE between the first two generations, after which, with the exception of generation 4, MAE stabilizes. The behaviour of RMSE is much more variable, in part, reflecting its sensitivity to extreme errors. Nevertheless, average RMSE is still small, just over 1.5 metres (1.75 per cent of the elevation range). The RMSE (and, to a lesser degree, the MAE) for generation 4 is strongly influenced by the presence of five absolute errors in excess of 50 metres (see Figure 4). The cause of these errors is the subject of on-going investigation.

Table 3. Mean Absolute Error and Root Mean Square Error by generation

Generation	Mean Absolute Error	Root Mean Square Error
1 to 5	0.5870	1.5630
0 (initial generators)	N/A	N/A
1	0.9903	1.2743
2	0.5319	0.7373
3	0.5571	1.3562
4	0.7354	2.7589
5	0.5413	0.9394

Further analyses focussed on absolute errors of at least 5 percent of the elevation range. We label these extreme errors. Figure 4 shows the distribution of extreme errors by generation.

Figure 4. Distribution of extreme errors by generation

Considering the entire set of points (generations 1 –5), extreme errors occur for only 0.53 per cent of the data points. Of the individual generations, generation 4 has the largest proportion of extreme error points, although even this is very small. Our anticipation was that the largest errors would be associated with the earliest generations since, on average, these points would be furthest from the points of the initial generator set. However, the largest errors occur for generation 4, which also has the largest number of extreme errors. Somewhat surprisingly, there are no extreme errors associated with data points generated during the first two generations. For the final two generations, negative errors (underestimates) are more common than positive ones. We suspect this result may be an artefact of the form of the trend surface (see below).

Next, we examined the relationship between errors and the actual values on the surface. Given that our procedure is constrained to generate values within the range of values for the generator points, we anticipated that the errors might be concentrated around the limits of this range. However, as Figure 5 shows this is not the case, errors occur throughout the range, with the largest cluster of extreme errors in the range of approximately 410-425 metres.

Figure 5. Extreme errors by trend surface elevation

This concentration can be explained as an artefact of the trend surface and, by extension, the source data points on which subsequent recursions were based. As Figure 6 shows, the number of source data points with these elevations is smaller than for other elevations.

Figure 6. Frequency of source data points by elevation

Figure 7 shows the spatial distribution of extreme errors. Note that despite the presence of our buffer zone, many of the errors are concentrated around the edges of the study region.

Figure 7. Spatial distribution of extreme errors

The remaining errors show no obvious spatial distribution although there is a tendency for them to be located in small local clusters. While some of these clusters appear to be associated with initial generator points, this does not appear to be a general tendency. Indeed, we had anticipated that there might be a positive relationship between error magnitude at a generated location and the distance of that location to the nearest initial generator point. However, as Figure 8 shows this does not occur. In this figure, the distance of each extreme error point to the nearest generator point (generation 0) is plotted. Since the generator points are located on an approximate square grid with spacing of approximately 20,000 metres, the furthest a generated point could be from a generator point would be approximately 14,000 metres. In contrast, the largest number of extreme errors occurs at distances of 7-9,000 metres.

Figure 8. Extreme error and proximity to initial generator points

5. Discussion and Conclusions

The primary aim of this paper was to introduce an automated method, based on the simple recursive point Voronoi diagram, for progressively increasing the density of an initial point set, thus providing a basis for interpolating surfaces of variable resolution from sparse samples of data sites. As we show, an attractive feature of this procedure is that it can be readily implemented in Arc/Info. In our illustrative example, absolute errors in excess of 5 per cent of the variable range occurred at less than 1 per cent of the generated data sites. Further, the lack of systematic relationships between error values and characteristics of either the underlying surface or the initial point set, suggests that the procedure may be robust with respect to these features. However, in order to confirm this much more testing needs to be done. For example, we need to explore the sensitivity of the results for a given surface to different spatial sampling designs (random, systematic, stratified, VIP points, etc.) for selecting the generator points. We also need to consider a range of surfaces with different characteristics, especially more complex surfaces. And, of course, we need to compare the performance of the procedure relative to other COTS interpolation procedures.

