Methodologies for Integrating data across geographic scales in a Data-rich Environment: Examples from Honduras

E.B. Knapp, W. C. Bell, G. Leclerc, H.M. Ravnborg, J.A. Cox, A. Nelson, P. Couillaud, S. Nath, G. Rosenberg, B.P. Verma, D. Nute[1].

1 Presentation of progress of a special-funded project funded by the Eco-Regional Trust Fund to Support Methodological Initiatives for a review workshop, "Methodological Research at the Ecoregional Level", ISNAR, The Hague, April 20-22, 1998.

Methodologies for integrating data across geographic scales in a data-rich environment: Examples from Honduras[1]

E.B.Knapp, W.C.Bell, G.Leclerc, H.Ravnborg-Munk, J.A.Cox, A.Nelson, P.Couillaud,[2]
S.Nath, G.Rosenberg, B.P.Verma, D.Nute [3]

Introduction

Throughout tropical American hillsides, the vast majority of inhabitants scratch out a living from small farms on infertile soils, often on steep slopes vulnerable to erosion. Shortages of just about everything – land, water, labor, inputs, cash, credit, schools, clinics, roads, transport and communications – frustrate their daily efforts to escape from poverty. It is fair to ask, in this difficult setting, how can problems, areas and beneficiaries be prioritized for development?

The CIAT Hillsides Project has a vision of future development efforts that are characterized by multi-institutional alliances for establishing consor*tia with* the analytical capacity to plan and support community watershed resource management. In cases of conflicting interests, a process of "deal making", in which costs of resource conservation are balanced by concrete incentives, can be catalyzed by a process of successive refinement of information and analysis.

The Problem

One of the greatest challenges for CGIAR researchers, government and non-government organizations and resource-poor farmers is the need to adopt perspectives that transcend field and farm boundaries and accept solutions necessitating some form of collective action among landscape users.

Agroecoregional research literature, too ample to cite properly but well represented in conference proceedings by Fresco, Stroosnijder, Bouma and van Keulen (1994), is in agreement that traditional definitions of research site and data collection must include different spatial and temporal scales targeted by different stakeholders. So far, with notable exceptions (Veldkamp and Fresco, 1996), "across-scale analysis" has resulted

[1] Presentation of progress of a special-funded project funded by the Eco-Regional Trust Fund to Support Methodological Initiatives for a review workshop, "Methodological Research at the Ecoregional Level", ISNAR, The Hague, April 20-22, 1998.
[2] Soil and systems specialist and project manager; CIO, GIS specialist, rural sociologist, research fellow, research fellow, research fellow, research fellow, respectively with the Centro Internacional de Agricultura Tropical (CIAT), Cali, Colombia.
[3] Post-Doctor, Post-Doctor and Professor respectively, Biological & Agricultural Engineering, and Director, Artificial Intelligence Center, University of Georgia, Athens, Georgia.

in little more than independent characterization of ever larger geographical areas in less and less detail - from plots through landscapes. Research needs to emphasize *the organizing principles and functional relationships that structure* multiple-scale systems. A key aspect of this strategy, based on hierarchical systems theory, is an increased emphasis on sample surveys and controlled prospective and retrospective studies vis-à-vis laboratory and field experimentation.

Partnerships in Ecoregional Research

At the initiation of work in Honduras and Nicaragua in 1993, the CIAT Hillsides Project organized a "Consultative Group" of local stakeholders[4] to give general guidance and technical consul to the project. The Consultative Group has met yearly since 1993 to comment on annual project workplans. More detailed planning is the responsibility of "Local Operating Committees" who ultimately have responsibility for developing community-scale action plans.

Significant partnerships in research exist in Central America. During March 1-3, 1995, scientists working on sustainable agriculture and economic development in the hillsides of Central America under the auspices of international and regional centers assembled in Trujillo, Honduras. Participants were from CATIE, CIAT, CIMMYT, EAP, IFPRI, IICA, PASOLAC, PRM, and PROFRIJOL[5].

The meeting complemented priority-setting workshops with national program and other partners. Bilateral discussions specified collaborative activities focusing on processes of institutional collaboration of international and regional centers in an ecoregion. Participants were seeking to improve their own research by clarifying the process of collaboration. In addition it was recognized that colleagues in national programs, NGOs, donors and research managers would appreciate greater clarity in these mechanisms.

In June 1996, the Swiss Agency for Development Cooperation (SDC) commissioned an external review of activities carried out in Honduras and Nicaragua by the CIAT Hillsides Project. To quote the report:

> "The detailed research at the site level needs to be placed in the geographic, biological and physical context that includes making an inventory of secondary information at local, national and regional level, and this requires compilation in digital databases to permit information and analysis of the inter-relationships that can exist.

---

[4] Current members include donor, university, NARS, private industry and project representatives.
[5] Centro Agronómico Tropical de Investigación y Enseñanza, Centro Internacional de Agricultura Tropical, Centro Internacional de Mejoramiento de Maiz y Trigo, Escuela Agricola Panamericana, International Food Policy Research Institute, Instituto Interamericano para Cooperacion Agricola, Programa para Agricultural Sostenible en Laderas de Centroamerica, Programa Regional de Maiz, Programa de Frijol.

The activities carried out are the result of notable inter-institutional cooperation by virtue of the expressed needs of the public consulted, especially among those who could better take advantage of the information.

It should be recognized that the compilation of information and its geographic referencing has resulted in the possibility to evaluate secondary information that was disperse and unknown and which is now accessible to be consulted and analyzed.

The time to carry out this activity is opportune because a clear demand not only exists, but isolated efforts are being developed.... Lack of coordination could mean costly investments and low efficiency.... "[6]

It was as a consequence of these formal activities that CIAT approached the Ecoregional Trust Fund for support.

Project Outputs

The goal of the CIAT Hillsides Project is to institutionalize community-led management of watershed resources for more productive, sustainable and healthy hillside agroecosystems. This is being accomplished through financial and research support from a number of initiatives including the Ecoregional Trust Fund for Methodological Research[7].

