Exploring Microclimate data – some issues

***Peter Whigham*[1]**

[1]Spatial Information Research Centre
University of Otago. Dunedin, New Zealand
Phone: +64 3 479-7391 Fax: +64 3 479-8311
Email: pwhigham@infoscience.otago.ac.nz

Presented at SIRC 2001 – The 13th Annual Colloquium of the Spatial Information Research Centre
University of Otago, Dunedin, New Zealand
December 2nd-5th 2001

ABSTRACT

Advances in data collection devices and monitoring stations have allowed microclimate data to be collected at high resolutions, both spatially and temporally. These increased data resources should imply that patterns in space and time can now be analysed with increased accuracy. This paper demonstrates some approaches to the visualisation of microclimate data, based on temperature measurements at 6 minute intervals, and raises issues regarding the definition, automatic interpretation and selection of spatio-temporal features in large datasets.

***Keywords and phrases*:** microclimate, visualisation, spatio-temporal analysis, feature selection

1.0 INTRODUCTION

The advent of increased data collection through the use of automatic data collection devices has meant that new techniques for the visualization and exploration of this data are required. This paper looks briefly at some of the issues, and describes a simple system that highlights some of the approaches and difficulties involved.

The lack of sufficient data at appropriate resolutions was once a basic problem for physical geographers and modelers. This has largely been eliminated with decreasing prices for automatic data collection devices, and the use of techniques such as satellite imagery, spatial information systems and large-scale relational databases. This paper will use, as an example, a small subset of the data collected during the Topoclimate south project, based in Southland, New Zealand.

1.1 The Southland Dataset

The dataset used in this paper is shown in Figure 1. It is composed of 39 sites over an approximate area of 344 km^2, and each site has temperature recorded at 6 minute intervals for a period from 1st July, 1998 to 10th November, 1998. This represents 133 days of data at 39 locations, giving 1,244,880 data points in total. Note that this is but a small subset of the total data generated by the Topoclimate south project, which has 2550 sites giving over 223 million points of data. This rich dataset has many applications, including the study of links between the topographic and climatic features in an environment (Oke,1987). For example, at the spatial and temporal scales that this data has been collected such temperature-based features as cold air ponding and radiation loading can be studied in detail. Additionally, other climate variables, such as growing degree days, can be accurately constructed from the data to produce models for agricultural and other environmental studies.

1.2 Interpolation Methods

Accurate data interpolation methods, such as kriging and co-kriging (Burrough and McDonnell,1998; Fortin,1999), have been previously develop to produce surfaces based on point data. Although these approaches are accurate, in the sense that they minimise a model error, they require the user to match a variogram model to produce a mathematical equation representing how the attribute being interpolated varies across space. Since this matching cannot be done automatically kriging is not appropriate when real-time interpolation is required. For our simple example there are over 31,000 different surfaces representing the snapshots of the temperature over the study area. Hence a manual approach to interpolation cannot be applied. As such, a simple local interpolation method, inverse distance weighting (IDW) (Lam,1983), has been used to construct the temperature surface in real time for the study area. There are several different variables that can be altered when using IDW, including how to select the points to include in the interpolation (by distance or number), and how these points are combined to give the interpolated value for the grid surface being constructed. For our study a simple weighted approach for each interpolated grid square has been constructed, based on Equation (1), where Z_0 is the estimated value,

$$Z_0 = \frac{\sum_{i=1}^{n} w_i Z(x_i)}{\sum_{i=1}^{n} w_i} \tag{1}$$

$Z(x_i)$ is the measured value, w_i is the weighting applied to this point and n is the number of points used in the interpolation. The weighting w_i is calculated as the inverse Euclidean distance from the sample site to the point. This calculation can be performed rapidly without user intervention, especially if the distance ordering between each grid point Z_0 and the sample points are pre-computed. Figure 1 shows an application of IDW to the 39 point locations to create a 100x100 grid cell surface. Each grid value (Z_0) used the nearest 6 points, weighted by the inverse distance from the grid cell. Note that the study area ranges in height from 121 to 351 metres, and that the 39 points are spread throughout the area. Although ideally when using IDW the points are regularly sampled in space this spread of points will be adequate for our purposes.

