C3P

Crop Crisis Control Project

THE USE OF GEOGRAPHIC INFORMATION SYSTEMS (GIS) IN CROP DISEASE MANAGEMENT - THE C3P EXPERIENCE

Christopher Legg

Brief 2

Introduction

The use of Global Positioning Systems (GPS) and Geographic Information Systems (GIS) was an integral part of the Crop Crisis Control Project (C3P). It was recognized that the urgent nature of the project, its wide geographic area, and the large amounts of diverse data required and planned for collection, dictated precise geographic positioning and objective synthesis, analysis and presentation. Much has been learned during the life of the project, and the purpose of this brief is to summarize the contributions of GPS and GIS to C3P, and to indicate where there is room for improvement in future projects

Global Positioning Systems

The technology of Global Positioning Systems (GPS) has been in existence since the 1970's, but became more widely accessible in the 1990's with the advent of more portable and affordable receivers. Wider use of GPS, and the advantages of large-scale production and intense competition, has further reduced unit prices, so that by the start of the C3P program hand held GPS receivers were available for slightly more than US$100 per unit. This can be compared with US$25,000 for a receiver which took four hours to determine a location in 1979, and US$1200 for a unit with 150 meter accuracy in 1994. The GPS receivers of the new generation are also very rugged and easy to use, and have become an essential item of field equipment for all kinds of fieldwork, from geology to agriculture, hydrology to archaeology.

Christopher Legg: C3P GIS Consultant

Main applications of GPS in C3P

a) Disease mapping.

GPS has been used in surveys for cassava mosaic disease (CMD) in East and Central Africa by Dr. James Legg and his associates for many years, and was thus already an established part of such disease surveys before the start of C3P. Typically, surveys for CMD and BXW select sample fields of cassava or banana on a random basis and assess plants in the fields for disease incidence and severity. The center or corner point of the field is recorded by GPS, and all disease data recorded in a spreadsheet together with the geographic coordinates from GPS. This spreadsheet can then be imported into a GIS for geographic processing. Table 1 shows a portion of such a spreadsheet, while the geographic point locations derived from this spreadsheet are shown in Figure 1.

b) Estimating the supply of CMD-resistant material

Inventories of locations and amounts of planting material of CMD-resistant cassava varieties were undertaken in all six C3P countries in 2006 and 2007.

Table 1 *Part of a spreadsheet with data from a CMD survey in Burundi*

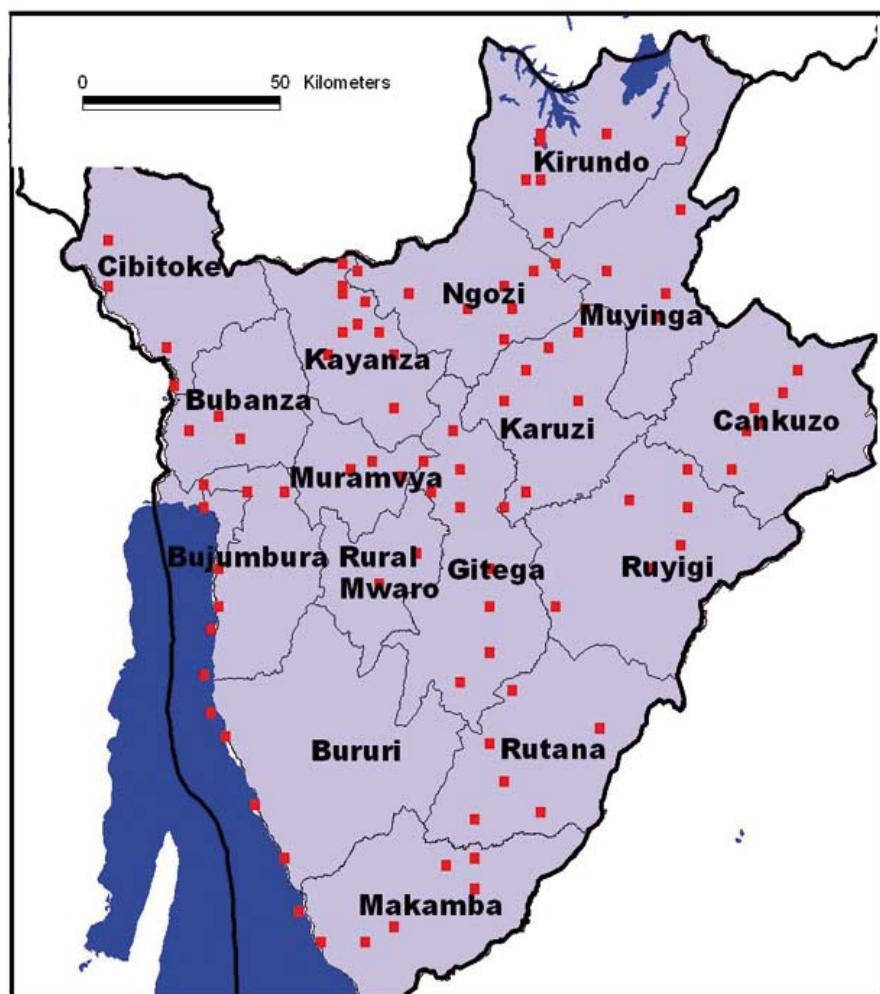| field	Commune	Var1	declong	declatt	CMDsev	totinc	virus
1	Matongo	Gitamisi	29.6	-3.0167	3.75	0.13	U
2	Kayanza	Gitamisi	29.6333	-2.9667	4.25	0.27	A+U
3	Kayanza	Rubona	29.6333	-2.8833	3	0.9	A+U
4	Kayanza	Gitamisi	29.6833	-2.9	4	0.17	A

Figure 1: *Locations of CMD survey sites in Burundi in 2006 from GPS data in Table 1*