The stated purpose of the Trust Fund Project is to develop and document principles and procedures for building a scale-consistent database and for performing multi-scale characterization of agroecosystems. Honduras was selected as the development site due to the relative availability and willingness of data "owners" to share data.

Project outputs are listed in the logical framework shown in Appendix 1 together with indicators of completion.

Output 1: Inventorying Resources and Database Development

It is not uncommon that that aspect least thoroughly understood rather than most thoroughly understood dictates the efficiency of the process of planning development projects. The consequence of lack of authoritative information can lead to two common problems or "information pathologies"; i) false analogies of ecosystem characteristics and ii) overgeneralization. Ecosystem characteristics refer here to potential production, resilience, sustainability, *et cetera*. Overgeneralization refers to lack of understanding of

[6] "Informe de la Evaluaciòn Externa", M.Jaurequi,F.O.Osorio, U. Scheidegger, 3-15 June, 1996.

[7] Other donors funding the CIAT Hillsides Project include the Swiss Agency for Development Cooperation (SDC). The International Development Research Centre of Canada (IDRC), the Inter-American Development Bank, and the Royal Danish Ministry of Foreign Affairs (DANIDA),

the nature and range of ecosystem variability. Therefore, the first critical element of success in the process of setting achievable outputs for community action, not to mention measurable milestones, is an accurate inventory of available resources. Without a resource inventory, no baseline condition can be described, no trends defined, no progress measured and gaining commitment of key individuals and organizations will be problematic.

During the first year of the Trust Fund Project, several tens of alpha-numeric and GIS databases were reviewed, edited, consolidated and georeferenced. Databases and coverages were exchanged and made available to numerous collaborators by means of our Web site and CD-ROM.

Data "mining" challenges

Data collection in tropical America is not as institutionalized to the degree it is in Europe and North America. Because of severe resource constraints, national agricultural and population censuses are not carried out regularly nor analyzed to the degree one might wish. Detailed climate and soils data are not usually available for more than the most intensively cultivated commercial areas and large scale (1:50,000) topographic maps, when digitized, need to be checked for common errors like: rivers overlapping "valley" edges, contour lines at map edges are not always congruent, contour elevations may be mis-coded and inconsistent digitizing of river direction (important for digital terrain model (DTM) algorithms). Not as obvious, perhaps as the aforementioned errors, which are identifiable when passed on to digital spatial coverages, is the issue of identification of original sources used in the creation of DTMs. In the process of checking DTMs, we have found some models are dominated by the grid of digitized points generated from the original map. This effect can be seen as spikes in the histogram of almost any DTM fitted to digitized or scanned contour data. We have found that some DTMs were quite likely not constructed from the original sources and resolution they were supposed to have been.

Enthusiasm, however, for new analytical software like GIS applications, without critical analysis easily lead to data misuse and false conclusions. Take the example of the large-scale map of watersheds in Fig 1. A rather startling revelation is that the areas of two typical watersheds, as determined by a routine analysis of the "flat" map representation, amounted to 11280 and 5130 hectares respectively (Fig 1a). The areas calculated from the three-dimensional representation of the DTM-GIS model are 13600 and 6115 hectares (Fig 1b), a difference of 20%!

Other data collection activities include successive refinement of remote sensed imagery using SPOT panchromatic, RADARSAT and digital air ortho-photographs. Work with digital ortho-photographs has focused on cost and accuracy tradeoffs for deriving terrain slope estimates. Terrain slope is one of the most fundamental variables in agriculture as it affects such variables as cost of preparing seedbed, erosion, runoff, mechanization and access. Despite its importance, slope remains elusive as estimates change continuously depending upon the distance over which it is averaged, grid cell size and scale of the original elevation data (Berry, 1993). A study was carried out for a watershed to compare accuracy that can be expected for slope and altitude values derived from low cost DTMs. Eight gridded DTMs were generated from digitized contour maps at a range of scales (1:10000, 1:25000, 1:100000 and 1:200000) and a range of contour intervals (25 m, 50 m and 100 m). A control DTM was produced from large-scale aerial photographs (1:28000) and field verified using 91 differentially measured GPS ground points. The control DTM showed a vertical RMSE well within USGS accuracy standards. In addition to cell size and slope relationships, cost of production of DTMs and accuracy of results were determined. Some of the conclusions reached were i) substantial savings in time and expense in DTM production accrue from digitizing every *n-th* contour as long as the new interval is not more than 25 m wider than the original interval when modeling altitude and/or slope, ii) regarding slope determination, contour interval has more influence than map scale, iii) cartographic data sources at scales equal or greater than 1:100000 and with contour intervals equal or greater than 100 m do not provide sufficient detail to usefully represent slope in community watersheds in hillside agroecosystems.

Output 2: Multi-scale characterization for targeting problems, priority areas and beneficiaries.

To date, census data and *ad hoc* household surveys have been primary mechanisms for inventorying a landscape resource base. Citation of census data can be found in many publications including those of IARCs. We have been particularly interested in what census/survey data really measure. In technical terms, what system is being quantified and where does the system fit in the hierarchy of biophysical and socio-economic systems from farm to country scales?

Significant progress has been made towards assessing the role of national agricultural census data of land use for inventorying the state of community-scale watershed systems. The studies depend upon quantifying relationships between data extracted from remote sensed images and data available from agricultural censuses. Very detailed work is most advanced for the Tascalapa watershed in Yoro, Honduras, but is being repeated for validation in three other targeted watersheds.

A land map of two *municipios* that contain the Tascalapa watershed, based upon field-validated air-photograph interpretation, was used as a base-line reference for generating spectral signatures for ten land cover/land use classes. The Landsat TM classification for the watershed study area showed good agreement with the high-resolution air photograph classification.

