Development of a Generic System for Modelling Spatial Processes

Andrew Marr
Richard Pascoe
George Benwell
Samuel Mann

The Information Science Discussion Paper Series

Number 98/09
March 1998
ISSN 1177-455X
University of Otago

Department of Information Science

The Department of Information Science is one of six departments that make up the Division of Commerce at the University of Otago. The department offers courses of study leading to a major in Information Science within the BCom, BA and BSc degrees. In addition to undergraduate teaching, the department is also strongly involved in postgraduate research programmes leading to MCom, MA, MSc and PhD degrees. Research projects in software engineering and software development, information engineering and database, software metrics, knowledge-based systems, natural language processing, spatial information systems, and information systems security are particularly well supported.

Discussion Paper Series Editors

Every paper appearing in this Series has undergone editorial review within the Department of Information Science. Current members of the Editorial Board are:

Assoc. Professor George Benwell
Dr Geoffrey Kennedy
Dr Martin Purvis
Dr Henry Wolfe

Assoc. Professor Nikola Kasabov
Dr Stephen MacDonell
Professor Philip Sallis

The views expressed in this paper are not necessarily the same as those held by members of the editorial board. The accuracy of the information presented in this paper is the sole responsibility of the authors.

Copyright

Copyright remains with the authors. Permission to copy for research or teaching purposes is granted on the condition that the authors and the Series are given due acknowledgment. Reproduction in any form for purposes other than research or teaching is forbidden unless prior written permission has been obtained from the authors.

Correspondence

This paper represents work to date and may not necessarily form the basis for the authors’ final conclusions relating to this topic. It is likely, however, that the paper will appear in some form in a journal or in conference proceedings in the near future. The authors would be pleased to receive correspondence in connection with any of the issues raised in this paper, or for subsequent publication details. Please write directly to the authors at the address provided below. (Details of final journal/conference publication venues for these papers are also provided on the Department’s publications web pages: http://divcom.otago.ac.nz:800/COM/INFOSCI/Publctns/home.htm). Any other correspondence concerning the Series should be sent to the DPS Coordinator.

Department of Information Science
University of Otago
P O Box 56
Dunedin
NEW ZEALAND
Fax: +64 3 479 8311
email: dps@infoscience.otago.ac.nz
www: http://divcom.otago.ac.nz:800/com/infosci/
In this paper is proposed a structure for the development of a generic graphical system for modelling spatial processes (SMSP). This system seeks to integrate the spatial data handling operations of a GIS with specialist numerical modelling functionality, by the description of the processes involved. A conceptual framework is described, the foundation of which are six defined modules (or services) that are considered a minimum requirement for basic system operation. The services are identified following description of the three key components to systems integration, and the examination of the preferred integrating structure. The relationship of the integration components to sample commentary on the future requirements of integration is discussed, and the benefits and deficiencies of an implemented system for modelling spatial processes are noted.

1. Introduction

“GIS is an increasingly sophisticated and widespread technology” (Davies and Medyckyj-Scott, 1994, p175). To meet the increasing technical demand of the user, system designers have had to incorporate high levels of complexity into developed systems to accommodate the range of functionality expected by the user.

Often the range of functionality is lacking in a standard GIS package, and the user may be forced to turn to specialist application software that may not accept the use of spatial data. Efforts to combine GIS with modelling systems have rarely been successful or offered the full functionality, interface design, and data handling demanded by the user. Abel et al. (1997, p5) argues that many examples of GIS and modelling systems integration “…are typically specific to the component subsystems and too narrow in the application focus of the integrated system”.

In this paper is proposed a conceptual framework for the development of a generic system for modelling spatial processes (SMSP). The structure of the conceptual framework is determined based on the analysis of key components that are used to
measure the level of integration between two systems. Consideration is given to sample commentary of the desirable features of such a system and how these relate to the key components of integration. An implemented spatial process modelling tool is also analysed that although deficient in terms of defined criteria, bears similarity to the perceived implementation of the conceptual framework.

The conceptual framework details six principle modules (or services) which are defined based on this analysis. It is argued that these services are the minimum required for system operation, and that if implementation were to occur, the use of this system would potentially eliminate many of the problems identified in existing tools.

