

6.4 Geographic information systems

by Jeffrey W. White, Adriana Rodriguez-Aguilar, David P. Hodson, A. Dewi Hartkamp⁴

Introduction

A geographic information system (GIS) is a computer-based system for storing, manipulating, and presenting data in a spatial (geographic) context. A GIS is much more than a computerized drafting system for producing paper maps. Spatial data can be analyzed with diverse methods, and outputs may be presented in formats ranging from summary statistics or tables to animation of maps varying with time. Although use of GIS shows explosive growth in such differing fields as marketing of consumer goods, tax collection, disease control, and forest management, the analytic capabilities of GIS hold special promise for agricultural research. Agriculture is one of the most spatially intensive activities of humankind, and spatial variation in production conditions, whether driven by climatic, edaphic, biotic, or socioeconomic factors, is of paramount concern in agricultural research. Furthermore, increasing concerns over the impact of agriculture on natural resources, both on-farm and in adjacent regions, dictate that agricultural research focuses even more on spatially varying processes.

This chapter first presents an overview of basic concepts of GIS, then examines applications of GIS to agricultural research and, finally, considers why GIS has not yet realized expectations in agricultural applications.

6.4.1 An overview of GIS

For a full introduction to GIS, readers should consult excellent texts such as Aronoff (1995) and Burrough and McDonnell (1998). This brief review will only attempt to highlight some concepts and issues that have particular relevance to agricultural scientists faced with a decision on whether to apply GIS in their research.

Data sources

Data may be entered into a GIS from sources including digitizing tablets, scanners, and satellite imagery. Historically, digitizing has been a major activity of many GIS laboratories, but this situation is changing rapidly as data become available through the Internet or in other digital forms. Examples of spatial data sets readily available for some regions include those derived from climate, soil, land cover, and elevation (table 6.3).

Use of the Geo-Positioning System (GPS) merits special attention since this system offers the prospect of adding a spatial component to research at a very low cost. Hand-held GPS units process signals from a system of navigation satellites maintained by the US government and calculate the user's position with an accuracy of approximately 100 m (Herring 1996). Larger, more costly units can provide accuracy at submeter levels. GPS units are now widely used by germplasm collectors, pa-

4. Natural Resources Group, International Maize and Wheat Improvement Center

Table 6.3: Examples of Data Sets of Spatial Factors Relevant to Agriculture

Variables	Region	Grid cell size or map scale	Source	Comments
Max/min temperature, precipitation, potential evapo-transpiration	Central America & Mexico	2.5 km	J. Corbett, personal communic. 1997	Available as monthly data and for various types of seasons (e.g., coolest three months, driest three months).
	South America	9 km	J. Corbett, personal communic. 1997	Available as monthly data and for various types of seasons (e.g., coolest three months, driest three months).
	Africa	5 km	Corbett (1995)	Available as monthly data and for various types of seasons (e.g., coolest three months, driest three months).
	Asia	4 km	IIMI (1997)	Global 16km IWMI (2000) available as monthly data.
Elevation	Global	1 km	USGS (1997a)	
Soil	Global	1:5 million	FAO (1995)	Includes pH, depth, water holding capacity, but small scale is a major limitation.
Land cover	Global	1 km	USGS (1997b)	Derived from remote sensing (very approximate).

thologists, soil scientists, and others, and one can anticipate that use of GPS will be as routine as taking photographs.

Data from satellite-based sensors or other forms of remote sensing are frequently cited as promising sources for data on land use, soil degradation, and climatic or weather conditions. The potential of these methods is illustrated by the effort of the United States of America Geological Service (USGS) to create a global land cover database, working primarily from weather satellite data (USGS 1997b), which provides data on over 200 types of vegetation formations or other land covers using an approximately 1 km-square grid cell size. Nonetheless, many projects have found that handling primary data from remote sensing is extremely difficult.

Spatial data are stored either in a vector or raster format. In the vector format, data are recorded as coordinates of points (e.g., locations of a town or site of a germplasm collection), lines (roads, rivers, or drainage paths) or polygons (boundaries of a field or a body of water) along with associated attributes. The spatial entities represented thus correspond roughly to objects that would be found on a map. In the raster, or gridded, format, a region is subdivided into regularly spaced cells in a rectangular array. Data values are assigned to specific row and column positions within the array.