To compare land cover of the watershed derived from TM imagery with agricultural census data, it was necessary to expand the area classified to the extent of multiple *municipios* - political boundaries within Departments. This is because published census data are only available aggregated for traditional political units. This severely limits there use, making it impossible to characterize non-traditional landscapes like community watersheds, using currently available government census and statistical data.

For the methodological study, the original Landsat TM image was re-sampled to produce a range of pixel resolutions, i.e., 32m, 50m, 100m, 250m, 500m and 1km. Land cover classification estimates beginning with different resolutions were compared with census estimates.

Results relating pixel resolution show estimates never vary by more than 10%, and, with the exception of "perennials" (CV=36), variation across pixel resolution within land cover class is small with coefficients of variation well below 10 (Table 1).

Table 1. Land cover characterization for the two *municipios* of Yorito and Sulaco, which contain the Tascalapa watershed, Department of Yoro, Honduras, showing affect of pixel resolutions.

Land Cover	1993 Agric. Census	TM Image 32m	TM Image 50m	TM Image 100m	TM Image 250m	TM Image 500m	TM Image 1 km
				ha			
Annuals	4,855	6,084	6,103	6,100	6,175	6,372	6,640
Perennials	2,455	481	471	473	456	393	113
Pasture	6,567	2,866	2,865	2,886	2,855	2,966	3,512
Fallow (>1yr)	2,074	3,251	3,527	3,251	3,617	3,493	2,939
Forest	954	13,766	13,219	13,752	13,510	12,743	13,393
Sparse conifer rangeland		8,843	8,876	8,851	8,843	9,251	9,537
TOTAL	16,905	35,291	35,061	35,555	35,456	35,218	36,134

Contrary to expectations, there is no obvious relationship between estimates from remote imagery and figures derived from the census. This was expected in the case of "forest-cover", as farmers do not privately own much of the forested land, although small *woodlots* managed for wood products are common.

A second study compared estimates for three degrees of differentiation of land cover type derived from remote imagery and published census data for the Department of Yoro. Secondary data for the Department of Yoro indicated it would have a significantly different geomorphological and land cover profile than the two *municipios* of Yoro and Sulaco contained within. It was decided that land cover estimates would be carried out using a 100m pixel resolution, given that decreasing pixel resolution resulted in only minor variation in estimates of major land cover classes. The comparison included three degrees of detail, i) agricultural land use/non- agricultural land use; the latter comprising settlement/bare land and sparse conifer range land, ii) cropland including bare fields/ pasture/ fallow/ forest and woodlots/ non-agricultural land, iii) ten land cover classes.

Table 2. Comparisions of areas for three degrees of detail of land cover characterization for the Department of Yoro, Honduras. Landsat TM image was sampled at 100m^2.

	1993 Agricultural Census			Landsat TM (composite of 3 dates)		
Land cover class	3rd Degree	2nd Degree	1st Degree	3rd Degree	2nd Degree	1st Degree
	ha			ha		
Annual crops and fallow <1yr	45,812			37,263		
Burned landscape	*			17,328		
Perennial crops	35,222			48,155		
Total cropped and prepared land		81,034			102,746	
Pasture	131,628	131,628		118,859	118,859	
Old (bush) fallow >1yr	38,159	38,159		119,972	119,972	
Total agro-pastoral land			250,821			341,158
Deciduous forest and woodlots				81,609		
Mixed forest and woodlots				84,965		
Conifer forest and woodlots				107,085		
Sparse conifer rangeland				62,322		
Total silvo-range land	18,560	18,560	18,560		335,984	335,984
Infrastructure, bare and eroded land				61,357	61,357	61,357
Unclassified [1.]	501,759	501,759	501,759	37,219	37,219	37,219
TOTAL	771,140	771,140	771,140	776,141	776,141	776,141

1. For the census, the class "unclassified" was calculated as the difference between published figures for total area and the sum of Census classes. * : In the Census, the class "Burned landscape" is assumed to be included within one of the classes of "cropping".

Table 2 summarizes results. As mentioned above, the agricultural census does not pretend to provide a complete estimate of forest cover. For technical reasons, a percentage of the forest/woodlot classified on the image represents perennial crops such as coffee and perhaps some banana and sugarcane. Excellent estimates of area in coffee, however, are available from the Honduran Coffee Institute (IHCAFE). The Project is sharing data with this organization and together we are looking for ways to improve image classification using IHCAFE annual survey data.

At the intermediate level of aggregation, there is a 27% discrepancy between the 1995 image classification and the 1993 census for "intensively" managed land. Future research must address this issue since this landscape component has critical implications for ecosystem health.

The image classification resulted in equal areas of pasture and old (bush) fallow, whereas the census is 'strongly biased' toward pasture with fallow contributing less than a third of the area reported for pasture. This discrepancy may be partly due to the spectral confusion between "fallow more than a year old" and "pasture", particularly regarding the younger fallow, where bush and shrub regeneration is minimal. It may be useful to apply a probability factor-weighting to the *maximum likelihood* classification based upon the census ratio of 3.4:1 for pasture: fallow > 1 year, in order to reconcile estimates of these two image classes with classifications by local stakeholders.

At the "third-degree" level, where ten land cover classes have been mapped, the census figures have again been compared with the areas derived from the spectral classifier. It is interesting to note that the class "perennial cultivation" for the Department of Yoro was overestimated by 13,000 ha, but significantly underestimated for the embedded units (*municipios*) of Yorito and Sulaco. This suggests that valley crops such as oil palm and plantain/ banana are easier to classify. Shade-grown crops such as coffee are much more difficult to detect using current remote sensing methods.

By comparison, the census reported on no more than 35% of the total geographical area of the Department of Yoro. We are looking at other Departments of Honduras to determine the relation between census data and image classes. This research is not meant to raise questions about the veracity of either census data or the usefulness of remote imagery. Clearly, census data characterize some sub-system in an agro-ecological hierarchy. The potential for using these data, gathered at great expense, to improve decision making remains to be seen and will be addressed in partial fulfillment of a PhD degree in collaboration with the University of Liecester, England.