The location of each site was recorded using GPS, and locations in Uganda are shown in Figure 2. These data were first of all important in revealing that some sources of planting material were far away from the areas where they were most needed, in terms of CMD incidence, food security and dependence on cassava. Cassava cuttings should not be transported over large distances due to the high transport cost, and also because of high mortality of cuttings during transport. The data highlighted areas where new planting material sites were required.

c) Inventory of cassava multiplication sites established under C3P

Country Program Managers (CPMs) and some of their partners were trained in the use of GPS during training workshops organized by C3P. After this training, GPS were used to define the locations of all multiplication fields established under C3P. The distribution of multiplication fields and varieties of each field within a project area can show where existing work is being done and highlight future needs. The distribution of farmers within a district, symbolized by the varieties planted and numbers of cuttings distributed, can show precisely where the main effort has been directed, and highlight gaps to be filled by future interventions. A map showing the distribution of four common disease-resistant cassava varieties in Eastern DRC is shown in Figure 3. It should be remembered that all these data, and similar data within the other five project countries, were collected by the CPMs and their partners, not by specialist GPS staff from IITA or CRS.
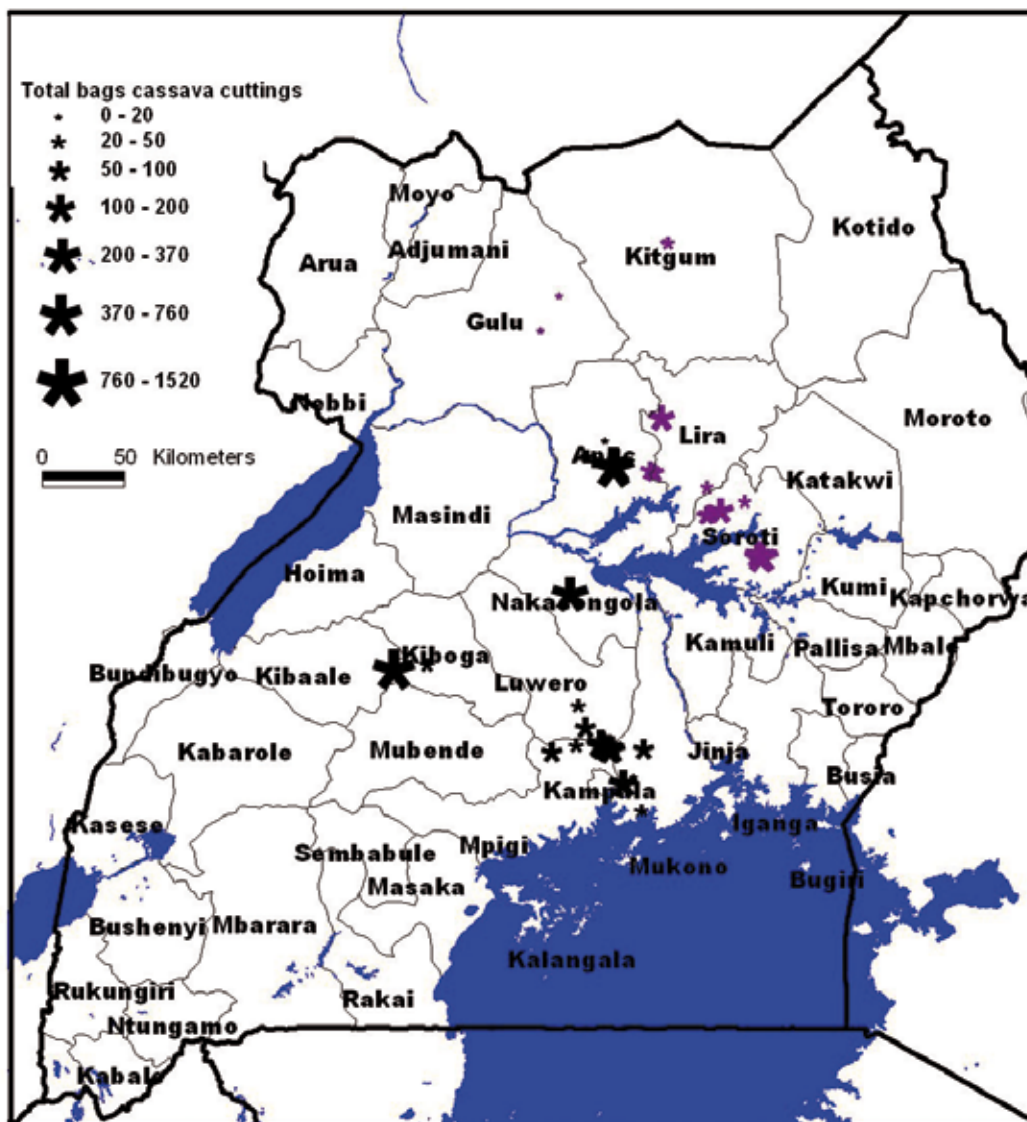

