Comparison of disease patterns assessed by three independent surveys of cassava mosaic virus disease in Rwanda and Burundi

H. Bouwmeester[a]*, G. B. M. Heuvelink[b], J. P. Legg[a] and J. J. Stoorvogel[b]

[a]*International Institute of Tropical Agriculture, Carolyn House, 26 Dingwall Road, Croydon CR9 3EE, UK; and* [b]*Land Dynamics Group, Wageningen University, Droevendaalsesteeg 3, 6708 PB Wageningen, The Netherlands*

Cassava mosaic disease (CMD) seriously affects cassava yields in Africa. This study compared the spatial distribution of CMD using three independent surveys in Rwanda and Burundi. Geostatistical techniques were used to interpolate the point-based surveys and predict the spatial distributions of different measures of the disease. Correlative relationships were examined for 35 environmental and socio-economic spatial variables of which 31 were correlated to CMD intensity, with the highest correlation coefficients for latitude (−0·47), altitude (−0·36) and temperature (+0·36). The most significant explanatory variables were entered in separate linear regression models for each of the surveys. The models explained 54%, 44% and 22% of the variation in CMD. The residuals of the regression models were interpolated using kriging and added to the regression models to map CMD across both countries. Significant differences were calculated in some areas after correcting for interpolation error. An important explanation of the differences is interaction between the CMD pandemic and the dates of the three surveys. Large relative prediction errors obtained in the regression kriging procedure show the need to improve the survey design and decrease measurement error. Improved maps of crop diseases such as CMD could aid targeting of control interventions and thereby contribute to increasing crop yields. This study validated the unique character of each of the survey approaches adopted and underlines the importance of specific interpretation of results for CMD management. The study emphasizes the need for optimization of sampling designs and survey protocols to maximize the potential of regression kriging.

Keywords: agriculture, cassava, GIS, regression kriging, sampling design, spatial distribution

Introduction

Agricultural production is often considered to be the main factor determining food security in eastern Africa and elsewhere in the tropics. Cassava is the main staple crop in eastern Africa in terms of production and third in terms of value (FAO, 2009). However, cassava yields are greatly reduced by cassava mosaic disease (CMD) (Thresh *et al.*, 1994). CMD is caused by *Cassava mosaic Geminiviruses* (Bock & Woods, 1983), which are disseminated through planting infected cuttings and spread by the whitefly vector, *Bemisia tabaci* (Storey & Nichols, 1938). CMD became an increasing problem in cassava-growing areas of the Lake Victoria Basin of East Africa during the 1990s (Legg *et al.*, 2006). This was a consequence of the region-wide spread of a severe virus variant and associated 'super-abundant' populations of the whitefly vector (Legg & Thresh, 2000; Legg *et al.*, 2006).

*E-mail: h.bouwmeester@solum.nl

First reports of the severe 'pandemic' of CMD were made from Rwanda and Burundi in the early 2000s (Legg *et al.*, 2001; Bigirimana *et al.*, 2004) following previous epidemics in Uganda and Kenya. Prior to the spread of the CMD pandemic into Rwanda and Burundi, CMD incidences were moderate to low, symptoms generally mild and *African cassava mosaic virus* (ACMV) was the predominant CMD-causing virus (Legg *et al.*, 2006). Following the beginning of the pandemic in these two countries, incidences increased, symptoms became more severe, and infections by the pandemic-associated *East African cassava mosaic virus-Uganda* (EACMV-UG) and mixed ACMV + EACMV-UG infections began to predominate. Severe CMD spread into Rwanda and Burundi from the northeast, entering Rwanda from neighbouring Uganda in 1997, and entering Burundi from Rwanda and Tanzania in 2003 (Legg *et al.*, 2006). By 2007, severe CMD had affected all regions of Rwanda (Kanyanga *et al.*, 2007) and all but the far south-west of Burundi (Bigirimana *et al.*, 2007). Large-scale initiatives to control the CMD pandemic were initiated in Uganda in the early 1990s using virus-resistant varieties (Otim-Nape *et al.*, 2000) and were subsequently extended throughout East and Central Africa (Legg *et al.*, 2005). Consequently, in parts of the

region, particularly Uganda and western Kenya, where control efforts have been underway for longest, CMD is no longer considered to be the major production constraint for cassava production (Fermont *et al.*, 2009). However, a recent study of six countries in the Great Lakes region revealed that most farmers still consider the disease to be the main constraint to cassava production (Kimetrica, 2008).

In order to design strategies to control the CMD pandemic, it is essential to gain insights into the spatial distribution of the disease. Over the past decade a number of surveys by socio-economists and biologists have attempted to describe the geographic spread of CMD and estimate its impact on overall production and food security (Sseruwagi *et al.*, 2004; Legg *et al.*, 2006; Abele *et al.*, 2007). Comparing point observations on the incidence and severity of the disease on individual farms is difficult because the observations recorded different facets of CMD and were taken at different times and at different locations. This problem is often dealt with by averaging farm data by administrative units (Bouwmeester *et al.*, 2009). In Uganda, disease incidence in banana has been mapped and linked to administrative units in order to describe the temporal spread of disease (Tusheremereirwe *et al.*, 2006). A similar approach was used to describe the early spread of the CMD pandemic in East and Central Africa (Legg, 1999). An alternative, more accurate method of transforming disease data from points to surfaces is geostatistical interpolation (De Smith *et al.*, 2007; Bouwmeester *et al.*, 2008). A wide range of geostatistical techniques is available (Goovaerts, 1997) that have been used in research areas of soil science (McBratney *et al.*, 2000), livelihood analysis (Cecchi *et al.*, 2010), climatology (Hijmans *et al.*, 2005) and epidemiology (Clements *et al.*, 2006). In crop science, geostatistics is used in describing patterns of pathogens and crop diseases (Chellemi *et al.*, 1988; Gandah *et al.*, 2000; Stonard *et al.*, 2010) and to describe CMD patterns in the Ivory Coast (Lecoustre *et al.*, 1989). However, these studies do not utilize auxiliary information on factors that could influence disease distribution. The accuracy of geostatistical interpolation can be improved by using relationships between the variable to be predicted and readily available predictor maps (Hengl *et al.*, 2004; Kempen *et al.*, 2009). So far, this so-called regression kriging technique has not been used to compare the spatial spread of crop diseases of different surveys across large areas. However, such comparisons could be useful for monitoring and implementing management strategies to prevent disease expansion because it makes possible upscaling and validation of survey information to large areas.

Different survey techniques are currently applied to help understand the spread of crop diseases through a population of farms and its effect on crop production. This study explored whether three very different survey approaches resulted in similar disease patterns, after standardization of the observations. The degree of divergence or similarity revealed the comparability of the approaches and hence determined whether the different assessments could be used interchangeably. The findings could have important implications for the design, implementation and interpretation of future disease surveys. The overall objective of this study was therefore to analyse the spatial comparability of different survey approaches. Three recent surveys that assessed the severity of CMD in Rwanda and Burundi and in neighbouring countries were selected for comparison.

Materials and methods

Study area

The landscapes of Rwanda and Burundi are both strongly influenced by the Albertine rift that stretches from north to south in the western part of the countries (Fig. 1). Although the countries are relatively small in area they have a diverse agro-ecology, dictated by the variability in topography and climatic conditions. For example, the countries have an average annual precipitation of 1176 mm with local minimum values of 851 mm and maximum values of 2178 mm (Hijmans *et al.*, 2005). The mean annual temperature in the countries is 19·2°C, with local minimum values of 5·6°C and maximum values of 25·5°C (Hijmans *et al.*, 2005). Soil fertility is highly variable due to the topography and volcanic activity in parts of the area. Because most of the farming in Rwanda and Burundi is low-input subsistence production, the region is particularly vulnerable to the effects of crop diseases like CMD.

Survey datasets

The data used in this study were collected in surveys of Rwanda and Burundi, and in adjacent parts of Tanzania, Kenya, Uganda and the Democratic Republic of Congo (DRC). The studies differed in their sampling density (Table 1) and distribution (Fig. 1). The non-uniform sampling designs were not spatially balanced, mainly as a result of financial and temporal constraints. The designs can be described as a mixture of purposive (targeting distinct areas), random (within targeted areas) and convenience sampling (along roads) (Binns *et al.*, 2006; De Gruijter *et al.*, 2007). All datasets have a limited coverage of northern Rwanda where the high altitude is unfavourable for cassava cultivation.