Transport accessibility

Some would argue that *transport accessibility* is the most important, short-term variable driving land-use change in lesser-developed countries. One activity of the Trust Fund Project was the creation of an accessibility GIS for Honduras (Fig 2) based on:

- existing transport maps,
- "most likely route" based on constraining terrain characteristics, e.g., mountain passes, forest, etc.,
- speed of transport estimates based on six road types, and where no roadways exist, travel times associated with land cover/use,
- village and farm site location data collected from interview and GPS readings.

Socio-economic data

In collaboration with Honduras INEC (Instituto Nacional de Estatistica y Censo) we have resurrected decade-old digital databases of past national population and agricultural censuses. For example, from household level 1974, 1988 and 1993 censuses, we have reconstructed digital databases allowing statistical characterization and aggregation at the village/*aldea* scale which is ten times more disaggregated (about 3000 records versus 300 records) than heretofore available. Until now, population and agricultural census data for countries like Honduras were only available in hardcopy. However, with the availability of reasonably inexpensive hardware and software and examples of the power of information and interactivity (meaning parameter ranges can routinely be defined, plotted, analyzed and interpreted) there is demonstrable enthusiasm for making digital databases available.

Targeting beneficiaries in Honduras.

Results of the work with the 1993 Agricultural Census supplied by the Honduran Institute of Statistics and Census (INEC), and re-aggregated to the geographic scale of village (*aldea*) is already affecting the way research is carried out in Honduras.

The CIAT Hillsides Program received funding from the Inter-American Development Bank to research the problem of linking local and regional perceptions of well-being and poverty. Methodological obstacles have prevented local perceptions of poverty from providing a basis for traditional poverty assessments by development organizations. First among these is the location-specific nature of peoples' perceptions which is claimed to make it difficult to compare and contrast assessments across locations and scales of analysis.

During 1997, methods developed in a Colombian study site were implemented at a regional scale in Honduras. The methodology involves a number of steps. First is to select sampling sites across a region with the explicit purpose of future interpolation of results. This strategy seeks maximum sample variation rather than average "representativeness".

Based upon data available in georeferenced databases described above, villages across Honduras were categorized according to predefined classes of altitude, accessibility, basic services, ethnicity, gender composition and estimates of population density. The six factors and classes gave rise to a possible 2,430 combinations. In Honduras, however, only 394 of these combinations are actually present.

Given availability of resources and existence of target watershed study sites, it was decided to conduct interviews in the Departments encompassing the study sites. This reduced potential sampling to 193 categories represented by 662 villages. A sample of 90 communities, 30 from each Department, was drawn from this population. Although the 90 communities comprise 3% of the total in Honduras, they comprise 20% of existing categories, i.e., variation, and closely follow the frequency distributions of the national population. This example of cost-effective and efficient research design could only be carried out as a result of the development of the databases by the project.

For the Honduran study, more than 300 descriptions of household well-being were translated into almost 400 indicators. Only about 100, however, were used by more than 5% of informants. Table 3 highlights a few examples of indicators extracted from descriptions.

Table 3. Some frequently used well-being indicators by well-being level

Description	Percent of *aldeas* where used	Number of times used to describe: highest level of well-being	middle level of well-being	lowest level of well-being
Day laborers	97	1	38	[redacted]
Farmers	53	[redacted]	[redacted]	4
Don't day labor	34	[redacted]	[redacted]	3
Farmers and day laborers	26	0	[redacted]	[redacted]
Merchants	69	[redacted]	39	7
Middlemen of agricultural products	36	[redacted]	6	2
Own Land	93	[redacted]	78	7
Don't own land	78	3	12	[redacted]
Have little own land	54	1	[redacted]	11
Some rent land	36	1	11	[redacted]
Share cropping	33	0	12	[redacted]
Own cattle	86	[redacted]	19	1
Don't own cattle	32	1	[redacted]	9
Lack resources to cultivate their land	25	0	7	[redacted]
Buy little inputs with difficulty	20	1	5	[redacted]
Have difficulties in getting sufficient food	63	0	5	[redacted]
Don't have difficulties in getting food	43	[redacted]	[redacted]	3
Need to buy grains	52	1	16	[redacted]
Harvest for home consumption	56	24	[redacted]	12
Have excess for sale	55	[redacted]	32	4
Don't own houses	53	0	5	[redacted]
Own good houses	56	[redacted]	15	0
Have houses of poor quality	43	0	10	[redacted]
Contract day laborers	56	[redacted]	12	0

Source: Well-being rankings conducted in Atlántida, Yoro and El Paraiso departments, Honduras, 1997.

Significance of spatial structure in hillside agroecosystem characterization

An outstanding methodological issue in ecosystem research and development relates to incorporating "space" or location as an explanatory variable equivalent to more traditional variables. Recent efforts by Veldkamp and Fresco (1996) using an approach based on multivariate statistics and factor analysis suggest dynamic modeling across geographic scale can be incorporated into more traditional dynamic temporal analysis. "Location", however is not a factor that has been routinely nor accurately described. Given that structure heterogeneity is widely accepted among ecosystem ecologists, acceptance of agricultural systems applications will require explicit identification of the role of spatial structure of socio-economic as well as traditional biophysical factors in examining future impacts of alternative interventions.

Two principles from the discipline of community ecology (synecology) are just beginning to appear as organizing principles for sustainable management of agricultural systems. The first, hierarchical systems theory (Müller, 1992) would seem to be very appropriate and useful for motivating stakeholders to be precise about limits of systems they wish to influence. Just that principle would go a long way towards promoting didactic solutions to seemingly intractable arguments among stakeholders about what problems exist, what consequences of proposed scenarios might look like and what verifiable indicators of progress make sense.

The second principle of community ecology is the function of non-environmental spatial structure as an important explanatory variable; as important, perhaps, as traditional environmental variables like climate and soil factors.