2. Integration of GIS with other Applications

Environmental modelling tools are among the most frequent applications requiring integration with a GIS (Parks,1993). The benefits of integrating GIS and Environmental Modelling are widely recognised (Abel et al, 1997; Bennett 1997). While this form of integration may be common, the work described in this paper attempts to remain as generic as possible, both in terms of the application for it is used, and the structure which is described. Literature describes many different ways to connect a GIS to a modelling system eg. tight, loose, embedded, etc. (Bennett 1997; Burrough 1997; Lilburne 1996; Fedra 1993). These descriptions treat the integration process as the joining of two separate systems. However, Fedra (1993,p46) argues that “...the challenge is in merging the respective paradigms to create a new field of integrated environmental information systems that goes beyond models and GIS”. This complete merging has typically not been achieved in the past due to technical difficulties. The data structures, modes of operation and types of user interface in these systems are typically very different and do not lend themselves to comprehensive integration with a GIS.

Lilburne (1996) visually depicts the level of integration between GIS and another system as a point within an Integration Cube (figure i). Points along the three axis’ are measures of integration in respect to three key components: user interface, functionality, and data access. The further away the point is from the origin, the higher the relative level of integration between the two systems for these components.
The *interface axis* represents a range from where two separate interfaces are used (the origin), to where a single interface is used for all activities. Intermediate points on the axis represent situations where two separate interfaces are used, or where the two interfaces have similar ‘look and feel’. The *functionality axis* is used to describe the range from minimal (the origin), to full functionality, of the systems being integrated. The *data axis* describes the type of data model being utilised. This ranges from two distinct data sources each unique to one of the systems (the origin), to a common data source.

Analysis using the integration cube was performed by Lilburne (1996) on 104 published case studies. The results of the research suggested that for the majority of cases, higher scores were obtained for the integration of data and interface, but these usually scored poorly in the assessment of functionality.

Lilburne further refines the level of integration between two systems into classifications: Standalone, Loose, Tight, Merged, Enhanced, Customised, Client/Server, and Framework. Of the case studies analysed, the highest level of integration was achieved by those belonging to the framework classification. Membership of this classification indicates that the two concurrently running systems are integrated using a third system. This third system manages data sharing, interfacing with the user, and combining functionality of the two systems being integrated.

In summary, the integration cube is a useful tool for assessing the level of integration between two systems. Furthermore, an assessment of several case studies using this cube has led to the conclusion that use of a framework is typically the preferred approach to integration. The next activity is to establish the generic elements of a framework.

### 3. Features of a Conceptual Framework for Integration

In the assessment of the generic elements of a conceptual framework, consideration should be given to the features requested in published literature and the functions available in example software tools. Fedra (1993) argues that research in the integration of GIS and other systems is often poorly defined. As the field evolves, the requirements of integrated systems in relation to essential functions and features become clearer.

#### 3.1 Sample Commentary

Many in the research community have documented proposals for current and future development related to this field. Table I represents a sample of this commentary as presented in the literature. To assist in assessing the proposals, each comment has been classified under one of the three components presented by the integration cube. The commentary presented is neither rigorous nor viewed as being all inclusive. It does however provided a highly indicative representation of the relationship between the requirements for full integration as described by the cube and the proximity of comments presented in the literature.
### Interface

- Clear identification of model components and assumptions. (Kemp, 1993)
- Prevent model calibration, validation, and investigation being neglected (Burrough, 1997)
- The development of higher level languages and toolkits. (Fedra, 1993)
- Interactive user interface, help and explanation included. (Fedra, 1993)
- Symbolic graphical representation of major problem components. (Fedra, 1993)
- Little user concern for technical computer details. (Fedra, 1993)
- User able to visualise ongoing spatial processes. (Bennett, 1997)
- Methods such as AI and expert systems to guide non-expert users in the appropriate handling of hybrid tools. (Parks, 1993)
- The user should not have to adapt to the interface. (Hix and Hartson, 1993).

