Netherlands Journal of Agricultural Science 43 (1995) 19-29

Linking GIS and models: structure and operationalisation for a Costa Rican case study

J.J. STOORVOGEL

Atlantic Zone Programme, Apartado 224-7210, Guápiles, Costa Rica

Received 12 October 1994; accepted 23 April 1995

Abstract

Often, users of geographical information systems require very specific, disciplinary operations on geo-information that are not supported by GIS. These operations can be made available to the GIS through links with external models. A general structure for the GIS-model interface is presented and identifies six consecutive steps: 1) geometry operations, 2) attribute operations, 3) data export from the GIS to the external model, 4) model run, 5) data import from the model into the GIS, and, 6) visualization or spatial analysis of the model results with the GIS. This structure is illustrated for a case study from Costa Rica, where a GIS is linked with a linear programming model for the analysis of alternative land use scenarios. The structure can be operationalised, using the abilities of many commercial software packages to develop user oriented applications.

Keywords: Costa Rica, GIS, land use scenarios, modelling

Introduction

Currently, many tools are being developed to support land use analysis and planning. One of these tools is the USTED methodology (*Uso sostenible de Tierras en El Desarrollo,* Sustainable Land Use in Development) (Stoorvogel *et al.* 1995), developed at the Atlantic Zone Programme (AZP), which integrates simulation models, linear programming models and geographical information systems (GIS). The AZP, a cooperation of the Wageningen Agricultural University with the Centre for Research and Education in Tropical Agriculture (CATIE, Costa Rica) and the Ministry of Agriculture and Animal Production (Costa Rica), applied the methodology for the Neguev settlement, comprising 4675 ha divided over 307 farms in the perhumid tropical lowlands in the northeast of Costa Rica. The need for a spatial analysis and a quick interpretation of scenario results required the use of GIS. A general structure for the link between models and GIS is included within a user shell to integrate the different models and database management systems.

Although GIS are powerful tools for the analysis of geo-information, commercial GIS packages do not always meet the specific needs of users (O'Kelly, 1994). User requirements often comprise very specific disciplinary operations and user oriented shells. To a limited extent, GIS software enables the development of applications that may include (simple) models. Implementation of additional procedures into GIS packages or modifications of GIS packages usually coincides with high costs due to the complexity of GIS software and additional modelling systems (Abel *et al.*, 1994). Therefore, GIS can be considered to be a closed system, i.e. no changes in the internal schemes of the software can be made. Specific disciplinary analysis like crop growth simulation need, therefore, external models, which work independently from the GIS and perform the analysis which the GIS package is unable to handle. For operationalisation, the GIS needs to be linked to these external models. Although the necessity to link GIS with models is generally recognised, many practical problems are known to occur (Burrough, 1989; Abel *et al.*, 1994). Part of the problems originate in the incompatibility of data formats, data organisation or semantics which respectively requires reformatting, restructuring and data analysis before the GIS database can be used in combination with external models. No GIS architecture has yet been developed that conceptualizes the link between GIS and external models. At present, therefore, the link between models and GIS is often established in an ad-hoc manner (e.g. Meijerink, 1989). Specially designed structures may facilitate this link and can be included in the GIS for operationalisation.

In the current paper, the structure and operationalisation of the link between a linear programming (LP) model and GIS as it is used in the USTED methodology is presented. Although the structure is developed for a specific case, it can form the basis for many different applications in which a GIS and external models are linked.

Materials and methods

Concepts

Large spatial databases are often organised in a layered structure, where each layer stores data on geographical features related to a specific theme and within a specific geographic area (Frank and Mark, 1991). In the present context, each layer is referred to as a map. A GIS database from a land use planning project may thus contain layers with data on soils, climate and land use. The geographical features in each of these layers are polygons, which represent areas with a specific soil type, climate and land use respectively. Combinations of layers can be analysed through map overlays. Van Oosterom (1991) indicates a hierarchical structure for the description of geographical features (Figure 1). Each feature is characterized by spatial data and thematic attributes, which are linked by an unique identifier. Spatial data are subdivided in geometric data and topological data. The geometric data comprise information on the position and shape of the features. On the basis of the geometric data of all the features in one map, the topological data for the individual features can be determined which indicate the spatial relationships (e.g. connectivity and adjacency) between the features. The thematic attributes include the characteristics of the geographical features (e.g. soil type and soil properties for a mapping unit in a soil map).

