INT. J. GEOGRAPHICAL INFORMATION SCIENCE, 2003
VOL. 17, NO. 6, 495–516

Taylor & Francis
Taylor & Francis Group

Research Article

Infometric and statistical diagnostics to provide artificially-intelligent support for spatial analysis: the example of interpolation

CLAIRE H. JARVIS[1], NEIL STUART[2] and WILLIAM COOPER[2]

[1]Department of Geography, University of Leicester, Leicester LE1 7RH, UK; e-mail: c.jarvis@le.ac.uk

[2]Department of Geography, University of Edinburgh, UK

(*Received 19 November 2001; accepted 23 August 2002*)

Abstract. The wider uptake of GIS tools by many application areas outside GIScience means that many newer users of GIS will have high-level knowledge of the wider task, and low-level knowledge of specific system commands as given in reference manuals. However, these newer users may not have the intermediate knowledge that experts in GI science have gained from working with GI systems over several years. Such intermediate knowledge includes an understanding of the assumptions implied by the use of certain functions, and an appreciation of how to combine functions appropriately to create a workflow that suits both the data and overall goals of the geographical analysis task.

Focusing on the common but non-trivial task of interpolating spatial data, this paper considers how to help users gain the necessary knowledge to complete their task and minimise the possibility of methodological error. We observe that both infometric (or cognitive) knowledge and statistical knowledge are usually required to find a solution that jointly and efficiently meets the requirements of a particular user and data set. Using the class of interpolation methods as an example, we outline an approach that combines knowledge from multiple sources and argue the case for designing a prototype 'intelligent' module that can sit between a user and a given GIS.

The knowledge needed to assist with the task of interpolation is constructed as a network of rules, structured as a binary decision tree, that assist the user in selecting an appropriate method according to task-related knowledge (or 'purpose') and the characteristics of the data sets. The decision tree triggers exploratory diagnostics that are run on the data sets when a rule requires to be evaluated. Following evaluation of the rules, the user is advised which interpolation method might be and should not be considered for the data set. Any parameters required to interpolate the particular data set (e.g. a distance decay parameter for Inverse Distance Weighting) are also supplied through subsequent optimisation and model selection routines. The rationale of the decision process may be examined, so the 'intelligent interpolator' also acts as a learning tool.

1. Introduction

1.1. *'Intelligent' GIS*

Facilitating the appropriate and efficient use of GIS by creating more 'intelligent' GIS to support decision makers has long been identified as a priority for basic

International Journal of Geographical Information Science
ISSN 1365-8816 print/ISSN 1362-3087 online © 2003 Taylor & Francis Ltd
http://www.tandf.co.uk/journals
DOI: 10.1080/1365881031000114099

research within the environmental modelling community (Densham and Goodchild 1989, Burrough 1992, Fischer and Nijkamp 1992). This goal is also congruent with the desire to provide 'easy-to-use' spatial analysis functions expressed during the late 1980s. Anselin for example argued that 'With the vast power of a user friendly GIS increasingly in the hands of the non specialist, the danger that the wrong kind of spatial statistics will become the accepted practice is great'. Ten years on, it is useful to reflect on the advances towards 'intelligence' in GIS, to question whether the original intentions remain desirable and, if so, what GIScience-specific research remains to be carried out. We approach this subject both by reviewing progress and by providing a conceptual design and prototype implementation of an 'intelligent' module for the interpolation of geographical data.

The last decade has seen several methods developed as part of research into artificial intelligence becoming used by researchers in GIScience. Followers of the GeoComputation series will be familiar with the introduction of expert systems (Leung and Leung 1993), neural networks (Fischer and Reismann 1999, Rigol *et al.* 2001), fuzzy logic (Stefanakis *et al.* 1999), artificial life and cellular automata (Câmara *et al.* 1996), genetic algorithms (Murnion and Carver 1996) and more recently genetic programming and autonomous agents (MacGill *et al.* 1999, Westervelt and Hopkins 1999). Broader establishment of some of these techniques in geography is marked by books by Openshaw and Openshaw (1997) and Hewitson and Crane (1994) writing from quantitative human geography and physical geography perspectives respectively. In general, we would argue that most of this research has viewed AI methods as *alternatives* to existing statistical tools. AI techniques have been shown as viable alternatives to methods such as maximum likelihood classification for satellite images (Jarvis and Stuart 1996), parametric regression (Cheesman and Petch 1999) and specific process models (Dawson and Wilby 2001). It is much less common, however, to see AI methods being used to encapsulate the broader domain knowledge of a GIScientist who uses this knowledge to select and order the use of tools from an existing collection although exceptions may be found (Morse 1987, Zhu 1996). We conclude from this that we presently have artificial intelligence *in* GIS, but not artificially intelligent GIS. AI methods are being deployed within the GIS toolbox rather than as a wrapper around the tools.

1.2. *Incorporating 'intelligence' within GIS*

Application specialists who use GIS commonly have high level knowledge of their wider task, and low-level knowledge of the specific system commands typically provided by help systems or user manuals. However, these users may not necessarily have the intermediate knowledge (Bhavnani and John 2000) that experts in GI Science have gained from working with GI systems over several years. This intermediate knowledge can be vital for ensuring that appropriate analysis is conducted (Anselin 1989). Indeed, one of the main goals identified for spatial DSS over the last decade was to make GIS tools available to users with different levels of expertise (Cowen and Shirley 1991).

Intermediate knowledge is acquired from 'rules of thumb' formalised and refined through the experience of processing real data sets on a case-by-case basis; this is the intelligence required of an expert GIS user that an AI module would be designed to provide. An example use of intermediate knowledge is when an expert GIScientist chooses a specific method of analysis. This can involve evaluating the desirability of using different techniques, given the observed characteristics of the data. For instance, an expert uses intermediate knowledge when considering whether a data set encoded using a particular data model is suitable for a certain type of spatial analysis or whether one should first transform the data to a different encoding, bearing in mind inaccuracies that might occur from this preparatory process. Intermediate knowledge is also inherent in the workflow chosen to solve a particular task (Jarvis and Stuart 2001b). As Lovett (1998) and Levinson and Wilkinson (1997) note in relation to intelligent instruction systems and data analysis tools in statistics and exploratory data analysis more generally, these systems rarely support the use of anything more than isolated components or concepts. Progress has been slow in this area across many disciplines, perhaps because of a more general shift within computing and AI away from knowledge representation towards systems created using inductive methods such as data mining and knowledge extraction.

When designing methods that capture and use intermediate knowledge, several questions need to be addressed. These include the following.

- Is it preferable to select methods according to the wider purpose and domain of the analysis, or should choice be based mainly on the results of quantitative analysis of the data to hand? In other words, at what level should context be introduced?
- In relation to the procedural 'intermediate' knowledge of the interpolation selection problem, how should one balance knowledge that reflects the 'general best practice' against specific knowledge that can be obtained from a diagnostic analysis of the particular data set in use?
- Where should intermediate knowledge be stored and accessed? For example, if this knowledge is task specific, can it be recognised according to defined types of analysis and associated with specific GIS functions, or should it be stored with the data set as part of the metadata?
- How far should a user be aware of the analyses being undertaken to support the decision processes? Should this be hidden, or at what level of detail and at what stages in the analysis should there be interaction between the user and the AI module?

1.2.1. *Should methods be selected according to purpose and domain, or the characteristics of the data?*

With the emergence of faster computers, a shift may be identified in more recent years towards the use of process-intensive data manipulation techniques in GIS, rather than knowledge-based methods. We argue however that this computationally intensive approach, advocated for example by Burrough (1992), fails to tap into higher processes of cognition that an expert would also employ. There are arguments for the use of inductive, data focused methods. Model selection methods and criteria for example, be they statistical or from the AI domain, are valuable tools for the selection of the most significant input variables for a particular task or for structuring or parameterising the optimal form of a particular mathematical function (Ellner and Seifu 2002, Paradis *et al.* 2002, Ragg 2002, Shi and Tsai 2002). They are, however, an arguably expensive means of resolving the steps needed to solve a *multi-stage* geographical analysis task where each stage requires the user to choose between several functions, both in terms of financial cost and computer time. Firstly, this would necessitate the licensing and linking of functions, some of which may be implemented as separate programs Secondly, the data driven approach requires that all permutations of options must be evaluated numerically for their performance to be ranked, regardless of their potential for success. Arguments for the data-driven approach are also least strong where there is already a body of established, theoretical knowledge or a base of empirical knowledge, as is often the case in geographical research. Such knowledge falls into a variety of categories, and figure 1 highlights the particular need to recognise both the expertise in the application domain and the intermediate knowledge of a GIS specialist to ensure that appropriate methods are selected.