### Functionality

- Splitting of functions into separate components. (Fedra, 1993)
- Embedded AI components. (Fedra, 1993)
- Built in collaboration with the users. (Fedra, 1993)
- The inclusion of modelbase management technologies. (Bennett, 1997)
- Minimising program complexity by structuring models as a set of distinct modules. (Maxwell and Costanza, 1995)
- Use simple logical operations to explore complex relationships. (Parks, 1993)

### Data

- Removal of idiosyncratic command languages and data transfer facilities of independently-designed software systems. (Abel et al, 1997)
- Determination of the effects of data resolution on the quality and propagation of error on numerical models. (Burrough, 1997)
- Take into account the effects of scale. (Burrough, 1997)
- Coupling one or several databases, local or remote (Fedra, 1993)
- Use data models that capture spatial change, allow error and spatial imprecision, and represent complex interacting objects. (Bennett, 1997)

<table>
<thead>
<tr>
<th>Table I: Sample Commentary for GIS and Modelling Integration, (Marr et al, 1997)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Analysis of the commentary suggests that the authors request improvements in all three components (interface, functionality, and data access) identified by Lilburne (1996).</td>
</tr>
</tbody>
</table>

The focus of the ‘interface’ commentary is on helping the user to construct complex systems with a toolbox style graphical user interface.

The theme of the ‘functionality’ commentary is on the need for module based systems for additional flexibility and capable of re-use in unrelated applications. The commentary is particularly important since this aspect was identified as the main failing in the systems analysed by Lilburne using the integration cube. Additional flexibility in functionality can be achieved by enabling the user to combine primitive operations to form a composite operation. Albrecht (1996) has to this end defined such primitive operations in the spatial data context.

For the data component, the focus appears to be on the need for coherent and seamless access to different data sources. This would appear to be consistent with the aims and
ideals of the move to interoperable GIS, the goal of the Open GIS Consortium (OGC, 1997).

3.2 An Example System

When a GIS is integrated with a modelling system, the design of the interface is particularly important due to the complexity the operations bring performed. The most promising interfaces to date are those that present the user with an interactive toolbox for functionality. This approach removes the limitations of fixed-menu and command-based systems and allows a greater amount of user adaptation. Two good examples of this approach are presented by the virtual GIS project (Albrecht et al, 1997), and the spatial process modelling system (SPMS) project (Mann, 1996).

Mann’s SPMS prototype (figure ii) allows users to build complex environmental models by drawing diagrams of the environmental systems. These diagrams consist of three components, spatial objects (maps), data objects and process objects. These components are linked together to form the model structure. The system interprets the diagram and performs the processing. In the SPMS, spatial processing is performed by a separate GIS package (Idrisi) though this is hidden from the user. After processing, the thumbnail representations of the spatial data are updated where new values are needed. As process models may include feedback loops and be defined for any time period, SPMS includes functionality for scenario development and prediction testing. This differs from VGIS (Albrecht et al, 1997) which, while having more advanced operations available, does not allow feedback and is therefore a workflow representation rather than a modelling tool. Because the system is domain independent and does not presume any model structure, the result is a flexible tool that can be rapidly applied in a wide range of situations.

![figure ii – Screenshot of the Spatial Process Modelling System (Mann, 1996)](image-url)
In figure ii, the example process model shown represents the effect of burning on vegetation growth. The user first built a simple model to predict vegetation growth. The vegetation map at the top left is joined to a growth component along with a data file which represents growth conditions over 15 years (this may be past or predicted). The user’s model has a time step of years and is set to run over 15 cycles (years). In this example the feedback from the growth component is fed back into the vegetation map which was updated for each cycle. To explore the effects of burning a part of the landscape, the model was adapted to include a burning component. The fire data icon represents a hypothesised burn response curve over the 15 years. In environmental management such a curve may be the subject of controversy so the data icon may be annotated with appropriate comments (as can the entire model). The response curve is combined with the affected area, returned to a multiplier and combined with the normal growth before being fed back into the vegetation map. The user may then explore the effects of changing the fire response curve or growth conditions or adapt the model to investigate, say, the effects of including altitude in the system. While this example contains ‘pseudo-spatial’ operations such as multiplier based overlay of raster data, the system does include spatial operations such as buffering.