A model is a formal relation between exogenous input parameters and derived endogenous output parameters. In the case of e.g. a crop growth simulation model, the

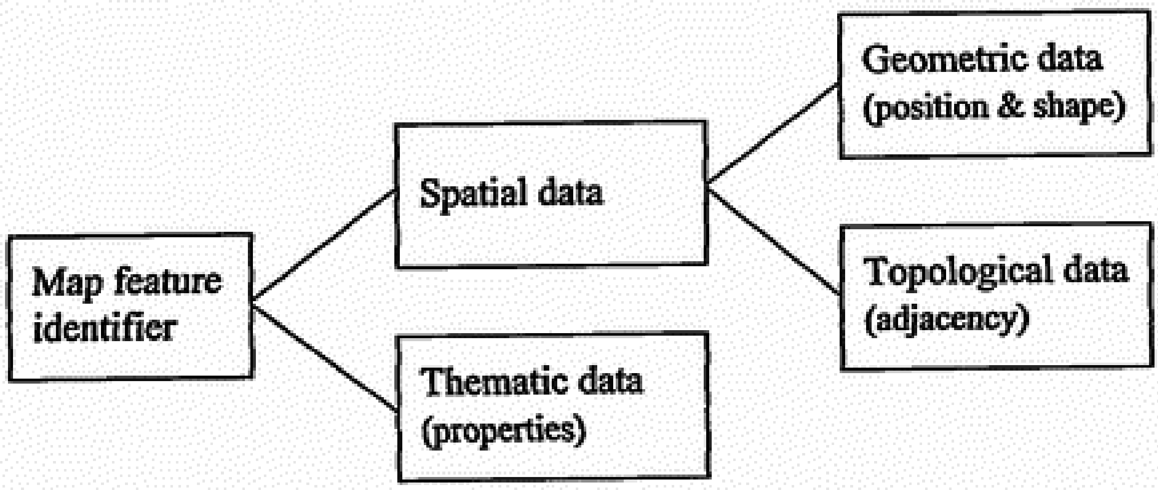


Figure 1. An hierarchical structure for data storage in a GIS (after van Oosterom, 1990).

exogenous input parameters comprise data on soil, crop and climate. On the basis of these input data, the model calculates e.g. the expected crop yield, which is the endogenous output parameter. Models, which are linked to a GIS, can use geometric, topological and thematic attributes as exogenous parameters and calculate new geometric data and/or thematic attributes as endogenous parameters. It is not necessary to import topological data in the GIS, as they are a function of the geometric data and, therefore, can be determined within the GIS. Variations in the type of model input and output can vary considerably as shown in the following three examples:

- Crop growth simulation models can simulate for polygons representing objects at different aggregation levels (e.g. field, farm, agro-ecological zone) the crop production on the basis of climatic and pedological characteristics (e.g. Van Keulen & Wolf, 1986). For the simulation, thematic attributes for the different geographical features (soil and climatic properties) are used. The simulated crop productions can be linked as new thematic attributes to the spatial objects represented in the map.
- A model for the infestation of a crop with pests requires data for the specific field which is being modelled (e.g. soil type, micro-climate). In addition, the occurrence of the host plants in surrounding fields may influence the risk of infestation. Hence, the input exists of topological data and thematic attributes of both the modelled and neighbouring fields. Although the model input exists of both topological and thematic attributes, the output only comprises thematic attributes, e.g. the risk for the infestation of a specific field.
- Models which simulate three-dimensional processes like ground water-flows, need spatial data (including geometric and topological data) as well as thematic attributes. During three-dimensional modelling a new geometry will be created by changing the geometry of existing objects and possibly the creation of new objects. When data are stored in raster format, spatial objects are represented by a set of fields. The spatial definition of the fields remains the same and, therefore, no changes in the geometry take place. In the case of vector-based maps, the changes in the geometry will lead in many cases to a new geometry. If the geometry changes, both geometric data and thematic attributes will be imported and a new map will be created in the GIS. However, if the geometry of the features is not al-

tered, the model results can be imported and added as thematic attributes to the original map from which the data where exported.