Consideration of which formats to use is an important early step in developing a GIS project. The vector format has advantages in terms of storage efficiency and the ease

with which map scales and projections may be adjusted (table 6.4). It is particularly suited for data that is inherently discontinuous, such as locations of political boundaries, roads, rivers, and other linear features. The raster format is better suited for handling quantitative data that will vary continuously over a region. In agricultural work, where researchers frequently wish to examine variation in quantitative data such as precipitation, soil pH, and slope, this format is especially useful. Fortunately, most GIS systems now allow both vector and raster data to be manipulated together in a single project.

Selecting an appropriate map scale often requires balancing research objectives that require a high resolution (large scale) against constraints of data availability, storage and processing. Table 6.5 provides some guidelines on how different map scales relate to research objectives.

Since the Earth is a spheroid, any attempt to present data on a flat surface requires that the spatial data be transformed. This transformation process basically involves trade-offs between conserving relative areas and preserving the shapes of regions. Depending on the expected final product or use, different projections such as Uni-

Table 6.4: Comparison of Vector and Raster Data Formats

Characteristic	Vector	Raster
Efficiency of data storage	High	Moderate
Handling of point and line data	Excellent	Poor
Handling of quantitative data showing continuous variation	Fair	Excellent
Handling of quantitative data with discontinuous variation	Good	Excellent
Ease of overlay operations	Low	High
Ease of representations of topological relations	High	Medium
Ease of re-projection	High	Low
Ease of re-scaling	High	Low
Ease of use with digital imagery	Low	High

Table 6.5: Comparison of Different Map Scales

Map scale	Arcs	Equivalent grid cell size	Useful for studies within
1:5,000,000	1.5°	25 km	continent
1:1,000,000	20'	5 km	country
1:250,000	4'	1 km	region within country
1:50,000	1'	200 m	province/state level
1:10,000	10''	40 m	watershed/municipality
1:500	0.5''	2 m	single field

° = degrees; ' = minutes; '' = seconds

versal Transverse Mercator (UTM) and Robinson will be required. For initial data processing, many groups prefer to manage data using latitudes and longitudes, rather than projecting them. Unfortunately, many data are supplied with incomplete or incorrect information on map projections.

Analytic tools of GIS

Aronoff (1995) classifies analytic functions of GIS as including data retrieval, overlaying, neighborhood analyses, and connectivity functions. The full classification is listed in table 6.6 with examples relevant to agriculture. Simple retrieval of data, classification and measurements are examples of retrieval functions. Layers of spatial data may be overlaid using either arithmetic or logical relations. Neighborhood analyses include such functions as searches based on distance criteria, determining whether points or lines are within polygons, and interpolation procedures. Connectivity functions emphasize the spatial relations among features and include tools for evaluating proximity and paths along networks of roads, streams or other linear features. While this framework has clear heuristic value, few GIS projects will select these functions explicitly. Rather, they will be invoked in highly dynamic and interactive fashion as part of the research process.

Presentation of results

While conventional paper maps are still an important product of GIS, results are increasingly provided in electronic formats. This may be as simple as a series of map images stored in electronic form or, as is increasingly the case, an intermediate software and data product that allows end-users to view and manipulate the maps to suit their particular needs. The Internet now has many examples of applications that allow users to view data on maps interactively. Advantages of this approach include reduced publication costs, ease of updating, and maximizing the utility of the data products.

6.4.2 Applications to agricultural research

It is difficult to imagine activities in field-oriented agricultural research where GIS could not be applied with benefit. In research planning, GIS can assist in identifying target regions, prioritizing sites either as “hot spots” or as representative of key regions, and defining efficient sampling frameworks. During actual research activities, GIS may provide the foundation for a data collection and management system, as well as allowing careful monitoring of research progress. And during analyses, the diverse analytic tools of GIS can be exploited as well as the ability of GIS to link spatially referenced experimental results to data on climates, soils, and other factors. In the sections that follow, three examples of applications of GIS to agricultural research are provided. For further examples, the report compiled by UNEP GRID-Arendal (UNEP 1997) contains 43 case studies on a wide range of topics and covering diverse regions and spatial scales.