Testing in scenario development, prediction and environmental management situations has shown that the approach of the SPMS prototype results in measurable benefits in decision-making (Mann, 1998). The benefits may be related to a number of factors. The ease of use and overall satisfaction suggest that the user’s model is matched by the system model (Pidd, 1996). It is also a definite move in the right direction (Davies and Medyckyj-Scott, 1994) in the aim to break down differentiation in terms of control and user display representations. Further, in (Woodmansee, 1988) terms, the user can not only envision several layers simultaneously, but also the links between layers are made explicit. The SPMS prototype performs well on Lilburne’s integration cube. For both the interface and functionality axis’ a production version of the SPMS prototype would rate very highly. The major limitation of the SPMS in its current form towards a higher position in the cube is an inability to handle different file formats and data structures.

Methods of testing developed models in terms of sensitivity (ie. to initial conditions and assumptions) is an unresolved issue. The current structure means that models are flexible and the user can ‘experiment’ with the structure; e.g. “does adding altitude have an effect?”, but, as described by Rothenburg (1991), this is a rather ‘naive perturbation’, just fiddling with all the variables!.

The SPMS prototype is ignorant of time constructs, so defining an action such as “graze for 3 weeks then spell over summer then graze lightly in autumn” is not possible. Neither is multi-temporal cycling available (that is having leaf fluxes changing rapidly while yearly processes tick over slowly or only occasionally). These issues potentially could be solved by a method to embed models in other models, though this would also require more consideration of error propagation. Such a structure would also facilitate the archiving and transfer of models in the modular manner proposed in the literature (Bennett 1997; Maxwell and Costanza 1995).

The SPMS developed by Mann, was shown to be a substantial improvement in assisting in the decision making process in the case of environmental managers.
However the system is not designed to be homogenous to this group of users, with potential uses in a variety of fields. Further, the research of Mann also demonstrated that with very little tuition, novice users were capable of completing relatively complex and sophisticated tasks with the assistance of the whiteboard type user interface.

Even though several deficiencies were identified in this initial prototype system, the ease of use in terms of modelling design and implementation, suggest many beneficial aspects worth incorporating in a more comprehensive design.

3.3 Summary
There have been many systems developed to solve individual problems, or groups of problems in a generic and reproducible manner, using GIS and modelling programs. However there is relatively little research into systems that support the construction and sharing of user defined algorithms and models requiring complex analysis of spatial data. This is one of the major limitations of the SPMS developed by Mann. Other deficiencies include the inability to process heterogeneous spatial data, incorporate a model as a sub-model to larger configurations, and handle variable time frames within a process model.

One of the main features of the SPMS is to make the operation of the system available to non-specialist users including complete novices. This usability aspect of SPMS has clearly proven to be successful, and a factor that should be reinforced in any subsequent system development. Part of the success of this system is the style of progressive problem development and description that allows the users to understand the procedures employed from the initial assumptions onwards.

4. Generic Elements of an Integration Framework

Figure iii represents a conceptual framework architecture for modelling spatial processes (Marr et al, 1997). The architecture is designed to solve the issues previously identified and is principally based on a logical breakdown of required features.
There are six main software components differentiated by function. These components (described as services) are:

- **Process model design**
  The process model design service is a software module responsible for facilitating the documentation of process models designed by the user.

- **Process model interpretation**
  The process model interpretation service is responsible for the re-construction of each model, the capture of required data sources, and consequent model execution.

- **Spatial data operations**
  A software module responsible for provision of common spatial data operators and functionality on demand.

- **Modelling operations**
  A software module responsible for provision of common modelling operators and functionality on demand.

- **Spatial data compatibility**
  A software module responsible for the provision of data format compatibility on demand.

- **Result presentation and visualisation.**
  A software user interface responsible for the display of process model execution results (textual, graphical, or a process model).

These services are considered the requirement for minimum system operation. It is suggested that other specialist modules (eg. terrain, network, and statistical analysis) may be attached to the core modules, on an as required basis. The points of contact with regard to direct user interaction are principally the process model design and process model interpretation services.

### 4.1 Process Model Design Service

The process model design service is a software module responsible for facilitating the documentation of process models designed by the user. Process models are described both in descriptive form (including meta-data and lineage) and in mathematical representation. Both these aspects must be captured as the process model is constructed interactively. While the information inserted by the user must be checked for errors and inconsistencies, the process model design service does not execute the model designs. To assist in reader comprehension, figure iv has been provided which represents a potential non-functional implementation of this service. The example illustrates the selection of suitable parachute drop sites given specific criteria relating to maximum ground slope, and proximity close to or away from air corridors, access roads,
and waterways. This problem requires the use of the buffer and overlay spatial operations but in this case does not include any specific non-spatial modelling operations. In addition, for simplicity, this example does not include any feedback loops requiring iteration.