Although models differ in their data requirements, they do not necessarily differ in the type of data which need to be imported in the GIS. All three models illustrated in the examples may yield only thematic attributes, which generally can be linked to an existing map in the GIS.

Structures to link a GIS with external models

When a GIS is linked to an external model, the GIS provides the input data and the model subsequently determines the derived parameters for the geographical features or for the map, in cases where the model changes the geometry of the objects. Afterwards the GIS can be used to visualize and analyse the model results. The link between GIS and external models is two-way and is primarily based on data interchange. Figure 2 gives a general framework for the GIS-model interface, based on six main steps which are distinguished. In most cases, available data in the GIS have to be translated to the specific input parameters of the model (Step 1–2). The data are exported to the model (step 3), and after the model run (step 4), the model outputs have to be imported in the GIS database (step 5). The GIS can subsequently visualize and analyse the model results (step 6). The steps will be described and illustrated with an example for the link of a crop growth simulation model with a GIS to calculate the regional distribution of potential production.

Step 1 deals with the formulation of the geometric data. When the geographical features are not yet defined, they have to be created on basis of one or several base maps. Step 1 results in a map with the proper geometry, i.e. the features are the basic elements for the model. The map, however, may still lack the proper thematic attributes necessary for the model. The geometry operations used to define the geographical features include overlay operations for the combinations of maps but also buffer operations to generate zones around geographical features within certain spatial proximities, e.g. the area within a certain distance from a road. For crop growth simulation models the map features should be characterised by a combination of soil type and climate. An overlay of the soil map with the climatic map yields polygons with a specific climate and soil.

In Step 2 the thematic attributes for the map are determined. A number of attribute operations, which may comprise mathematical and statistical calculations, have to be performed to acquire the correct variables. Queries to the database are necessary for the appropriate data structure. The result is a map with the appropriate map features and thematic attributes. In the case of the crop growth simulation model, queries can select the required climatic and soil parameters for the different polygons.

Step 3 deals with the export of geometric, topological and/or thematic attributes from the GIS to the external model. In general, standard formats for data interchange (e.g. 'comma separated value' files for thematic attributes and 'digital line graph' files for geometry data) can be used. A well structured data exchange will enable automatization and thus operationalisation of the GIS-model interface. In the case of a crop growth simulation model, the identifier of the geographical feature with different climatic and soil parameters are organized in a table and exported to the simulation model.

Step 4 comprises the actual model run, where on basis of the exogenous parameters, the endogenous parameters are calculated. Depending on the type of model, runs are carried out for the whole map or for the individual features. In the case of a crop growth simulation model the simulation will be carried out for each individual feature, or, if different features with the same thematic attributes occur, groups of features. During the model run the identifiers to the original geographical features should be preserved to link the model results to the original map.

The results of the model are imported into the GIS in step 5. This is the reverse procedure from step 3 and in most cases similar formats for data interchange can be used. In cases when the geometry of the geographical features is not changed, the endogenous model parameters become new thematic attributes for the features of the base map. In the case of models which change the geometry, a new map is created on the basis of the model outcome.

Finally, in step 6, the model results are further analyzed. Subsequently, the model outcome may be analysed in the GIS, through e.g. aggregation of geographical features with similar model results, or overlays with other maps. In the case of alternative production systems, the analysis may e.g. include a comparison with actual land use and production levels.