The Kenya Maize Database Project

In 1992, the Kenya Agricultural Research Institute (KARI) and CIMMYT established the Kenya Maize Database Project to harness GIS in assessing the successes and fu-

Table 6.6: Summary of GIS Analytic Functions and Agricultural Applications

Function	Description	Examples of agricultural application
Retrieval		
Retrieval	selecting data without modifying spatial relations	determining soil pH of a polygon or grid cell
Classification	grouping a set of features based on retrieved attributes	defining land-use suitability classes in a watershed
Measurement	measuring lengths or areas or counting features in a given class	determining the area of a region identified as suitable for zero tillage
Overlay		
Arithmetic	combining layers through arithmetic operations	creating a layer representing total annual precipitation by combining 12 monthly precipitation surfaces
Logical	combining layers through logical operations	creating a layer showing regions with soil pH > 6 and annual precipitation ≤ 1200 mm
Neighborhood		
Search	assigning a value to target features based on criteria related to a search area	identifying farms within a 10 km radius of a fertilizer distribution center
Point-in-polygon	determining whether a point is contained in a given polygon	determining which germplasm accessions were collected from sites with soils of volcanic origin
Line-in-polygon	determining whether a line passes through a given polygon	determining whether an irrigation network crosses all farms in a district
Topographic functions	defining traits such as slope and aspect	creating a surface showing variation in slope within a watershed
Thiessen polygons	defining "areas of influence" around points	in the absence of other information, interpolating data from various weather stations over a region
Interpolation	predicting values at specific locations based on known values at neighboring locations	interpolating data from various weather stations over a region but assuming certain statistical relations and including effects of elevation
Contour generation	creating contour lines from point data	creating a contour map of soil depth based on point samples
Connectivity		
Contiguity	evaluate traits of features that are connected	determine the total area of uninterrupted buffer zone (e.g., not crossed by roads) along a river
Proximity	measure the distance between features according to specific rules	locating buffer zones for nonagricultural use along a river margin
Network	analyzing traits of paths defined by interconnected lines	finding the shortest route for distributing seed samples to multiple farms
Spread	a more complex form of proximity analysis where that evaluates a function that accumulates with distance	finding a cost-effective route for distributing seed samples by accounting for road quality, terrain, and distance
Seek	performs a directed search according to specific criteria	trace the probable flow of water based on a digital elevation model
Intervisibility	identifying areas that can be seen from a specific view position	locating a food processing facility where it will not be visible from an important archaeological site
Illumination	portraying the effect of illuminating a surface from a nonvertical position	representing a surface of expected soil lost in a visually dramatic manner for policymakers
Perspective view	portraying a surface from a non-vertical viewing position	representing a surface of expected soil lost in a visually dramatic manner for policymakers

Note: Based on Aronoff (1995)

ture needs of KARI's maize research (Lynam and Hassan 1998). GIS was applied both to define an efficient sampling network to assess KARI's impact and to characterize maize systems in Kenya. Climate surfaces were created, and a cluster was conducted to define eight maize production zones (Corbett 1998). The project was successful not only in documenting the impact of KARI's maize germplasm, but also served in highlighting many opportunities to increase impact both through germplasm and crop management. In particular, it identified a dry mid-altitude to highland region where farmers lacked adequate germplasm (Hassan et al. 1998). This led to changes in maize breeding priorities and a revision of the initial climatically based zone classification.

African Country Almanac—Putting GIS technology within reach of NARS

Information systems such as GIS are often unavailable to NARS due to lack of the appropriate software/hardware, data in digital format, or technical expertise. This section describes a new research resource for several sub-Saharan Africa countries: the African Country Almanac, which overcomes these problems but still delivers powerful GIS functionality. The African Country Almanac is a joint development project between John Corbett's group at the Integrated Information Management Laboratory, Texas A&M University, and CIMMYT's Natural Resources Group. The Almanac arose from the UNIX-based Spatial Characterization Tool (SCT), developed by the Texas group (Corbett and O'Brien 1997). Unlike the SCT, however, the Almanac is a completely stand-alone, easy-to-use system that requires no specialized GIS software or expertise. The Almanac was created using MapObjects (ESRI, Inc.) and Visual Basic (Microsoft, Inc.) and is distributed on CD-ROM. Users of the Almanac only need a PC running Windows 95 or NT, and a CD drive. Almanacs are currently available for Angola, Ethiopia, Kenya, Liberia, Nepal, Sierra Leone, Tanzania, Uganda, Zambia, and Zimbabwe, and several other countries are in preparation.

