A GIS-Based Decision Support Tool for Targeting Biophysical and Socio-Economic Niches in Tropical Agriculture

***Rachel O'Brien*[1,2] , *Robert Corner*[2], *Michael Peters*[1] & *Simon Cook*[1]**

[1]International Center for Tropical Agriculture
Cali, Colombia
Phone: 57-2-445-0000
Email: r.obrien@cgiar.org

[2]Department of Spatial Sciences
Curtin University of Technology, Perth, Western Australia
Phone: 61-8-9266-7565
Email: rcorner@vesta.curtin.edu.au

Presented at SIRC 2002 – The 14th Annual Colloquium of the Spatial Information Research Centre
University of Otago, Dunedin, New Zealand
December 3-5th 2002

ABSTRACT

The aim of this paper is to describe a methodology for recommending crop species for smallholder farmers in the tropics, targeted to biophysical and socio-economic niches, incorporating both available data and expert knowledge. Although it is important that species are matched to biophysical environments, it is equally important that they are matched to the unique socio-economic situation and management practices of a farmer. Agriculture in the tropics and subtropics is complex and heterogeneous, both biophysically and socio-economically. Bayesian modelling has been identified as the most appropriate method for incorporating sparse and uncertain data with expert knowledge to predict the probability of a given species being suitable for a given environment or niche. As a case study, a spatially enabled decision support tool designed to assist in the decision-making process of targeting forages in Central America is under development. It will use existing data and expert knowledge to mimic the decision-making process involved in selecting forage varieties for specific socio-economic and biophysical environments. Validation and verification of the model are discussed, as well as the technological specifics of the tool development and deployment.

Keywords and phrases: GIS, decision support tool, tropical agriculture, Bayesian modelling, socio-economic factors, expert knowledge, forages, Central America

1.0 INTRODUCTION

Research in tropical agriculture faces the challenge of developing crops, technologies and methodologies applicable not only to the unique biophysical environments of the tropics but also to socio-economic niches. Many farmers in tropical countries are smallholder farmers struggling to maintain and improve their livelihoods whilst remaining or becoming self-sufficient. At the same time environmental pressures dictate that technologies are needed which allow farmers to meet these needs in environmentally sustainable ways, through intensification and other appropriate technologies.

64% of rural dwellers in Latin America live below the poverty line and are constrained in development by the risk of living in marginal and isolated hillsides. Approximately 77 million people live in the rural areas of Latin America, so for this part of the tropical world alone there are about 47 million rural poor (IFAD, 2001).

Biophysically there is huge variation within the tropics and subtropics, with Holdridge for example classifying 15 lifezones for the tropics and subtropics alone (Leemans, 1990). However tropical countries are mostly developing countries without large research budgets or government subsidies. Most farmers cannot afford inputs such as irrigation or fencing, let alone yield monitors or expensive fertilizers. Therefore it is important that crops are well suited to biophysical niches without the need for large inputs such as fertilizer or irrigation.

Biophysical niches however form only part of the picture. Equally important are socio-economic characteristics of the farmer and the community. There is no point for example in recommending coffee as a crop to a farmer, even if the soil and climate are ideally suited, if there is no market available for selling the product, or if the required labour force for planting and harvesting is not available. Other important factors include the amount of land available for planting, purpose of the crop (e.g. household consumption, cut and carry crop for cattle, local market crop or export crop), available time investment (e.g. a crop that matures within three months to plant in between maize crops, or a tree crop requiring a time investment of 20 years or more), financial resources, and cultural factors such as whether a farmer is traditional or progressive (level of risk-averseness).

A number of organisations are dedicated to research and development in tropical agriculture, including international research centres, non-government organisations, ministries of agriculture and local agricultural extension agencies. A large amount of data has been collected throughout the tropical world on plant habitats, adaptation, characteristics and management. However in addition to data in formal databases and literature, a wealth of unrecorded knowledge exists inside experts' heads. This knowledge is often tapped into when an expert is contacted to make recommendations for a certain crop or a certain area but obviously often this expert knowledge is not readily available to most farmers or their advisors.

Lack of information and associated risks form a major constraint to development, so tools that capture data and expert knowledge are extremely valuable. This paper outlines a methodology for accomplishing this goal, and describes a spatially enabled decision support tool based on this methodology, applied to forage selection in Central America.

The following section outlines Bayesian modelling, as a method for combining factors to predict the probability of an outcome. The case study of selecting forage species in Central America is then introduced and used to illustrate the conceptual model. This is followed by a section describing the implementation of the decision support tool and the process of validation and verification. Finally the specifics of the technology used to develop the tool are discussed as well as the deployment process.