The structure as presented in Figure 2 mainly comprises a series of consecutive operations in the GIS environment, which is specific for each GIS-model link. Most

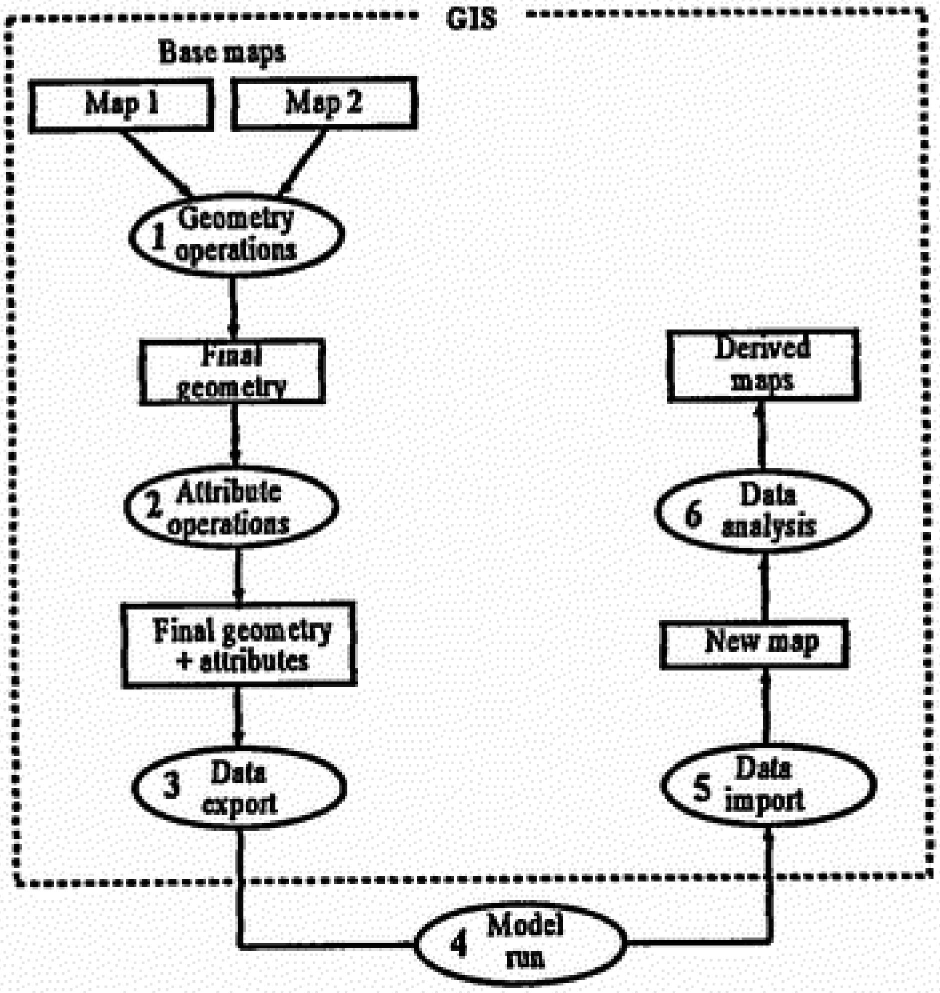


Figure 2. A six step approach (indicated by the ovals) for the GIS-model interface.

GIS packages enable the development of simple applications. Although in many cases the models can not be defined in these applications, they can automate or support the GIS-model interface, resulting in the operationalisation of a series of standard GIS commands through a macro language.

Results

Land use scenarios can be analyzed by linear programming models (e.g. Schipper *et al.*, *1995*; Veeneklaas, *1990*). Typically, GIS does not provide tools for linear programming and no commercial LP software includes GIS facilities. Although linear programming models are not spatial (Chuvieco, 1993), they can be linked to a GIS to relate the analysis to certain geographical features (e.g. farms, fields). The spatial presentation of the model results enables a quick interpretation of the linear programming results and a spatial analysis.

The USTED methodology has been developed for the analysis of land use scenarios. The LP model maximizes total net farm income for a sub-region (e.g. settlement or municipality), through a simultaneous selection of alternative land use systems for different farm types in the region. The farms are grouped into farm types according to size and soil type distribution. Alternative land use scenarios represent changes in the socio-economic or bio-physical environment of the region affecting the goal function, the constraints, and/or the alternative activities (land use systems and technologies, denominated LUSTs) of the LP model. The constraints indicate the availability of resources like land and labour, but may also comprise restrictions on other parameters for e.g. sustainability. The USTED methodology is operationalised for the Neguev settlement.