Each survey recorded the relevant disease parameter and the geographic coordinates of the observation using a hand-held geographic positioning system (accuracy approximately 20 m). From the original surveys a subsample that covered Rwanda and Burundi was selected because these countries were spatially best represented in all surveys (Fig. 1; Table 1). All observation sites in DRC, Uganda and Tanzania that were within 50 km of the national borders of Rwanda and Burundi (area within the dotted line in Fig. 1a) were included to limit border effects and favour interpolation over extrapolation. Relatively many sites were located on the shores of Lakes Tanganyika and Kivu. Due to errors in geo-referencing of sites or by imprecise mapping of the lakeshore, some of the sites appeared to be within the lakes and were removed from the datasets. Some sites had identical geospatial references. If the CMD parameter
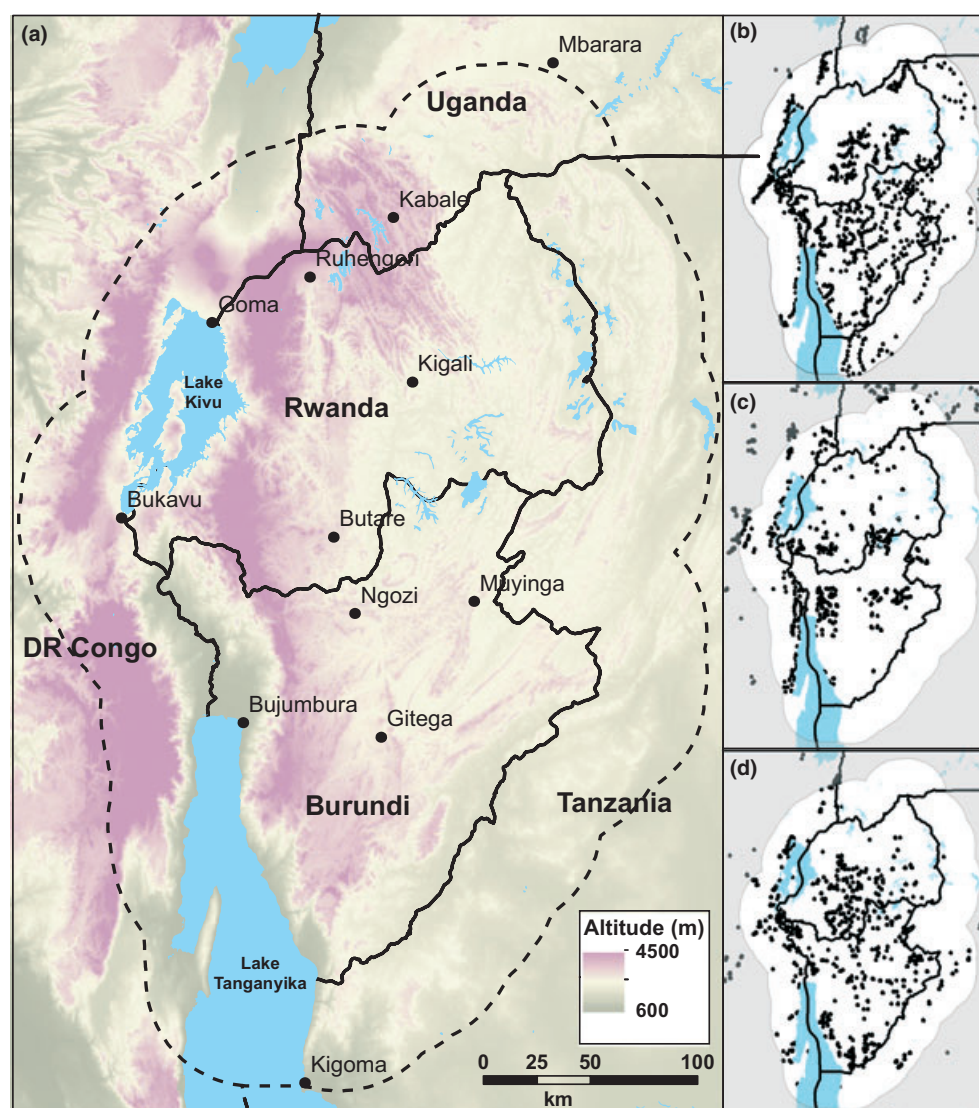
Figure 1 Study area (Rwanda and Burundi) with altitude (a) and sampling sites (dots) of three surveys: C3P-D (b), C3P-S (c) and GLCI (d). Dotted line in (a) represents area in the adjacent countries of DRC, Uganda and Tanzania within which observations were included in the analysis.

Table 1 Number of sites sampled in the three cassava mosaic disease surveys. The buffer zone corresponds to the area of the adjoining countries Democratic Republic of Congo, Uganda and Tanzania that lie within 50 km of the national borders of Rwanda and Burundi

Description of observation sites	C3P-D	C3P-S	GLCI
In initial survey (raw data)	1967	2279	7624
Inside Rwanda, Burundi and buffer zone	786	942	3754
Observation sites used in this study	775	871	3667
Inside Rwanda	199	367	1354
Inside Burundi	306	207	1508
Outside Rwanda and Burundi but inside buffer zone	270	297	805

at those locations was identical then one of the two observations was removed, otherwise both observations were removed. All further analyses were based on the survey

observations within this subsample. The three datasets were initially geo-referenced in WGS84 coordinates and were transformed to a UTM projection zone 36S.

The first survey (C3P-D) was conducted between January 2006 and February 2007. It aimed to assess the spatial distribution of CMD and of the pandemic-associated EACMV-UG. Only the main cassava-growing areas were surveyed (Bigirimana *et al.*, 2007) with a preference for those along roadsides. Young fields, 3–6 months old, were selected for sampling. CMD incidence and severity were recorded by visually examining 30 plants at each sampling site that were selected along representative transects along two diagonals in the form of an 'X'. CMD incidence was defined as the percentage of infected plants (0–100). CMD severity was the average of the severity scores for the 30 plants. Severity scores were categorical and ranged

from 1 for no symptoms to 5 for the most severe damage (Sseruwagi *et al.*, 2004). Symptomless plants (severity score 1) were excluded when calculating average severity. As cassava production losses are greatest when both incidence and severity are high (Thresh *et al.*, 1994), a measure of disease 'intensity' was calculated as the product of CMD incidence and CMD severity. The parameter used in this study was CMD intensity as it relates better to socio-economic measures of CMD than incidence or severity alone.

The second survey (C3P-S) was conducted between May 2006 and August 2007 (Abele *et al.*, 2007). Cassava and/or banana-growing areas were targeted in an attempt to characterize the socio-economic status of farmers. At each sampling site, farmers were asked how much of the previous season's cassava production had been lost as a result of CMD. The parameter used in this study was the farmer estimate of the percentage production loss (0–100).

The third survey (GLCI) was conducted between May and August 2008 to assess the impact of the viruses CMD and cassava brown streak disease (CBSD) (Kimetrica, 2008). The sites were randomly selected within areas where project partners (Catholic Relief Services and Food and Agricultural Organization of the United Nations) were active. While the sites of the C3P-D and the C3P-S surveys were typically scattered over the area of interest, the GLCI sites were more clustered, with groups of about 10 observations per village assessed. Each farmer was shown pictures of symptoms associated with CMD and asked if these were seen on their farm. The parameter used in this study was the positive or negative response of the farmer which was interpreted as prevalence of CMD.

The three assessments of CMD were made with contrasting methodologies. It is difficult to quantify the accuracy and confidence of the variables that were recorded without doing additional validation surveys. However, it is likely that C3P-D had the greatest accuracy, because each record of incidence and severity was based on 30 sampled plants assessed by trained observers. In contrast, the CMD assessed in the socio-economic C3P-S and GLCI surveys reflected the views of individual farmers. In such situations, elements such as the phrasing of the question or the knowledge of the farmer can significantly influence the response. However, it is significant that visual aids were used during the socio-economic surveys, as standard images of cassava plants infected by CMD were shown to farmers to aid recognition.