It has been observed previously in relation to spatial analysis methods (Openshaw and Alvanides 1999) that 'The ultimate aim is to develop an intelligent partnership between user and machine, a relationship which currently lacks balance.' Andrienko and Andrienko (1997) too make the important point, in the context of intelligent data visualisation, that even where data-model approaches may be able to take into account the characteristics of the data they fail to consider whether the given representation meets the objectives of the particular task.

1.2.2. *Should intermediate knowledge be associated with GIS functions, or specified as metadata?*

To provide a measure of intelligence to GIS, 'appropriate' methods for analysing particular data sets have been encapsulated within their accompanying metadata structures (Vckovski and Bucher 1996, Stefanakis *et al.* 1999). The argument has been that implementing a particular pre-defined model of the subject under consideration both saves analysis time and maintains consistency of use. Vckovski and Bucher, for example, associated interpolation methods with data sets in the form of metadata, suggesting that their '*VDS* [Virtual data sets] can serve for many applications without need for conversions and transformations.' However, these arguments run in contradiction to an increasing realisation of the importance of context at a variety of interacting levels of knowledge when developing intelligent information systems (Lovett 1998).

We argue that a simplistic binding of 'appropriate method' to data is fraught with problems, for at least two reasons. Firstly, using GIS data can be aggregated or partitioned easily. This alters their characteristics considerably in a manner that

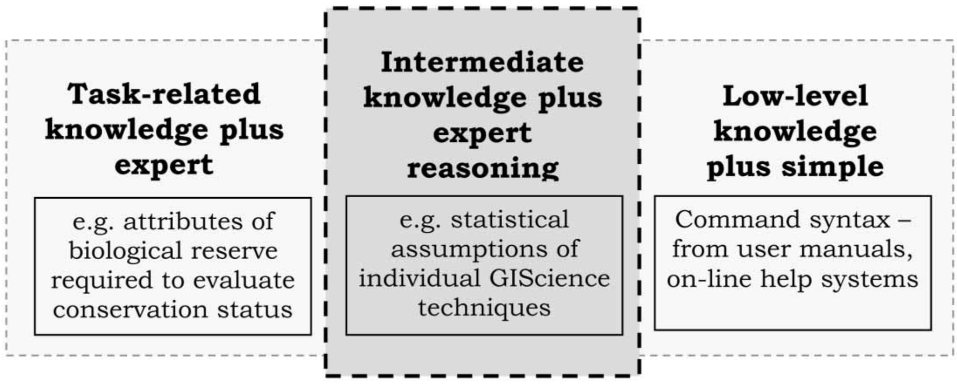

Figure 1. Knowledge and reasoning required when adopting a GIS approach to problem solving ('Intermediate knowledge' concept after Bhavnani and John (2000)).

has the potential to render such encapsulated methods inappropriate. Since the same data set can be re-used for many different purposes, and the appropriateness of different spatial analysis methods varies according to purpose, metadata needs to be re-evaluated for each occasion. Encoding intermediate knowledge as metadata restricts our ability to choose methods according to each purpose or domain.

Secondly, early object oriented research discovered considerable difficulties when attempting to develop standard ontologies for geographical objects. For example, while one person may define a hydrological ontology based on the local watershed, another may conceptualise hydrological thinking in terms of larger drainage basins. Rather, as Fonseca and Egenhofer (1999) note in the context of inter-operability, object profiles need to be constructed dynamically such that multiple ontologies can be mapped to system classes. Research expressing spatial queries in natural language (Shariff *et al.* 1998, Wang 2000) has also revealed the ontological complexity and multiplicity of representations when describing a geographical data set. These findings suggest that an appropriate selection of certain methods of analysis requires a matching with both data set characteristics and purpose, on each occasion.

1.2.3. *Should a user be aware the decision making process, or should this be hidden?*

Rather than *hiding* the complexity of GIS methods from the user, as has previously been suggested of intelligent GIS, we advocate a shift towards supporting the client to learn to use GIS software more appropriately. As with later decision support systems, it treats the user as the final decision maker, acknowledging that superior thinking and intuition are not, as yet, realistic components of a GIS.

Pragmatically, this 'supportive' philosophy allows the development of a number of function-related modules that are not bound to a particular GI system but, rather, can be used in support of many. This module-by-module approach reflects the finding that artificial intelligence works best where the domain of the application can be tightly specified. It is also more congruent with the recent modularisation of software, and some increase in the accessibility of individual GIS functional components, by some vendors of GIS (Sondheim *et al.* 1999).

Exploring the means by which the knowledge of GIScience specialists (intermediate knowledge) can be captured and used to assist applied users of GIS forms the rationale for this paper. The requirement to have a confined domain for AI methods encapsulating knowledge suggests that the example task should be focused upon a particular class of GIS functions. The choice of appropriate methods for interpolating point-based observations of environmental variables to create continuous representations at the accuracy required for modelling or data visualisation can be particularly time consuming and depends greatly on the data set. For many applied users, interpolation is a preliminary task, not the sole purpose of their analysis using GIS. Given this situation, we explore the advances necessary to assist such users by designing and prototyping an 'intelligent' module that sits loosely between task and tool, using interpolation as an example class of GIS function. That is, we express intermediate knowledge in relation to the functionality of a GIS rather than as part of a data model. This approach assumes that a user is able to specify, in broad terms, what they wish to do with GIS to achieve their purpose. The properties of the actual data set may mediate the final selection of a method, and the user can be assisted in querying the data using GIS or standard statistics to explore and to understand this. In order to re-create the flexibility of a human expert, the module draws initially and influentially on the cognitive knowledge of the user and a theoretical rule base, but balances this with hidden, low-level and statistical data analyses (figure 2).

2. The task of interpolation

To focus our analysis, we examine the common scenario of a GIS user who wishes to construct continuous surfaces from scattered point observations. The fact that there is no universally 'best' interpolation method poses a potential problem, despite a number of useful overviews of different methods (Lam 1983, Burrough and McDonnell 1998, Mitás and Mitásovà 1999). This situation arises since the 'best' method varies according to the purpose envisaged of the interpolated surface and the unique characteristics of a particular data set. Moreover, the largely separate development of different groups of interpolation techniques (e.g. kriging and splining) has resulted in a historically dispersed and poorly connected literature base.

Expert geostatisticians know that poor results arise, for example, where ordinary kriging is applied with highly skewed or bimodal data or when a variogram cannot be modelled realistically. Similarly, the use of high order trend surfaces in x and y is undesirable where the trend can more simply be eliminated by the inclusion of collateral data such as elevation. Choosing an appropriate interpolator is important, both for maintaining visual realism and as part of a strategy to control error. Inaccuracies arising through the use of a less appropriate interpolator may subsequently propagate throughout subsequent modelling applications (Burrough 1992, Heuvelink 1998, Jarvis and Stuart 2001b).

Incorporating intelligence within interpolation has attracted some limited attention over the past fifteen years (table 1). The earliest work (Maslyn 1987, Dutton-Marion 1988,) incorporated a rule–based rationale for the choice of method. Necessarily therefore, the domain of interest required focus, in these cases on geological applications (Maslyn 1987, Dutton-Marion 1988) or a sub-class of interpolation functions (Dimitrakopoulos 1993). However, the use of rule-based methods alone places arguably too high a burden of interaction and prior exploration of data upon the user. Within table 1, a shift may be identified in more recent years towards the use of intensive statistical analysis rather than rule-based diagnostic methods to support the choice of interpolator. For example, an 'Extended' Exploratory Data Analysis (EEDA)' method for selecting an appropriate interpolator is presented by Bucher (1998). A data set is characterised statistically and an appropriate interpolator then selected using the statistics to evaluate a structured rule-set.