The link between LP model and GIS is established by the six step approach presented in figure 2. The results are given in Figure 3a (Step 1–3) and 3b (Step 4–6).

Step 1: two base maps, a 1:20.000 map of the farms (Anonymous, 1981) and a soil map at the same scale (De Bruin, 1992), were combined by an overlay procedure to yield a map with the farms and the soil limits. The thematic data of the map yield the soil types and the size of each of the 307 farms.

Step 2: the thematic attributes of the combined map comprise the farm identifier and a reference to a specific soil type. The LP model, however, does not deal with individual farms but considers farm types. Consequently, a farm classification was carried out by means of a cluster analysis (Schipper *et al.*, 1995), resulting in five different farm types, each with a specific size and soil type distribution.

Step 3: the input parameters for the linear programming model, comprising the number of farms in each farm type and its average size and soil types, are exported to a file which can be read by the model.

Step 4: during the optimisation with the linear programming model, LUSTs are selected for the soil types on each of the farm types. The selection of the LUSTs is based on the maximisation of the total net farm income, given the constraints which are defined for the model. The results of the linear programming model indicate the selected LUSTs for each soil type on the different farm types.

Step 5: The output of the LP model presents the selected land use systems for the different farm types. This data is linked to the map with the farm types, which was the result of Step 2. For each of the polygons (defined on the basis of farm type and soil type) the LP model can select several LUST. If more than one land use system is selected, LUSTs with high labour requirements are considered to be cultivated closer to the roads than crops with a low labour consumption. Using several buffer operations, the polygons of the map are subdivided in different zones, each between two distances to the road. The land use systems can now be distributed over the polygons. On the basis of the map with the optimal LUST distribution according to the scenario definition, a quick interpretation can take place.
Step 6: the analysis of the results is user dependent. Additional LUST characteristics can be linked to the map (e.g. biocide use in Figure 3b). This enables a spatial analysis of e.g. the sustainability of the scenario results. It may also yield data on spatial concentrations of specific productions, which may be important for the planning of specific services.

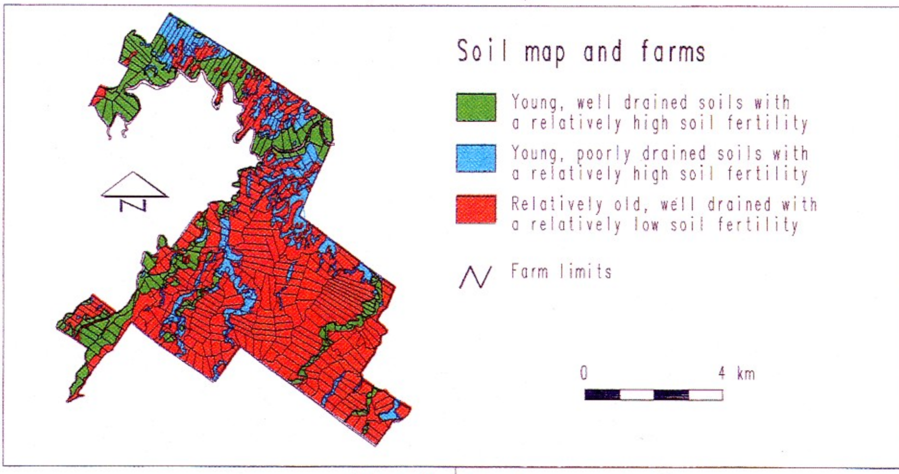
One of the advantages of scenario based studies as in the USTED methodology is the interactive way that users can analyse the effect on changes in the socio-economic and the bio-physical environment. Interactivity often results in a large number of alternative model runs and, therefore, requires a rapid interpretation of the results through the visualization of the scenario results and spatial analysis. The link between the GIS and the LP model determines the degree of interactivity. In the case of the Neguev settlement, the GIS allows for the development of applications through a macro language. The different steps described above are included as an application in the GIS. Therefore, although the LP model is not included within the GIS software, a highly interactive procedure is developed.