The current study aims to predict and compare the spread of CMD across the two countries using the three different methods of CMD assessment. Inevitably the imperfect accuracy of the methods of assessment will account for at least part of the differences between the resulting predictions. Despite this, the value of the comparisons is not diminished because all approaches have their intrinsic advantages and disadvantages. Socio-economic variables may be less accurate, yet they provide a vital indication of farmer perception, which is an important measure of the overall impact of a crop disease.
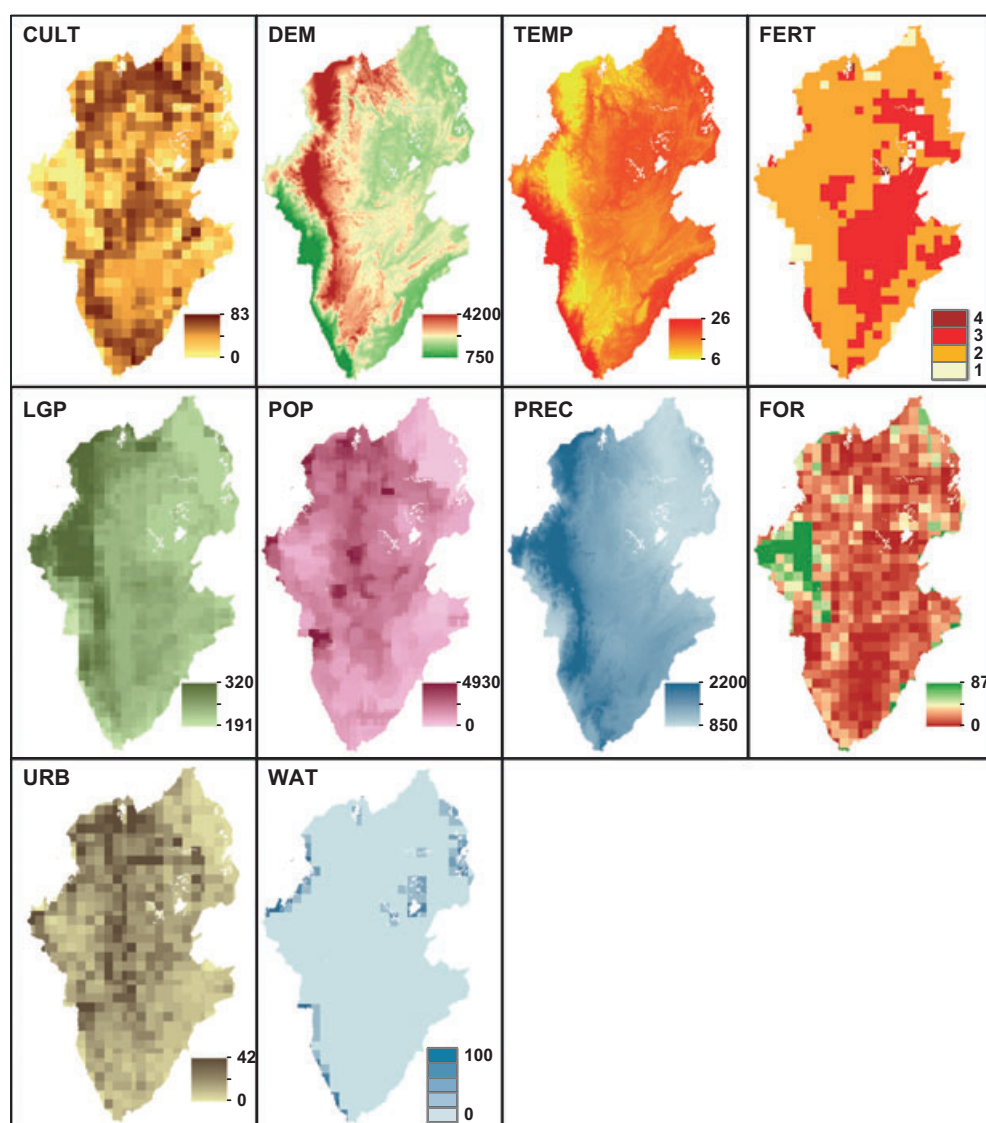
Predictor maps

To improve the spatial interpolation of CMD, 10 publicly available predictor maps were selected (Fig. 2; Table 2). These predictor maps describe the environmental and socio-economic conditions of the study area and potentially explain part of the variability in CMD. All predictor maps were resampled to a 1 km resolution using ArcGIS (Version 9·3). Resampling is a common GIS technique which is used to convert a raster map from one resolution to another. Most of the predictor maps were originally at a much coarser resolution, which may cause large differences between neighbouring cells at the borders of the coarse grid cells. To minimize these border effects and take into account that the influence of some predictor maps on CMD can carry over longer distances than the resolution, these were smoothed with a local sample mean within circles of 2, 5, 10 and 20 km radius. On half of the predictor maps smoothing with a radius of 2 km had limited effect, as their original resolution was 5 arc min (approximately 10 km). Nevertheless, the smoothed versions were adopted to maintain uniformity. Most predictor maps were measured on a continuous scale, with the exception of nutrient availability (FERT) that consisted of four classes indicating constraints for crop suitability (Fischer *et al.*, 2009). FERT was represented in the regression analysis by binary dummy variables for each of the classes. All predictor maps were initially in the geographic WGS84 coordinate system and were transformed to a UTM projection zone 36S to ensure equal areas of all cells. Because of the proximity to the equator, the distortion caused by this transformation was limited.

Data analysis

The CMD observations were interpolated with regression kriging, which is a two-step procedure, using the R software (Version 2·7·2; Pebesma, 2004). First, the relationship between CMD and the predictors was quantified using multiple linear regression. The regression explained part of the variation of CMD, which in the geostatistical literature is often referred to as 'drift' (Hengl *et al.*, 2004). Secondly, the residuals from the regression analysis were interpolated with simple kriging. Finally the drift and the interpolated residuals were summed, to calculate a prediction map (Odeh *et al.*, 1994). The interpolated maps for each of the surveys and associated prediction error variance maps were compared visually and quantitatively.

A multiple linear regression analysis was applied to assess the relationship between the response variable CMD and all predictor maps using the SP module of the R software (Dalgaard, 2002). Pearson's correlation coefficients were calculated between CMD and 35 predictor maps that were all treated as deterministic quantities. These included the original predictor maps (Table 2), their smoothed descendants, the 'dummy' variables derived from the nutrient availability map and latitude and longitude. Initially, an overlay was created of the survey data and the 35 maps. This resulted in a dataset with the CMD scores and the

Figure 2 Predictor maps used in regression kriging. CULT, cultivated land as percentage of total land area per cell; DEM, elevation (m.a.s.l.); TEMP, average annual temperature in period 1950–2005 (°C); FERT, nutrient availability (scale of 1–4); LGP, length of growing period (day/year); POP, human population density (number km^{-2}); PREC, average annual precipitation in period 1950–2005 (mm); FOR, forested land as percentage of total land area per cell; URB, urbanized land as percentage of total land area per cell; WAT, water surface as percentage of total land area per cell.

values of the 35 predictor maps for each observation location. The extreme values were inspected within the GIS system and found to result from logical extremes in the predictor maps (e.g. highly populated areas, densely forested areas, upland areas, etc.). In the next step, scatter plots were made illustrating the correlation between CMD and each of the predictors. At first, all 35 predictors were entered into the regression model. Next, stepwise regression was used to simplify the model and reduce the number of predictors. Predictors that insufficiently increased the Akaike Information Criterion were removed (Crawley, 2007). Further reduction of predictors was necessary, because from a multi-colinearity perspective it was not desirable to retain both the original predictor and its smoothed descendants in the same model. No automated process was available that disqualified predictors that originated from the same map. Consequently, the individual correlation and significance of each predictor was taken into account in a manual selection process. Only the most correlated (Table 3) and significant predictors were included in the final regression model (Table 4). Finally, the combined fit of the regression models could be increased by including interactions between predictors. Because no automated method was available to select the interactions between the many combinations, pairs were selected on subjective grounds (i.e. using expert judgment). A trial and error approach was used to determine if the combined fit could be improved by adding different pairs of predictors.