In the model of the selection of suitable parachute drop sites, there are four spatial data inputs, *Slope*, *Airspace*, *Road*, and *Hydro*. Using the interface the user does not explicitly define the data to be utilised but more the generic characteristics of the data to expect (e.g., Raster or Vector). The desired spatial data output is also shown as, *Suitable Parachute Drop Sites*, as are the required spatial operations, *Buffer* and *Overlay*. The menu on the left (from which the user selects required objects and drags them on the ‘whiteboard’) in this instance shows the available spatial operations. The spatial operations depicted are defined by Albrecht (1996, p36) as the derivation of a “conclusive list of universal GIS operations”, from which more complex operations may be constructed. The maths operator menu option is used to provided mathematical functionality to model designs (e.g., +, -, x, /, Sin, Cos, Tan, Squared, Square Root, etc).

A more important menu option is data and models. From this menu the user can select where inputs or outputs are required. Inputs may be spatial or non-spatial in nature. Outputs are similar to inputs, but there is potential for the development of an output that is a process model in itself. Besides standard input and output options there is a facility to insert existing process models that appear on the whiteboard as a single icon, but that can be expanded if required. The insertion of existing process models into the current model may bring with it addition requirements for data that will need to be met in the eventual execution.
One feature of this design is that if a ‘time icon’ is included as part of (but not associated with the other objects) an individual model then there exists the ability to have differing temporal events occurring among elements of the same overall process model. This is achieved by including sub-process models as part of the main model, but with each occurring in different time steps. This aspect was a major limitation of the SPMS system developed by Mann (1996), in which all activities had to occur simultaneously (e.g. daily, monthly, yearly, etc).

4.2 Process Model Structure Files
As previously discussed, the purpose of the process model design user interface is to document user process model designs (both descriptive and numerical representations). The process model structure file, while not a member of the six services, is a crucial component of the system. Benz (1997) has completed extensive work in the documentation of ecological and environmental mathematical models which should approximate the contents of the structure file.

In the context of spatial process modelling as defined in this paper, the process model structure file is the instrument that facilitates free distribution of the technical specification among researchers. This instrument may be used in a stand-alone scenario, or as a sub-component of another process model. To permit this multiple use the eventual structure of the file requires considerable research and testing. The structure of the design file is also important because it is the only mechanism by which the original assumptions and limitations of the model constructed by the designer may be conveyed to the eventual user. Some observers (PC, 1997) have conveyed concern over this aspect in respect to the lack of control by the designer over the purpose the user may eventually put a process model to. However, if the analogy of the spreadsheet is used, assuming the integrity of the software is maintained, the obligation for accurate calculation remains with the user, and not with the software vendor. This is applicable to spatial process modelling.

4.3 Process Model Interpretation Service
The Process Model Interpretation Service is responsible for the re-construction of each model, the capture of required data sources, and consequent model execution. As before, to assist in reader comprehension, figure v has been provided which represents a potential non-functional implementation of this service. The graphical user interface is dynamically created based on the interpretation of the design file. The details of each model (both descriptive and mathematical) are presented to the user. The middle section of the diagram is created based on the number and type of data sources and outputs as determined by the design file. In this case using the example of the selection of suitable parachute dropsites, there are four spatial inputs (slope, airspace, road, and hydro) and one output, the resulting spatial data set of the indicated format. For each of these data insertion points, detailed discussion of the criteria, intentions, limitations with regard to the data are included under the ‘discussion’ buttons.
The interface attempts to check aspects of the execution process as indicated by the user in their checkbox selection. Figure v is not considered comprehensive in terms of the type of checks that may be performed, and some of the desirable checking may rely on the development of suitable meta-data standards in the future.

### 4.4 Spatial Data Operations Service

In meeting the expectations of the user, a series of common spatial data operators and functions should be made available by default. As previously discussed, a suitable starting point in the development of this service is work by Albrecht (1996) on the derivation of a conclusive list of universal GIS operations (table ii).