Figure 2 *Locations of CMD-resistant cassava planting material in Uganda recorded in surveys undertaken in 2006 (black) and 2007 (purple).*
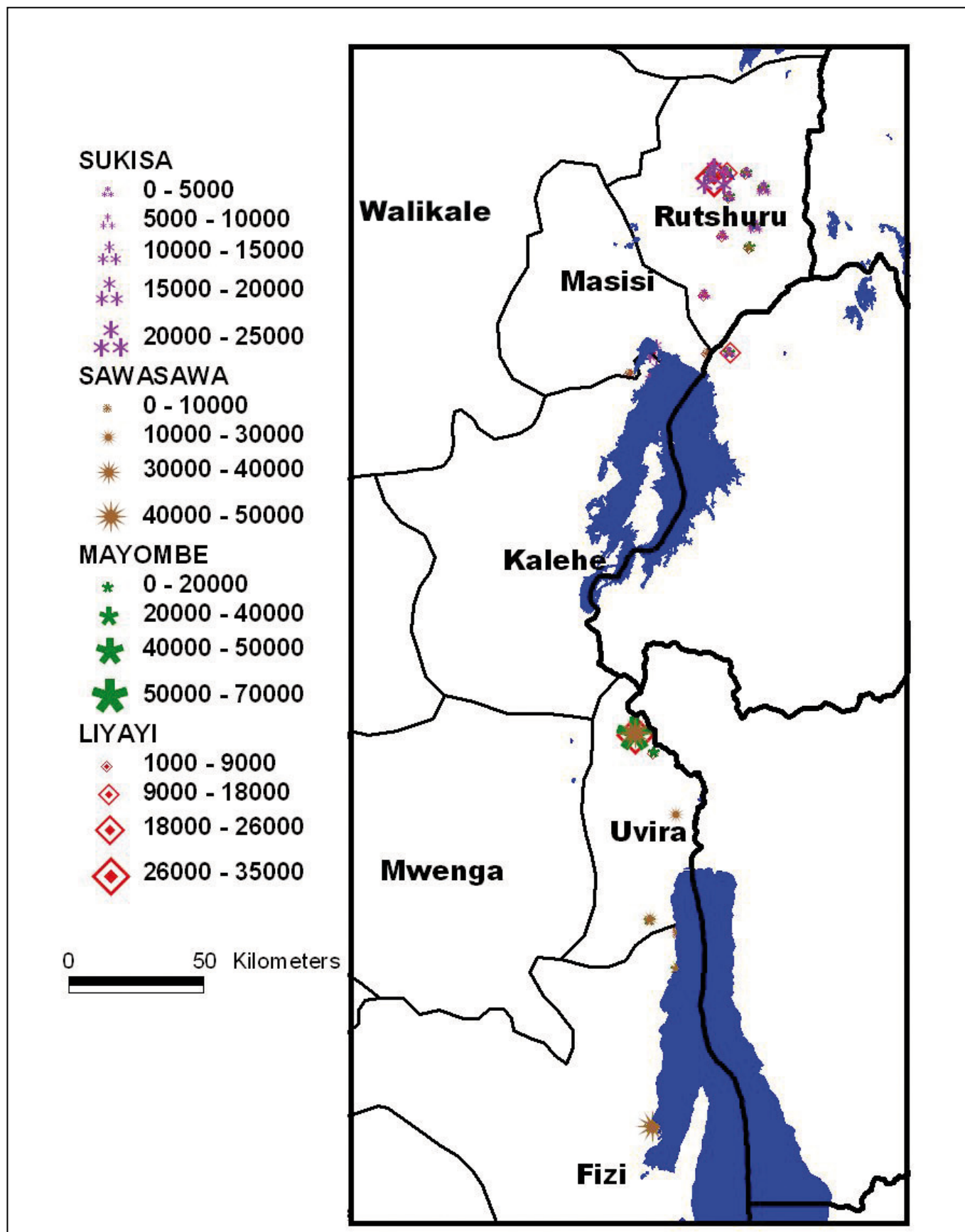
Figure 3 *Cassava planting material (number of cuttings) available in multiplication fields established through C3P in Eastern DRC. Locations of fields measured by GPS. Four main resistant varieties symbolized in this map, with size of symbol indicating numbers of cuttings. (10 000 cuttings of planting material will plant on average one hectare).*

d) Planting of CMD-resistant varieties by farmers.

All surveys for CMD include recording of up to four varieties of cassava observed at the sample sites. Since these are geo-located by GPS, this enables time series maps to be generated from the GPS data showing the progressive adoption of resistant cassava varieties by farmers in the study areas. Figure 4 shows data collected between 2004 and 2006 in western Kenya, showing the gradual adoption of CMD-resistant varieties by farmers in the area, mainly before the additional activities of C3P. This map is an example of the GIS being used to add a fourth dimension, that of time, to the maps, incorporating data from many different dates of GPS surveys.