Table 2 Characteristics of the predictor maps included in the regression analysis; note that not all were used in the final models

Code	Description	Original resolution	Smoothing	Source
CULT	Cultivated land as percentage of total land area per cell	5 arc min	Circles with radius 2, 5, 10 or 20 km	Fischer *et al.* (2009)
DEM	Elevation (m.a.s.l.)	3 arc s	NA	CGIAR, (2008)
TEMP	Average annual temperature in period 1950–2005 (°C)	30 arc s	NA	Hijmans *et al.* (2005)
FERT	Nutrient availability[a]	5 arc min	NA	Fischer *et al.* (2009)
LGP	Length of growing period[b] (day/year)	3 arc min	NA	ERGO (2005)
POP	Human population density (number km^{-2})	2·5 arc min	Circles with radius 2, 5, 10 or 20 km	SEDAC (2009)
PREC	Average annual precipitation in period 1950–2005 (mm)	30 arc s	NA	Hijmans *et al.* (2005)
FOR	Forested land as percentage of total land area per cell	5 arc min	Circles with radius 2, 5, 10 or 20 km	Fischer *et al.* (2009)
URB	Urbanized land as percentage of total land area per cell	5 arc min	Circles with radius 2, 5, 10 or 20 km	Fischer *et al.* (2009)
WAT	Water surface as percentage of total land area per cell	5 arc min	Circles with radius 2, 5, 10 or 20 km	Fischer *et al.* (2009)

[a]Nutrient availability classes based on soil texture, structure, pH and total exchangeable bases, with class value 1 = no constraint, 2 = moderate, 3 = severe, 4 = very severe.
[b]Length of growing period is based on the ratio evapotranspiration/precipitation.

Regression kriging was used to compare patterns of CMD. This does not use the observations directly but rather the residuals from the regression analysis. The residuals are defined as the real observed CMD parameter minus the value predicted by regression. Kriging predicts at unobserved locations by taking a weighted average of the observations, whereby the weights are derived from the degree of spatial correlation. The spatial correlation is characterized by a semivariogram, which shows the magnitude of spatial variation as a function of distance (Goovaerts, 1997). A semivariogram is characterized by the nugget, sill, range, number of lags and lag size. The nugget represents the variation at a spatial infinitely small scale and can be caused by a true variation in the measurement and/or by measurement error. The sill represents the maximum semivariance that can be obtained from the measurements. The range represents the distance beyond which there is no spatial autocorrelation. The number of lags and the lag size refer to the distances between the measurements that are taken into account when fitting the semivariogram model. For each survey, a nested structure of two spherical models was applied to fit the semivariogram model. The parameters of the semivariogram models were obtained with weighted least squares fitting of nested semivariogram structures (Bivand *et al.*, 2008). A nested model was used to describe the autocorrelation of the observations to account for the correlation between observations at greater distances. All observations were included in the neighbourhood definition. The semivariogram fitting and kriging was done with the GSTAT module of the R software (Pebesma, 2004). Negative values were predicted at some locations in the C3P-D and C3P-S prediction maps. As these values were not realistic they were set to 0. Likewise, some predicted values were greater than the theoretical maximum of the measured values (CMD scores above 100% in C3P-S and above 1 in GLCI). These overestimates were set to the maximum values of the respective surveys.

The accuracy of the spatial predictions was determined by calculating kriging variance maps. Kriging predicts the response variable CMD for every cell. It also calculates the kriging variance as an indicator of the prediction error. A prediction with a large kriging variance has poor accuracy. The kriging variance is influenced by the sampling density, the sampling pattern and by the semivariogram (Goovaerts, 1997). The overall quality of the prediction map was expressed by relating the mean kriging variance to the mean prediction. Kriging standard deviation maps were derived by taking the square root from the kriging variance maps. This process facilitated comparison between the maps because the predictions and standard deviations have the same units.

Comparison between surveys

The CMD prediction maps yielded CMD scores in different units. To make the maps comparable the CMD scores were standardized. This was justified from a qualitative point of view, as in this study the main interest was the relative CMD scores and not so much their absolute values. Standardized values were calculated by subtracting mean values of all predictions from the predicted values and dividing the outcome by the standard deviation of the predicted values.

The standardized predictor maps allowed two comparisons to be made between the prediction maps. First, the standardized maps were pair-wise subtracted from one another to reveal the areas with differences in predicted values. Secondly, by invoking the normality assumption, it is

Table 3 Pearson correlation between the cassava mosaic disease parameters in the C3P-D, C3P-S and GLCI datasets and the predictors

Predictor[a]	C3P-D	C3P-S	GLCI
Longitude	−0·15***	−0·02	−0·08***
Latitude	−0·10**	−0·47***	−0·11***
CULT	−0·04	−0·06	−0·02
CULT02	−0·04	−0·05	−0·02
CULT05	−0·04	−0·05	−0·02
CULT10	−0·03	−0·07*	−0·03*
CULT20	−0·04	−0·04	−0·05**
DEM	−0·36***	−0·08**	−0·08***
FERT1	0·12***	−0·18***	0·08***
FERT2	−0·02	0·04	−0·05**
FERT3	−0·12**	0·10**	−0·01
FERT4	0·06	0·06	0·01
FOR	−0·02	−0·07*	0·07***
FOR02	−0·02	−0·08*	0·06***
FOR05	−0·01	−0·08*	0·07***
FOR10	0·01	−0·07*	0·07***
FOR20	0·05	−0·08*	0·07***
LGP	0·00	−0·17***	0·09***
POP	0·02	−0·15***	−0·10***
POP02	0·01	−0·14***	−0·10***
POP05	0·02	−0·13***	−0·11***
POP10	0·02	−0·09*	−0·11***
POP20	0·01	−0·03	−0·12***
PREC	−0·05	0·01	0·07***
TEMP	0·36***	0·10**	0·05**
URB	−0·02	−0·08*	−0·07***
URB02	−0·02	−0·08**	−0·09***
URB05	−0·02	−0·08**	−0·10***
URB10	0·00	−0·04	−0·10***
URB20	−0·03	−0·01	−0·13***
WAT	0·14***	−0·01	0·05**
WAT02	0·15***	−0·02	0·06**
WAT05	0·15***	0·00	0·07***
WAT10	0·14***	−0·02	0·08***
WAT20	0·09**	0·00	0·09***

[a]CULT, cultivated land as percentage of total land area per cell; DEM, elevation in m.a.s.l.; TEMP, average annual temperature in period 1950–2005; FERT, nutrient availability; LGP, length of growing period; POP, human population density; PREC, average annual precipitation in period 1950–2005; FOR, forested land as percentage of total land area per cell; URB, urbanized land as percentage of total land area per cell; WAT, water surface as percentage of total land area per cell.
The numbers added to the abbreviations: 2, 5, 10 or 20 indicate the smoothing radius in km that was applied to the original predictor map.
FERT class values: 1 = no constraint, 2 = moderate, 3 = severe, 4 = very severe.
Significance levels: *P ≤ 0·1; **P ≤ 0·01; ***P ≤ 0·001.

Table 4 Predictors used in the multiple regression analysis and corresponding regression coefficients

Predictor[a]	C3P-D	C3P-S	GLCI
Intercept	3·34	2363***	6·99***
Latitude	–	−0·000215***	−7·37E−07***
CULT05	−3·19***	–	–
CULT10	–	−0·466***	–
DEM	−0·0932***	–	–
FOR02	–	−1·80***	–
FOR10	−4·08***	–	–
LGP	–	–	0·00326***
PREC	–	−0·180***	–
TEMP	1·44***	−0·980***	0·00150***
URB10	–	–	0·00988*
WAT05	–	–	0·00293***
WAT10	1·12**	–	–
PREC:CULT05	0·00257***	–	–
DEM:FOR10	0·00230***	–	–
DEM:FOR02	–	−0·000782***	–
LGP:POP05	–	−7·78E−05***	–
PREC:TEMP	–	0·000948***	–
URB02:WAT02	–	−0·0791***	–
DEM:FOR20	–	–	−1·57E−06***
POP20:PREC	–	–	−2·73E−07***
Multiple R^2	0·193***	0·290***	0·049***

[a]CULT, cultivated land as percentage of total land area per cell; DEM, elevation in m.a.s.l.; TEMP, average annual temperature in period 1950–2005; LGP, length of growing period; POP, human population density; PREC, average annual precipitation in period 1950–2005; FOR, forested land as percentage of total land area per cell; URB, urbanized land as percentage of total land area per cell; WAT, water surface as percentage of total land area per cell.
The numbers added to the abbreviations: 02, 05, 10 or 20 indicate the smoothing radius in km that was applied to the original predictor map.
Significance levels: *P ≤ 0·1; **P ≤ 0·01; ***P ≤ 0·001. The sign ':' indicates the interaction between two terms.