2.0 BAYESIAN MODELLING

Many spatial methods and algorithms have been developed to solve the problem of predicting spatial habitat distribution. Commonly used techniques include simple statistical methods such as multiple linear regression, rule-based classifications such as Classification and Regression Trees (CART), spatially explicit models such as habitat envelopes and artificial intelligence techniques such as Bayesian modelling, neural networks and multi-agent systems.

Preliminary research has identified Bayesian modelling as the most appropriate way of incorporating uncertain and sparse data as well as expert knowledge. Bayesian methods provide a "formalism for reasoning under conditions of uncertainty, with degrees of belief coded as numerical parameters, which are then combined according to rules of probability theory" (Pearl, 1990). A simple Bayesian model codes prior probabilities for each factor, then combines these to form joint probabilities. The prior probabilities may be derived from data, set by experts, or defined from a combination of data and expert opinion. The factors may then be combined with equal weighting, or weighting can again be calculated from available data, set by experts, or defined by a combination of the two. This process of combining prior probabilities produces joint probabilities for each possible outcome.

Bayesian modelling does not seek to predict exact outcomes, but rather the probabilities of various outcomes, given the effects of the input factors on each outcome. Therefore if two factors strongly support a given outcome, then the combined effect of these two factors will produce a high probability for that outcome. Conversely if one factor strongly supports an outcome, but another factor does not, then the probability of that outcome occurring will be lessened. Uncertain data can be accounted for by reducing the probability that the data will support a given outcome. Expert opinion can be used to update prior probabilities defined by data, or to define prior probabilities where data is lacking. One of the advantages of Bayesian modelling is that it is easier to incorporate human reasoning and is thus very useful when the problem emerges in parallel with modelling.

Figure 2.1. shows the concept of combining multiple evidence layers to produce a multi-level hypothesis based on joint probabilities.

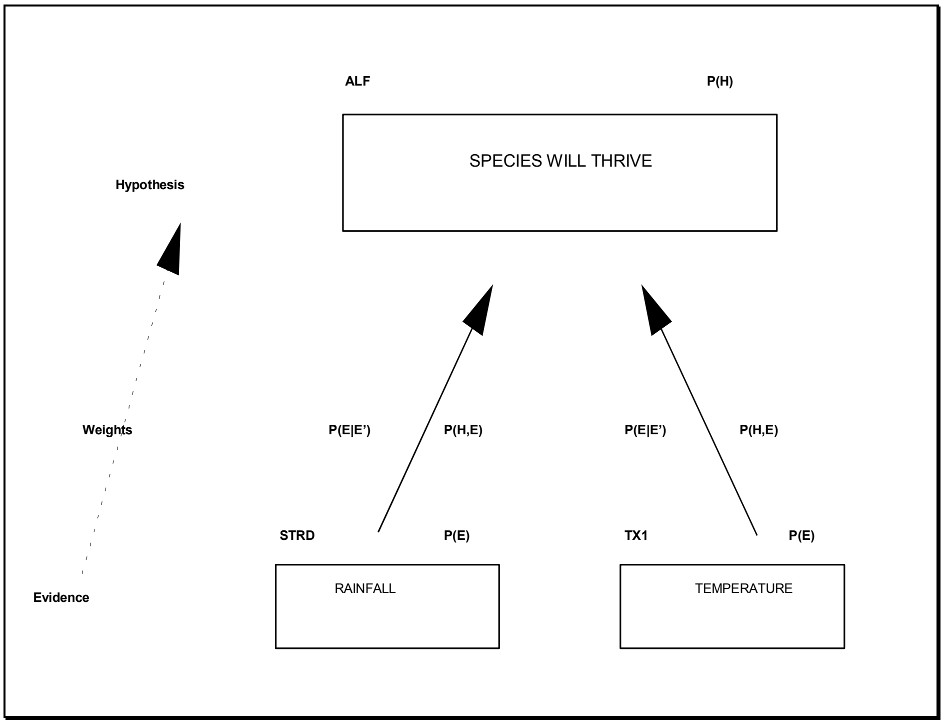

Figure 2.1 Bayesian schema.

3.0 CASE STUDY – FORAGE SELECTION IN CENTRAL AMERICA

3.1 Tropical Forages in Central America

Forages in Central America are used mainly as feed for cattle, either through grazing or for cut and carry, and play an important role in agriculture dominated production systems. Demand for livestock products in the developing world is increasing (Delgado et al, 1999), and farmers are responding by increasing milk and meat production. Improved forages can be introduced into tropical farming systems to improve livelihoods for smallholder farmers, reducing hunger and poverty. The International Center for Tropical Agriculture's (CIAT) experience in Latin America and Asia, amongst other institutes, has demonstrated the effectiveness of new forage-based technologies for intensifying meat and milk production on small farms, as well as for other uses. Forages therefore can have a positive impact on development in the tropics.