Discussion

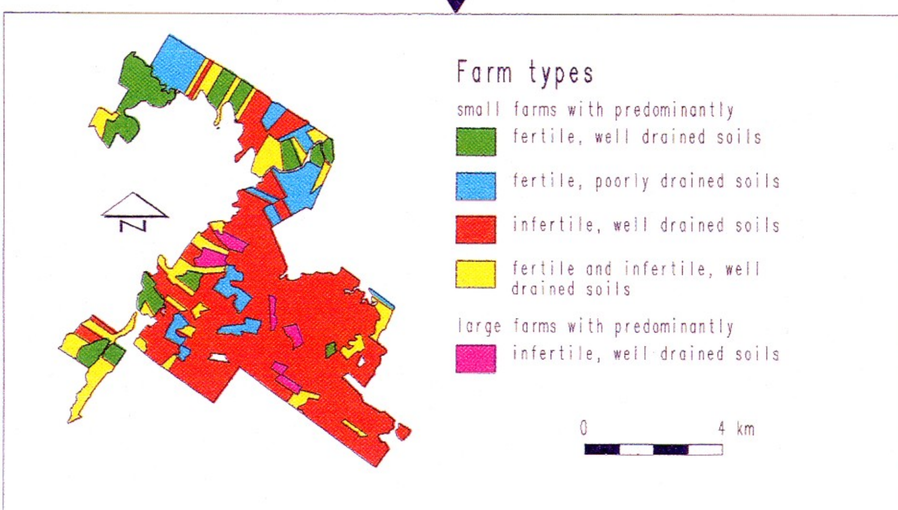
GIS can play an important role in land use planning (Sharifi, 1992; Despotakis, 1991) and assessment of environmental projects (Campbell *et al.*, 1989). Many applications need external modelling in combination with GIS. Nijkamp and Scholten (1993) emphasize a consistent use of GIS and models. Although the use of GIS and models for many applications is clear, linking is often a problem. Abel *et al.* (1994) provide a general structure for system architecture and a general set up for systems integration. They stress the importance of user interfaces for the GIS. However, with the use of commercial software packages, the user is restricted to the tools that come along with the package and, therefore, in the possibilities to link external models.

To avoid operational difficulties to implement specific disciplinary models in commercial software packages, the models can be linked to the GIS. Most GIS systems allow for the development of applications through e.g. macro-languages and can thus be used for simple models. The present article shows the importance to focus the applications on the link between models and GIS. This can easily be operationalised for models which use topological and/or thematic attributes as input. In cases, where the model uses and changes geometric data, the model internally will

Step 1: Overlay



Step 2: Cluster analysis



Step 3: Data export

Farm type	Number	Area/farm (ha)	Soil type (in %) SFW	SIW	SFP
1	35	14.1	91	6	3
2	33	15.7	12	28	60
3	189	13.1	6	88	6
4	46	13.5	52	41	7
5	4	32.1	10	78	12

→ Step 4: LP model

Figure 3a. The link between the GIS and the LP model for the Neguev case study, step 1–4.