possible to delineate those parts of the study area where the absolute difference between two standardized predictions is significantly greater than the interpolation error standard deviation. These are areas where the magnitude of the differences is unlikely to be explained by interpolation error (i.e. $P < 5\%$) and therefore highlight the areas of 'real difference'. If interpolation errors were the only source of differences and the standardized maps represented the same phenomenon (Z_S), then the expected squared difference of two standardized maps (A and B) satisfies:

$$
\begin{aligned}
&\mathrm{E}[(\hat{Z}_{S,A}(x) - \hat{Z}_{S,B}(x))^2] \\
&= \mathrm{E}[(\hat{Z}_{S,A}(x) - Z_S(x) - (\hat{Z}_{S,B}(x) - Z_S(x)))^2] \\
&= \mathrm{Var}(\hat{Z}_{S,A}(x) - Z_S(x)) + \mathrm{Var}(\hat{Z}_{S,B}(x) - Z_S(x)) \\
&\quad - 2\,\mathrm{Cov}(\hat{Z}_{S,A}(x) - Z_S(x), \hat{Z}_{S,B}(x) - Z_S(x)) \\
&= \mathrm{Var}(\hat{Z}_{S,A}(x) - Z_S(x)) + \mathrm{Var}(\hat{Z}_{S,B}(x) - Z_S(x)) \\
&= \sigma^2_{K,A}(x) + \sigma^2_{K,B}(x)
\end{aligned}
$$

(1)

where $\sigma^2_K(x)$ is the kriging variance at location x for the standardized maps. The covariance term was zero because interpolation errors for the separate surveys were independent.

Results

Figure 3 shows the distribution of the CMD parameters that were assessed in the three surveys. In the C3P-D survey
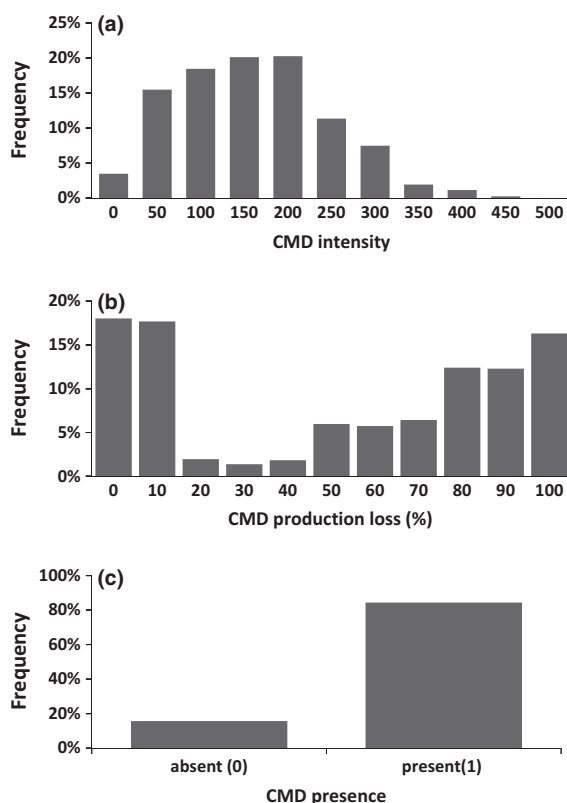
Figure 3 Histograms of CMD intensity in the C3P-D survey (a), perceived production loss as a result of CMD in the C3P-S survey (b) and presence of CMD symptoms identified by farmers in their fields in the GLCI survey (c).

CMD intensity was widespread, with about 97% of all fields being affected by the disease (Fig. 3a). In the C3P-S survey a bimodal CMD distribution was present, where about one-third of the farmers stated that they had experienced <20% production loss, whereas about half of the farmers perceived a loss >50% (Fig. 3b). In the GLCI survey about 85% of farms were affected by the disease (Fig. 3c).

Correlations between CMD and most predictors were predominantly weak (Table 3). In the C3P-S survey, CMD was negatively correlated to latitude, indicating that CMD increased from north to south. In the other two surveys the correlation with latitude was weaker but remained negative. A negative correlation was also identified with longitude, indicating that CMD increased from east to west. In the C3P-D survey CMD was negatively correlated with altitude. Consequently, as expected, CMD was positively correlated with temperature for all surveys and most strongly for C3P-D.

A selection of predictors was used in the regression models (Table 4). In many cases the models' combined fit could be improved by adding interactions between predictors. Only two-way interactions were taken into account. In the C3P-D regression model, for example, precipitation (PREC) was important only in combination with cultivation level (CULT05).

The coefficients of variation demonstrated that the best fit was achieved for the C3P-S survey, where 29% of the variation of CMD could be explained by the predictors. The other R^2 values were 19% for the C3P-D and only 5% for the GLCI survey. The R^2 also indicated the magnitude of the residuals that are passed on to the kriging process (i.e. if R^2 is smaller, kriging might be more important in the final prediction). Thus, the regression had more influence on the final predicted CMD in the case of C3P-S than in the case of GLCI.

Kriging

The semivariograms (Fig. 4; Table 5) demonstrated spatial dependency between the residuals of the three datasets. All semivariograms had a large nugget, indicating large variation over short distances. In the C3P-D and C3P-S surveys the nugget to sill ratio (56% and 59%) indicated moderate spatial autocorrelation. In the GLCI survey the nugget to sill ratio of almost 80% indicated a weak spatial autocorrelation.

Results from the regression kriging were depicted as prediction maps of the response variable CMD (Fig. 5). In the C3P-D map the disease severity appeared to be more concentrated in border areas, while the interior of Burundi had relatively small values. This interior coincides with higher elevation and lower temperatures in the mountains that stretch from north to south. Altitude and temperature had less predictive power in the C3P-S and GLCI predictions (Table 3; Fig. 5). In Burundi the C3P-S and GLCI maps appear to be fairly similar as these show larger concentrations of the disease in the south. The spatial pattern of CMD appears to be most homogeneous in the C3P-S map, where the whole of Burundi, apart from a few areas along the shore of Lake Tanganyika, had large CMD values. This confirmed the correlation between the disease and latitude in the C3P-S survey that was calculated in the regression analysis (Table 3). The GLCI map is patchier with a block structure inherited from the predictor maps with a coarse resolution and relatively high correlation coefficients (e.g. URB10 and POP20). The prediction results of the GLCI map may be of limited value as the minimum kriging standard deviation (0·35 in Fig. 7) was large compared to the spread in the predicted values (92% of the predictions are in the range 0·7–1·0; Fig. 6c). In all maps CMD pressure appeared to be less intense in northern Rwanda, which incidentally, was sparsely sampled (Fig. 5).

The predicted values (Fig. 6) deviated from the measurements in the original surveys (Fig. 3). The predicted values in the C3P-D (Fig. 6a) shifted to the left where the mean CMD value was 120 compared to 137 in the original observations. The C3P-S predictions (Fig. 6b) still showed the same bimodal distribution as the original survey although, as expected, a substantial smoothing occurred and the lower peak was clearly displaced to the right. Whereas originally about a third of the farmers experienced losses <20%, only 4% of the predicted values were so low. While the original GLCI showed binary values (presence or absence) the GLCI predictions (Fig. 6c) consisted of values in between 0 and 1, which may be interpreted as the
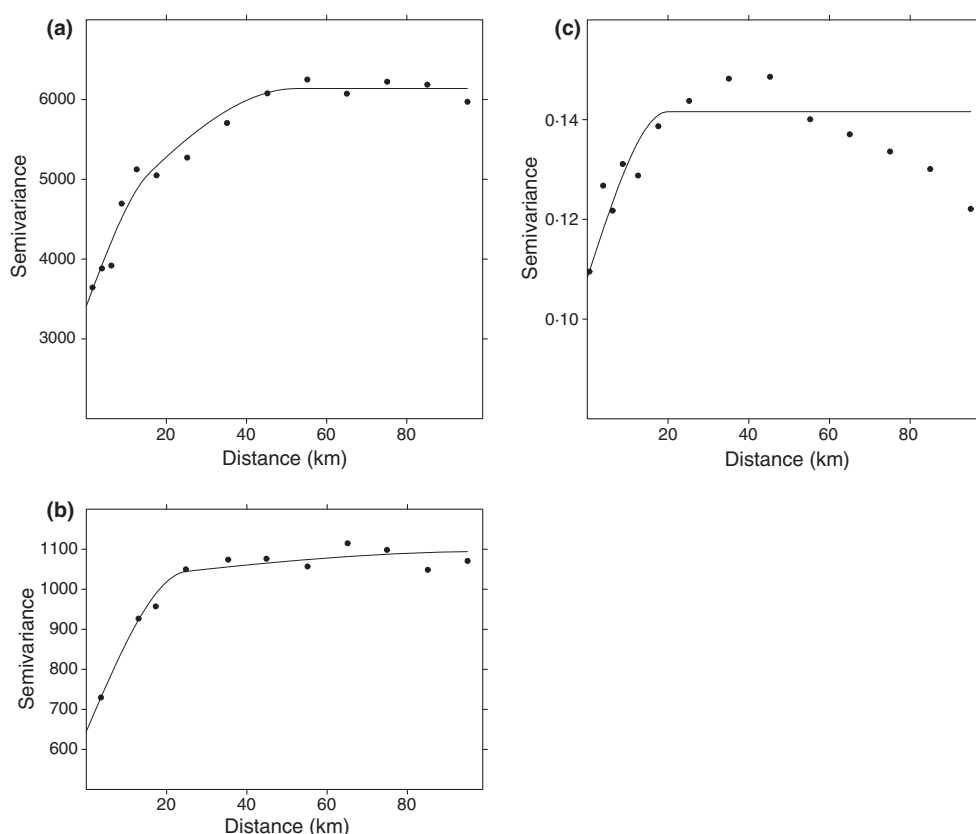
Figure 4 Semivariograms of the regression residuals for C3P-D (a), C3P-S (b) and GLCI (c).

