Thoughts on Representing Spatial Objects

William A. Huber
Quantitative Decisions
Rosemont, PA
Overview

1. Some Ways to Structure Space
2. What to Put into a Grid
3. Objects and Fields
4. Hybrid Structures
Part 1
Some Ways to Structure Space
A Square Grid
A Rectangular Grid
A hexagonal grid
The Cell Centers Form a Regular Array
An Irregular Rectangular Grid

Finite element models—e.g., for groundwater flow—routinely use such grids as a simple way to adapt their resolution to the rate of change of input parameters (e.g., hydraulic gradient).
A Confocal Conic Grid

I once used these for quantum mechanical calculations. See W.A. Huber and C. Bottcher, J. Phys. B 13, L399 (1980)
A Grid with Free Boundaries

Application: Spatial Simulated Annealing.

Row 22:
(Left, right) =
{(9, 13); (14, 17); (25, 39)}
A Quadtree
Application: Minimize Average Nearest Neighbor Distance

Colors are soil types.

**Challenge**: Find 10 points that minimize the *average* distance to the nearest.

(If you like, weight the average according to soil type. More importantly: re-express the distance as a semivariance. The solution tells us where to sample to obtain the most information when interpolating from the sample.)
A Solution Method

These locations are found via spatial simulated annealing (SSA). During this randomized algorithm, the points bounce around. As they bounce, their Thiessen polygons (restricted to this convoluted study region) change. Because several hundred thousand “bounces” are needed, it is imperative to update the Thiessen polygon boundaries as quickly as possible.
Illustration of SSA

**Objective:**
Situate 20 points to minimize average distance within the gray square to the nearest point; i.e., minimize

$$ \int_{x \in \text{Square}} \min_{y \in P} \{\|y - x\|\} \, dx $$

where $P$ is the set of 20 points.

**Constraint:**
The points must lie within the large square but not within the central black square.

**Solution:**
Adjust the points randomly, accepting new positions with controlled probabilities.
SSA Steps Have to be Fast!

**Naïve algorithm** for one SSA step:

- Move a point. Check that it satisfies the constraints.
- Recompute the distance grid.
- Estimate the integral with a global average.

**Improved algorithm**:

- Move a point *anywhere*.
- Recompute *only the part of the grid affected by the move*.
- *Update* the integral based only on the recomputed cells, adding a *penalty* for a point located in a forbidden area.

The improved algorithm is made possible by using a free-boundary representation of the Voronoi cells (Thiessen polygons) around each point.
Ideas

• Spatial *structure* and spatial *content* are different. Structure is a *container* for the content.

• Good structures:
  – Allow rapid identification of immediate neighbors.
  – Are homogeneous enough to permit efficient computation of local properties such as slopes.

• As we move down the ladder from highly regular, static structures to less regular, adaptive, or dynamic structures,
  – Calculations become more difficult and slower.
  – Storage can go down.
  – Accuracy (for a given amount of storage) can go up.
Application: Air Deposition Modeling (TNO, 2001)
“Kernels” Spread All Sources Over the Land Surface
How a Kernel is Represented

Data consist of numeric values provided at all nodes of this irregularly-spaced radial grid.

This value is interpolated logarithmically or cubically in the radial direction and linearly in the angular direction.
A Hybrid, Adaptive Data Structure Makes Optimization Possible

The large grid is computed within the region occupied by very low kernel values.

This value is estimated by the approximate value in the center of the large grid cell.
Consequences of Adaptive Data Structures

• Typically, the expensive computation (log-linear interpolation) has to occur only within 0.1% of the kernel’s support region.

• Typically, the large grid cells contain 100-200 model grid points. Thus, values at 100-200 points are computed with one calculation, reducing computation in that region by 98 to 99%.