Robust multivariate statistical methods are just becoming known outside the academic community studying community and landscape ecology. Even though these analytical techniques are largely untested in tropical agro-ecological systems studies found in the literature, their relevance and potential value must not be underestimated. When one considers the number and lack of precision and quantification of factors that potentially influence the fate of stakeholders in hillside agroecosystems, optimism rises with the statement:

> *"On a larger spatial scale landscape ecology focuses on spatial patterns and processes related to them; it considers th edevelopment of spatial heterogeneity and spatial and temporal interactions across heterogeneous landscapes. It attempts to answer questions about land-use and land management."* (Jongman, 1995)

Legendre (1993) has suggested analytical approaches that specifically address issues of spatial dependence and problems it presents for statistical testing of hypotheses. Legendre demonstrates that total variance of a target ecological georeferenced data table (three-matrix case) may be partitioned, mapped and modeled separately into:

1. Non-spatial environmental variation, (perhaps prices and taxes)
2. Spatially structured environmental variation, (perhaps temperature and pH)
3. Spatial variation of the targeted variable(s) that is not shared by the environmental variables, (perhaps ethnicity and "burning")
4. Unexplained, non-spatial variation.

Legendre explores two approaches that could be used to explicitly incorporate spatial structure into agroecosystem analysis and interpretation; the "raw data" approach and a "matrix approach" followed by more traditional causal analysis modeling (Asher, 1976).

CIAT does not have any expertise in community and landscape statistical techniques and would like to collaborate with advanced academic institutions to explore this area of research. We have, however, conceptualized a generic, interactive, GIS-based methodology for across scale analysis. The key property of the technology is a focus on data-driven rather than a model-driven analysis. For example, typical model-driven, decision-support systems (DSS) require the user to supply prerequisite, formatted data, whereas a "data-driven" DSS will select/create variables from a database and propose a "best representation" followed by an intelligent classification and display of the dataset.

Components of the "data-driven" DSS are:

- A powerful database engine and query builder.
- A reliable data regionalization process.
- An adaptable method for reconstructing data across scales.
- The capacity to analytically compare "single-scale" and "across-scale" representations.
- User-defined protocols to classify and represent data.

Table 4 gives an example of data reduction from the 1998 Honduran population census. The query screens for "method of water supply for households with more than eight members". The database is to be aggregated to the village (*aldea*) scale from individual census responses. In this example, results for four villages (*aldeas*) are extracted from more than 5000 responses to two separate questions. Where data does not add to 100%, the missing percentage is an invalid response from the census.

Table 4. Typical output from data-driven, census database query for four villages.

Query: Method of water supply for households with more than 8 members					
Aldea Code	**Responses**	**By Pipe**	**By Pump**	**Shared Pipe**	**River**
180901	1566	51.2%	38.3%	9.4%	0%
180902	814	41.4%	39.6%	10.2%	1%
181001	1845	50.3%	35.5%	12.2%	0%
181002	642	48.4%	40%	10.1%	1.5%

An important property of this method of aggregation is that no data are lost from the interviewee to the village (*aldea*) scales. This is unusual for a GIS databse. The advantage of storing data at its fundamental scale and generalizing in an *ad hoc* query is the dramatic reduction in volume of data with no loss of quality for very little overhead in computation.

The second step is to "regionalize" point data across a range of scales in a manner that accounts for the property of "location". Computationally this is accomplished by successive summing of point data within increasing larger polygons or circles as illustrated in Fig 3. Advantages of this technique are that i) no data are created in locations where no original data exist as occurs in a grid-based technique, ii) the area of study is centered on the point location rather than an ambiguously positioned window or lattice, iii) there are no imposed limits of "rate of change", i.e., resolution, as in grid-based techniques, iv) the vector polygon can be any shape.

The result of this technique is a new "point" coverage with each new point having values for multiple-scales. This completes the stage of analysis of scale dependence of data, i.e., the assessment of change of a variable across scales of aggregation. More importantly, it allows us to choose the scales and concomitant variances for selected variables of interest. This new across-scale data-set can be analyzed i) on its own merits, e.g., does the frequency distribution change across geographic scale, ii) by comparison with single-scale representations, iii) by comparison with other scale-dependent variables, iv) by comparison with physical land metric data, e.g., slope, land cover, accessibility.

The 1988 population census was the last conducted in Honduras – although there was a 1993 agricultural census. Although dated, we feel it is relevant for Honduran stakeholders, and others, for establishing data collection standards, and analysis and presentation protocols. This work, in its initial stages, relates specifically to issues addressing "minimum" quantity and quality of data sets.

Output 3: Capacity to supervise and guide change using multi-scale spatial analysis.

An important milestone for this, the second year of the Trust Fund Project, will be three community workshops, managed by project personnel, to be held in May, August and September in Honduras and Nicaragua. The objective of these workshops is to present databases and planning technologies developed so far by the project, and to document user satisfaction and suggestions for finalizing the work. Not all intended end-users will attend these workshops. Intended end-users that have high technical research skills and are in positions of regional institutional planning have been part of the development process from the beginning and have had input through formal GIS networks[8], training and visits to CIAT headquarters. The three workshops for 1998 will specifically focus on community-level stakeholders, primarily farmers, merchants and community leaders, many belonging to the Local Organizing Committees mentioned above. It is our conviction that these stakeholders have traditionally been marginalized during development planning processes, and unless our decision aids can raise the level of participation by these stakeholders, our mission will have failed.

Assuming outcomes of the evaluation workshops are generally favorable, we envision emphasizing two final project products; i) interactive, group decision- support technology, and ii) computer-assisted training and workshop management technology.

Future technology to elicit perspectives and facilitate negotiation

Decision-support tools are often associated with mechanistic models or goal optimization models. However, many land use decision requirements may be satisfactorily and economically addressed through a strategy of successive refinement of data on demand. Our vision calls for an "interactive group decision-support system"

---

[8] The CIAT Hillsides Project has been a founding partner in the organization of national GIS networks in both Honduras and Nicaragua.