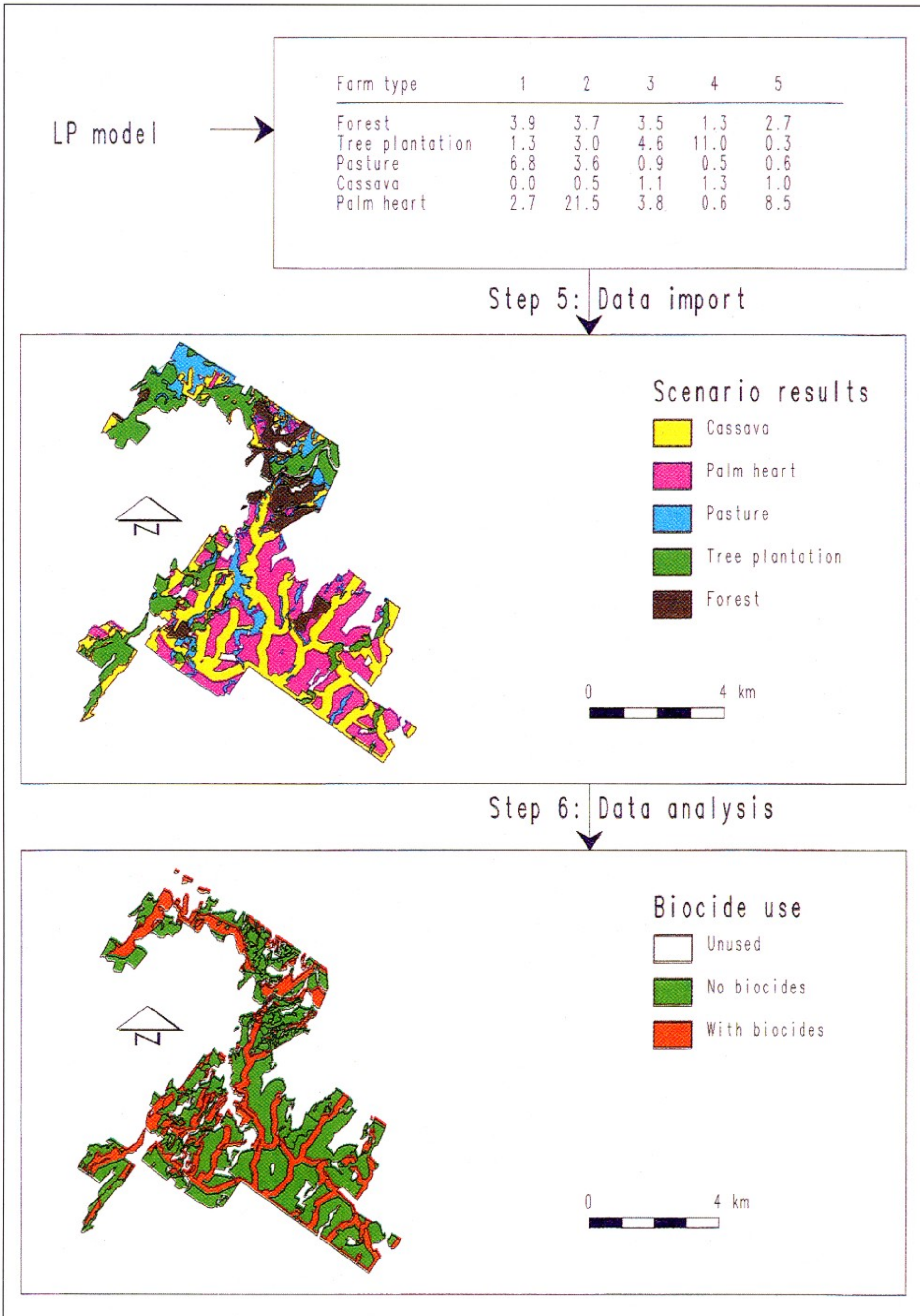


Figure 3b. The link between the GIS and the LP model for the Neguev case study, step 4–6.

make use of a GIS and integration may be necessary. The link between GIS and models of any kind may be useful to extend the standard possibilities of a GIS with additional operations. The structure presented in the present article organizes the GIS-model interface in a number of relatively simple operations.

The six-step structure for the link between GIS and external models is generally applicable when users deal with one model and a GIS. In cases where additional database systems are involved the structure becomes more complex and more specific. Due to the clear definitions of the different steps, it can function as a good basis for the development of applications. Specific requirements for the operationalisation of the GIS-model link are:
- the availability of a GIS which allows for the development of applications,
- the availability of formats for data interchange which can be used by both GIS and model, and,
- users, who are aware of the limitations and the assumptions of the different procedures.

When these requirements are fulfilled, the operationalisation can be realized and in most cases automated. The efficiency of the operationalisation depends strongly on the number of user decisions which are necessary during the model run. Considering a relatively simple model, which determines the endogenous parameters on the basis of a set of exogenous parameters without any user decisions, the application can function as a new command within the GIS environment. The user does not have to be aware that an external model is being used. The risk of automatization of analysis and applications is that users unaware of the procedures may use the application as a black box and for datasets outside the range of validation. When users are relatively unexperienced with the models the link between GIS may be supported by a well designed interface to clearly show the interaction instead of full automatization.