Smallholder cattle farmers in Central America typically have a very small amount of land with 10-50 ha being typical. In some regions over 70% of cattle farmers have less than 20 head of cattle (Cruz Flores, 2002, pers. comm.) Paddocks may be only 2-3 ha each.

CIAT has a number of databases with data on forage adaptation, establishment and production trials (Barco et al, 2002). The RIEPT database (International Network for Tropical Pasture Evaluation) contains data on 2539 trials of 929 accessions in 28 locations in Central America (Figure 3.1). This database includes data about trial locations, such as elevation, temperature and rainfall, as well as data relating to the trials such as level of adaptation, plant height and cover. Although this data is undoubtedly useful, it contains biases as well as many inconsistencies, errors and gaps. This is where expert knowledge can be used to complement data.

Figure 3.1 RIEPT database trials in Central America. Source: Barco et al (2002)

This project is linked to SoFT (Selection of Forages for the Tropics) in a complementary way. SoFT is an international project started in 2002, in which CIAT is involved, that aims to collect and make accessible expert knowledge about tropical forages. SoFT expert knowledge will include adaptation of forages in different biophysical niches in the tropics, and will be published as a database. The RIEPT database, expert knowledge gathered by SoFT, and expert knowledge and data available through CIAT make up the inputs to the case study for forages.

The premise behind the development of a GIS-based decision support tool for targeting tropical forages is that many appropriate forage species are not being recommended to farmers, through lack of knowledge or experience of the farmers' advisors. A tool that suggests appropriate forage species based on local biophysical factors, a farmer's management practices and available markets and labour, can fill these gaps in knowledge, provide expert information, and ultimately benefit the farmer by suggesting species to trial that best fit the unique situation of the farmer.

3.2 Conceptual Model

Data and knowledge on tropical species exist as databases, in literature and as expert knowledge, and a computer-based decision support tool is an effective way of bringing these together and making them available to farmers' advisors. Because the success of each crop varies spatially, depending on factors such as climate, soil and management practices, a spatially enabled decision support tool can take spatial factors into account and can display intermediate and final results for improved interpretation.

The model begins by identifying the biophysical characteristics of the location of interest. These biophysical characteristics include climate and soil descriptors, and are either derived from GIS maps, or requested from the user. In addition the user is required to input socio-economic and management information, including the required purpose of the species, the amount of land available for planting and whether labour is available. Access to markets is calculated from a GIS surface produced by CIAT, identifying travel time to closest populated areas.

The management factors reduce the number of species under consideration, especially the purpose. However space and time available will also introduce limitations, e.g. there is no point planting grasses in a very small area.

Prior probabilities for biophysical and socio-economic factors related to each species will have been previously calculated from databases and updated by experts, or, where data is not available, determined by experts. These prior probabilities are then used to calculate joint probabilities for each species for the specified location. These probabilities are then ranked and the top species are then suggested to the user. Maps created dynamically can be examined as well as reasons for these species being suggested or for alternative species not being suggested. A flowchart showing the conceptual model is presented in Figure 3.2.

The user may choose to rerun the model to examine the effects of changes in management practices, addition of irrigation, or future climate change. In the case of a species expert using the tool they may wish to update certain prior probabilities.