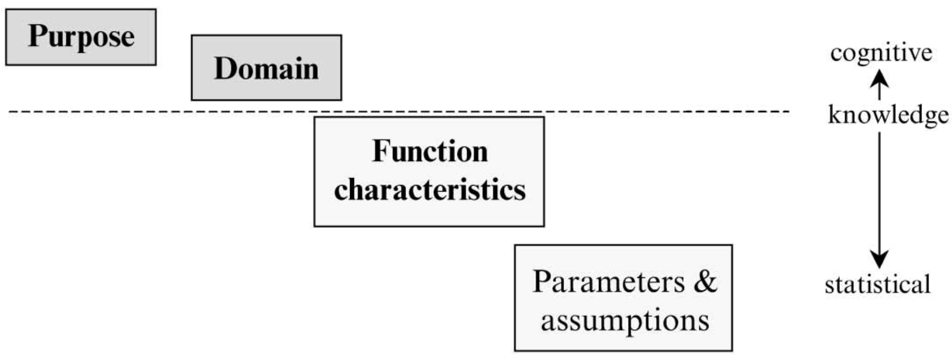

Figure 2. Elements of knowledge: from cognitive to statistical.

Table 1. Interpolation with 'intelligence': previous research.

Reference	Rule-led	Statistics-led		Interpolation methods	Status
Bucher (1998)		Yes		Kriging family	Design
Bucher and Vckovski (1995)				Kriging family	
Maslyn (1987)	Yes		Mining/geology	Polynomial, kriging, double fourier, inverse-distance squared, polynomial trend surface	Implementation
Dutton-Marion (1988)	Yes		Geological contouring	Trend surface analysis, kriging family, smoothing splines, triangulation, distance weighted average	Design
Dimitrakopoulos (1993)	Yes		Mining/geology	Kriging family	Design

We propose a flexible method of assisted interpolation which combines the findings of previous researchers, acknowledging that a simple rule based approach (Maslyn 1987, Dimitrakopoulos 1993) is computationally efficient for selecting methods within a specified domain, that exploratory analysis of the data is an essential pre-requisite to making an informed choice (Bucher 1998) and that the ability to meet a user's accuracy requirements is a further factor that often influences the choice of method (Heuvelink *et al.* 1989). In our approach, users are guided through the choice of process, allowing them to express their knowledge of the data and any expected relationships while being helped to answer questions of a more complex spatial nature through a facility to interrogate their underlying data or see explanatory examples where necessary (table 2).

3. Module design

In designing the 'intelligent' module, we found it useful to adopt some principles of scenario-based software design originally intended to assist with human-computer interactions (Carroll 2000). Designing for particular scenarios of use helps one to understand and formalise an activity (in this case interpolation) and emphasises the value of the prototype as an initial, usable tool for allowing a process of reflection and learning that can iteratively improve the design.

In our case, we evolved a set of initially limited interpolation scenarios where

Table 2. Statistical diagnoses fired by rules within knowledge net.

Method evaluated	Associated statistical diagnoses conducted
General	Trend
Kriging	Spatial correlation, normality of data, anisotropy (see Dimitrakopoulos (1993) and Bucher (1998) for examples)
Partial thin plate splines	Spatial correlation, linearity of trend, order of derivative of spline, normality of data

the term ‘scenario’ captures both the construction of multiple ‘views’ from the problem domain and a limited group of methods known to have been applied to such data (§3.1).

3.1. *Interpolation scenario*

The most effective work in developing ‘intelligent’ spatial analysis to date has focused on common tasks such as line generalisation or land evaluation for which there are a well-defined set of choices (Zhu 1996, Edwardes and Mackaness 2000). We shall define our scenario as the context of providing continuous estimates of primary environmental data sets, as we believe this topic typifies an area of increasing usage of GIS by a wide range of users whose main specialism is not GIScience.

Many models in ecology, agriculture and entomology are primarily driven by meteorological variables such as maximum and minimum temperature, potential evapo-transpiration or rainfall. At the daily temporal scale required, such data are found only at point sources. For an understanding of how the processes being modelled apply throughout the landscape, for example as a precursor to modelling dispersion and movement, it is necessary to interpolate input point data to create continuous, mappable data at a variety of resolutions. Given the large numbers of users of interpolated temperature data in particular, the intelligent module was designed for this initial application.

3.2. *Interpolation methods*

In the examples we present, the user is assisted in selecting between partial thin plate splines (Hutchinson 1991), simple, ordinary and universal kriging (Deutsch and Journel 1998), trend surface analysis that incorporates linear regression and an automatic form of inverse distance weighting. Encoding a wider variety of interpolation algorithms was not considered feasible or essential for illustrating the benefits of this prototype. The selection of techniques was determined by those most often applied previously for interpolating meteorological data, methods of greatest generic utility and those most commonly found in existing GIS. This drew on a comprehensive review of interpolators used both for temperature data at multiple spatial and temporal scales and locations, and more widely within environmental modelling applications (Oliver and Webster 1990, Hutchinson 1995, Jarvis and Stuart 2001a).

3.3. *Module overview*

To help users gain the necessary knowledge to complete their task and reduce the possibility of methodological error, we demonstrate an adaptive knowledge base that combines both infometric, or cognitive, knowledge (§3.4.1) with statistical knowledge (§3.4.2). Our approach uses a rule base that merges task related knowledge obtained by requests made to the user with both expert procedural knowledge regarding the interpolation selection process and knowledge about the data set obtained by automatic diagnostic analysis of spatial data. In this case, the cognitive knowledge has been extracted from the literature and verified by users experienced in the subject of interpolation. This knowledge is supplemented and made context specific by requesting data from the user regarding the particular application task. The network then triggers exploratory diagnostics *when required by the rule base*. That is, some rules are evaluated using results from subservient statistical diagnostics. The module therefore becomes more than a standard expert system that contains a static knowledge base and inference engine (Durkin 1994), since the evidence upon which rules are evaluated is made specific to the context of the analysis and the nature of the given data properties. We summarise the components of knowledge that could ideally be encoded within an 'intelligent' module within figure 3, indicating the initial proportion of rules from the different sources that contribute to the rule base.