Problems with GPS

Problems experienced with GPS data sent from field surveys for further processing in the GIS can be divided into two main categories – those problems which create extra work for the GIS staff, but are eventually soluble, and problems which cannot be resolved without additional fieldwork or at least careful checking of field notebooks.
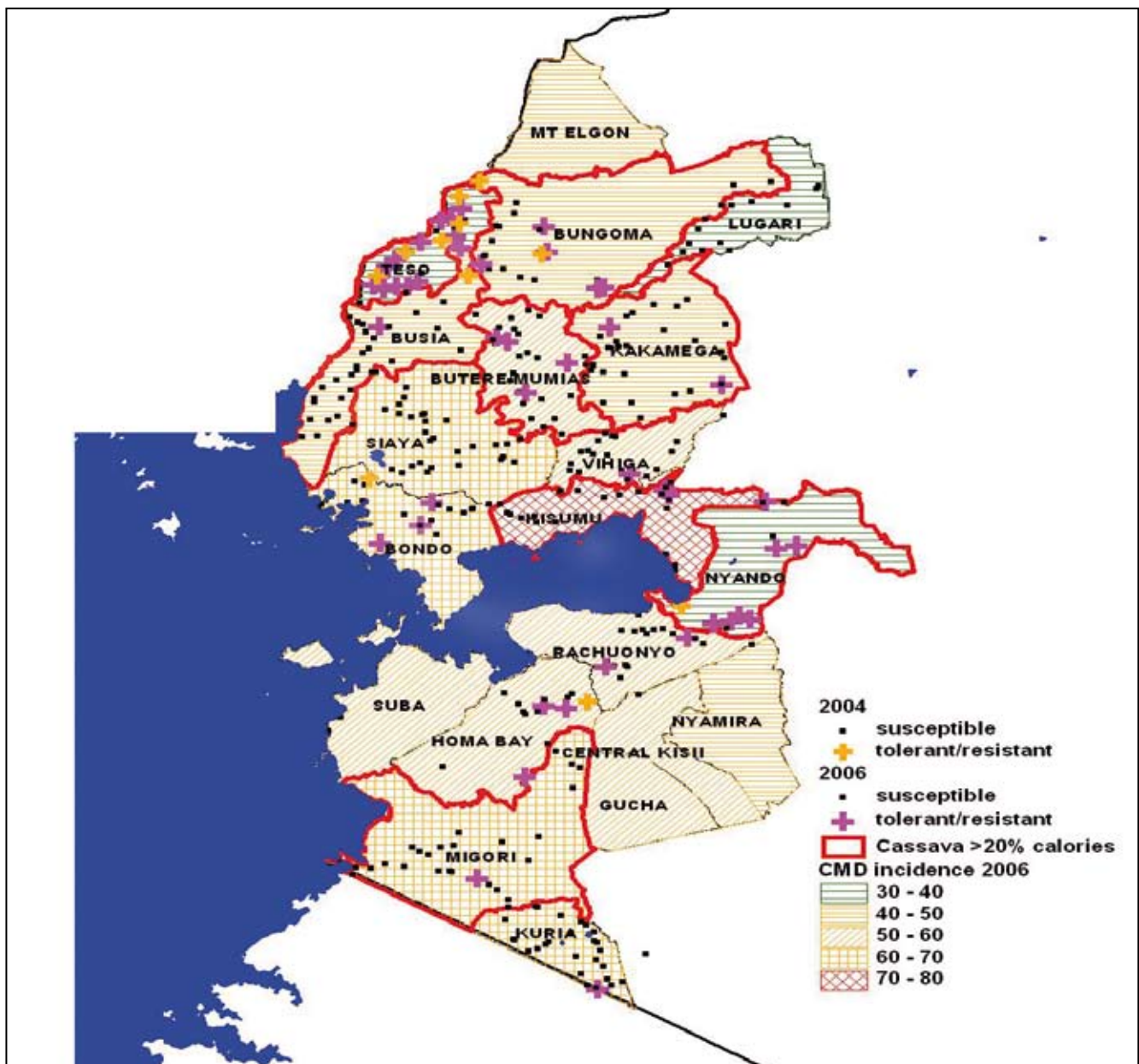

Figure 4 *Locations of CMD-tolerant cassava varieties in western Kenya recorded during CMD surveys between 2004 and 2006.*

In the first category, the most common errors are incorrect format of the numerical latitude-longitude data and omission of any indication whether the latitude values are north or south of the equator. The data format required for entry of globab positioning system (GPS) points into a global information system (GIS) is decimal degrees (for example 23.46578), but most GPS units are pre-set to a default format of degrees, minutes and decimal minutes (for example 23 deg 28.315 min). The GPS user can set the output format to decimal degrees, but if the locations are sent to the GIS in the degree, minute, decimal minute format, it is a fairly trivial matter for the GIS technician to convert this to the correct format. In latitude-longitude projections, it is the convention that locations north of the equator are positive, south of the equator negative. Alternatively, the decimal degree latitude values can be preceded by an "N" or an "S". This is important in the C3P countries, since Uganda, Kenya and DRC all straddle the equator, and the project has worked north and south of the equator in all countries. If the GPS user omits an indication of whether the point is north or south of the equator, it is usually possible to correct this using other information in the data, for example district names, although when a single district contains the equator this may not be possible. Longitude values are always east of the Greenwich meridian in the C3P areas, and so, by convention, are always positive, therefore this is not a source of confusion (as it is in some countries of West Africa).

The second category of GPS errors, those which are usually not correctable by the GIS staff, are almost always caused by GPS field staff not transferring GPS data electronically to a computer using the connecting cables provided by the project. GPS locations are read from the GPS unit in the field, and copied into a field notebook or onto a field reporting form. Later (sometimes much later) these numbers are copied once again into a computer. GPS locations are long numbers, the screens of GPS units are small and often difficult to read in bright sunlight, and errors in transcription are very common, especially where the GPS operator is not sufficiently experienced to recognize that the number just entered is improbable in the context of other recent readings. If the transcription error is simple one of transposing numbers (for example 34.251 instead of 43.251) the error can be easily detected and corrected at the GIS stage, but other errors may go undetected, or be impossible to correct.

Recommendations for Future GPS

It should be noted that the majority of GPS users in C3P did not receive any specialized training in the use of GPS. The quality of GPS data in future projects in East Africa can be greatly improved by more training, decentralized to country or even district level. This should include special emphasis on the data formats required for easy conversion of global posioning systems (GPS) points to global information systems (GIS), and on the importance of indications of north or south in countries crossing the equator. Any future projects should ensure that enough data transfer cables are available to service the planned number of field crews, and training should emphasize the importance of transferring GPS data electronically. The typical GPS system can store at least 500 points, and memories of GPS units are very rugged, so unless data are deliberately deleted, this electronic transfer does not need to be done every day, or even every week, under most working situations. It is suggested that field notes should still contain GPS locations read from the instrument in the field, linked to waypoint numbers and dates corresponding to the stored data in the GPS. This will ensure that all manually recorded data on disease, availability of planting material, locations of farmers, and so on, will always be linkable to the GPS locations at a later date.

It is recommended that a GPS training officer should be appointed for each country, to provide training as required by field staff during the life of the project, and to upgrade skills by regular refresher courses. These training officers should themselves be trained by an experienced GPS/GIS scientist at the outset of the new project.