(IGDSS) as a technological tool to facilitate the mutual understanding of perspectives held by different stakeholders. The tool is an effective way to match goals with resources and judge the use of research results.

The multiple-stakeholder planning process we envision will use a limited number of land use scenarios (landscape "scenes" or demonstrated situations) to which stakeholders can react. Stakeholders will compare and contrast their own visions and expectations for the future by criticizing the scenarios. Examples of issues that stakeholders have raise in past planning workshops we have held include, i) underestimation of productive capacity of the watershed, ii) dependence upon a limited number of agricultural crops, iii) contribution of erosion of agricultural land to deterioration of aqueducts (non-point source pollution), iv) contribution of agro-industrial residues to contamination of waterways (point source pollution), deficit of forest cover due to encroaching agriculture and production of charcoal.

The IGDSS is an extension of traditional group decision support system techniques that already have been shown to increase effectiveness of group decision making (Karan et al, 1996; George et al, 1990). Technologically the IGDSS employs computational "virtual" agents as participants in the decision making process. Highly specialized agents within the system will fill a variety of roles including accessing data from the GIS database system described above, accessing knowledge bases, and suggesting implicit and explicit conflicts in stakeholders' goals and values. The entire process will aim at helping users take generic concerns, like those mentioned above, and turn them into concrete statements of values and goals in a way that allows stakeholders to see their significance, and that gives them a chance to enact compromises within a virtual environment. In this way, the system will facilitate dialogue and compromise. This work is being carried out in collaboration with the University of Georgia, Athens, Georgia, USA.

Computer-assisted training in use of project outputs

Mention was made of three technology-evaluation workshops to be held during 1998. Aside from assessing relevance, usability and suggestions for improvements of the methodological research, an important goal will be to identify criteria for computer-assisted materials for facilitating workshop management and mass education in information technology-based planning. Our vision is that planning workshops will always require real human facilitators. However, there is a need for training materials that address data inventorying and database development protocols for a variety of spatial and non-spatial data, as well as basic skills required during the planning workshops and advanced skills required for facilitating workshops. This work is being carried out in collaboration with the Royal Agricultural College, Cirencester, England.

Asher, Herbert. 1976. Causal Modeling. Sage University Paper Series on Quantitative Applications in the Social Sciences, series no. 07-003, Beverley Hills and London, EU.K:Sage Publications.

Berry, J. 1993. There's more than one way to figure slope. Beyond mapping, concepts, algorithms and issues in GIS. GIS World Books, Forth Collins. Co. pp147-151.

Fresco, Stroosnijder, Bouma and van Keulen (1994), The Future of the Land: Mobilizing and Integrating Knowledge for Land Use Options, eds. Louise O.
Fresco, Leo Stroosnijder, Johan Bouma and Herman van Keulen, John Wiley & Sons, Chichester, England.

George, J. F., J.F. Easton. 1990. A study of collaborative group work with and without computer-based support. Information Systems Research, 1:4.

Jongman, R.H.G.1995. Introduction. In Data analysis in community and landscape ecology. Eds. R.H.G. Jongman, C.J.Fiter Braak and O.F.R. van Tongere. New edition University of Cambridge, Cambridge, U.K.

Karan, Vijay. 1996. Information technology support for collaborative decision making in auditing: An experimental investigation, Decision Support Systems, 21:3.

Legendre P. 1993. Spatial autocorrelation: Trouble or new paradigm. Ecology, 7(6):1659-1673.

Müller, F. 1992. Hierarchical approaches to ecosystem theory. Ecological Modelling, 63:215-242.

Veldkamp A and Fresco, L.O. 1996. CLUE: a conceptual model to study the conversion of land use and its effects. Ecological Modelling 85:253:270.

Fig. 1. One of a number of alternative boundary delineations representing an area of central Honduras. Also illustrated is the difference in routine determination of areas for two of the community watersheds using standard GIS application software. There is a 20% difference in calculated area depending on whether the base representation is 2 or 3 dimensional.