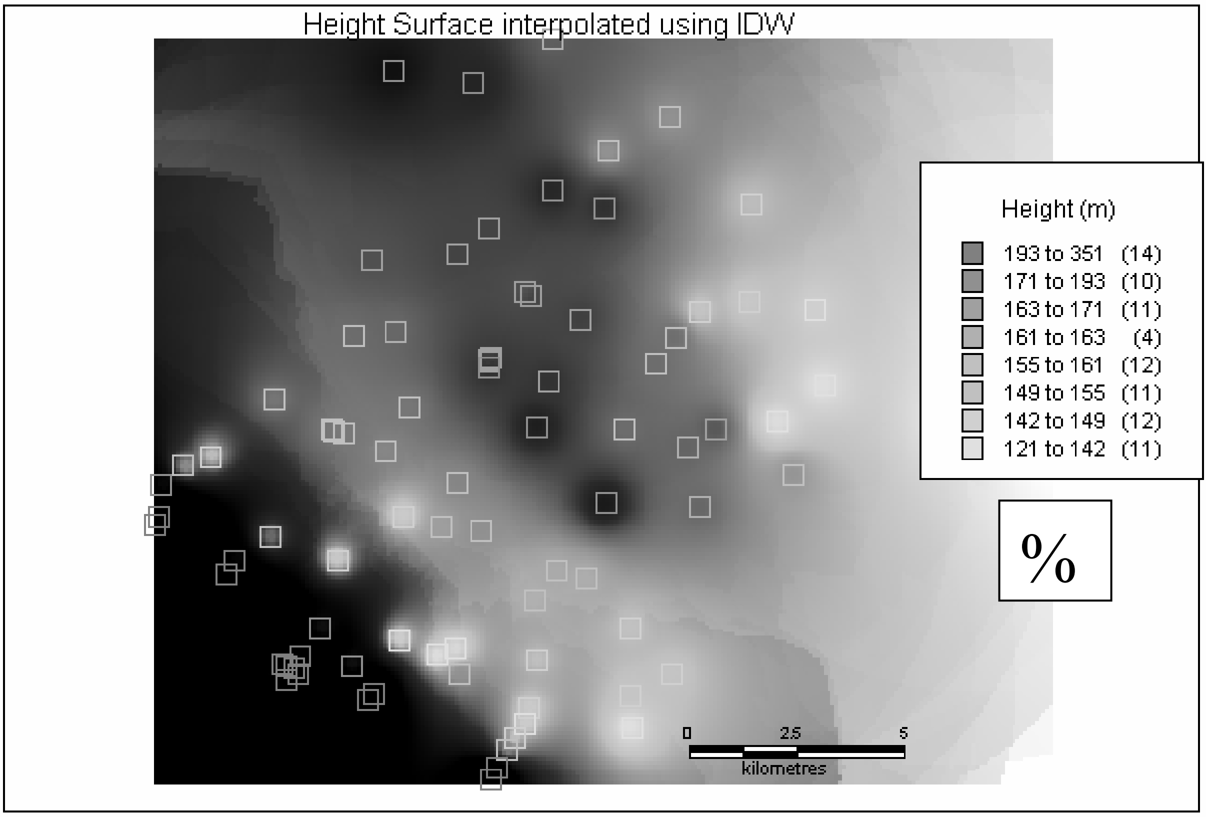

Figure 1. The interpolated height surface, showing the 39 point locations.

2.0 VISUALISATION OF TEMPERATURE

The system described in this section was created using Delphi V, using a graphical component GLScene for the 3D surface construction. GLScene is an OpenGL based 3D library for Delphi (further information may be found at http://glscene.org). The main issue related to the visualisation and exploration of a rich dataset is that it is often not possible to closely examine all points in the data. Therefore, methods are required to help discover patterns, either through additional visualisations that can be used to support the data, aggregation methods that reduce the number of data points, or approaches to defining in a generalised manner the patterns that are of interest to be discovered. The visualisation issues regarding the single variable, temperature, will be used to demonstrate some of the issues regarding these approaches.

The main window for the visualisation of the temperature data is shown in Figure 2. The basic features are a map window, showing the surface and the interpolated temperature at a selected time, a key range indicator, a gradient and difference view, and the current date and time. Additional tools include slider bars to control how rapidly the animation is shown, the temporal resolution of the display and a slider to allow a small time period to be moved back and forth. Apart from the gradient and difference features these controls are straightforward.