Notwithstanding the need for further testing, there are several ways the procedure may be refined and extended. For example, it is possible, and probably desirable, not to add all the Voronoi vertices of a given generation to the generator set for the next generation. There has already been some initial consideration of such a Generator Constrained Recursive Point Voronoi Diagram (GCRVPD) (Boots et al., 2002). In the present context, we may wish to constrain the new generators to those within a specific distance range of existing generators. If the potential new generators are very close to existing ones, we can expect that they may add little new information and thus may be discarded. Similarly, if a potential new generator is more than some specified distance from an

existing generator it may be unwise to extend interpolation to that generator. One means of identifying such an upper limit would be to use the characteristics of the semi-variogram of the initial data points, perhaps limiting new generators to those whose distance to existing generators is less than half of the range of the semi-variogram. A similar strategy has been used successfully in the choice of control points in classification of spatial imagery into classes (Shine and Wakefield, 1999).

Although we illustrated our procedure using quantitative data, it is also applicable to categorical data. In this case, instead of a single value being interpolated at each generated point, a vector of Fuzzy Membership Values (FMVs) for classes can be interpolated (Lowell, 1994) and equation [1] will need to be adjusted accordingly. FMV surfaces can then be generated for each class. Alternatively, at each recursion, each polygon can be labelled with the value of its generator and these values mapped to create a piecewise continuous surface (choropleth map).

6. References

- BOOTS, B., FEICK, R.D., SHIODE, N., and ROBERTS, S. 2002, Investigating recursive point Voronoi diagrams. In M.J. Egenhofer and D.M. Mark (eds.), *Geographic Information Science: Second International Conference, GIScience 2002, Boulder CO, USA, September 2002. Lecture Notes in Computer Science*, 2478. (Berlin: Springer-Verlag), pp. 1–21.
- BOOTS, B., and SHIODE, N., 2003, Recursive Voronoi diagrams, *Environment and Planning B: Planning and Design*, **30**, 113–124.
- DE FLORIANI, L., 1989, A pyramidal data structure for triangle-based surface description. *IEEE Computer Graphics and Applications*, **9**, 67–78.
- DE FLORIANI, L., FALCIDIENO, B., and PIENOVI, C., 1985, Delaunay-based representation of surfaces defined over arbitrarily shaped domains. *Computer Vision, Graphics, and Image Processing*, **32**, 127–140.
- DE FLORIANI, L., MARZANO, P., and PUPPO, E., 1996, Multiresolution models for terrain surface description. *The Visual Computer*, **12**, 317–346.
- DE FLORIANI, L., and PUPPO, E., 1992, A hierarchical triangle-based model for terrain description. In Frank, A., Campari, I., Formentini, U., eds., *Theories and Methods of Spatio-temporal Reasoning in Geographic Space. Lecture Notes in Computer Science*, 639. (Berlin: Springer-Verlag), pp. 236–251.
- DE FLORIANI, L., and PUPPO, E., 1995, Hierarchical triangulation for multi-resolution surface description. *ACM Transactions on Graphics*, **14**, 363–411.
- HUTCHINSON, M.F., and GALLANT, J.C., 1999, Representation of Terrain. In Longley, P.A., Goodchild, M.F., Maguire, D.J., and Rhind, D.W., eds. *Geographical Information Systems: Principles, Techniques, Applications, and Management. Volume 1*. Second edn. (New York: John Wiley & Sons), pp. 105–124.
- LOWELL, K.E., 1994, A fuzzy surface cartographic representation for forestry based on Voronoi diagram area stealing. *Canadian Journal of Forest Research* **24**, 1970–1980.
- OKABE, A., BOOTS, B., SUGIHARA, K., and CHIU, S.N., 2000. *Spatial Tessellations: Concepts and Applications of Voronoi Diagrams*. Second edn. (Chichester: John Wiley & Sons).

- SHINE, J.A., and WAKEFIELD, G.I., 1999, A comparison of supervised imagery classification using analyst-chosen and geostatistically-chosen training sets, *Geocomputation 99* CD-ROM.
- SIBSON, R., 1978, Locally equiangular triangulations. *The Computer Journal*, **21**, 243-245.
- VOIGHTMANN, A., BECKER, and L., HINRICHS, K., 1994, Hierarchical surface representations using constrained Delauney (Sic) triangulations. In *Advances in GIS Research: International Symposium on Spatial Data Handling. Volume 1 & 2*, pp. 848–867.
- WANG, K., LO, C.P., BROOK, G.A., and ARABNIA, H.A., 2001, Comparison of existing triangulation methods for regularly and irregularly spaced height fields. *International Journal of Geographical Information Science*, **15**, 743–762.
- WATSON, D.F., 1992. *Contouring: A Guide to the Analysis and Display of Spatial Data* (Oxford: Pergamon Press).