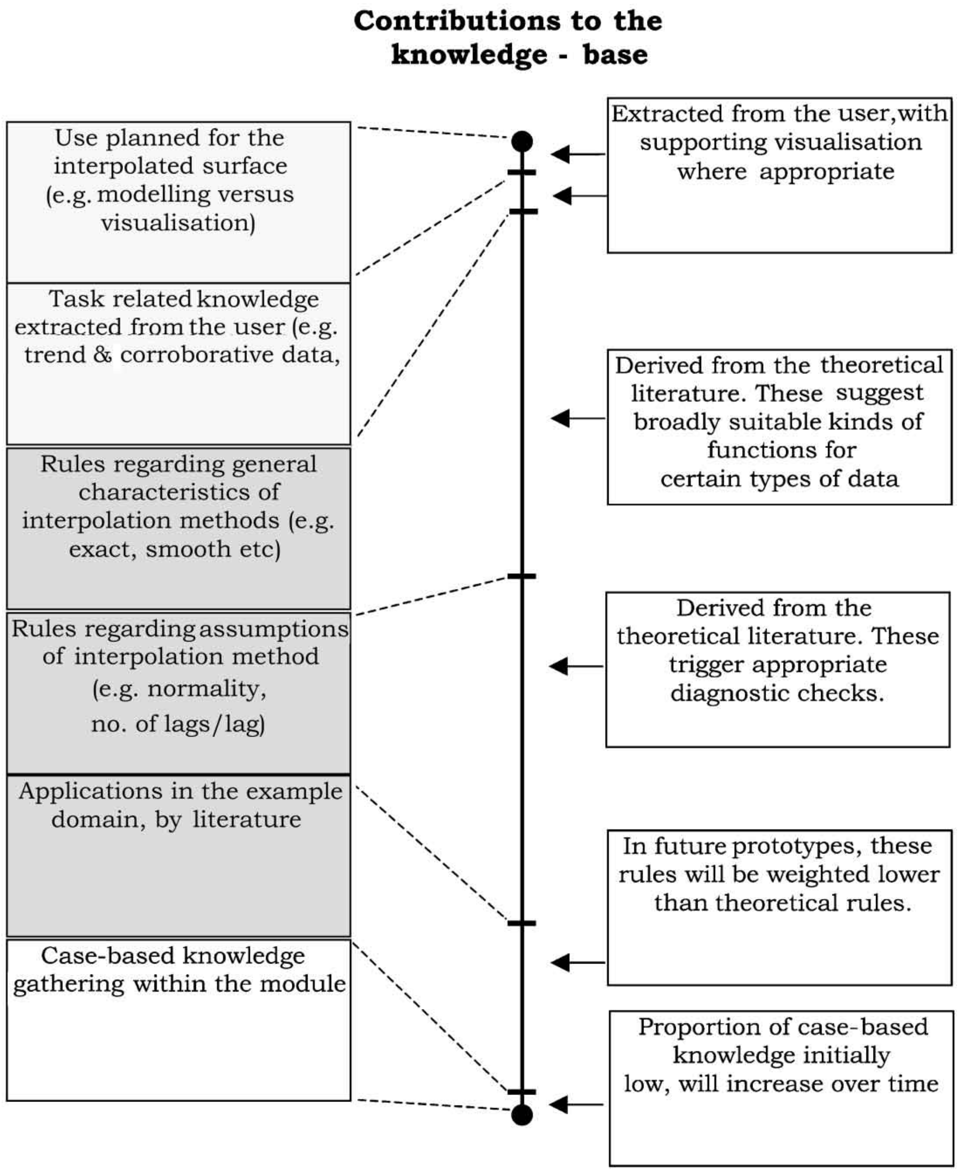

Figure 3. Infometric diagnostics: components within the knowledge network, by proportion of rules.

3.4. *Module diagnostics*

3.4.1. *Infometrics*

We use the term infometrics to include both information about the reasoning process and qualitative knowledge about associations in the data (e.g. temperature and elevation are linearly related). Both of these cognitive principles are used to structure the knowledge base.

- *Expectations and uses of the interpolated data*

When considering how the hierarchy of rules should be structured, we chose to emphasise the responses from the user regarding their use for the data. This is because we believe the 'acceptability' of an interpolated surface can differ markedly, depending for instance on whether the surface is intended for use in further modelling, or the purposes of visualisation, as demonstrated by Declercq (1996). Additionally, scientists using an intelligent module are likely to have expectations for their interpolated surface, and potentially an understanding of information that may be used to guide the process of interpolation. For example, prior knowledge of whether a surface is expected to be smooth or rough, whether there is known trend in the data or whether corroborative data is available (e.g. a linear relationship between elevation and temperature) will often influence which interpolation methods are first investigated. A user may also have preferences as to whether the values of the interpolated surface should always remain within those of the original data, or occasionally exceed them. To assist the user in understanding the questions asked, we provide visual support relating to the issues raised (figure 5). Overall, the proportion of questions asked of the user compared to the number of rules evaluated using statistical metrics is intended to be small, but the user's responses will strongly influence the order in which the knowledge-based diagnosis and subsequent statistical analyses will be carried out.

- *Technical rules regarding interpolation methods*

Given the large volume of literature covering both the theory and application of interpolation methods, the expertise to build the inference engine for this project was taken from previous research papers and books, together with practical knowledge gained from previous research projects. The use of the literature base was important since expertise is scattered worldwide, often fragmented in coverage and difficult to cross-relate. Most researcher theoreticians for example have specialised in using either spline or kriging methods, but rarely both, despite their acknowledged similarities (Hutchinson 1993).

The cognitive knowledge about interpolation tasks and operations was encapsulated by dividing the rule base firstly into rules describing the general characteristics of interpolators (Rule type 1), and secondly into rules about the specific assumptions of the interpolators that need to be met by a data set (Rule type 2) (figure 3). For example, splines are generally considered to smooth data (Rule type 1), and the expert user of partial thin plate splines would typically need to establish that the fit of the data to the spline surface was statistically valid (Rule type 2), for example using the 'trace diagnostic' (Hutchinson and Gessler 1994) or similar. Rules falling into these two categories form the bulk of the knowledge within the intelligent module. In general, type 1 rules are evaluated through direct questions to the user, while type 2 rules are evaluated 'behind the scenes'.

For completeness, figure 3 also refers to a third main group of rules, the case-based/experience-based group. Establishing a methodology to encompass the abstraction and encoding of these rules is an area for future research.

3.4.2. *Statistics*

Because of the inherent assumptions behind the many interpolation algorithms, each method performs differently for data sets with different properties. As Burrough (1986) noted 'It is unwise to throw one's data into the first available interpolation technique without carefully considering how the results will be affected by the assumptions inherent in the method'. Regardless of which exploratory techniques are used, 'the time taken to explore, understand, and describe the data set should be amply regarded' (Isaaks and Srivastava 1989). Often, a user is unaware of the characteristics of the input data being offered for interpolation.

Subservient to the infometrics, but critical to success, statistical methods trigger only when required by the structured rule set to assist with the characterisation of the data set. The results of the statistics can inform the choice of the most suitable interpolation method, and may be used further to carry out the necessary pre-processing of a data set prior to its interpolation.

The set of statistical methods used for diagnosing the appropriateness or otherwise of a data set for certain methods of interpolation have been chosen by analysing the work of experienced GIS users. For example, it is well known that different interpolation methods will perform better or worse depending on the spatial association measurable within the data. As Bucher (1998) identifies, exploring the data set for this characteristic is often a paramount consideration. Additionally, testing for stationarity, and subsequently for spatial association is a principle that goes beyond good geostatistical practice. The identification and where necessary extraction of 'trend' from a data set has been shown to be a basic step that significantly improves the accuracy of many interpolation methods (Jarvis and Stuart 2001a).

Rather than attempting to enumerate or rank all feasible alternatives to the interpolation problem, following the diagnosis, we propose that the user is advised of one or two interpolation methods most likely to satisfy their requirements. Additionally, where an interpolation method is clearly unsatisfactory for the task, this is also identified. As Cameron and Abel (1997) note, enumerating and ranking all alternatives to a decision making process is impractical, but providing tangible boundaries to the set of likely solutions provides what may be termed a 'satisficing' result. Once the main type of method has been diagnosed, the system can also assist the user in setting any parameters (e.g. decay parameter for IDW, variogram model and parameters) that may be required to apply a particular interpolation method (§3.5).

3.5. *Parameter setting*

Following the initial phase in which the most 'appropriate' methods are diagnosed, further statistics and data preparation are often required to ensure that the subsequent interpolation process is as straightforward as possible for the user. Not only the choice of method, but also its parameter settings, can be critical if one is to avoid misleading results (Hodgson 1993). The majority of GIS leave the selection of parameters for interpolators entirely to the user. Wherever possible, we presently approach the parameter setting problem as an optimisation problem, using statistical model selection procedures to find numerically acceptable values for parameters such as the variogram range. Further development of a more explicitly AI-based approach to model selection might include model fitting using techniques such as genetic algorithms (table 3).

3.6. *Teaching tool*

In its present form, the intelligent module acts primarily as an interpolation assistant. Since expert systems allow the rules that led to a certain decision being reached to be provided back to the user, the 'intelligent interpolator' has the potential to be developed into a learning or tutoring tool. The reporting of rules that have been fired has two roles. Initially it provides a means by which an expert may verify the actions of the intelligent module, and subsequently it allows the less experienced user to learn the questions that they should be posing and evaluating to make an appropriate choice of method. Work developing embedded training systems for complex information systems develops this rule analysis further by investigating the 'gulf of execution' between the expected workflow and that followed by the trainee (Cheikes *et al.* 1998).

4. Implementation of a prototype intelligent module

To this point, we have considered broadly what is to be produced, and the types of knowledge that will be incorporated. We now report on the implementation of a prototype 'intelligent interpolation' module, focusing upon the management of cognitive knowledge and the communication of the diagnostic process to the user.

We considered that a stand-alone module could better serve users of the many GIS systems (Arc-Info, Erdas GIS, Arc View, Info-Map, Spans, Genamap, etc.) and statistical packages (Minitab, SPSS, S-Plus) currently commercially available. This was because:

- Each package implements a different subset of interpolation algorithms;
- Each package has different scripting or interfacing capabilities such that a closely coupled module would only serve a subset of users unless re-implemented multiple times;
- There is general lack of efficient macro-language capabilities in commercial GIS packages, forcing at least some measure of external coupling between an expert-system shell and a GIS during the implementation process.