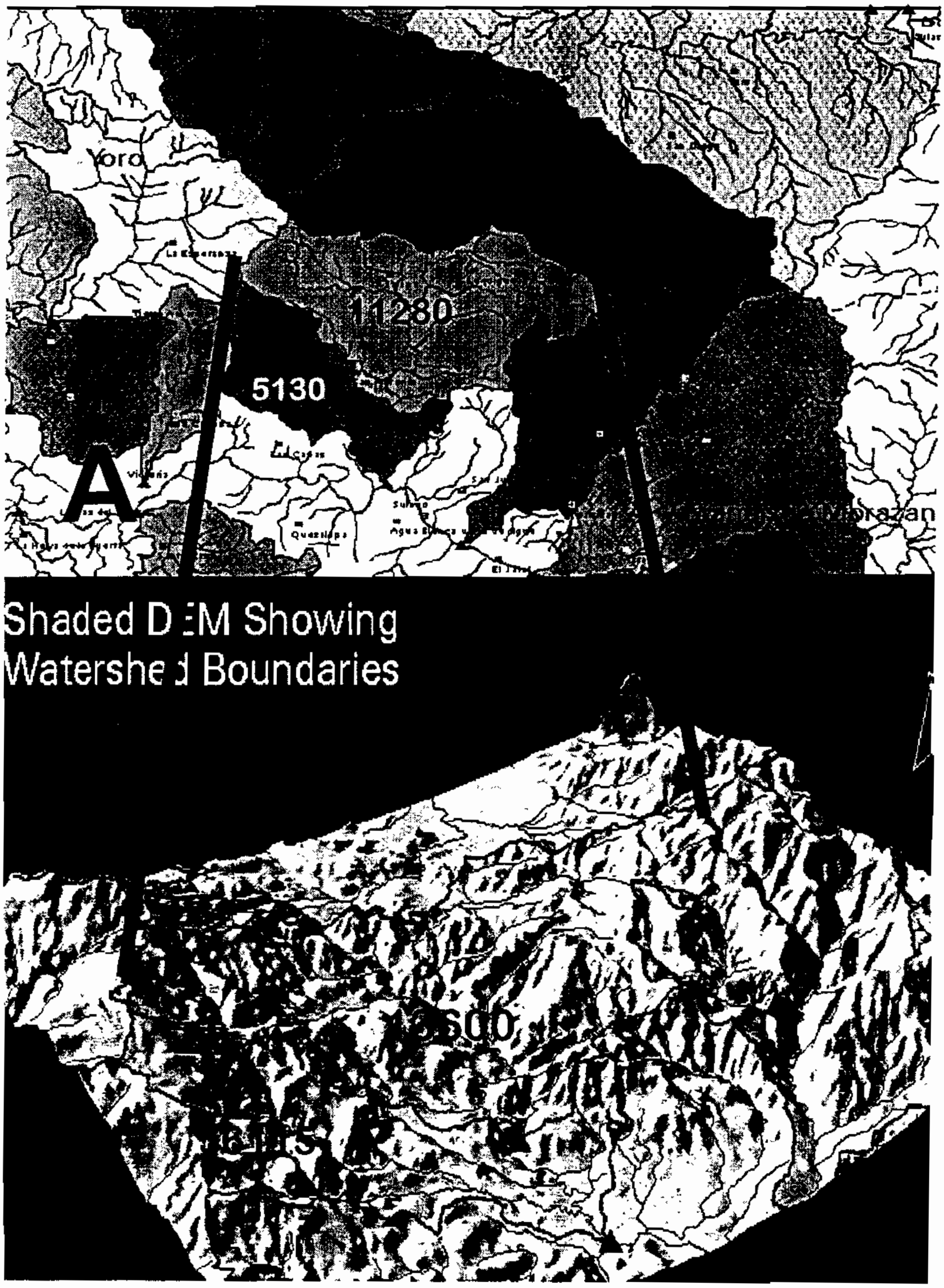


Fig 2. From a number of spatial coverages prepared by the project, a transport accessibility map for Honduras is now available. Some would argue that transport accessibility is the most important, short-term variable driving land-use change in tropical America.

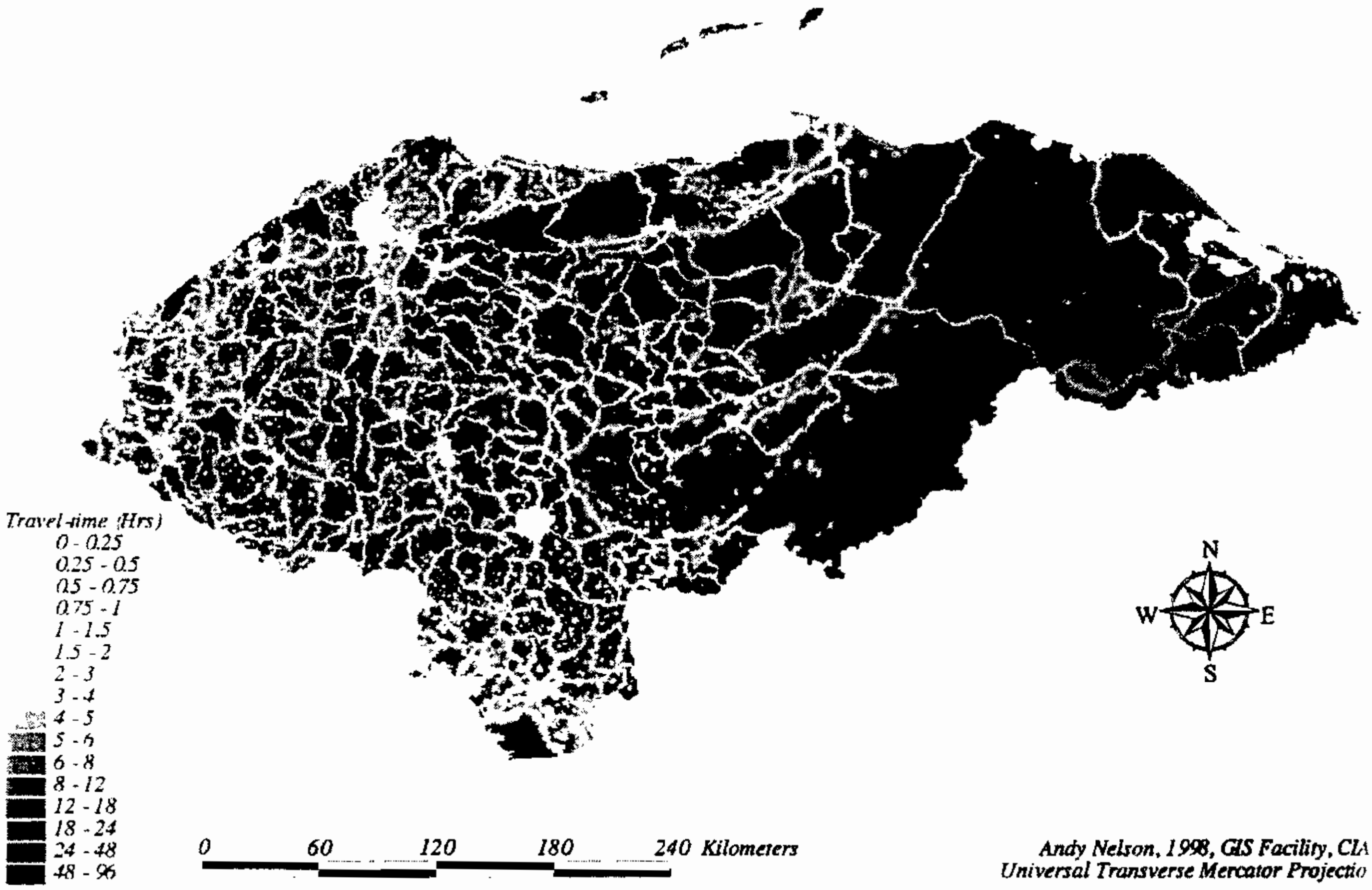
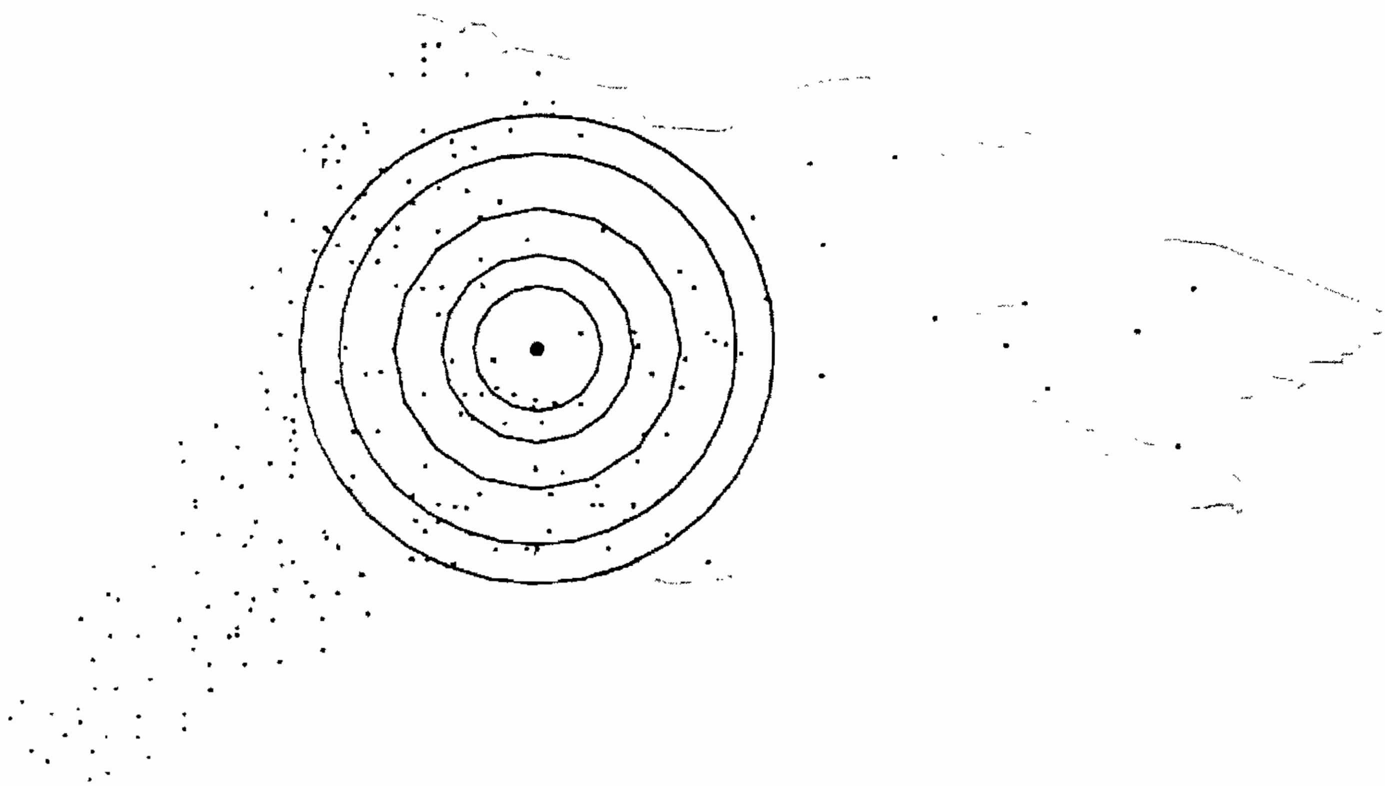