Geographic Information Systems

Geographic Information Systems (GIS) provide powerful tools for combining and displaying different data sets, for extrapolating from data samples to the whole study area, for defining targets based on mixtures of criteria, and for temporal modeling in a geographic framework. All these activities have taken place within C3P, and are likely to continue in follow on projects.

Main GIS applications in C3P

a) Visualization of data
Any data which have a geographic distribution (where each data point, line or area has a specific and

different geographic location) can be much better understood when represented in map form rather than in a table. Patterns of distribution of high and low data values which may not be obvious in a table become apparent in a map. Different data sets can be displayed in a single map, allowing visual recognition of the relationships between them. Data from different countries can be displayed simultaneously to produce a map of the region.

An example of data visualization through GIS maps is shown in Figure 5, which combines three different data sets for Rwanda and Burundi. The background color for each district shows the percentage calorific dependence on cassava. Grid line designs show the

incidence of CMD at a district level. Points show the locations of CMD-resistant cassava planting material. A composite map such as this can present important data required for decision-making to a non-specialist audience.

b) Extrapolation of disease and food security results

Surveys for diseases such as CMD and BXW are usually carried out using a semi-random sampling method, where cassava or banana fields are selected at random along roads which effectively cover the whole area of interest. This results in a large series of individual data points with different disease statistics which are hoped to be representative, in their variance
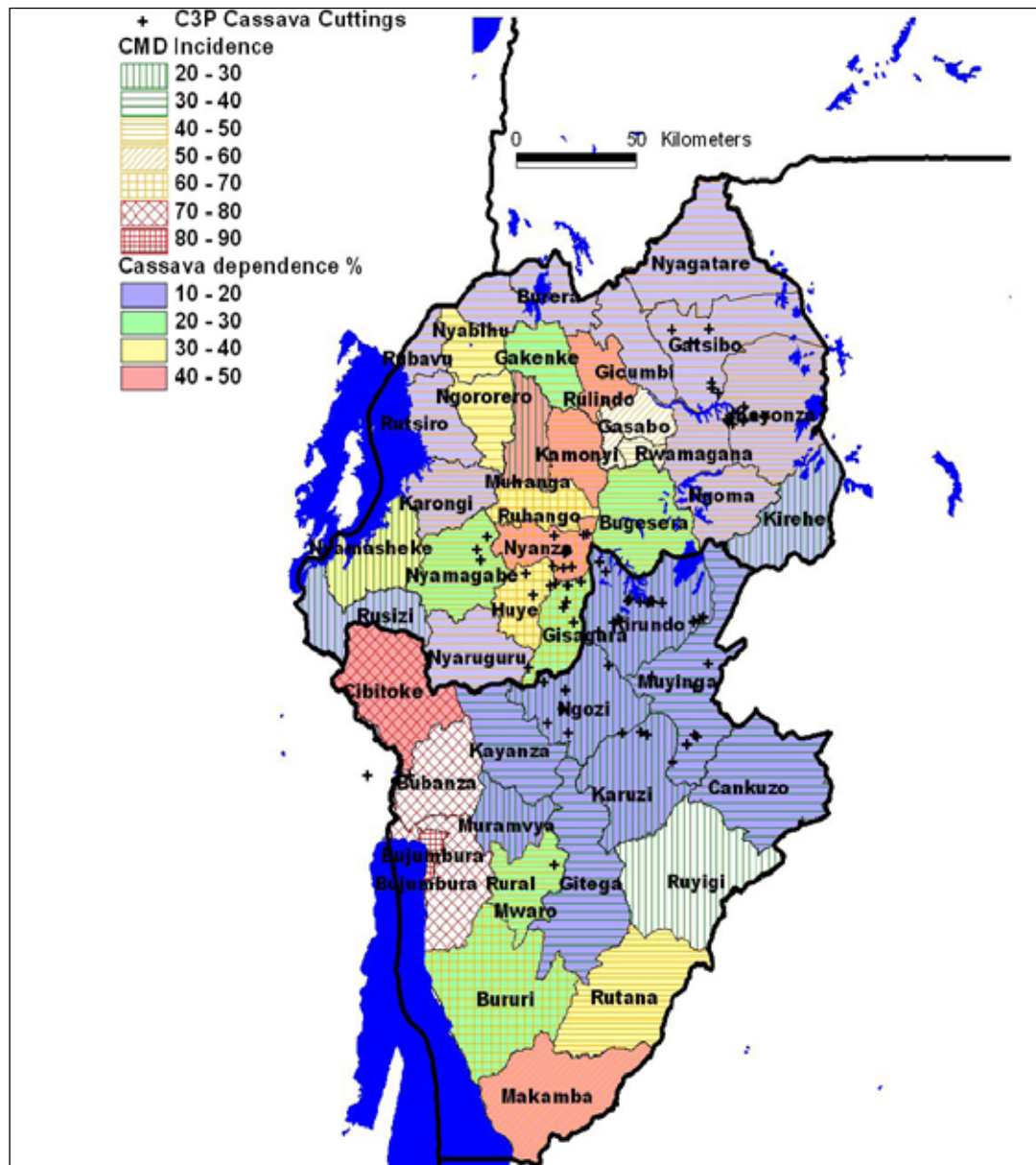

Figure 5: *Cassava calorific dependence, CMD incidence and availability of cassava cuttings in Rwanda and Burundi*

as well as their means, of the whole area sampled. The first stage in GIS processing of these randomly located points is to estimate the values at each point of a regular grid, a process of interpolation, which can use algorithms of varying degrees of complexity. Following experimentation with geostatistical techniques such as kriging, a simple Inverse Distance Weighting interpolator was found to be fast and effective. Average values for incidence, severity or any other numeric descriptor of the disease can then be calculated for irregular polygonal areas such as districts.