4.1. *Software environment*

A combination of Java and the Jess expert system shell were used to implement the prototype software. *Java*© from Sun is a platform independent object orientated language, and *Jess*© from E. J Freidman Hill (Sandia Corporation) is an expert system shell written in Java and based on the widely established *Clips* expert system. Java was chosen for its ease of use across different operating systems and because it

Table 3. Parameters required of the user for interpolation tasks by proprietary GIS and statistical tools.

Method	Associated parameters required by other GIS/Statistical tools
General	Trend element (e.g. regression parameter, order of xy trend)
Kriging	Variogram model type, variogram parameters, anisotropy, number of neighbourhood points
Inverse distance weighting	Number of neighbourhood points, decay parameter

was suitable for developing an explanation facility, for calling external programs (interpolation algorithms) and for making a graphical user interface (§4.3). Jess was chosen as it accommodates a rule based approach, so facilitating the encoding of the knowledge relatively easily, and incorporates an inference engine.

4.2. *Knowledge management*

4.2.1. *Knowledge acquisition*

We term the approach that we used for knowledge acquisition a 'Phase Teaching' technique. This concept works on the basis of phased refining of the knowledge gained from interviewing experts and from literature research. The phase teaching method is composed of two phases, the first being a broad and shallow examination and gathering of the intelligent (expert) knowledge to construct a fast prototype. At this stage, broad background reading on interpolation literature gives a solid understanding of the problem at hand and helps to clarify the different terminology used in the domain. This is then transferred to the rule base of the intelligent system and a fast prototype developed. The second phase is 'hands-on' interaction with the prototype and refinement of the rule base and system, following more detailed reading and further 'in-depth' interviews with experts. Binary decision trees were used to structure this knowledge refinement process, based on their ability to express knowledge in a formalism that is often easier to interpret by experts and ordinary users (Skidmore *et al.* 1996, Lagacherie and Holmes 1997, Janikow 1998). In effect, this process became a form of what is referred to in the AI literature as a 'cognitive task analysis' (Lovett 1998).

4.2.2. *Structuring knowledge*

In considering a control strategy by which the knowledge is structured and interpreted, a review of the various tools for building intelligent (expert) systems was necessary. The majority of these are expert system shells, high level programming languages and mixed programming environments. A rule-based system was preferred for this research as it best matches the methodology of deciding an interpolation method, by capturing 'the global problem solving approach used by the expert' (Durkin 1994, p. 627).

As mentioned in §4.2.1, decision trees were used for knowledge elucidation, encoding and refinement. A decision tree structure, in this case a binary structure, was also used as a mechanism for knowledge control. This choice was made since decision trees represent a logical reasoning system, which derive sound conclusions from formal declarative knowledge. Additionally, the shortest route to making a decision is always taken. For example, when following a path down a tree, rules that are not relevant to the case in hand are effectively bypassed. Finally, the resulting models are easy to understand, as they can be directly expressed as a set of IF … THEN rules.

A binary decision tree is one where each node of the tree has only two transition branches and are typically used to implement knowledge based on a sequence of yes/no questions. Each decision node has associated with it a question and an answer node (Giarratano 1998). Inside the nodes, some decision occurs that transfers control via any of its branches to other nodes or leaves. In this type of structure, the inference process starts at the top of the decision tree, i.e. the root and follows one and only one of the sub nodes, either a [yes] or [no] branch. The heuristics dictate that one question leads to consequences that in turn lead to other questions until a conclusion is arrived at an answer node. This approach allowed us to emphasise the purpose for which the interpolated surface was intended, and the user's knowledge of their domain, by placing these rules near the root of the tree.

For example, in figure 4, the root question is 'Do you wish to interpolate a surface for the purpose of visualisation?' Questions relating to the expected nature of the surface follow, and finally the analytical rules or statistical tests that are evaluated directly upon the data. In the figure, user-driven rules have been placed in rectangular boxes, and analytical rules/tests that are evaluated directly from the data have been placed within oval units. The decision tree structure required a depth of eight levels before a recommendation of feasible/infeasible methods from the eleven variant techniques of the interpolation families considered could be determined. From this point, selection of the most optimal of the target methods could be made following careful parameter (e.g. inverse distance rule) and structure setting (e.g. order of the spline, variogram model) using mathematical model selection methods in combination with visualisation.

4.3. *User interaction*

Given the growing diversity of user populations and application domains of GIS, different levels of assistance to answer the questions and feedback from these are provided by the module as the diagnostic process progresses. Yet, as Egenhofer (1997) notes, a serious disadvantage of textual questioning implicated in figure 4 'is that it forces users to translate a spatial image they may have in their minds about

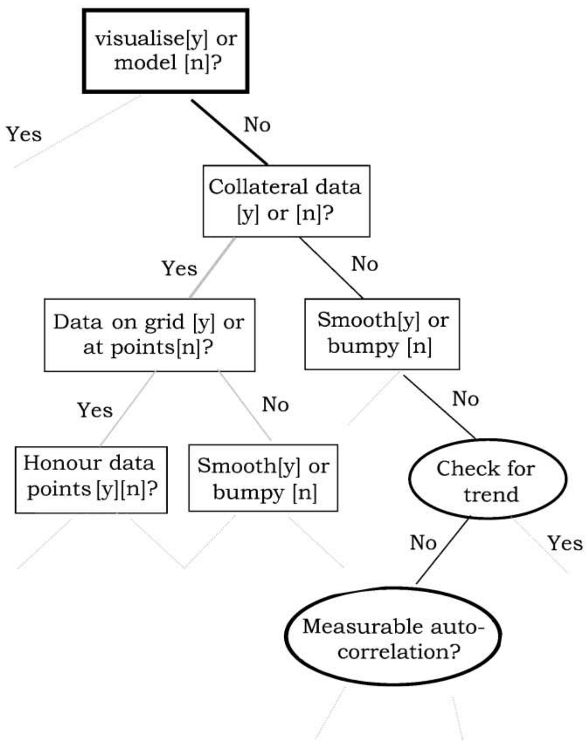

Figure 4. Selected upper-level decision nodes for interpolation from the eight-level decision tree.

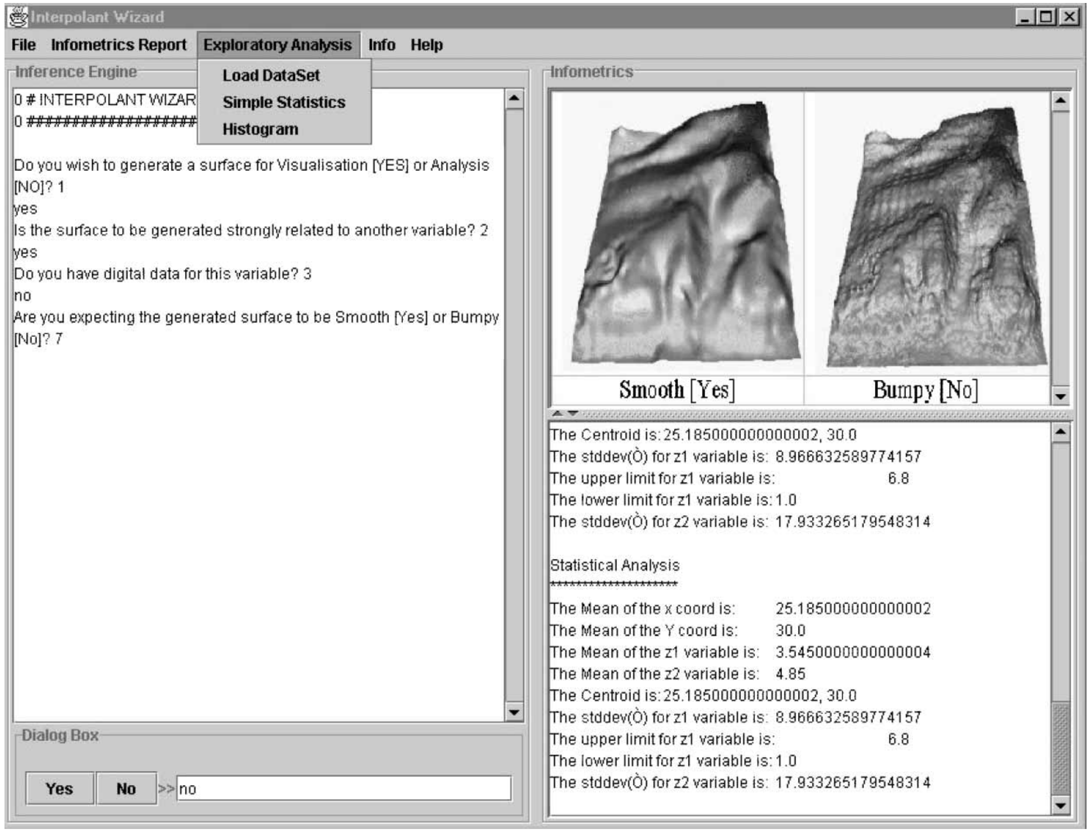

Figure 5. Prototype user interface.