• The resulting computation time can be 1/50 of the original (fully accurate) computation time without substantial loss of accuracy.
Accuracy Assessment

“Modeling the model”
OPS Simulation

The reference
OPS Convolution
In our case, accuracy is spatially variable: it is best in the areas that drive the management decisions.
Part 2
What to Put into a Grid
ESRI Puts in Indexes and Floats

<table>
<thead>
<tr>
<th>sfid</th>
<th>pop</th>
<th>hi_occ</th>
<th>hu</th>
<th>p_black</th>
<th>p_native</th>
<th>p_</th>
</tr>
</thead>
<tbody>
<tr>
<td>577623</td>
<td>418</td>
<td>131</td>
<td>133</td>
<td>65.07</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>577624</td>
<td>1448</td>
<td>520</td>
<td>524</td>
<td>73.89</td>
<td>0.14</td>
<td></td>
</tr>
<tr>
<td>577625</td>
<td>131</td>
<td>37</td>
<td>37</td>
<td>64.89</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>577626</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>577627</td>
<td>583</td>
<td>182</td>
<td>182</td>
<td>65.07</td>
<td>0.17</td>
<td></td>
</tr>
<tr>
<td>688388</td>
<td>464</td>
<td>64</td>
<td>64</td>
<td>62.99</td>
<td>0.00</td>
<td></td>
</tr>
</tbody>
</table>
For ESRI-Watchers: a Short List of What Spatial Analyst Needs

**Basic Features**
- Focal, zonal, and global percentiles and ranks. These are essential for exploratory and robust procedures.
- Global statistics as grids. (A convenience.)
- Complex-valued grids, supported by suitable functions (arg), modulus), complex-valued numerical functions, etc.) and displays (as vector fields, e.g.). Plus creation of complex grids by (a) combining real and imaginary parts, (b) combining polar coordinates, (c) representing gradients of other grids.
- Fourier transforms.
- Zonal stats (tables) for “region” layers (overlapping polygons).
- Point declustering for conversion to raster format, with a variety of declustering options for combining attributes.
- Argmin. (A convenience.)
- Option to treat NoData as zeros (or as some constant) in focal stats.
- Spherical distance grids and calculations.
- Enhanced focal stats and block stats. Maybe flexible scripted stats? E.g., identify all cells that are unique within their neighborhoods.
- Improved focal min and block min, focal max and block max, with methods to resolve ties instead of yielding NoData.
- Scatterplots, scatterplot matrices, and cross-correlations of grids.

**Specialized or Advanced Features**
- Generic dynamic programming to solve calculus of variations problems (Bellman).
- Streamlined and more flexible path distance and particle tracking procedures.
- Topographic feature identification and extraction.
- An automatic root finder to solve inverse problems. (E.g., cut and fill calculations could easily be done this way.)
- Extended viewshed calculations, such as finding horizon points.

**Performance**
- FFT for convolution (most focal stats).
- Improved capacity for zonal stats.
What Else?

Vector fields

Fields on a sphere

And tensor fields, etc. (These are hard to picture but are important for analysis of movement on a surface. E.g., a (1,1)-tensor describes the actual motion at a location ensuing from an attempted motion there.)
Why Do We Care?

Parallel transport of a vector field on a surface produces a holonomy group reflecting its curvature.

We need this kind of calculation when doing global GIS! Applications include meteorology, climate analysis, and plate tectonics.

This vector, when carried around the path PQR, ends up as the red vector. The amount of change is proportional to the total curvature of the ellipsoid within the triangle PQR. Tensors exhibit holonomy too.
General Objects

We should be able to associate *any* kind of objects with points on a surface, curved or flat. *E.g.*, spectra, time series, photographs, videos, sound, touch, arbitrary text, *etc.*

This becomes interesting when the object changes in continuous ways with change in position on the surface. *E.g.*, a video shows a continuity of scene as one travels around the earth. Continuity implies the ability to *interpolate* and suggests *spatial correlation*. 
"Objects refer to things in the world whilst a field refers to a single valued function of location in two-dimensional space."