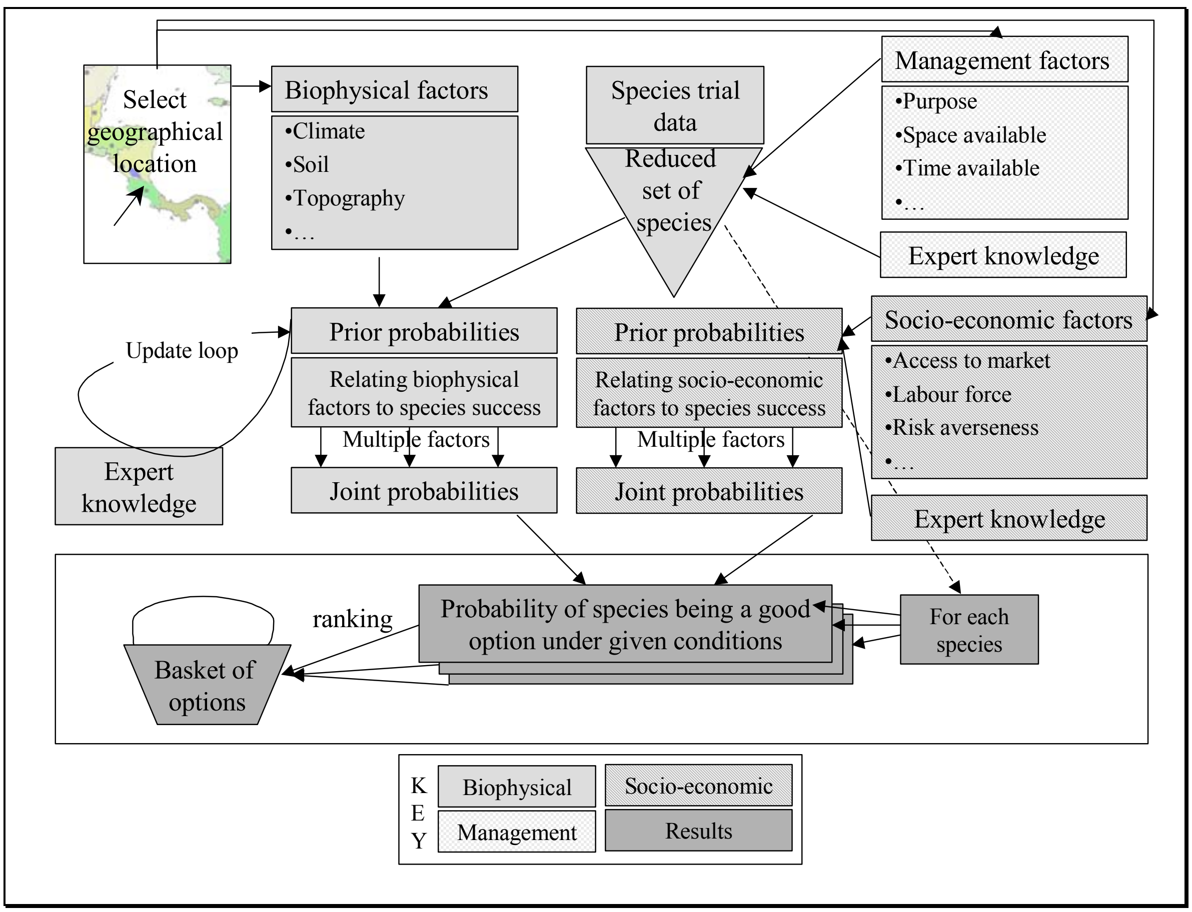

Figure 3.2 Conceptual model showing process for combining biophysical, management and socio-economic data, expert knowledge and trial databases to produce a basket of options based on Bayesian probabilities

The conceptual model has been developed with forages and the factors affecting them in mind. A combination of available data and expert opinion determined which factors should be included in each part of the model. For other species different factors may be relevant, indeed not all factors are relevant for all forage species. For example soil pH has no direct bearing on the success of the legume *Arachis pintoi*. The implementation of the model should remain flexible enough that new factors can be included as data becomes available or as experts determine their importance. The following section on implementation shows a specific example of determining the probability of success of three forage species in San Dionisio, Nicaragua.

4.0 IMPLEMENTATION OF A GIS-BASED DECISION SUPPORT TOOL FOR TROPICAL AGRICULTURE

4.1 Implementation

The first step as outlined in the conceptual model in the previous section is to select a location. Say we are aiming to recommend forage options to a farmer located near San Dionisio in the Matagalpa department of Nicaragua (Figure 4.1). This farmer has 7 ha available for planting and wishes to plant a long-term forage crop suitable for cut and carry. In this part of Nicaragua the average is less than one head of cattle per ha, so a paddock of 7ha could be expected to support around 5 cows. The aim is to reduce reliance on feed supplements at the same time as improving milk quality and quantity. He has access to labour for establishment and is open to suggestions for managing the crop. However he does not have irrigation installed or access to a tractor, and therefore needs a forage biophysically well adapted to his land.