The Almanac provides three main categories of analytic tools: visual display and querying, site similarity analyses, and zonal analysis. Over 50 layers of spatial data are available for visualization and query with the Almanac. These include climate, soils, land use, population, elevation, infra-structure, political boundaries, and environmental data. Datasets in the public domain are used in nearly all cases. In addition, users may add their own existing spatial data. The interpolated climate surfaces, created by Corbett and coworkers (Corbett 1995) using tri-variate thin plate splining techniques (Hutchinson 1995) are an integral part of the Almanac. These provide long-term (30-year) monthly averages for a range of climate variables (precipitation, evapotranspiration, maximum and minimum temperature) and incorporate elevation as a covariable at a resolution of 3 arc-minutes (approximately 5 x 5 km square). Various climate models have been derived from these climate surfaces including: "Optimal season": the five consecutive months with the highest precipitation to potential evapotranspiration ratio (P/PE); "Trigger season": the longest run of consecutive months where the P/PE ratio is greater than 0.5; "Dry season": the longest run of consecutive months where the P/PE ratio is less than 0.5; and "Quarters": the three wettest, driest, warmest, and coolest months.

The Almanac allows rapid display of any of these layers of data and also the creation of customized maps (e.g., figure 6.2), which can then be exported and incorporated

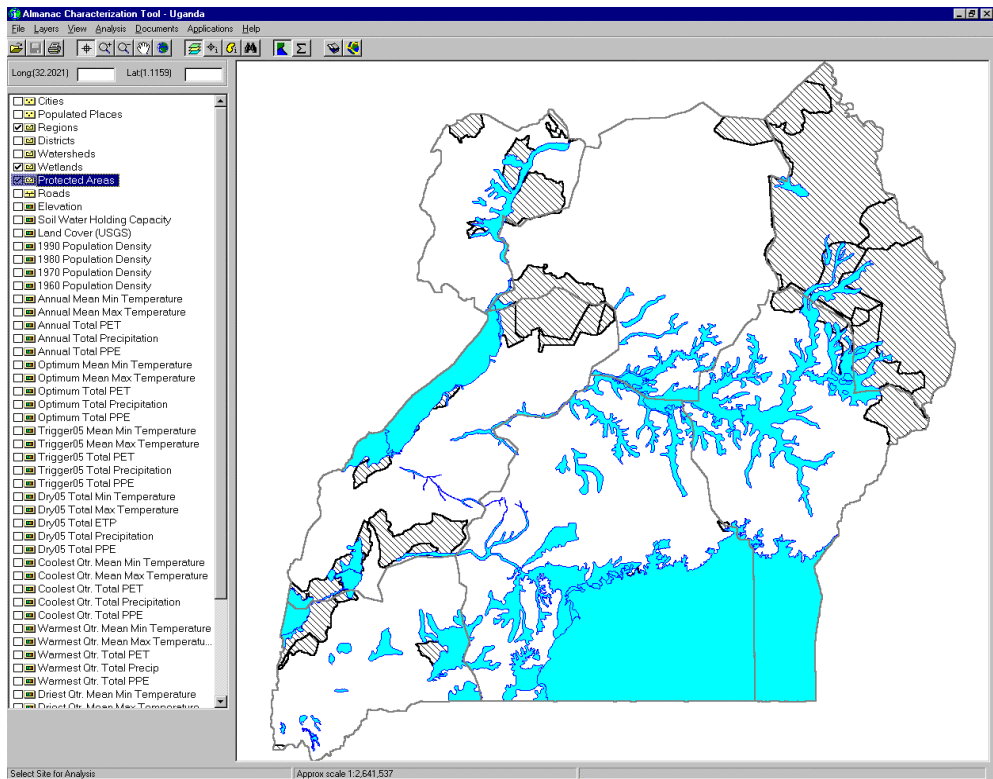


Figure 6.2: Almanac screen display illustrating protected areas, wetlands, and regional boundaries for Uganda

into other software packages, including word processing and presentation packages. Data from a specific location, either obtained by entering a latitude and longitude or simply by pointing and clicking with a mouse, can also be viewed and exported.