Acknowledgements

The present study is part of the USTED methodology (Stoorvogel *et al.*, 1995), which is the result of a multi-disciplinary study in close cooperation with the working group Costa Rica of the Wageningen Agricultural University, which is greatfully acknowledged for many fruitful discussions. Hans Jansen, Don Jansen (PZA), and two anonymous reviewers are acknowledged for their useful comments on earlier drafts of this article.

References

Abel, D.J., P.J. Kilby & J.R. Davis, 1994. The systems integration problem. *International Journal on Geographical Information Systems* 8: 1–12.

Anonymous, 1981. Plano mosaico del proyecto Neguev. Instituto de Tierras y Colonización, San Pedro, 1 map.

Burrough, P.A., 1989. Modelling land qualities in space and time: the role of geographical information systems. In: J. Bouma & and A.K. Bregt (Eds.). Land qualities in space and time. Pudoc, Wageningen, 317–320 pp.

Burrough, P.A., 1986. Principles of geographical information systems for land resources assessment. Clarendon Press, Oxford, 194 pp.
Campbell, W.G., M.R. Church, G.D. Bishop, D.C. Mortenson & S.M. Pierson, 1989. The role for a geographical information system in a large environmental project. *International Journal on Geographical Information Systems* 3: 349–362.
Chuvieco, E., 1993. Integration of linear programming and GIS for land-use modelling. *International Journal on Geographical Information Systems* 7: 71–83.
De Bruin, S., 1992. Estudio detallado de los suelos del asentamiento Neguev. Working Document No. 25 (Phase 2). Atlantic Zone Programme, CATIE, Turrialba, 35 pp + map.
Despotakis, V.K., 1991. Sustainable development planning using geographical information systems. PhD thesis, Free University, Amsterdam, 274 pp.
Frank, A.U. & D.M. Mark, 1991. Language issues in GIS. In : D.J. Maguire, M.F. Goodchild, and D.W. Rhind (Eds.). Geographical information systems. Longman Scientific & Technical, Harlow, 147–163 pp.
Meijerink, A.M.J., 1989. Modelling in the land and water domain with a versatile GIS, ILWIS; experiences from a large tropical catchment. In: J. Bouma, J. & A.K. Bregt (Eds.). Land qualities in space and time. Pudoc, Wageningen, 73–87 pp.
Molenaar, M., 1991. Status and problems of geographical information systems. The necessity of a geoinformation theory. *ISPRS Journal of Photogrammetry and Remote Sensing* 46: 85–103.
Nijkamp, P. & H.J. Scholten, 1993. Spatial information systems: design, modelling, and use in planning. *International Journal on Geographical Information Systems* 7: 85–96.
O'Kelly, M.E., 1994. Spatial analysis and GIS. In : S. Fortheringham & P. Rogerson (Eds.). Spatial analysis and GIS. Taylor & Francis, London, pp. 65–79.
Schipper, R.A., D.M. Jansen & J.J. Stoorvogel, 1995. Sub-regional linear programming models in land use analysis: a case study of the Neguev settlement, Costa Rica. *Netherlands Journal of Agricultural Science*: this issue.
Sharifi, M.A., 1992. Development of an appropriate resource information system to support agricultural management at farm enterprise level (ARIS). PhD Thesis, Wageningen Agricultural University. Wageningen 217 pp.
Stoorvogel, J.J., R.A. Schipper & D.M. Jansen, 1995. USTED, a methodology for land use planning including the sustainability. *Netherlands Journal of Agricultural Science*: this issue.
Van Oosterom, P.J.M., 1990. Reactive data structures for geographic information systems. PhD thesis, Leiden University, Leiden, 197 pp.
Van Keulen, H. & J. Wolf (Eds.), 1986. Modelling of agricultural production: weather, soils and crops. Pudoc, Wageningen, 478 pp.
Veeneklaas, F.R., 1990. Compétition pour des ressources limitées: le cas de la cinquième région du Mali. Rapport 3. Description formelle du modèle d'optimisation MALI5. CABO, Wageningen & ESPR, Mopti, 64 pp.