Table 5 Settings of the semivariogram lags and fitted semivariogram model parameters

Setting	C3P-D	C3P-S	GLCI
Number of lags	14	11	12
Lag size	variable[a]	variable[b]	variable[c]
Nugget	3409	645	0·113
1st Sill	4259	1016	0·139
1st Range	15·9	25·0	20·0
2nd Sill	6138	1094	0·140
2nd Range	52·6	100·0	70·0

[a]Lag boundaries at 2·5, 5, 7·5, 10, 15, 20, 30, 40, 50, 60, 70, 80, 90, 100 km.
[b]Lag boundaries at 10, 15, 20, 30, 40, 50, 60, 70, 80, 90, 100 km.
[c]Lag boundaries at 5, 10, 15, 20, 30, 40, 50, 60, 70, 80, 90, 100 km.

probability of CMD presence. Because CMD was prevalent in the original data (Fig. 3c), the predicted CMD values for GLCI are predominantly large, indicating that large parts of the study area were affected by the disease.

Quality of prediction

The kriging standard deviation maps (Fig. 7) demonstrated that the standard deviation was large when compared to the range of the predicted values (Fig. 5). This is mainly caused by the large nugget effect that resulted from relative large differences in CMD observations at short distances. The standard deviation increased with distance to the observation sites and maximum values occurred in areas that were sparsely surveyed. This is particularly apparent for the C3P-S study, where observations were clustered and large parts of the study area were not sampled (Fig. 1). In the GLCI assessment the range in standard deviations is small, indicating that distance to observation sites had little effect.

Survey comparisons

The standardized difference maps revealed large differences, shown by increasing colour intensity (Fig. 8). C3P-D predicted higher values than the other two surveys in most of the border areas, particularly along Lake Tanganyika in south-western Burundi. In contrast, CMD intensity in the interior of Burundi was remarkably low according to C3P-D when compared to those predicted by C3P-S and GLCI. Conspicuous are the areas of large differences in south-west Rwanda that were caused by low CMD values in the GLCI predictions. The blocky structure in the maps whenever GLCI predictions were included results from the coarse predictor maps.

The area of 'real difference' is represented by category A in Figure 9. These areas often coincided with the areas
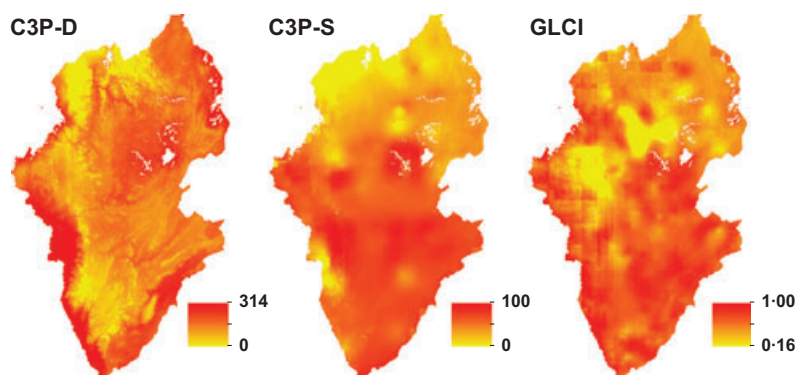
Figure 5 Maps of predicted CMD of the three surveys. C3P-D relates to CMD intensity on a scale of 0–500, C3P-S relates to farmer perception of production loss as a consequence of CMD on a scale of 0 to 100% and GLCI relates to the probability of occurrence of CMD on a scale of 0–1.
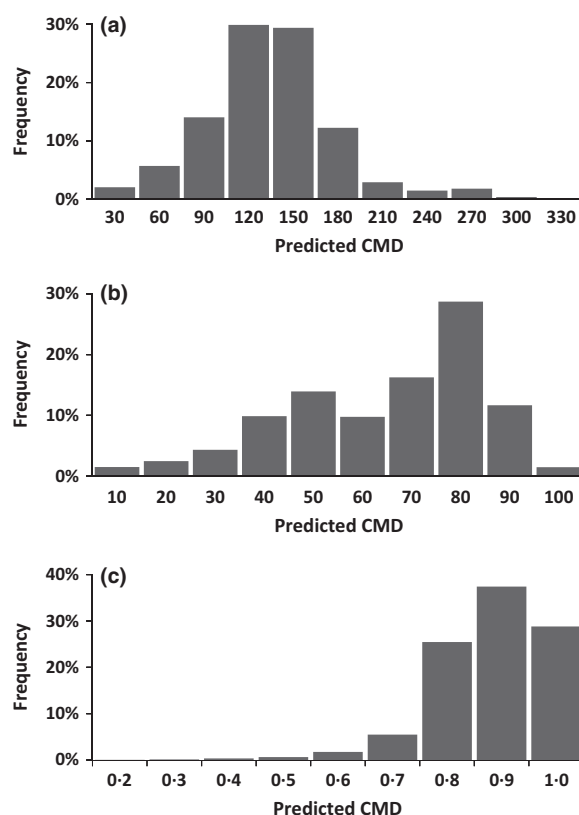

Figure 6 Histograms of CMD predictions of the three surveys: C3P-D (a), C3P-S (b) and GLCI (c).

of absolute high differences in CMD (Fig. 8). However, this is not always so, because the variance of the prediction error maps was also considered. The C3P surveys showed several areas where CMD differed significantly between the two surveys. Fewer significant differences occurred for comparisons with GLCI predictions.

The absolute values of the difference grids show that prediction maps of the C3P-S and GLCI were most similar, as was the case when the absolute difference was scaled by the maximum absolute difference (Table 6). The differences between the surveys were relatively small, meaning that none of the comparisons stood out with an exceptionally small or large difference.

Discussion

Regression analysis showed that part of the variation in CMD could be explained by environmental and socio-economic predictor maps, although the influence was weaker for the GLCI dataset than for the other two. This is not surprising as CMD is spread by viruses transmitted by a white-fly vector that is highly sensitive to the environmental conditions (Gerling *et al.*, 1986; Fargette *et al.*, 1994). Moreover, cassava is grown in agro-ecosystems that are themselves greatly influenced by the environment. For example, large parts of Rwanda and Burundi are at altitudes too high and too cold for the effective cultivation of cassava, and the intensity of cassava cropping decreases with increasing altitude. This is indirectly highlighted by the predictor maps of elevation and temperature that were significantly and negatively correlated with CMD. The correlation with latitude was demonstrated by all maps implying that the impact of CMD was greatest in the southern part of the study region. The progression of the CMD pandemic in this region from north-east to south-west has been well documented, and south-west Burundi was the last to be affected (Legg *et al.*, 2001, 2006; Bigirimana *et al.*, 2004). The correlation with longitude was weaker but also follows this progression pathway. Interpolation with kriging did not greatly improve the prediction of CMD. This was a direct result of the high nugget: high variation in observations at close range and random measurement error. Both C3P surveys showed moderate autocorrelation but in the GLCI survey the nugget dominated. Kriging resulted in a considerable smoothing effect of the predictions (Fig. 6) when compared to the original surveys (Fig. 3). This is because kriging smooths predictions towards the mean (Goovaerts, 1997).