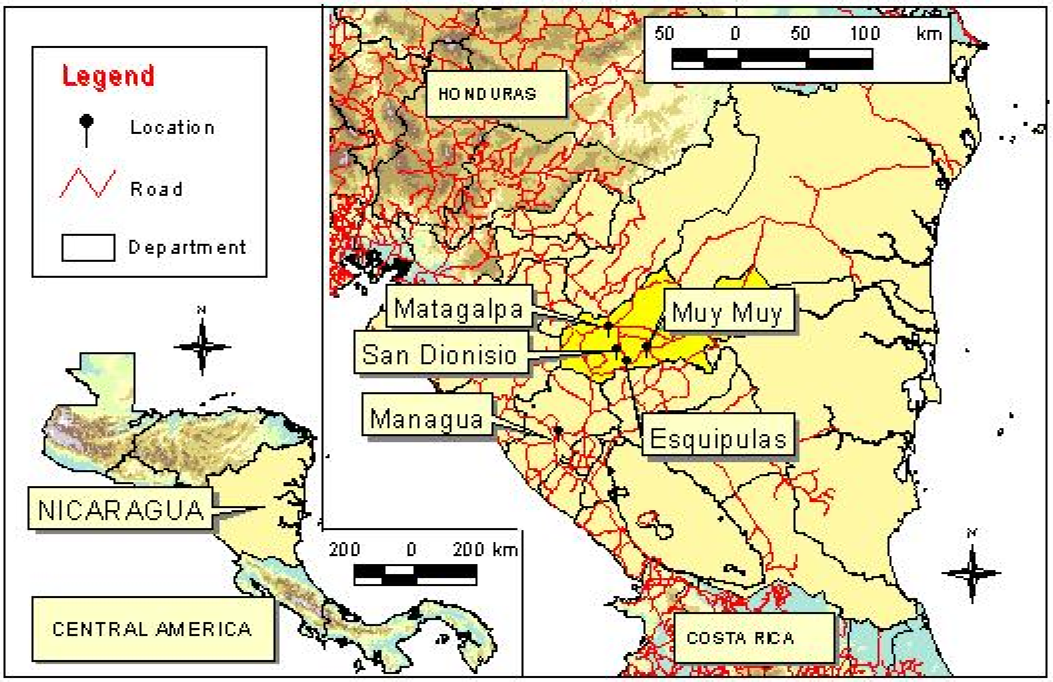

Figure 4.1 Location of San Dionisio in Nicaragua.

Table 4.1 shows the biophysical, socio-economic and management characteristics of our hypothetical farmer in San Dionisio. The figures are derived from GIS data and from the farmer.

Elevation	430masl	Mean Annual Temperature	24C
Mean Annual Rainfall	1217mm	Minimum Monthly Temperature	17C
Dry Season	5 months	Maximum Monthly Temperature	32C
Soil pH	5.5	Soil Texture	heavy
Soil Fertility	medium	Land Available	7 ha
Time Available	permanent	Purpose	Cut and carry
Access to Market	10-20 min	Labour	Available
Risk-averseness	low		

Table 4.1 Biophysical, socio-economic and management characteristics - farmer in San Dionisio

Because the purpose of the crop is cut and carry as opposed to grazing or other uses, a number of forage species are immediately filtered out of consideration. We will consider here three species that are suitable for cut and carry: the legume *Stylosanthes guianensis*, the shrub *Cratylia argentea* and the legume *Centrosema pubescens*. In the full implementation of the tool all species suitable for cut and carry would be included in the analysis.

Prior probabilities can be derived from data and expert knowledge for each of these species with relation to adaptation. Adaptation is classed as poor, regular, good or excellent and is recorded for trials in the RIEPT database. This is just one possible indicator of success, others are ease of establishment, time to full production, height and cover, resistance to pests and diseases, and suitability for desired purpose. Prior probabilities are shown for *Stylosanthes guianensis* for rainfall and soil pH in Figure 4.2. These can be interpreted as follows: If annual rainfall is 1200-1600mm, then adaptation is most likely to be excellent. If annual rainfall is below 900mm then adaptation is most likely to be poor or regular, and also in Central America a small proportion of the land has this rainfall class. Similarly for pH we can see that most of Central America has pH between 4 and 6, and adaptation in these classes will most probably be good.