<table>
<thead>
<tr>
<th>Search</th>
<th>Interpolation; Search-by-region; Search-by-attribute; (Re-)Classification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Locational Analysis</td>
<td>Buffer; Corridor; Overlay; Voronoi/Thiessen</td>
</tr>
<tr>
<td>Terrain Analysis</td>
<td>Slope/Aspect; Catchment/Basins; Drainage/Network; ViewShed</td>
</tr>
<tr>
<td>Distribution/Neighbourhood</td>
<td>Cost/Difintegration/Spread; Proximity; Nearest-Neighbor</td>
</tr>
<tr>
<td>Spatial Analysis</td>
<td>Multivariate analysis; Pattern/Dispersion; Centrality/Connectedness; Shape</td>
</tr>
<tr>
<td>Measurements</td>
<td>Measurements</td>
</tr>
</tbody>
</table>

*table ii – Universal GIS Operators from (Albrecht, 1996)*

Albrecht defined algebraically 20 basic spatial operations based on the analysis of common GIS software that could form the fundamental building blocks for more complex operations. This work is intended to form the initial basis of the spatial data
operations service in the form of an ‘on demand’ module based on the users process model design.

Potentially there is value in facilitating user defined spatial functionality beyond the initial set of 20 operators. This aspect requires substantial research and in particular resolving standardisation issues between the software systems of the developer and user.

4.5 Modelling Operations Service
In the same manor as with Spatial Data Operations Service, this service provides on demand modelling functionality consisting of a standard set of common mathematical operations.

4.6 Spatial Data Compatibility Service
The Spatial Data Compatibility Service is software model responsible for managing heterogeneous spatial data formats. “Data are the raw facts entered into the computer” (Shore, 1988, p10). In GIS terms, data has traditionally been viewed as the ‘raw facts’ in the structure of fixed proprietary vendor formats. These formats have resulted from the general evolutionary nature of GIS development itself. Because of the ‘barriers’ (Glover, 1995) created by use of different non-interchangeable vendor formats, efforts to overcome these differences have traditionally been time consuming, difficult and resource intensive.

Recent developments, possibly spurred on by the Open GIS initiative (OGC, 1996) have seen some software vendors starting to tackle this problem (Strand, 1996).

One solution to the current proprietary format exchange problem, is the use one of a growing number of spatial data interchange software package such as (FME, 1997) or (BlueMarble, 1997).

‘The Feature Manipulation Engine (FME) is a sophisticated configurable spatial data processor and translator. The FME facilitates powerful interoperability between diverse systems, and can be used as the backbone of an on-demand mapping system.’ (FME, 1997)

It is proposed that such spatial data interchange software could perform the required functions of the spatial data compatibility service.

4.7 Result Presentation and Visualisation Service
The result presentation and visualisation service permits the user to review executed outcomes. This could be a spatial data viewer or a spreadsheet depending on whether the expected output was spatial or non-spatial in form. It is conceivable that in certain instances, the anticipated outcome of the analysis could be to produce a new process model that describes interrelationships between process model variables. In this case, the design service would be used to investigate the results and design further experimentation.
5. Conclusion

In this paper, a conceptual framework for the development of a generic system for modelling spatial processes (SMSP) has been proposed. The structure of the conceptual framework is determined based on the analysis of key components that are used to measure the level of integration between two systems (Lilburne, 1996). These components include: user interface; functionality; and data access. Lilburne was able to determine the characteristics of successful integration, the results of which suggested the use of the framework approach in resolving the integration issues. Consideration is given to sample commentary of the desirable features of such a system and how these relate to the key components of integration. An implemented software tool (SPMS) is also analysed. While this particular system has identified deficiencies, there are appreciable similarities to the perceived implementation of the conceptual framework.

The conceptual framework details six principle modules (or services) which are defined based on this analysis. These services include: process model design; process model interpretation; spatial data operations; modelling operations; spatial data compatibility; and result presentation and visualisation. It is argued that these services are the minimum required for system operation, and that if implementation were to occur, the use of this system would potentially eliminate many of the problems identified in existing tools, including the SPMS.

6. References