Fig 3. A GIS-based method for exploring the significance of spatial structure, i.e., scale and location, in hillside agroecosystem characterization is illustrated below. Village-level point data are successively "regionalized" in a manner that accounts for "location". Computationally this is accomplished by successive summing of point data within increasingly larger polygons or circles as is the case illustrated. Results are tabulated and a new "across-scale" database is made explicit and can be analyzed from the unique perspectives of different stakeholders.



Appendix 1. Logical Framework for the project, “Methodologies for integrating data across geographic scales in a data-rich environment: Examples from Honduras”, funded by the Ecoregional Trust Fund for Methodological Research.

Narrative Summary	Measurable Indicators	Means of Verification	Important Assumptions
Goal To institutionalize community-led management of watershed resources for more productive, sustainable and healthy hillside agroecosystems.			
Purpose To develop and document principles and procedures for building a scale consistent data base and for performing multi-scale characterization of agro-ecosystems.	* Number of requests for publications, software consultations, training and collaboration from stakeholders.	* Acknowledgements of collaboration in publications, software, etc., by project external review.	* Development funds are available at international, national and local level. * Enabling political policy.
Output 1. Quality controlled multi-scale spatial database for Honduras with associated methodology and training workbooks.	* Databases reviewed, edited, georeferenced. * Written workbooks for database development. * CD-ROM of databases and procedures.	* Workbooks and and CD-ROM available on request.	* Socio-economic and biophysical databases and remote imagery available. * Interest groups agree to share data.
Output 2. Multi-scale characterization of Honduran agro-ecosystem s for targeting problems, priority areas and beneficiaries.	* Databases reviewed, edited and georeferenced. * Written guides for database analysis.	* Publications available on request.	* Graduate students are available. * Viable CIAT computer systems group.
Output 3. Institutional capacity to supervise and guide change using multi-scale spatial analysis.	* User organizations operational. * User planning workshops. * Trained people.	* Annual reports of project and organizations. * Lists of trained people. * Donors continue to fund project.	* Collaboration with other Ecoregional projects. * Institutions identify individuals for training.
Output 4. Administration, management and monitoring project.	* Documentation and reports.	* Annual reports of project and organizations. * Lists of trained people. * Donors continue to fund project.	