The reasons for the limited accuracy of the prediction maps could be attributed to the low sampling density of the
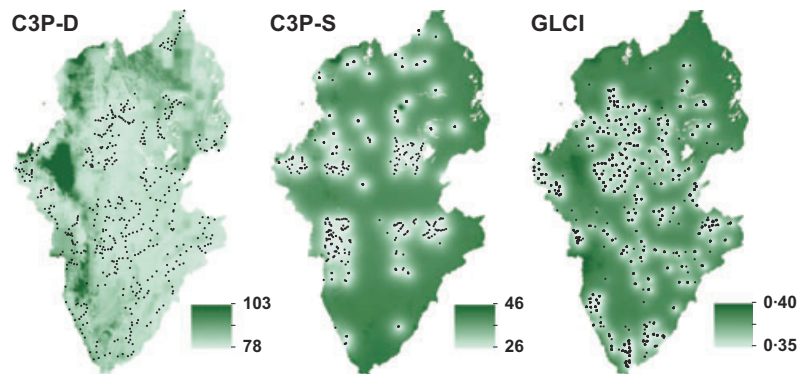
C3P-D   **C3P-S**   **GLCI**

103 — 78

46 — 26

0·40 — 0·35

Figure 7 CMD kriging standard deviation maps. Black dots depict the observation locations.
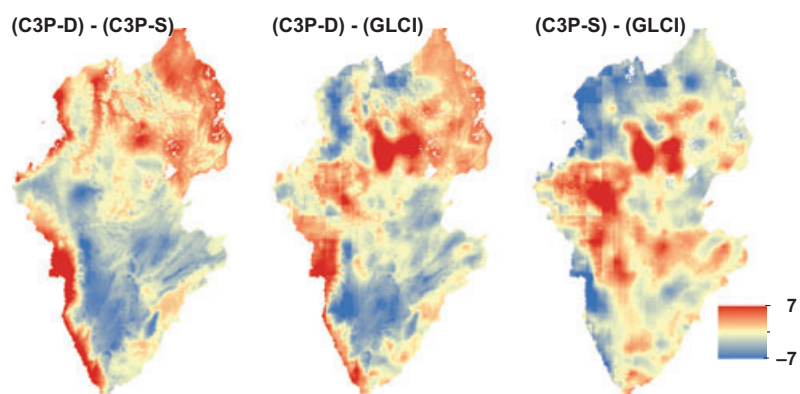
(C3P-D) - (C3P-S)   **(C3P-D) - (GLCI)**   **(C3P-S) - (GLCI)**

7 — −7

Figure 8 Difference of the standardized prediction maps between surveys, calculated by subtracting one map from the other.

(C3P-D) - (C3P-S)   **(C3P-D) - (GLCI)**   **(C3PS) - (GLCI)**

A

B

Figure 9 Category A represents the area of 'real difference' and category B represents the area where the difference might be caused by interpolation error.

surveys and the high variation of measurements at close range. These causes should be identified first so that they can be incorporated in future sampling designs which would result in more accurate maps. In general, increasing the sampling density and/or applying a more uniform distribution of the observation sites could improve the

accuracy of the prediction maps (Stein & Ettema, 2003). However, in this case a more systematic survey design would be hindered by the large size of the study area and the practical difficulties that this raises. In addition, simply increasing the sample size and density would probably not improve accuracy much, as is highlighted by the large

Table 6 Spatial averages of differences in cassava mosaic disease standardized predictions

Formula	Description	C3P-D and C3P-S	C3P-D and GLCI	C3P-S and GLCI
abs (B−A)	absolute difference	1·01	0·98	0·85
$\frac{abs\,(B-A)}{\max\,(abs\,(B-A))}$	scaled absolute difference	0·15	0·14	0·13

minimum kriging standard deviation in the GLCI map (Fig. 7). The high variation of measurements at close range (i.e. large nugget) means that there are important factors that vary on a local scale and/or large measurement errors. High variability in itself is not surprising as the ecology of CMD suggests that there are many factors that cause local differences in the impact of CMD, such as the relative CMD susceptibility or resistance of the cassava cultivars being grown, the virus species present, the abundance of the whitefly vector, the cropping systems and farmer's management practices (Fargette & Thresh, 1994). Unfortunately, predictor maps that potentially describe these factors were not available. Measuring CMD at larger spatial supports (averages over larger areas) may be a more adequate solution to tackling this problem, because such averaging will lead to measurements from which local variability has partly been removed (Goovaerts, 1997). Hence, these measurements would likely be correlated more strongly with the coarse predictor maps (i.e. improved regression results) and would result in a smaller nugget to sill ratio (i.e. improved kriging results). However, a survey must be tailored to this and hence it was not possible to do this with the available surveys. Measurement errors may be reduced by harmonizing protocols and improved training of assessors. The graded scale used in both C3P surveys was probably better suited for predicting CMD than the binary scale used in the GLCI survey.

The prediction maps of the different surveys revealed areas with large differences in CMD (Fig. 8). Despite these large differences, only a small portion of the differences was significant (Fig. 9) because the kriging interpolation standard deviation (Fig. 7) was large in most of the study area. The differences in predicted CMD were simply not large enough to compensate for this spatial interpolation error. Despite these shortcomings, two main areas were identified with differences in CMD that could not be explained by interpolation error and therefore must be explained by other causes. The first area was along the shore of Lake Tanganyika in the south-western part of Burundi. It was characterized by significantly larger CMD values in the C3P-D survey than in the other two surveys. This might be explained when considering the nature of the data types collected in the three surveys. Farmer responses (GLCI and C3P-S) were based on historical recall, in contrast to field assessments (C3P-D) in which CMD was recorded in 'real-time'. Thus, a new severe disease outbreak would be recorded by direct field assessments (C3P-D) before growers begin to recognize its presence (GLCI) or experience its impact on yields (C3P-S). The second area

included locations east and west of Kigali in central Rwanda where the GLCI survey indicated very low values as opposed to high values in both C3P surveys. This might be explained by a systematic multiplication and distribution campaign of a CMD-resistant variety in Central Rwanda in recent years (G. Gashaka, Institut des Sciences Agronomiques du Rwanda, Butare, Rwanda, unpublished data). The effects of this campaign are likely to have been more evident in the GLCI survey, which was the last of the three surveys to be conducted.

Regression kriging has the potential for widespread use in the mapping and analysis of crop disease epidemics. Predictor maps of environmental and socio-economic conditions explained a significant part of the variance of CMD. Kriging successfully standardized different disease observations at a regional level and allowed comparisons to be made. Areas of divergence were identified and could be explained. However, most areas were broadly comparable and not significantly different because of the limited accuracy of the predictions. Therefore, the need for optimization of sampling designs and survey protocols should be emphasized, to decrease prediction error and maximize the potential for the application of regression kriging.

The results of this study suggest that it is risky to take action based on the outcome of only one survey. As demonstrated, different responses might be appropriate, depending on the survey approach adopted. By comparing the results of different surveys in a spatially explicit way, this study validates the unique character of each of the survey approaches. It highlights the importance of choosing a survey approach that is appropriate for the specific research question to be addressed.

References

Abele S, Twine E, Legg C, 2007. *Food Security in Eastern Africa and the Great Lakes Region*. Dar es Salaam, Tanzania. International Institute of Tropical Agriculture. Final report submitted to USAID on October 15th 2007.

Bigirimana S, Barumbanze P, Obonyo R, Legg JP, 2004. First evidence for the spread of East African cassava mosaic virus-Uganda (EACMV-UG) and the pandemic of severe cassava mosaic disease to Burundi. *Plant Pathology* 53, 231–45.

Bigirimana S, Bizimana S, Barumbanze P, Busungu C, Legg JP, 2007. *Monitoring and Diagnostic Survey of Cassava Mosaic*

Virus Disease (CMD) in Burundi. Dar es Salaam, Tanzania: International Institute of Tropical Agriculture. Internal Report.

Binns MR, Nyrop JP, Van Der Werf W, 2006. *Sampling and Monitoring in Crop Protection.* Wallingford, UK: CABI Publishing.

Bivand RS, Pebesma EJ, Gomez-Rubio V, 2008. *Applied Spatial Data Analysis with R.* New York, USA: Springer.

Bock KR, Woods RD, 1983. Etiology of African cassava mosaic disease. *Plant Disease* **67**, 994–5.