However, we can represent things in the world by means of their *indicator functions*; *i.e.*, as special kinds of fields. Using an object or a field is a *choice* about how best to represent data for a particular need, whether it be storage, cartography, or analysis.
A Mathematical Picture

Object: Domain → Space

Field: Space → Range

The following ideas generalize easily to 3D and time-varying GIS.
Objects Are Represented as Simplicial Complexes

Object:

- **Domain**
  - Multipoint: $\mathbb{R}^0$
  - Polyline: $\mathbb{R}^1$
  - Polygon; surface: $\mathbb{R}^2$
  - Solid: $\mathbb{R}^3$

- **Space**
Fields Are Represented by Gridded Data

**Field:**

- **Space**
- **Range**

- \{Numbers\}, \{Tensors\}, \{Whatever…\}
In All Cases, the *Start* of the Arrow is Discretized
Indicator Fields Represent Objects

Blue indicates presence.
Gray indicates absence.
1-0, true-false, present-absent: all are forms of binary indicators.
Discretizing the Range Creates Contours

\[ f : \text{Space} \rightarrow \text{Range} \]

\[ \text{Space} \leftarrow f^{-1} \text{ Levels} \quad \text{(a discrete subset)} \]

Contours effectively represent real-valued fields as collections of objects.
Part 4
Hybrid Structures
Why Not Put the Two Together?

We can simultaneously represent objects and fields.

The trick is to connect the two representations in useful ways.

What do we gain?
Here are 20,000 blocks with people in them.
This *subset* of the road network contains over 100,000 features represented by a million total edges and vertices.
Comparison of Two Geometries

Euclidean distance
(converted to travel time by using the average road speed)
Extending Travel Times Off-Network

- We create a discrete grid to represent off-network points.
- The grid actually is a network where each cell is connected to some of its neighbors (usually eight of them).
  - The cost to travel left-right or up-down is fixed.
  - The cost to travel diagonally is $\sqrt{2}$ times greater.
  - Barrier cells are disconnected from all neighbors: one can neither leave them nor arrive at them.
- The network travel times are rasterized into the grid to establish initial conditions. The cost of this operation is proportional to the total length of the network and inversely proportional to the grid cell size.
- The computational cost and RAM requirements for finding travel times on the grid are inversely proportional to the square of the grid cell size, so there is a limit to the resolution we can use.
  - In practice, the resolution is about 25 to 100 meters, or up to one minute in travel time at a walking pace.
Extending Travel Times: Illustration

- Compute travel times at every street node.
- Interpolate times along edges and rasterize them onto this grid.
- Mark the barrier cells.
- Extend travel times off-grid at a walking speed.
  - These cells are 50 m across.
  - The speed is 2 miles/hr.
  - Notice the effect of the barrier in the lower left corner: the time at one cell is 4.47 minutes, although it is diagonally opposite (70 meters from) a cell just 1.24 minutes from the site.
Evaluation of this Hybrid Data Format

• Advantages:
  – The network representation lets us apply Dijkstra’s algorithm to get accurate on-network travel times.
  – The grid representation handles off-network travel times well and copes with barriers easily.
  – Once a grid is computed, obtaining a travel time to any location incurs a computational cost of $O(1)$.
  – This setup is ideal when a large number of travel times from a small set of given locations is needed.
  – Analyzing travel-time differences among competing suppliers is easy.

• Disadvantages:
  – Grids can occupy a lot of RAM. For example, to cover a 200 km region at 50 m resolution requires a 4000 x 4000 array of floats, occupying 64 MB.
Dreaming…

Why not let the system store spatial entities both as objects and fields?

We can automate geoprocessing without needing to know the data structure: the system can choose the appropriate storage method and resolution.

We can improve resampling, visualization, and analysis by using relevant parts of a (redundant) data structure.

The casual user does not need to know or understand how spatial entities are represented.

Accuracy can be arbitrarily high.
Conclusions

Musing about the fundamentals of our profession—how space is represented and how spatial objects are stored—suggests creative ways to enhance the capabilities of GIS and the performance of spatial algorithms.