2.1 Gradient and Difference Measures

Two measures that may be of use when visualising temperature are the first derivatives in space and time. These are shown in Figure 2 as the gradient (derivative in space) and difference (derivative in time) maps. For any grid cell, the derivative in space is calculated by taking the square of the difference of the temperature on the left and right, and the square of the difference of the temperature on the top and bottom cells. These are then added together to give a single value. Finally, the square root of the value is taken to give a final value >= 0 indicating how rapidly the temperature is changing in an isotropic sense surrounding the cell. This is basically a simplified version of the Laplacian (Rosenfeld and Weszka,1980). This measure is useful for detecting edges (the Laplacian is a traditional edge detection filter used for grey-scale images) that represent rapid changes in temperature across space.

The calculation of the difference is simply the absolute difference between each cell from time *t* to time *t*+*k*. Note that the time step *k* is based on the selected time step period. This measure shows locations where the temperature is changing from *t* to *t*+*k*. Often locations where the temperature is either changing rapidly or very slowly will be of interest.

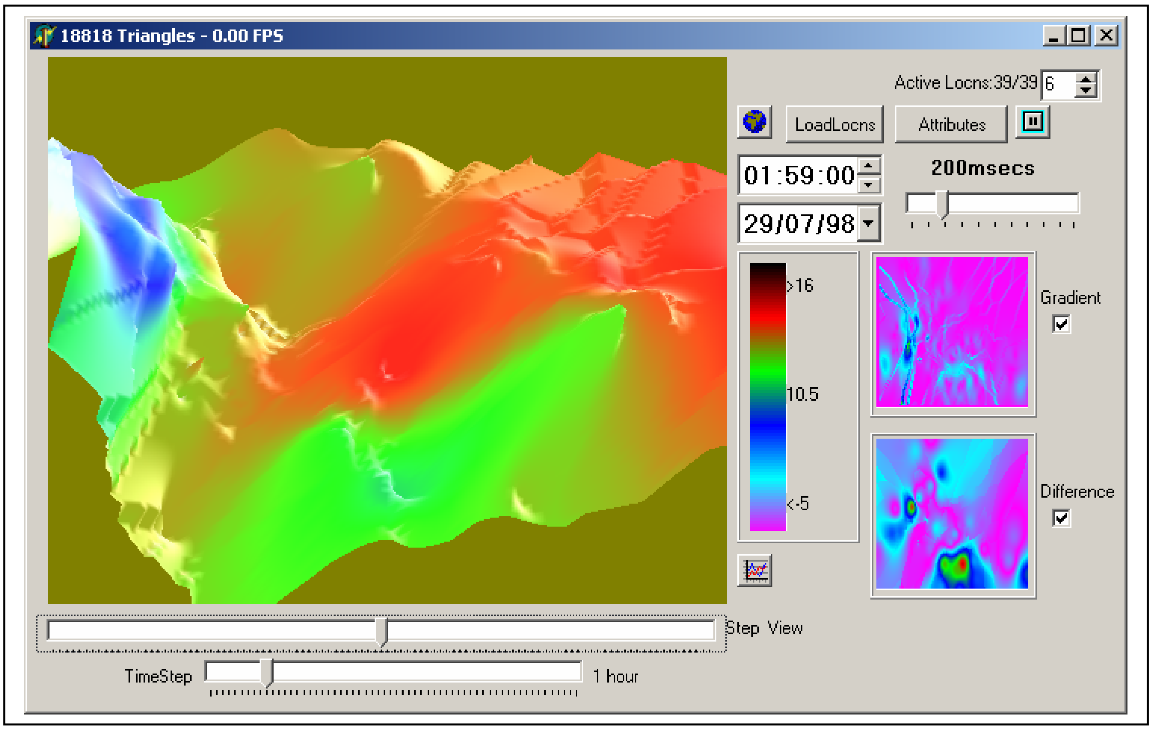

Figure 2. Main window for the temperature visualisation model.

2.2 Altering the Visual Output

The visual output of the animation can be varied in several ways, and each tends to alter the output to a large extent. This supports the statement that automated methods are required to explore these climatic pattern variables. For example, the spread of colours for the displayed output values can be altered to focus on a particular range of values. Small adjustments in these values can greatly change the colours displayed, and therefore the likely interpretations of the output.