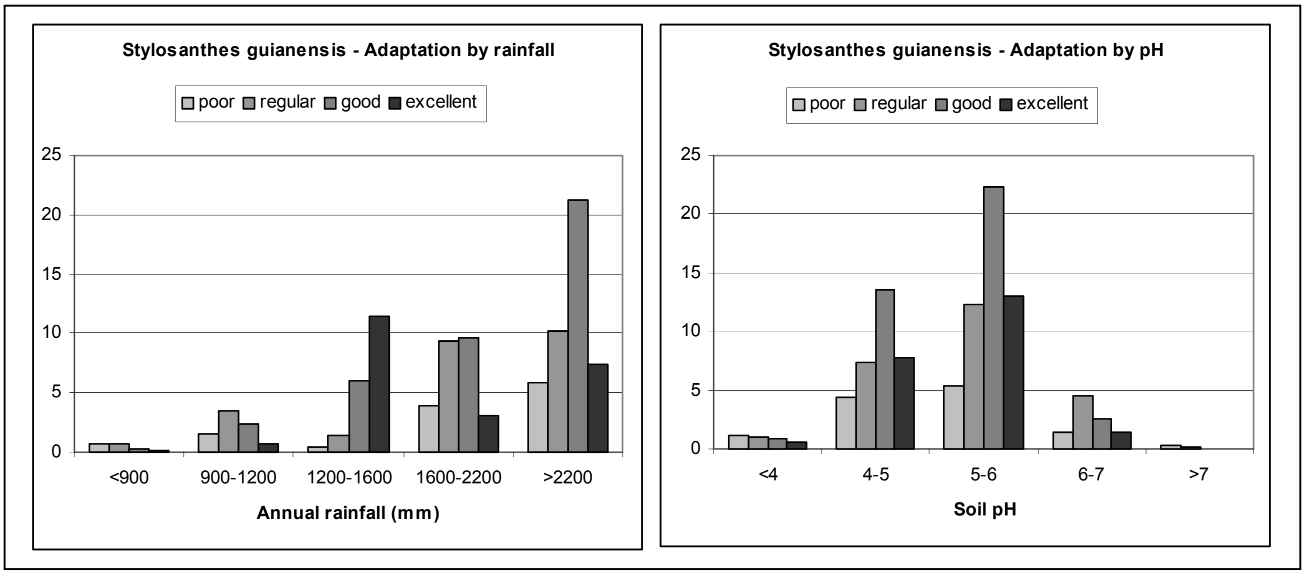

Figure 4.2 Prior probabilities for S. guianensis – adaptation by rainfall and by pH

Similar figures can be created for the other factors for *S. guianensis*, and for all factors for *C. argentea* and *C. pubescens*. Prior probabilities are derived from the RIEPT database and updated using expert knowledge. These can be used to produce maps where GIS data is available, such as probability of good or excellent adaptation based on annual rainfall (Figure 4.3).

Once prior probabilities have been defined for all factors for all species, they can be combined using Bayesian modelling with equal weighting or using data or expert knowledge to define weightings. In this example equal weightings are used to produce joint probabilities for each species for the farmer in San Dionisio. Figure 4.4 shows joint probabilities for *Stylosanthes guianensis* calculated from prior probabilities for the biophysical factors in Table 4.1. Those factors that can be reliably mapped (all biophysical factors except soil properties) are mapped as joint probabilities for the area surrounding San Dionisio (Figure 4.5). With all biophysical factors, the land under consideration in San Dionisio has a 98% probability of producing good or excellent adaptation for *S. guianensis*. With the soil factors removed this figure remains the same but a larger proportion of the 98% is good rather than excellent. From the map in Figure 4.5 it can be seen that the probability of good or excellent adaptation reduces to the west of San Dionisio – this is mainly due to a reduction in annual rainfall.