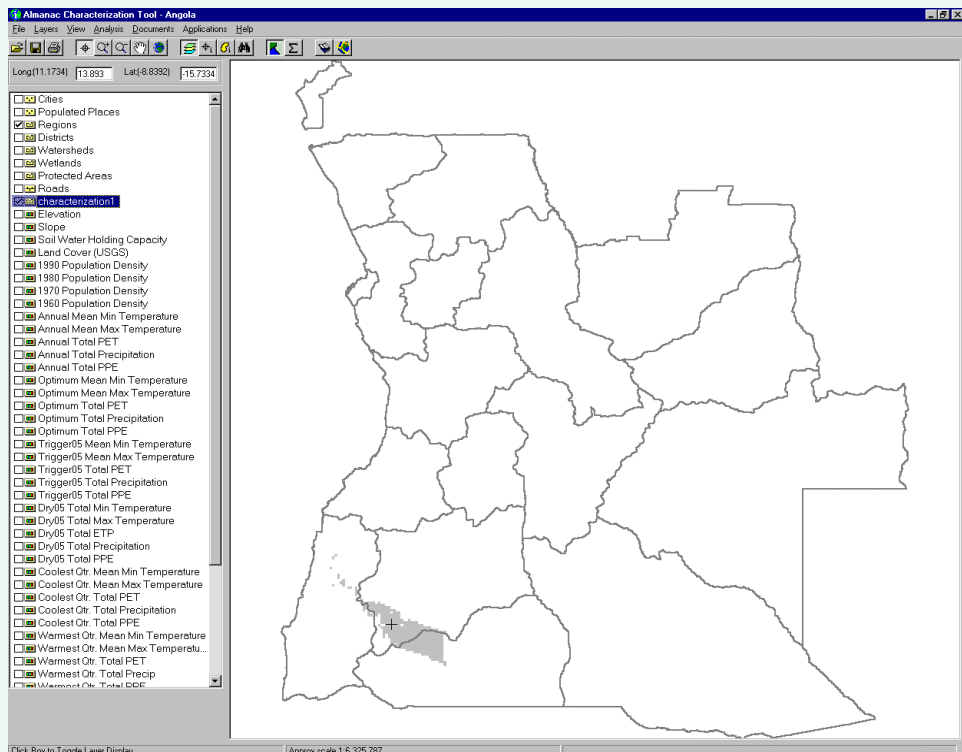
The site comparison function allows users to determine regions of similarity (e.g., in terms of climate during a particular season) to a user-specified site (see box 6.1.)

The zonal analysis function allows users to define agroecological zones by specifying upper and lower limits for climatic, edaphic, demographic, ecological, and topographic variables. Zone characterization may be as complex or simple as the user requires. Using the Almanac, it is just as easy to determine either of the following agroecological zones: (1) where in Uganda has favorable season precipitation less than 1500 mm, an evapotranspiration range of 500–1000 mm, medium or high soil water-holding capacity, landuse that is not forest, and an elevation range of 500–1500m; (2) where in Angola has an annual precipitation less than 2000mm and an elevation less than 1000m.

Box 6.1: The African Country Almanac: A Hypothetical Case Study Illustrating the Application of the Site Comparison Function

Examination of data from a CIMMYT International Maize Trial at Chiange, Angola, (searchable data contained within the Almanac documentation section) reveals that the three highest yielding varieties in this trial were Poza Rica 8022, Guaymas (1) 8022, and El Paraiso 7929. As a result of this trial, Angolan maize researchers might want to determine where else in Angola these varieties may perform well. The site comparison tool of the Almanac enables them to select the location Chiange and then generate a zone of similarity, in this case $\pm 10\%$ of values for favorable season climatic conditions [but users have total flexibility in terms of how much variance and which climate model to use]. The result of this analysis is illustrated in figure 6.3. In addition to the zone being mapped, the Almanac generates a range of statistics relating to the zone (e.g., total zone area; mean, min., max., and standard deviation for all variables) in exportable text file format. The inclusion of ArcExplorer software (ESRI, Inc.), enables users to display and query such a similarity zone in many different ways. For example, the climate similarity zone for Chiange could be displayed in terms of differing population density, which may assist researchers in determining not only where climatic conditions may favor the new varieties, but also where they may have maximum impact.