Most of the food security surveys for C3P have also been sample surveys, in that a random sample of households in selected districts were surveyed. This, when analyzed by the IITA agricultural economists, provided food security indicators for the sampled districts, but not for the remainder of the country. Extrapolation of cassava and banana percentage calorific dependency from sampled districts to the whole country is possible if agricultural production statistics are available at the same level of administrative unit as the food security survey. For the districts sampled in the food security survey, the percentage calorific content of cassava and banana in total staple food production is calculated, and a regression equation derived between calorific dependence from the food security survey and calorific contribution from production. In most countries the correlation was very good (R-squared > 0.5), indicating that most food consumption was from own production or produced in the same district. This regression equation was then applied to the districts not covered in the food security survey in order to estimate the cassava and banana calorific dependence. The same method could not be reliably used for other food security indicators. Attempts were made to relate FS1 (calorie availability from own production) and FS2 (potential calories from production and off-farm income)indicators to agricultural production, climatic factors and distance to markets, but no significant correlation was seen.

c) Targeting for CMD and BXW interventions
This was intended to be one of the main applications of GIS within C3P. Target areas were to be selected interactively by objective analysis of available data, presented for group discussion using the GIS, and target priorities were to be updated regularly based on new data on disease incidence and severity, as well as on food security. In practice, as will be discussed

further below, information was not available for many countries in the planning stage of the project, making initial target selection difficult, and updates of disease and food security data were not sufficiently frequent to allow adaptive adjustment of targets to respond to a dynamic disease and food security situation.

Despite these limitations, GIS targeting was still important within the project, and provided the basis for much discussion at planning meetings. It was also important in planning the follow on CMD project to C3P. The simplest GIS targeting technique is to attempt to display all the relevant information in a single map, so that someone knowledgeable of the area would then have a clear view of the interactions of different factors, and could make an informed choice of targets.

Figure 6 is a map of northwestern Tanzania combining cassava calorific dependence with food security, as extrapolated from survey data. This highlights districts which are both cassava-dependent and food-insecure, and, when combined with CMD data, would enable for a more explicit discussion on targeting.

Another targeting technique is illustrated in Table 2 and Figure 7. A series of factors considered most important in selecting areas in highest need of assistance are first chosen, usually because of discussion amongst people with knowledge of the area. For this example, CMD incidence, the number of farming households, and the annual cassava production in the district are selected as being the critical factors. Others could also have been used, for example the calorific dependence on cassava, per capita calorific intake per day or land available per household. Each of these chosen variables is then assigned a rank from 1 to 4, with rank = 1 being the most important in terms of C3P aims and rank = 4 least important. A total ranking is then derived from the individual ranks. This can be a simple sum of ranks, or one of the variables could be assigned a greater weight than others. In this example, the number of households has been assigned a weighting of 1 while cassava production and CMD incidence are assigned weights of 2. The districts can then be classified according to their total ranks, as indicated by the color-coding in Table 2. Figure 7 is a map of these target districts in northwestern Tanzania.

This targeting technique was used as an aid to planning for the follow on Great Lakes Cassava Initiative.

d) Modeling disease spread

One of the most powerful applications of GIS in epidemiological studies is to model the probable spread of disease and thus permit planning of pre-emptive action. Funding within C3P did not permit full exploration of this potential, but this should be an important part of any follow on programs.

Problems with GIS

One of the first problems delaying the full utilization of GIS in a project of this type is the problem of two-way communication between the GIS specialist and the other project participants, who may not yet be familiar with the ways in which GIS can help them in their work. Project staff may not know what they can reasonably request from the GIS, and the GIS specialist may not fully understand the problems and constraints of project implementation. The result may be that the

GIS specialist provides inappropriate or misleading maps, while the project staff does not understand enough about the technology to make clear to the GIS specialist what it is that they really want.