Bouwmeester H, Abele S, Manyong VM *et al.*, 2008. The potential benefits of GIS techniques in disease and pest control: an example based on a regional project in Central Africa. *Acta Horticulturae* **879**, 333–40.

Bouwmeester H, Van Asten PJA, Ouma E, 2009. *Mapping Key Variables of Banana Based Cropping Systems in the Great Lakes Region, Partial Outcomes of the Baseline and Diagnostic Surveys.* Kampala, Uganda: Consortium for Improving Agriculture-based Livelihoods in Central Africa. Technical Report no. 11.

Cecchi G, Wint W, Shaw A, Marletta A, Mattioli R, Robinson T, 2010. Geographic distribution and environmental characterization of livestock production systems in Eastern Africa. *Agriculture Ecosystems & Environment* **135**, 98–110.

CGIAR-Consortium for Spatial Infrastructure, 2008. *Shuttle Radar Topographic Mission (SRTM) Digital Elevation Data.* [http://srtm.csi.cgiar.org]. Accessed 15 April 2009.

Chellemi DO, Rohrbach KG, Yost RS, Sonoda RM, 1988. Analysis of the spatial pattern of plant pathogens and diseased plants using geostatistics. *Phytopathology* **78**, 221–6.

Clements ACA, Lwambo NJS, Blair L *et al.*, 2006. Bayesian spatial analysis and disease mapping: tools to enhance planning and implementation of a schistosomiasis control programme in Tanzania. *Tropical Medicine & International Health* **11**, 490–503.

Crawley MJ, 2007. *Statistics – An Introduction using R.* Chichester, UK: John Wiley & Sons Ltd.

Dalgaard P, 2002. *Introductory Statistics with R.* New York, USA: Springer.

De Gruijter JJ, Brus DJ, Bierkens MFP, Knotters M, 2007. *Sampling for Natural Resource Monitoring.* Berlin, Germany: Springer.

De Smith MJ, Goodchild MF, Longley PA, 2007. *Geospatial Analysis – A Comprehensive Guide to Principles, Techniques and Software Tools.* Leicester, UK: Matador.

Environmental Research Group Oxford, 2005. *Agro-Ecological Distributions for Africa – Length of Growing Period.* [http://www.ergoonline.co.uk]. Accessed 7 August 2006.

FAO, 2009. *Production Statistics.* [http://faostat.fao.org]. Accessed 10 October 2009.

Fargette D, Thresh JM, 1994. The ecology of African cassava mosaic geminivirus. In: Blakeman JP, Williamson B, eds. *Ecology of Plant Pathogens.* Wallingford, UK: CAB International, 269–82.

Fargette D, Jeger M, Fauquet C, Fishpool LDC, 1994. Analysis of temporal disease progress of African cassava mosaic virus. *Phytopathology* **84**, 91–8.

Fermont AM, Van Asten PJA, Tittonell P, Van Wijk MT, Giller KE, 2009. Closing the cassava yield gap: an analysis from smallholder farms in East Africa. *Field Crops Research* **112**, 24–36.

Fischer G, Nachtergaele F, Prieler S, Van Velthuizen HT, Verelst L, Wiberg D, 2009. *Harmonized World Soil Database* v1.1. [http://www.iiasa.ac.at/Research/LUC/External-World-soil-database]. Accessed 12 July 2009.

Gandah M, Stein A, Brouwer J, Bouma J, 2000. Dynamics of spatial variability of millet growth and yields at three sites in Niger, West Africa and implications for precision agriculture research. *Agricultural Systems* **63**, 123–40.

Gerling D, Horowitz AR, Baumgaertner J, 1986. Autecology of *Bemisia tabaci. Agriculture Ecosystems & Environment* **17**, 5–19.

Goovaerts P, 1997. *Geostatistics for Natural Resources Evaluation.* New York, USA: Oxford University Press.

Hengl T, Heuvelink GBM, Stein A, 2004. A generic framework for spatial prediction of soil variables based on regression-kriging. *Geoderma* **120**, 75–93.

Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A, 2005. Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology* **25**, 1965–78.

Kanyanga MC, Night G, Nyirahorana C *et al.*, 2007. *Monitoring and Diagnostic Survey of Cassava Mosaic Virus Disease (CMD) in Rwanda.* Dar es Salaam, Tanzania: International Institute of Tropical Agriculture. Final report submitted to USAID in March 2007.

Kempen B, Brus DJ, Heuvelink GBM, Stoorvogel JJ, 2009. Updating the 1:50 000, Dutch soil map using legacy soil data: a multinomial logistic regression approach. *Geoderma* **151**, 311–26.

Kimetrica, 2008. *Baseline Study –Great Lake Cassava Initiative & FAO. Regional Cassava Programme.* Washington, USA: Kimetra International Ltd. Report submitted to CRS and FAO on September 3rd 2008.

Lecoustre R, Fargette D, Fauquet C, Dereffye P, 1989. Analysis and mapping of the spatial spread of African Cassava mosaic-virus using geostatistics and the kriging technique. *Phytopathology* **79**, 913–20.

Legg JP, 1999. Emergence, spread and strategies for controlling the pandemic of cassava mosaic virus disease in east and central Africa. *Crop Protection* **18**, 627–37.

Legg JP, Thresh JM, 2000. Cassava mosaic virus disease in East Africa: a dynamic disease in a changing environment. *Virus Research* **71**, 135–49.

Legg JP, Okao-Okuja G, Mayala R, Muhinyuza JB, 2001. Spread into Rwanda of the severe cassava mosaic virus disease pandemic and the associated Uganda variant of East African cassava mosaic virus (EACMV-Ug). *Plant Pathology* **50**, 796–804.

Legg JP, Whyte J, Kapinga R, Teri J, 2005. Special topics on pest and disease management: management of the cassava mosaic disease pandemic in East Africa. In: Anderson PK, Morales F, eds. *Whiteflies and Whitefly-borne Viruses in the Tropics: Building a Knowledge Base for Global Action.* Cali, Colombia: Centro Internacional de Agricultura Tropical, 332–8.

Legg JP, Owor B, Sseruwagi P, Ndunguru J, 2006. Cassava mosaic virus disease in East and Central Africa: epidemiology and management of a regional pandemic. *Advances in Virus Research* **67**, 355–418.

McBratney AB, Odeh IOA, Bishop TFA, Dunbar MS, Shatar TM, 2000. An overview of pedometric techniques for use in soil survey. *Geoderma* **97**, 293–327.

Odeh IOA, McBratney AB, Chittleborough DJ, 1994. Spatial prediction of soil properties from landform attributes derived from digital elevation model. *Geoderma* **63**, 197–214.

Otim-Nape GW, Bua A, Thresh JM *et al.*, 2000. *The Current Pandemic of Cassava Mosaic Virus Disease in East Africa and its Control.* Chatham, UK: Natural Resources Institute.

Pebesma EJ, 2004. Multivariable geostatistics in S: the gstat package. *Computers & Geosciences* **30**, 683–91.

Socioeconomic Data and Application Center, 2009. *Gridded Population of the World 2005 (GPWv3)*. [http://sedac.ciesin.columbia.edu/gpw]. Accessed 12 October 2009.

Sseruwagi P, Sserubombwe WS, Legg JP, Ndunguru J, Thresh JM, 2004. Methods of surveying the incidence and severity of cassava mosaic disease and whitefly vector populations on cassava in Africa: a review. *Virus Research* **100**, 129–42.

Stein A, Ettema C, 2003. An overview of spatial sampling procedures and experimental design of spatial studies for ecosystem comparisons. *Agriculture Ecosystems & Environment* **94**, 31–47.

Stonard JF, Marchant BP, Latunde-Dada AO *et al.*, 2010. Geostatistical analysis of the distribution of *Leptosphaeria* species causing phoma stem canker on winter oilseed rape (*Brassica napus*) in England. *Plant Pathology* **59**, 200–10.

Storey HH, Nichols RFW, 1938. Studies on the mosaic of cassava. *Annals of Applied Biology* **25**, 790–806.

Thresh JM, Fargette D, Otirn-Nape GW, 1994. Effects of African cassava mosaic geminivirus on the yield of cassava. *Tropical Science* **34**, 26–42.

Tusheremereirwe WK, Okaasai O, Kubiriba J *et al.*, 2006. Status of banana bacterial wilt in Uganda. *African Crop Science Journal* **14**, 73–82.