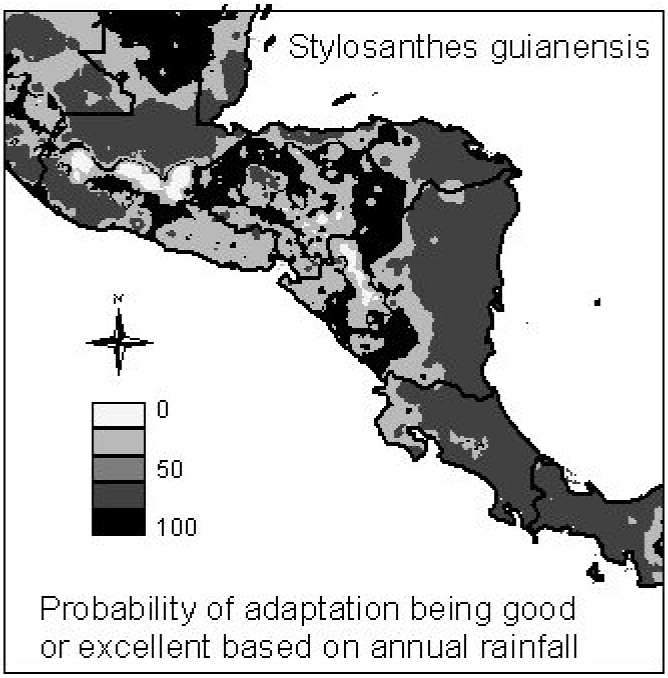

Figure 4.3 Prior probabilities map - rainfall

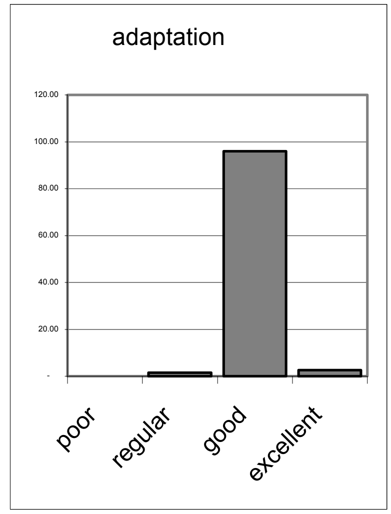

Figure 4.4 Joint probabilities for San Dionisio

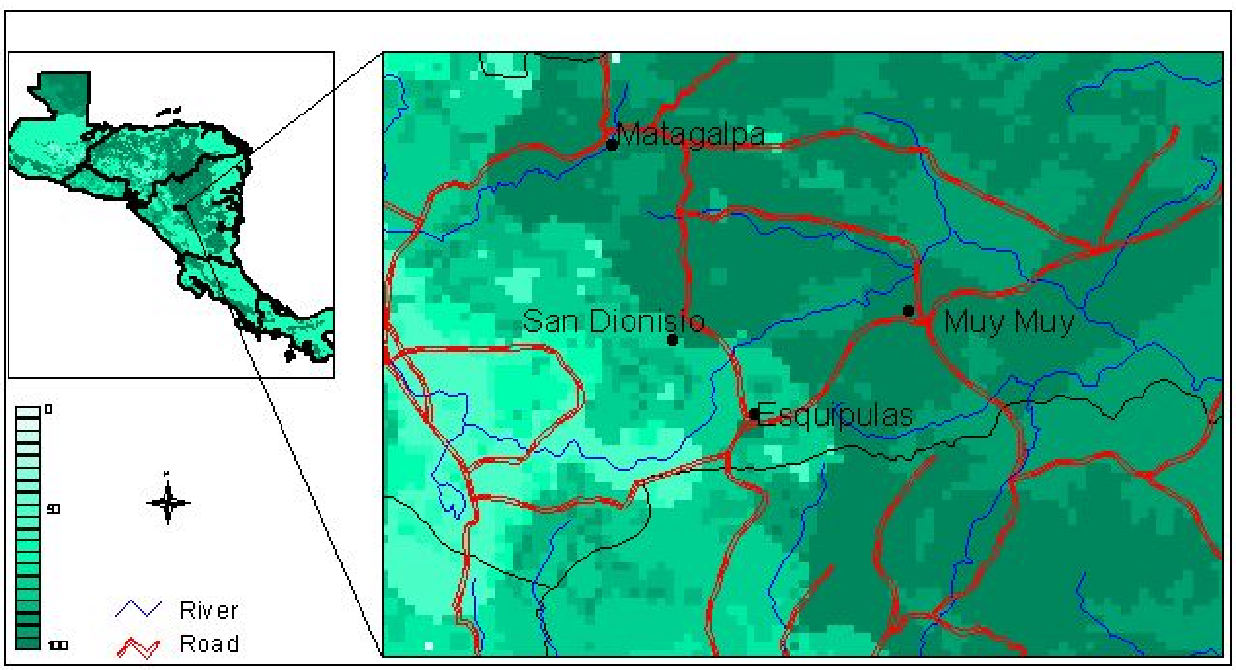

Figure 4.5 Joint probabilities: probability of S. guianensis adaptation being excellent or good based on prior probabilities of biophysical factors – San Dionisio region

The result is a ranking of species according to the probability of each having adaptation that is excellent or good, and of being appropriate to the farmer's needs (Figure 4.6).

Species	Ranking	Favourable factors	Unfavourable factors
S. guianensis	1 - **3**	Elevation, Rainfall, pH, Soil pH, Soil fertility, Temperature	Soil texture, Dry months
C. argentea	3 - **1**	Elevation, Temperature, **Dry months, soil fertility**	~~Rainfall~~, ~~Soil pH~~
C. pubescens	2 -**2**	Elevation, Rainfall, Soil pH, Soil fertility	Dry months, Minimum temperature

Figure 4.6 Basket of options

Favourable factors are those that contribute towards a 50% or higher probability of adaptation being excellent or good; unfavourable factors are conversely those that contribute towards a 50% or higher probability of adaptation being regular or poor. Those not listed do not contribute strongly either way. In the final decision support tool, users will be able to examine these factors and their contributions to the final ranking.

The above outlines the process for combining various biophysical factors to produce joint probabilities. Introducing socio-economic factors becomes a little more problematic. Firstly socio-economic GIS data is usually aggregated and sometimes fairly coarse, making it impossible to deduce meaningful information about an individual or even group of individuals. Secondly socio-economic behaviour varies much more than biophysical behaviour (**?**), making any efforts of prediction much less reliable. Thirdly there is the problem of linking socio-economic knowledge to success of forages.

Risk averseness, for example, could be incorporated into the model in the following way. When a species is recommended it is given a probability score based on what data and expert knowledge suggest the impact of biophysical factors on adaptation will be. The level of confidence that can be attributed to this probability could be derived from metadata or expert knowledge. For example, if there are many trials and expert knowledge all indicating that adaptation under certain conditions will be excellent, then the farmer can be confident that under similar conditions the species is likely to have excellent adaptation. However if data or knowledge is scarce or uncertain, then even though available data and knowledge suggests excellence, the level of confidence the farmer can have in this information is lower. A highly risk-averse farmer may opt for a species with worse adaptation, but higher confidence, whilst a farmer willing or able to take more risk may opt for a species with excellent predicted adaptation, but about which little is known, meaning that confidence in predicting the adaptation of the species is lower.