GIS technology has now evolved to the point where it has become accessible, in terms of simplicity and cost, to non-specialist users. Most of the basic tasks, such as importing and displaying GPS points against a background of roads, villages, rivers and topography, and generating maps of agricultural production, food security, rainfall and population density (to give a few examples from many possibilities), are within the ability of any computer literate scientist or technician. The data sets are all available, or have been generated by GIS specialists working for the project. Most project staff and partners now have access to computers of sufficient power and speed. Freeware and open source
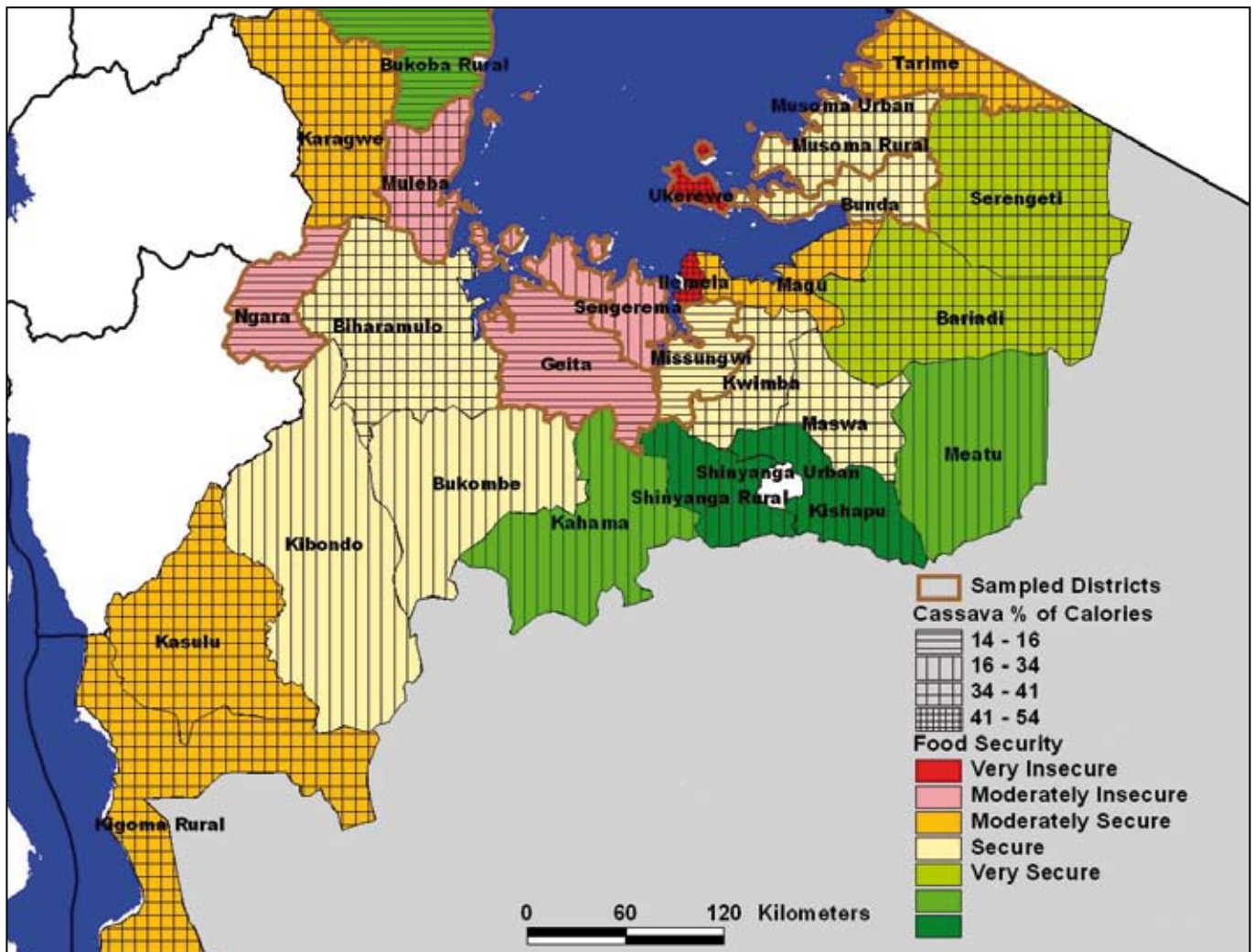

Figure 6. *Map of NW Tanzania showing food security and cassava contribution to diet.*

Table 2 *Targeting districts in Tanzania, based on CMD incidence, numbers of farming households and cassava production.*

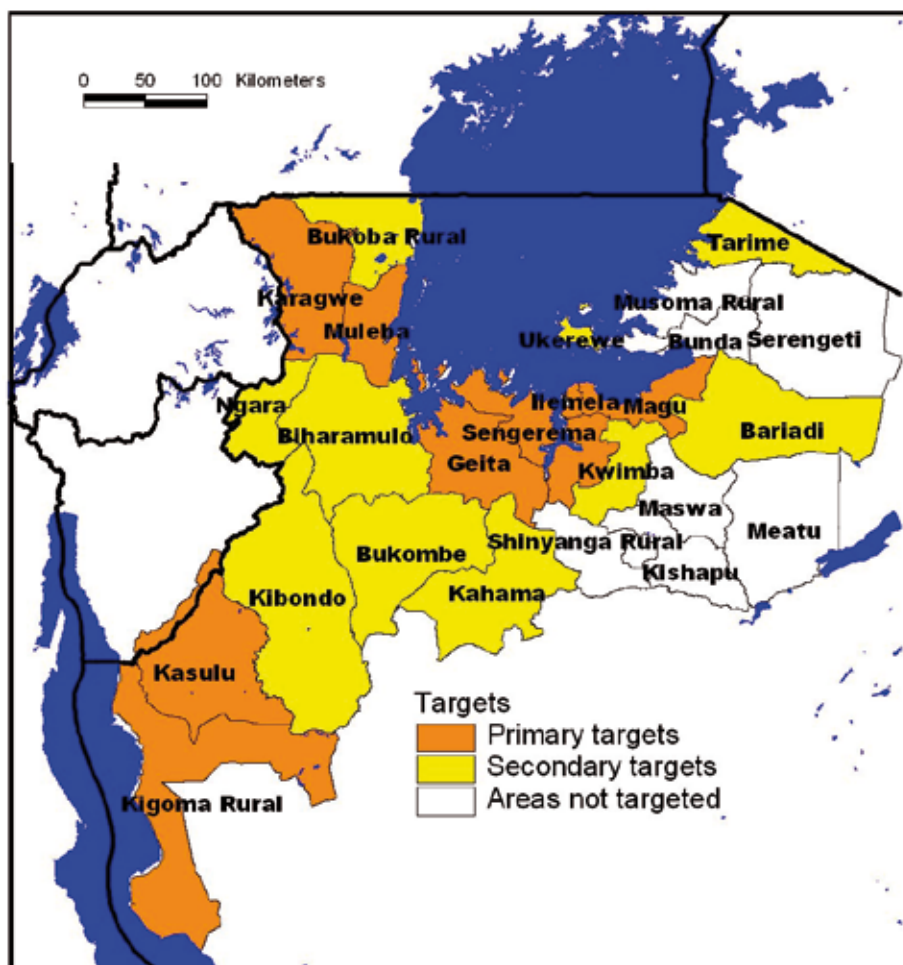| DISTRICT	CMD incid	# farm hh	Cassava tonnes	Rank hh	Rank cass	Rank CMD	Total rank
Kigoma Rural	57	58712	222200	1	1	1	5
Geita	67	85089	290400	1	1	1	5
Sengerema	67	59879	290400	1	1	1	5
Magu	59	49800	290400	2	1	1	6
Ilemela	69	31784	290400	3	1	1	7
Muleba	57	46222	112800	2	2	1	8
Karagwe	46	50914	112800	1	2	2	9
Kasulu	40	76409	112800	1	2	2	9
Missungwi	45	30735	290400	3	1	2	9
Biharamulo	60	49126	56800	2	3	1	10
Bukombe	65	47435	96100	2	3	1	10
Ngara	42	40129	112800	2	2	2	10
Bukoba Rural	44	47282	112800	2	2	2	10
Kibondo	39	49653	222200	2	1	3	10
Kwimba	28	37791	290400	3	1	3	11
Tarime	39	58887	82100	1	3	3	13
Bariadi	27	72432	96100	1	3	3	13
Kahama	22	71386	96100	1	3	3	13
Ukerewe	12	31299	290400	3	1	4	13
Maswa	21	36528	96100	3	3	3	15
Meatu	20	29785	96100	4	3	3	16
Bunda	22	31071	1300	3	4	3	17
Musoma Rural	12	39578	82100	3	3	4	17
Shinyanga Rural	10	33167	96100	3	3	4	17
Serengeti	16	21126	82100	4	3	4	18
Kishapu	14	28716	96100	4	3	4	18