The surface can be rotated, zoomed, panned and each axis (X,Y,Z) can be independently scaled. This allows the surface to be stretched to emphasize patterns that may be of interest. Normally this is used to extend the Z (height) variable so that the shape of the surface, including valleys and hilltops, are exaggerated. This can help to visualise the temperature changes occurring over the surface when focussing on the topographic and temperature relationships.

Varying the time step dramatically changes the visual output. Since the minimum time step is 6 minutes, increasing this step to 12, 18, 24, etc. changes the appearance of the gradient and difference filters, and shows how rapidly the microclimate data changes.

2.3 Other Visual Aids

Exploring temperature patterns often leads to focusing on one particular location in the landscape to study in greater detail. The window shown in Figure 3 allows this more focused approach to visualizing temperature, by plotting the temperature at a single cell location. The temperature values are based on the time step period selected on the main window. In addition, the left-right and top-bottom profiles for the surface are displayed so that the correct location in the landscape (for example, the center of a valley) can be selected. Note that the surface is shown in 2D with the standard colours of the key used to indicate height values.

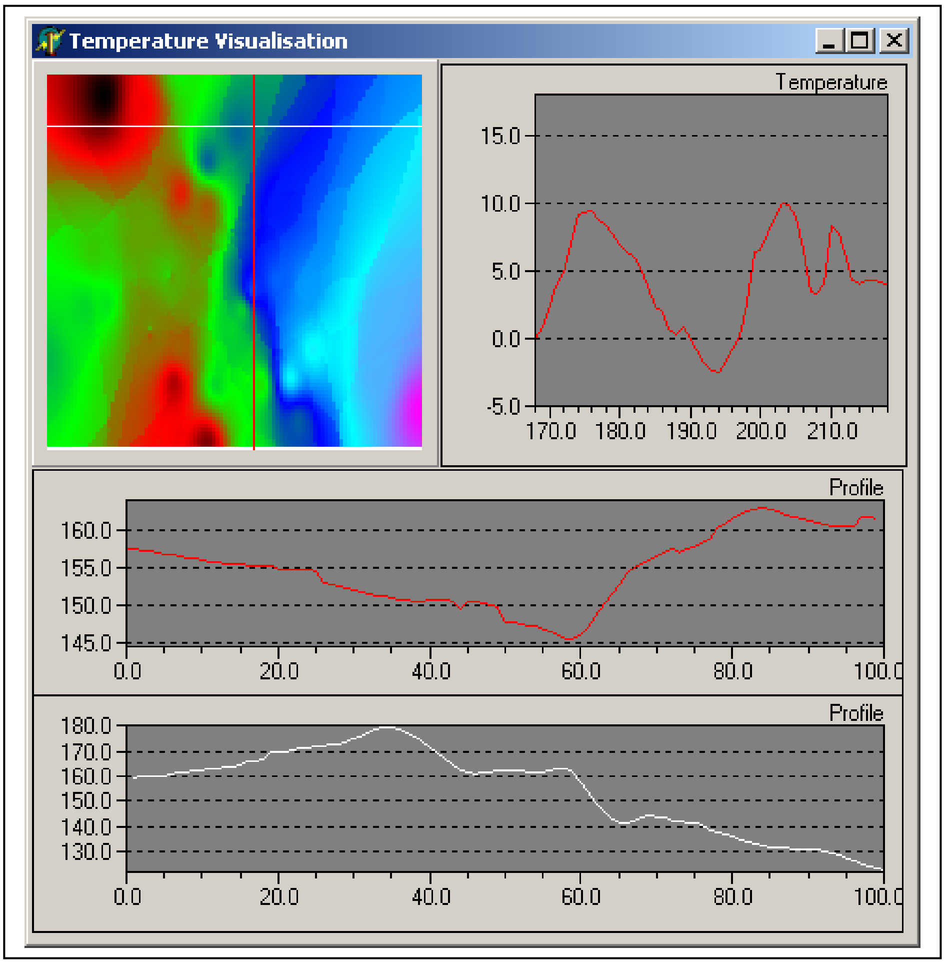

Figure 3. Surface profiles and temperature at a location

3.0 A CASE STUDY: COLD AIR PONDING

To study cold air ponding requires detecting the pattern of cold air movement down a slope and settling in the valleys and pockets of the landscape. This requires a number of different spatial and temporal patterns to be combined and detected. How, if at all, can this be achieved?