Figure 6.3: Almanac screen display illustrating the favorable season climate similarity zone ($\pm 10\%$ of climate variables) for Chiange, Angola (+), and district and regional boundaries



Additional examples of applications of the Almanac tools include the following:

- assisting in the selection of research sites that are most representative of particular regions within a country
- determining where promising new agronomic practices or germplasm may be deployed with the highest probability of adoption
- locating agroecological zones which may be suitable for germplasm adapted to certain environmental conditions
- setting breeding program priorities in cases where it is discovered that large agroecological zones exist for which there is no suitable germplasm currently available

The Almanac is being further improved to expand the number of countries covered and increase the functionality and types of data. Subsets of major databases being developed at CIMMYT—International Wheat Information System (IWIS), Sustainable Farming Systems Database (SFSD), and International Maize Information System (IMIS)—will be incorporated into the Almanac. Additional analytical functions may include enhanced querying facilities, within-polygon statistical capability, transect analysis, and a batch tool to extract data from multiple sites simultaneously.

Interfacing agronomic models to GIS

Agronomic models, particularly crop simulation models, are logical partners of GIS since their point-based predictions can be usefully extended over regions. Typically, a GIS is used to provide input data for the models, the models are then run over the appropriate regions, and then GIS is again used to display the simulation outputs. Since a large amount of data are exchanged in this process (often many megabytes), data input and output structures have to be carefully defined. The International Consortium for Agricultural Systems Applications (ICASA) has proposed standards to facilitate such work (Ritchie 1995). Two examples of software systems that use ICASA standards to link GIS and models are Spatial Analysis Tool (Thornton et al. 1997) and AEGISWIN (Engel et al. 1997).

We have used a similar tool developed at Texas A&M (J. Corbett and S. Collis, personal communication) to simulate the potential yields of rainfed maize for the state of Jalisco in Mexico (figure 6.4). Climate data from approximately 200 locations were first interpolated using the spline method of Hutchinson (1995). The interpolated data were then subject to cluster analysis to identify sets of environments with similar climates. These data were then overlaid with soil profile descriptions linked to soil data from the FAO Digital Soil Map of the World to use as inputs to the CERES Maize model. Although only yield limited by possible water deficits was simulated, future work will include effects of tillage practices and residue management both on crop yields and on runoff.

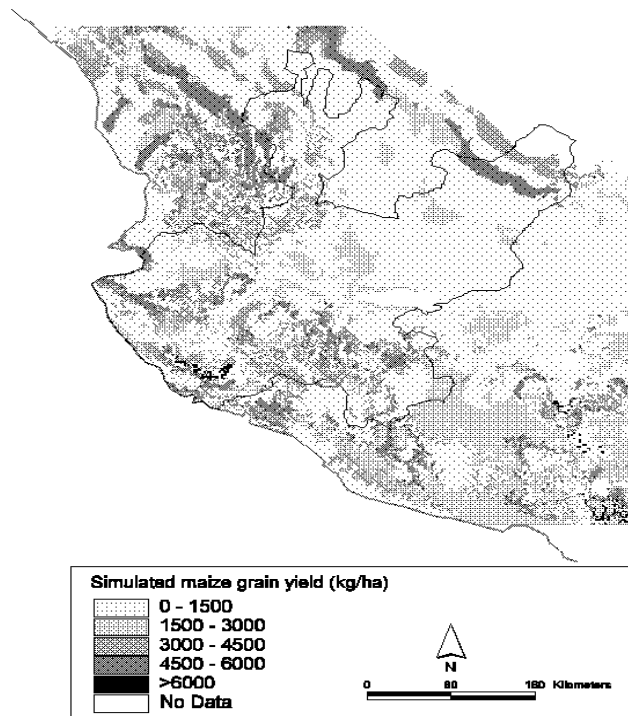


Figure 6.4: Variation in yield potential of summer rain-fed maize in Jalisco, Mexico

6.4.3 Limitations to effective use of GIS

Notwithstanding some notable successes, there is widespread sentiment that the impact of GIS in agricultural research has fallen well below expectations. We cite four major factors that influence this:

1. GIS awareness. Because GIS is a recent development, many agricultural scientists simply lack an awareness of how GIS can benefit their research activities. Experience at CIMMYT is that “awareness building workshops,” of one or two days duration, are invaluable in orienting scientists.
2. Data availability. Although the availability of agronomically useful data is increasing rapidly, there are many surprising deficits. Most notable are the lack of more detailed soil data than provided in the FAO Digital Soil Map and of data on crop distributions and production levels.
3. Hardware and software costs. Hardware costs are decreasing rapidly, especially storage devices and computers per se. However, specialized equipment such as plotters and digitizers are still too expensive for many groups. Software vendors have generally preferred to increase the functionality of GIS packages while maintaining prices at relatively high but constant levels. Nonetheless, products such as ArcExplorer and the Country Almanacs, which are distributed free of

charge, offer the possibility of minimizing the number of full GIS packages that a research project must maintain.