the situation they are interested in, into a non-spatial language'. Multi-modal communication, the use of complementary media to communicate the same idea in parallel cognitive 'channels', was therefore adopted in specific instances to support users to understand more fully the spatial context of the questions posed. Within the software interface (figure 5), the interplay between the left hand 'text-questioning' panel and right hand statistical and visual panels follows this concept. The upper area is used to display visualisations that support the user in understanding the questions asked in the left hand pane, on a rule-by-rule basis.

In the example of figure 5, the images are bound statically with the appropriate rule, and represent abstract notions of the degree of spatial continuity. Other images are used to convey what is meant by 'exact' interpolation, the averaging or extrapolating tendencies of different techniques and the typical shapes of different variogram models. Following ideas from the literature on exploratory data analysis in GIS and geostatistics (Gunnink and Burrough 1996, Pannatier 1996), a further set of dynamic 'exploratory' images are constructed by the software. These are used for example to compare the actual distribution of the user's data with a normal distribution, provide experimental variograms in different directions to assist with the exploration of isotropy, and to allow the verification of modelled variograms against experimental variograms in the case of kriging-related questions. The images are intended to assist the user in responding to particular questions as (or if) they arise. For the less experienced user, these images may be the most important first step in considering the many facets of spatial data and also the reasons for choosing a particular interpolator.

The lower pane is used to report the results of the statistical tests that the module carries out that contribute to the knowledge base but which the user is otherwise unaware of. These results support users with more expertise, and for the most advanced user might form the basis for choosing a method in their own right. This pane also reports, at the end, a bounded solution set and suggested parameters (e.g. variogram model, distance decay parameter, order of trend) that the user could apply using stand-alone packages. On request, this pane may subsequently be used to report which rules have been fired and why. This provides information on how decisions are reached by interpolation experts from which the less experienced user may learn.

5. Discussion

The decision to implement the intelligent interpolation module in a combination of Java and Jess languages allowed the system to be prototyped more rapidly. The use of Java also provided a means for distributing the module over the Internet for use by multiple users. The use of a decision-tree method both for knowledge capture and structuring is expected to allow the future expansion of the system to more diverse domains with the development of additional object oriented classes in Java. The modularity of rules, the separation of control from knowledge and the ability provided by Jess to check for consistency will be beneficial if the module is developed for further techniques and in other domains. While the use of decision trees to structure the approach facilitated the identification of component tasks and ordering of processes, the method also forced the experts to attribute ranked levels of significance to individual tasks. This top-down approach has the advantage of avoiding computer intensive analyses to determine matters that are of low priority and relative value to the overall problem. However, the relative importance of some conditions or assumptions may not in reality be differentiable while the linearity imposed on the model of diagnosis prevents recursion. Furthermore, the use of binary decision trees in particular limits our ability to weight rules, for example according to the level of experience of the expert or the perceived strength of rules derived from literature. Additionally, decision trees do not allow node sharing, which reduced the compactness of the structure. Future work may therefore focus on the use of more flexible structures such as for example Petri nets (Peleg *et al.* 2002), Bayesian networks (Brezillon *et al.* 2000) or contextual graphs (Brezillon *et al.* 2000) as more powerful means of encoding knowledge and appropriate workflows.

While verification has been defined as '... the process of ensuring that the intelligent system conforms to specifications, and that its knowledge base is consistent and complete within itself' (Gonzalez and Barr 2000), we view the verification process rather more broadly. We considered whether the structure and content of the rules conformed to those expected by local experts, an important process since automatic techniques to check if a knowledge base is accurate and complete are in practice not very useful. This evaluation initially involved checks and refinements of the individual rules, their semantics and ordering by the local experts from whom knowledge was gleaned. Secondly, by using both the decision-tree presentation and through 'hands-on' experience with the prototype, experts independently identified whether the rules and structures implemented were consistent with their experience and minor differences of opinion were resolved. The modular implementation of the rules also enabled those with knowledge of individual domains of interpolation to assess sub-components of the knowledge base and structure. Finally, the prototype was assessed informally by GIS users with little experience with interpolation to determine the level of usability of graphical and semantic aspects of the software that we considered would influence its future uptake. Further details regarding this verification and the subsequent validation process, where the module was applied to data sets on which extensive interpolation experiments had previously been performed, are reported in Cooper and Jarvis (2002).

The combination of infometric and statistical diagnostics within this knowledge-based module provides a flexibility that goes beyond a typically brittle standard rule base, allowing the capacity for rules to be applied to interpolation in further application domains, with different data characteristics and needs of users. Meeting diverse application needs is a particular problem in more traditional knowledge–based systems. We have adopted a scenario-based approach when developing this prototype module, both to ensure that the rules are not over-stretched and also to make the system easier to verify. Rather than hiding the complexity of the interpolation methods from the user, as has previously been suggested of intelligent GIS, we advocate a philosophical shift towards supporting the client to use GIS software. The rational behind this shift is to instil the important scientific principles of 'know your data', and to support the wisdom of 'parameterising your methods'. Extending the provision of information from statistical and non-interactive reports to incorporate interactive ESDA-style multiple 'views' of the data should assist communication with the user, and enhance their ability to provide an informed response.

The intelligent module, intended to sit between 'tool' and 'task', may be valuable for users of multiple GI systems, therefore extending its potential scope to a wider range of interpolation tasks than would be achievable if a closely coupled design were to have been adopted. Loose coupling also affords other benefits. For example, the wisdom of automatically fitting variogram models meets with some caution within the geostatistical literature (Webster and Oliver 1990, Deutsch and Journel 1998), and the de-coupled design of the module allows the possibility of interactive comparisons with, for example Variowin (Pannatier 1996) or similar tools.

Currently, the intelligent module draws most heavily on the first category of knowledge, the theoretical characteristics of different interpolators. The challenge is how to encode application-specific information, which is both qualitative and quantitative. Applied approaches reported in the literature may be biased by availability of software or the manner in which methods were implemented, and the information reported may be considered 'incomplete' if too complex a matching process is designed. Issues regarding knowledge re-coding and means to overcome representational rigidity need to be attended to in future work expanding the rule base.

There is no lack of books and papers on the subject of interpolation, but there are few consolidated accounts of how and when to apply the guidelines they present. Moreover, the number of statistical analyses that should be undertaken if an interpolation problem is to be approached well is high, and commonly involves multiple software packages. The intelligent interpolator module addresses these problems through the application of infometric and statistical diagnostics that are combined within a knowledge-based reasoning tool. Critically, the use of both types of diagnostics draws on the wider realisation of the importance of context at a variety of levels when designing flexible intelligent systems (Bainbridge 1997, Pomerol and Brezillon 1999).

While many examples of artificial intelligence methods applied within particular GIScience applications may now be found (Artificial intelligence in GIScience), we argue that it is rarer to find artificial intelligence used to support new users to perform an appropriate sequence of tasks ('workflow') and avoid pitfalls with basic GIS functionality (Artificially-intelligent GISystems). Interpolation provides an example of just one of several domains within GIScience, for example error propagation (Ehlschlager 2002), where complex sequences of analyses are required of the user. In a broader context of the emerging market for GI web processing services across the Grid (Foster and Kesselman 1999), the selection of workflow tools according to both established practice, the particular data set and goal of analysis is an approach that provides the consistency and quality control that will be required for distributed spatial analysis tasks.