Incorporating other socio-economic factors such as distance to market, economic status, farm size and labour force is still under investigation, and is dependent on addressing some of the issues mentioned above.

4.2 Validation and Verification

Because a large part of the inputs to the model are based on expert knowledge, it is difficult to validate the model using data-based techniques such as bootstrapping and jack-knifing. Expert opinion is necessary to judge the accuracy and effectiveness of the model. Applying the model to locations where forage trials are underway, but not included in the databases used to specify the model, allows for an assessment of the validity of the model. One such location is San Dionisio, with data currently being collected on trial sites for a number of species. Because CIAT is active in research in the area, there is also a large amount of expert knowledge on how forages can be expected to adapt locally, and this will be used to verify the model for this particular location. Similar verification will be carried out in other locations where CIAT has a lot of knowledge.

5.0 TECHNOLOGY AND DEPLOYMENT

The technology used to develop the tool is Borland Delphi and ESRI MapObjects LT (Borland 2001; ESRI 2000). Delphi allows a user-friendly graphical user interface to be developed, and MapObjects allows the inclusion of GIS functionality without the requirement of a proprietary GIS platform such as ArcGIS. MapObjects allows easy inclusion of mapping functionality such as display, zooming and feature identification. However because the model is grid-based, raster functionality is also required. This means that all data is stored in a regular grid rather than in polygons or points, and overlays and grid calculations can be easily carried out between layers because each grid has the same cell-size and extent. Although MapObjects 2.0 allows grid display it does not cater for map algebra, which is necessary for updating map displays dynamically based on probability calculations. Therefore MapObjects LT has been used with bitmaps for grid display, and databases for map algebra.

The aim is to deploy the tool via CD and the internet. The intended users are farmers' advisors, who for the most part will have access to a computer with a CD drive and/or with internet access, but not necessarily access to GIS packages, databases or the training to use them. The tool will initially be deployed via CD to key potential users in Central America, with their feedback used to improve subsequent versions of the tool. It is intended that ongoing training and support will be offered and that the tool will continue to be developed as more data and improved software becomes available.

6.0 CONCLUSIONS

Spatial modelling can be incorporated into an agricultural decision support tool to target species to biophysical and socio-economic niches. Biophysical GIS data combined with trial databases and expert knowledge can remove uncertainty around whether a given species is suitable in a given location, thereby making very efficient use of sparse data in a situation which is desperate for information. By also introducing socio-economic and management knowledge, the probability of a given species being a good recommendation at a given location under certain circumstances can be estimated. Socio-economic factors may be incorporated into the model or included in a subsequent evaluation phase, in discussion with farmers' advisors. Validation is obviously needed, but this can proceed in parallel with model development. Deploying this information as a computer based decision support tool allows farmers' advisors to consider the suitability of many species simultaneously. The GIS component also allows visualisation and interpretation of the difference in species suitability over space.

ACKNOWLEDGEMENTS

The research was funded by the Bundesministerium für Wirtschaftliche Zusammenarbeit und Entwicklung (German Federal Ministry for Economic Cooperation and Development, BMZ). The authors also acknowledge the support from the Deutsche Gesellschaft für Technische Zusammenarbeit (German Agency for Technical Cooperation, GTZ). The work is being undertaken as part of a Doctor of Philosophy degree in Spatial Sciences at Curtin University of Technology, Perth, Western Australia, in conjunction with CIAT.

BIBLIOGRAPHY

Barco, F., Franco, M.A., Franco, L.H., Hincapie, B., Lascano, C.E., Ramírez, G. and Peters, M. (2002) Base de Datos de Recursos Genéticos Multipropósito. Version 1.0. Centro Internacional de Agricultura Tropical (CIAT) Serie CD ROM.

Borland (2001) Delphi Professional computer software, Borland Software Corporation.

Cruz Flores, H. (2002). CIAT Yorito, Honduras. Personal communication 30 Jan 2002.

Delgado, C., Rosegrant, M., Steinfeld, H., Ehui, S. and Courbois, C. (1999) Livestock to 2020: The Next Food Revolution, 2020 Brief 61, http://www.ifpri.org/2020/briefs/number61.htm, accessed 9 October 2002.

ESRI (2000) MapObjects LT computer software, Environmental Systems Research Institute, Redlands, California.

IFAD (2001) Regional Strategy Paper. Programme Management Division, IFAD.

Leemans, R., (1990) Global data sets collected and compiled by the Biosphere Project, Working Paper, IIASA-Laxenburg, Austria

Pearl, J. (1990) Bayesian Decision Methods, *in* Shafer, G. and Pearl, J. (*eds*) Readings in Uncertain Reasoning, Morgan Kaufmann, San Mateo California, pp. 345-352.