4. The institutional setting of GIS capabilities. Some degree of centralization makes sense for efficient management of base data and the more expensive software and hardware. However, rapidly declining costs and increased availability of data argues for increasingly placing GIS capabilities directly into the hands of researchers. The role of GIS specialists thus should evolve to one of developing new tools and key data sets and of providing support for particularly difficult GIS applications. In this case, access to specialists might include problem specific collaborations with other research institutions or outsourcing to commercial services.

References

- Aronoff, S. 1995. Geographic information systems: A management perspective. Canada: WDL Publications.
- Burrough, P.A. and R.A. McDonnell. 1998. Principles of geographic information systems. Oxford: Oxford University Press.
- Corbett, J.D. 1995. Monthly evapotranspiration surfaces for Africa, v1.1. Climate coefficients created using ANUSPLIN (CRES, ANU, Canberra). (CD-ROM). Nairobi: International Centre for Research in Agroforestry.
- Corbett, J.D. 1998. Classifying maize production zones in Kenya through multivariate cluster analysis. In *Maize Technology Development and Transfer - A GIS Application for Research Planning in Kenya*. Edited by R.M. Hassan. Wallingford: CAB International.
- Corbett, J.D., and R.F. O'Brien. 1997. The spatial characterization tool. Texas A&M University System, Blacklands Research Center Report No. 97-03. (CD-ROM).
- Engel, T., G. Hoogenboom, J.W. Jones, and P.W. Wilkens. 1997. AEGIS/WIN - A program for the application of crop simulation models across geographic areas. *Agronomy Journal* 89(6).
- FAO. 1996. Digital soil map of the world and derived soil properties. (Version 3.5, November 1995). Rome: Food and Agriculture Organization of the United Nations.
- Hassan, R.M., J.D. Corbett, and K. Njorge. 1998. Combining geo-referenced survey data with agroclimatic attributes to characterize maize production systems in Kenya. In *Maize Technology Development and Transfer - A GIS Application for Research Planning in Kenya*. Edited by R.M. Hassan. Wallingford: CAB International.
- Hutchinson, M.F. 1995. Interpolating mean rainfall using thin plate smoothing splines. *International Journal of GIS* 9: 305-403.
- Herring, T., 1996. The global positioning system. *Scientific American*. Feb.: 32-38.
- Koontz, H., C. O'Donnel, H. Weihirch. 1994. *Management*. McGraw-Hill.
- IIMI. 1997. *World Water and Climate Atlas: Asia Version 1.0 CD-ROM*. Colombo: International Water Management Institute.
- Lynam, J. and R. Hassan. 1998. A new approach to securing sustained growth in Kenya's maize sector. In *Maize Technology Development and Transfer - A GIS Application for Research Planning in Kenya*. Edited by R.M. Hassan. Wallingford: CAB International.

- Ritchie, J.T., 1995. International Consortium for Agricultural Systems Applications (ICASA): Establishment and purpose. *Agricultural Systems* 49: 329-335.
- Thornton, P.K., H.W.G. Bootink, and J.J. Stoorvogel. 1997. DSSAT Version 3.1 Spatial Analysis. In DSSAT version 3. Draft volume 4. Edited by G.Y. Tsuji and G. Uehara. Honolulu, Hawaii: IBSNAT, University of Hawaii.
- UNEP. 1997. GIS in agricultural research: Awareness package. Edited by N. Denisov, C. Heberlein, L. Czarán and O. Simonett. UNEP/DEIA/TR.97-9. (Also available at <http://www.grida.no/prog/global/cgiar/htmls/awpack.htm>)
- USGS. 1997a. GTOPO30 Documentation.
<http://edcwww.cr.usgs.gov/landdaac/gtopo30/README.html>.
- USGS. 1997b. Global Land Cover Characterization.
<http://edcwww.cr.usgs.gov/landdaac/glcc/glcc.html>