6. Conclusions

This article reviewed some means to assist users who may lack intermediate knowledge about GIScience, to select appropriate methods when they apply GIS tools to tackle problems in their application domain. AI methods were found to be used mainly as substitutes for existing, often statistical approaches in the GIS toolbox. There were fewer examples of AI methods being used to assist the user in choosing methods that were appropriate to their purpose and to the nature of the data sets at hand, yet this is precisely where knowledge may be lacking amongst an expanding group of applied users of GIS.

By designing and implementing a simple prototype system to assist users in choosing appropriate methods for the interpolation of spatial data, many of the main conceptual challenges in applying rule-based methods to a type of geographical problem solving were revealed. On the problem of where to encode knowledge, we conclude that because there will always be a balance between the purpose of the analysis as well as the fixed characteristics of a given data set, knowledge cannot only be encoded with the data set as metadata. Rather, we preferred to encode knowledge as a network of rules, some of which related to the data set characteristics, but others which related to the specific purpose of the analysis and to previous theoretical or empirical knowledge about the validity of different methods. In a pilot implementation, this knowledge network was implemented as a binary decision tree using a particular software shell that, whilst illustrating concepts had some limitations that would need to be overcome for an operational system.

The implementation confirmed the need for the assistant software to be a separate from any specific GIS and for it to support interaction with the user through multiple channels of communication. The ability of the prototype to help users learn about their data and the functions they can sensibly apply, suggests that such systems can contribute to the goal of developing GIS with at least some intelligence that help researchers in other disciplines to apply GIScience tools and techniques appropriately.

References

Andrienko, G. M., and Andrienko, N. V., 1997, Intelligent cartographic visualization for supporting data exploration in the IRIS system. *Programming and Computer Software*, **23**, 268–281.

Anselin, L., 1989, *What is special about spatial data?* NCGIA Technical Report 89/4, National Centre for Geographical Information and Analysis, Santa Barbara.

Bainbridge, L., 1997, The change in concepts needed to account for human behaviour in complex dynamic tasks. *IEEE Transactions on Systems, Man and Cybernetics*, **27**, 351–359.

Bhavnani, S. K., and John, B. E., 2000, The strategic use of complex computer systems. *Human–Computer Interaction*, **15**, 107–137.

Brezillon, P., Pasquier, L., and Pomerol, J.-C., 2000, Representing operational knowledge by contextual graphs. *Lecture Notes in Computer Science* **LNAI 1952**, 245–258.

Bucher, F., 1998, Using extended exploratory data analysis for the selection of an appropriate interpolation model. *Geographic Information Research: Trans-Atlantic Perspectives,* edited by M. Craglia and H.Onsrud (London: Taylor & Francis), pp. 391–403.

Bucher, F., and Vckovski, A., 1995, Improving the selection of appropriate spatial interpolation methods. *Lecture Notes in Computer Science*, **988**, 351–364.

Burrough, P. A. 1986, *Principles of Geographical Information Systems for Land Resources Assessment* (Oxford: Clarendon Press).

Burrough, P. A., 1992, Development of intelligent geographical information systems. *International Journal of Geographical Information Systems*, **6**, 1–11.

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

Câmara, A. S., Ferreira, F., and Castro, P., 1996, Spatial simulation modelling. *Spatial analytical perspectives on GIS*, edited by M. Fischer, S. H. J. Unwin and D. Unwin (London: Taylor & Francis), pp. 201–212.

Cameron, M. A., and Abel, D. J., 1997, A problem model for spatial decision support systems. In *Proceedings of the Seventh International Symposium on Spatial Data Handling*, edited by M.-J. Kraak, M. Molenaar and E. M. Fendel (Delft: Taylor & Francis), pp. 89–99.

Carroll, J. M., 2000, *Making Use: Scenario-Based Design of Human–Computer Interactions* (Cambridge, MA: The MIT Press).

Cheesman, J., and Petch, J., 1999, Interpolation of severely non-linear spatial systems with missing data: using kriging and neural networks to model precipitation in upland areas. In *Geographic Information Research: Trans-Atlantic Perspectives*, edited by M. Craglia and H. Onsrud (London: Taylor and Francis), pp. 175–188.

Cheikes, B. A., Geier, M., Hyland, R., Linton, F., Rodi, L., and Schaeffer, H.-P., 1998, Embedded training for complex information systems. In *Intelligent Tutoring Systems*, edited by B. P. Goettl, H. M. Halff, C. L. Redfield, and V. J. Shute, Lecture Notes in Computer Science **LNCS 1452** (Berlin: Springer Verlag), pp. 36–45.

Cooper, W., and Jarvis, C. H., 2002, Selecting an interpolation method: supporting the user with an intelligent system. In *Proceedings of GIS Research UK (GISRUK) 2002*, edited by S. Wise (Sheffield, UK: University of Sheffield).

Cowen, D. J., and Shirley, W. L., 1991, Integrated planning information systems. *Geographical Information Systems*, edited by D. Maguire, M. F. Goodchild and D. Rhind (London: Longman), pp. 297–310.

Dawson, C. W., and Wilby, R. L., 2001, Hydrological modelling using artificial neural networks. *Progress in Physical Geography*, **25**, 80–108.

Declercq, F. A. N., 1996, Interpolation methods for scattered sample data: accuracy, spatial patterns, processing time. *Cartography and Geographical Information Systems*, **23**, 128–144.

Densham, P. J., and Goodchild, M. F., 1989, Spatial decision support systems: a research agenda. *Proceedings of GIS/LIS'89* (Virginia: ACSM/ASPRS/AAG), pp. 707–716.

Deutsch, C. V., and Journel, A. G., 1998, *GSLIB Geostatistical Software Library and User's Guide* (Oxford: Oxford University Press).

Dimitrakopoulos, R., 1993, Artificially intelligent geostatistics: a framework for accommodating qualitative knowledge-information. *Mathematical Geology*, **25**, 261–279.

Durkin, J., 1994, *Expert systems: Design and Development* (New Jersey: Prentice-Hall).

Dutton-Marion, K. E., 1988, Principles of interpolation procedures in the display and analysis of spatial data: a comparative analysis of conceptual and computer contouring. PhD Thesis, University of Calgary, Calgary.

Edwardes, A., and Mackaness, W. A., 2000, Modelling knowledge for generalisation of categorical maps. *GIS and GeoComputation: Innovations in GIS 7*, edited by P. M. Atkinson and D. Martin (London: Taylor & Francis), pp. 161–173.

Egenhofer, M. J., 1997, Multi-modal spatial querying. In *Proceedings of the Seventh International Symposium on Spatial Data Handling*, edited by M.-J. Kraak, M. Molenaar and E. M. Fendel (Delft: Taylor & Francis), pp. 785–799.

Ehlschlager, C. R., 2002, Can we construct error buttons for every application? In *Proceedings of 5th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences (Accuracy 2002)*, edited by G. Hunter and K. Lowell (Melbourne, Australia: University of Melbourne), pp. 134–144.

Ellner, S. P., and Seifu, Y., 2002, Using spatial statistics to select model complexity. *Journal of Computational and Graphical Statistics*, **11**, 348–369.

Fischer, M. M., and Reismann, M., 1999, Parameter estimation in neural spatial interaction modelling by a derivative free globalization method. In *Proceedings of the 4th International Conference on Geocomputation*, Virginia: CD-ROM.

Fischer, M. M., and Nijkamp, P., 1992, Geographic information systems and spatial analysis. *Annals of Regional Science*, **26**, 3–17.

Fonseca, F. T., and Egenhofer, M. J., 1999, Ontology-driven geographic information systems. *7th ACM Symposium on Advances in Geographic Information Systems*, Kansas City, http://www.spatial.maine.edu/~max/ACMGIS99.pdf.

Foster, I., and Kesselman, C. (editors), 1999, *The Grid: Blueprint for a New Computing Infrastructure* (New York: Morgan Kaufmann).

Giarratano, J. C., 1998, *Expert Systems: Principles and Programming* (Boston: PWS).

Gonzalez, A., and Barr, V., 2000, Validation and verification of intelligent systems—what are they and how are they different. *Journal of Experimental and Theoretical Artificial Intelligence*, **12**, 407–420.