Figure 7. *Target areas in northwestern Tanzania, from data in Table 2.*

GIS software packages are now easy to use, flexible and powerful. The task of a GIS specialist supporting a project such as C3P should not only include the collation of basic data sets and their re-processing as necessary to make them more accessible, but also the training of project and field staff to undertake many basic GIS activities themselves in the same way that they use word processing and spreadsheet software. This has not happened within C3P. A training program was organized at a late stage in the project, and a few Country Program Managers (CPMs) produced their own maps from GPS data, but most GIS processing and map production was undertaken by the GIS specialist, not by other project staff at their own desks. The ultimate aim should be that the project professionals produce most of the basic maps themselves, while the GIS specialist provides continuing support to them, updating and revising data sets as necessary and providing email support, while at the same time carrying out the more complex data integration, analysis and modeling tasks for the project.

A recurrent problem with C3P has been that maps have not been delivered in time to be effectively used for planning the next step in the project. It was initially hoped that all targeting of inputs within C3P would be guided by GIS analysis, based on frequently updated and easily understood maps. In practice, the required maps were rarely available when decisions had to be made. Targeting maps had to be based on the latest available knowledge on food security, agricultural production and disease status. Particularly at the beginning of the project, baseline data on these three variables was critical for objective decision making. Food security surveys take time to plan, undertake and analyze, and only two of the six countries were surveyed shortly before the start of the project. CMD and BXW surveys require highly trained and experienced staff, and the small teams available could not cover all the countries frequently enough to provide up-to-date information when it was needed. Agricultural production statistics are collected, processed and released through government agencies, and delays of as much as two years are common before public access to the data is permitted.

In some of the project countries, detailed agricultural production statistics (by 3rd level administrative area: district or commune), were still not obtained by project staff by the end of C3P, even in cases where the data

were known to exist. In the case of DRC, data were very out of date, and at too coarse a resolution, but this was not surprising considering the recent history of the country. In Burundi, only production data for cassava and bananas were provided at a district level, and other food crops were not available. For Tanzania, production data for a whole range of crops were available at the Region (province) level, but district-level data were only obtained right at the end of the project, and then only for cassava. This meant that extrapolation of data from food security surveys, which normally only sampled selected districts within a country, to other un-sampled districts, was difficult or impossible. Provincial food production statistics can be disaggregated to district level using population and land-use data (AfriCover) but not at a very high level of confidence.

A conceptual problem with GIS targeting against dynamically spreading crop diseases in C3P has been that targets have been identified based, in part, on the actual incidence or severity of BXW or CMD within an area. Apart from the fact that the "actual" situation may be different from that indicated in a survey of up to two years old (in the most extreme cases), it might have been better to target those areas most likely to be affected by the diseases in the near future. This would probably do little to arrest the spread of the virus, but could have a dramatic affect on food security. While it is conceded that it might be difficult to convince farmers of the danger of a pandemic which has not yet arrived, the project could still have prepared the ground by raising awareness and establishing primary planting sites for disease resistant varieties prior to the arrival of the disease.

Targeting should be based partly on the known current disease situation, but partly also on dynamic models which predict the possible directions and speed of disease invasion.

Recommendations for future GIS
- Future projects to combat CMD and BXW in Central and East Africa will have the benefit of extensive data collected during C3P, especially on food security, disease prevalence and availability of resistant planting material, which will provide a reliable baseline on which to base project planning. This will enable more objective initial targeting for any follow on projects.

- Future projects should budget for extensive training of staff to undertake food security and disease surveys, and should also budget for equipment (including GPS units and base computers) and transport and per diem costs for, at the minimum, annual surveys in all project areas. This will enable regular GIS analysis and map production and promote dynamic management of the project.

- Project staff should improve their communication with government departments responsible for collecting and distributing agricultural statistics in order to achieve privileged access to this data at district level as soon as possible after they are collected.

- Project staff should be actively encouraged to undertake their own basic GIS processing. Freeware GIS software should be made widely available, and some kind of incentive such as quarterly competitions and prizes arranged to encourage participation.

- GIS staff responsible for supporting follow on projects should produce dynamic models of the spread of plant diseases based on historic data and on assumed mechanisms of dissemination, in order to predict the direction and speed of future spread.

Conclusion:

GIS and GPS are important tools to be used in managing crop diseases. They promote disease mapping, the location and establishment of planting material that is clean. These technologies can be used to model disease spread and extrapolate data to target new interventions.