The location of valleys and pockets in the landscape can be determined by using the surface profiles as shown in Figure 3. However, a more direct method would be to detect automatically the low points in the surface and focus on the patterns at these locations. This approach was built into the system, by locating all cells that were the lowest value in their local (8 cell) neighbourhood. Unfortunately this mainly highlighted the weakness of a single point-based feature as a definition of a surface property, and reflected the properties of the IDW algorithm. Since IDW uses a weighted mean of values from sampled points, it is not possible to compute a cell with a value lower than the values of any sample point used to compute the weighted average. Hence, the local lowest point was always the sample points with the lowest local values. Clearly the surface locations for cold air ponding were not only the local points, but also the shape of the landscape in a context. Although the approach of Figure 3 can be used to detect these patterns (and of course could be used to select the areas), no automatic approach could be defined.

Irrespective of the issues surrounding site selection, detecting cold air ponding was attempted by observing the selected cells at the lowest local point, and notifying the user when this point was colder than its 8 neighbours. Unfortunately, a similar issue to the one raised above occurs when looking locally at a surface where an interpolation has occurred with temperature. The system tends to find that almost all locally lowest locations were colder than the surroundings, and any more selective ability in terms of limiting the number of times the patterns were detected was only produced by lowering the threshold definition of ‘cold’. Clearly there is some problem with using local definitions combined with surface interpolation when detecting patterns of interest.

4.0 DISCUSSION

Large datasets are attractive resources for the investigation of theories relating to how variables relate in space and time. However, once the data becomes large enough that no person can be expected to closely examine every surface or datapoint, then automatic, or semi-automatic methods for interesting pattern detection must be developed. The use of gradient and differencing operators are useful visual aids to highlight when a pattern may be of interest, however unless they can be constructed into some form of language that expresses when a pattern should be notified they add little to the interpretation. For example, there are many different types of filters, and combinations of filters, that may be of use in detecting and representing certain patterns. A language that allowed these filters to be combined, and used to notify the user when a certain pattern was detected, could be a useful semi-automatic aid to pattern detection.

The use of profiles, such as Figure 3, could be extended to other variable patterns (such as temperature). For example, cold air ponding may be defined by a temperature profile that is required for this type of feature. The profile would relate temperature change over the surface, and therefore be a contextual definition of the pattern. A graphical interface could allow the user to define this profile, and this would be used to search over a surface to detect these profiles. This profile approach could also be used to define the ‘local minima’ location.

The system currently has no way of aggregating data over time, however this is clearly a method for detecting patterns at different scales and to allow the gradual discovery and focusing of interesting patterns. Note that the aggregation could be over time and/or space, and therefore allow scale to be manipulated as part of the exploration process.

These approaches are currently being studied and implemented to allow a more formal system for the detection and interpretation of microclimate patterns based on large datasets.

5.0 CONCLUSION

The visualization and detection of patterns with large datasets is a complex and difficult task. Not only is the interpolation procedures used with space of importance, but the methods to allow context to be incorporated into the space and time definitions is fundamental. This paper has merely highlighted these issues and outlined some possible approaches to help in the detection of some basic patterns of interest.

ACKNOWLEDGEMENTS

The author thanks the Topoclimate south project, and Gary Hutchinson, for access to the data. In particular, thanks must be given to Katrina Richards of the University of Otago for her support with the interpretation and preparation of the dataset used in this study. Desley Anthony must also be thanked for her help in producing this paper.

REFERENCES

Burrough, P. A. and R. A. McDonnell (1998). Principles of Geographical Information Systems, Oxford University Press.

Fortin, M.-J. (1999). Spatial Statistics in Landscape Ecology. Landscape Ecological Analysis: Issues and Applications. J. Klopatek and R. Gardner, Springer-Verlag, New York.

Lam, N. S. (1983). “Spatial Interpolation methods: a review.” The American Cartographer **10**(2): 129-149.

Oke, T. R. (1987). Boundary Layer Climates.

Rosenfeld, A. and J. S. Weszka (1980). Picture Recognition. Digital Pattern Recognition. K. S. Fu. Berlin, Springer-Verlag**:** 135-165.