Gunnink, J. L., and Burrough, P. A., 1996, Interactive spatial analysis of soil attribute patterns using exploratory data analysis (EDA) and GIS. In *Spatial Analytical Perspectives on GIS*, edited by M. Fischer, H. J. Scholten and D. Unwin (London: Taylor & Francis), pp. 87–110.

Heuvelink, G. B. M., (1998) *Error propagation in environmental modelling* (London: Taylor & Francis).

Heuvelink, G. B. M., Burrough, P. A., and Stein, A., 1989, Propagation of errors in spatial modelling with GIS. *International Journal of Geographical Information Systems*, **3**, 303–322.

HEWITSON, B. C., and CRANE, R. G., 1994, *Neural Nets: Applications in Geography* (Dordrecht: Kluwer Academic Publishers).

HUTCHINSON, M. F., 1991, The application of thin plate smoothing splines to continent-wide data assimilation. *Data Assimilation Systems, BMRC Research Report No. 27*, edited by J. D. Jasper (Melbourne: Bureau of Meteorology), pp. 104–113.

HUTCHINSON, M. F., 1993, On thin plate splines and kriging. *Computing and Science in Statistics 25*, edited by M. E. Tarter and M. D. Lock (Berkley: Interface Foundation of North America), pp. 55–62.

HUTCHINSON, M. F., 1995, Interpolating mean rainfall using thin plate splines. *International Journal of Geographical Information Systems*, **9**, 385–404.

HUTCHINSON, M. F., and GESSLER, P. E., 1994, Splines—more than just a smooth interpolator. *Geoderma*, **62**, 45–67.

ISAAKS, E. H., and SRIVASTAVA, R. M., 1989, *An Introduction to Applied Geostatistics* (New York: Oxford University Press).

JANIKOW, C. Z., 1998, Fuzzy decision trees: issues and methods. *IEEE Transactions on Systems, Man and Cybernetics Part B—Cybernetics*, **28**, 1–14.

JARVIS, C. H., and STUART, N., 1996, The sensitivity of a neural network for classifying remotely sensed imagery. *Computers and Geosciences*, **22**, 59–967.

JARVIS, C. H., and STUART, N., 2001a, A comparison between strategies for interpolating maximum and minimum daily air temperatures: b. The interaction between number of guiding variables and the type of interpolation method. *Journal of Applied Meteorology*, **40**, 1075–1084.

JARVIS, C. H., and STUART, N., 2001b, Uncertainties in modelling with time series data: estimating the risk of crop pests throughout the year. *Transactions in GIS*, **5**, 327–343.

LAGACHERIE, P., and HOLMES, S., 1997, Addressing geographical data errors in a classification tree for soil unit prediction. *International Journal of Geographical Information Science* **11**, 183–198.

LAM, N. S., 1983, Spatial interpolation methods: a review. *The American Cartographer*, **10**, 129–149.

LEUNG, Y., and LEUNG, K. S., 1993, An intelligent expert-system shell for knowledge-based geographical information systems. 1. The tools. *International Journal of Geographical Information Systems*, **7**, 189–199.

LEVINSON, R., and Wilkinson, J., 1997, Meta-reasoning for data analysis tool allocation. *Lecture Notes in Computer Science*, **LNCS 1280**, 89–100.

LOVETT, C. M., 1998, Cognitive task analysis in service of intelligent tutoring system design: A case study in statistics. *Lectures in Computer Science*, **LNCS 1452**, 234–243.

MACGILL, J., OPENSHAW, S., and TURTON, I., 1999, Web-based multi-agent spatial analysis tools. In *Proceedings of the 4th International Conference on Geocomputation*, Virginia: CD-ROM.

MASLYN, R. M., 1987, Gridding advisor: an expert system for selecting gridding algorithms. *Geobyte*, **2**, 42–43.

MITÁS, L., and MITÁSOVÀ, H., 1999, Spatial interpolation. *Geographical Information Systems: Principles, Techniques, Applications and Management: Volume 1*, edited by P. A. Longley, M. F. Goodchild, D. J. Maguire and D. W. Rhind (New York: Wiley), pp. 1–492.

MORSE, B., 1987, Expert interface to a geographical information system. *Proceedings of AutoCarto-8*, 535–541.

MURNION, S., and CARVER, S., 1996, Automatic analysis rule generation using genetic algorithms. *Proceedings of the 2nd International Conference on GeoComputation*, http://www.ashville.demon.co.uk/gc1996/.

OLIVER, M. A., and WEBSTER, R., 1990, Kriging: a method of interpolation for geographical information systems. *International Journal of Geographical Information Systems*, **4**, 313–332.

OPENSHAW, S., and ALVANIDES, S., 1999, Applying geocomputation to the analysis of spatial distributions. *Geographical Information Systems: Principles, Techniques, Applications and Management: Volume 1*, edited by P. A. Longley, M. F. Goodchild, D. J. Maguire and D. W. Rhind (New York: Wiley), pp. 267–282.

OPENSHAW, S., and OPENSHAW, C., 1997, *Artificial Intelligence in Geography* (Chichester: Wiley).

PANNATIER, Y., 1996, *VARIOWIN: Software for spatial data analysis in 2D* (New York: Springer-Verlag).

PARADIS, E., BAILLIE, S. R., and SUTHERLAND, W. J., 2002, Modeling large-scale dispersal distances. *Ecological Modelling*, **151**, 279–292.

PELEG, M., YEH, I., and ALTMAN, R. B., 2002, Modelling biological processes using workflow and Petri Net models. *Bioinformatics*, **18**, 825–837.

POMEROL, J.-C., and BREZILLON, P., 1999, Dynamics between contextual knowledge and proceduralized context. In *Modeling and Using Context*, edited by P. Bouquet, L. Serafini, P. Brezillon, M. Benerecetti, and F. Castellini. Lecture Notes in Artificial Intelligence **LNAI 1688** (Berlin: Springer Verlag), pp. 284–295.

RAGG, T., 2002, Bayesian learning and evolutionary parameter optimization. *Ai Communications*, **15**, 61–74.

RIGOL, J. P., JARVIS, C. H., and STUART, N., 2001, Artificial neural networks as a tool for spatial interpolation. *International Journal of Geographical Information Science*, **15**, 323–343.

SHARIFF, A. R. B. M., EGENHOFER, M. J., and MARK, D. M., 1998, Natural-language spatial relations between linear and areal objects: the topology and metric of English-language terms. *International Journal of Geographical Information Science*, **12**, 215–245.

SHI, P. D., and TSAI, C. L., 2002, Regression model selection—a residual likelihood approach. *Journal of the Royal Statistical Society Series B—Statistical Methodology*, **64**, 237–252.

SKIDMORE, A. K., GAULD, A., and WALKER, P., 1996, Classification of kangeroo habitat distribution using three GIS models. *International Journal of Geographical Information Systems*, **10**, 441–454.

SONDHEIM, M., GARDELS., K., and BUEHLER, K., 1999, GIS Interoperability. *Geographical Information Systems: Principles, Techniques, Applications and Management: Volume 1*, edited by P. A. Longley, M. F. Goodchild, D. J. Maguire and D. W. Rhind (New York: Wiley), pp. 347–358

STEFANAKIS, E., VAZIRGIANNIS, M., and SELLIS, T., 1999, Incorporating fuzzy set methodologies in a DBMS repository for the application domain of GIS. *International Journal of Geographical Information Science*, **13**, 657–675.

VCKOVSKI, A., and BUCHER, F., 1996, Virtual data sets—smart data for environmental applications. In *Proceedings of 3rd International Conference on GIS and Environmental Modeling* (Santa-Fe, New Mexico: National Centre for Geographical Data and Analysis), CD-ROM.

WANG, F. J., 2000, A fuzzy grammar and possibility theory-based natural language user interface for spatial queries. *Fuzzy Sets and Systems*, **113**, 147–159.

WEBSTER, R., and OLIVER, M. A., 1990, *Statistical Methods in Soil and Land Resource Survey* (Oxford: Oxford University Press).

WESTERVELT, D. J., and HOPKINS, L. D., 1999, Modelling mobile individuals in dynamic landscapes. *International Journal of Geographical Information Science*, **13**, 191–209.

ZHU, X., 1996, A knowledge-based approach to the design and implementation of spatial decision support systems. PhD Thesis, University, Edinburgh.