

# Spatial Analysis of Temporal Changes in the Pandemic of Severe Cassava Mosaic Disease in Northwestern Tanzania

A. M. Szyniszewska, C. Busungu, S. B. Boni, R. Shirima, H. Bouwmeester, and J. P. Legg†

First author: Rothamsted Research, Harpenden, Hertfordshire, AL5 2JQ, United Kingdom; second author: United Graduate School of Agricultural Sciences, Kagoshima University, 1-21-24 Korimoto, Kagoshima 890-0065, Japan; third author: PO Box 21026, Dar es Salaam, Tanzania; fourth and sixth authors: International Institute of Tropical Agriculture, PO Box 34441, Dar es Salaam, Tanzania; and fifth author: Geospace, Roseboomlaan 38, 6717 ZB Ede, The Netherlands.  
Accepted for publication 14 July 2017.

## ABSTRACT

To improve understanding of the dynamics of the cassava mosaic disease (CMD) pandemic front, geospatial approaches were applied to the analysis of 3 years' data obtained from a 2-by-2° (approximately 222-by-222 km) area of northwestern Tanzania. In total, 80 farmers' fields were assessed in each of 2009, 2010, and 2011, with 20 evenly distributed fields per 1-by-1° quadrant. CMD-associated variables (CMD incidence, CMD severity, vector-borne CMD infection, and vector abundance) increased in magnitude from 2009 to 2010 but showed little change from 2010 to 2011. Increases occurred primarily in the two westernmost quadrants of the study area. A

pandemic "front" was defined by determining the values of CMD incidence and whitefly abundance where predicted disease gradients were greatest. The pandemic-associated virus (*East African cassava mosaic virus-Uganda*) and vector genotype (*Bemisia tabaci* sub-Saharan Africa 1-subgroup 1) were both present within the area bounded by the CMD incidence front but both also occurred ahead of the front. The average speed and direction of movement of the CMD incidence front (22.9 km/year; southeast) and whitefly abundance front (46.6 km/year; southeast) were calculated, and production losses due to CMD were estimated to range from US\$4.3 million to 12.2 million.

Cassava mosaic disease (CMD) is one of the most important constraints to cassava production in sub-Saharan Africa, and causes more than US\$1 billion of losses annually (Legg et al. 2006; Thresh et al. 1997). The disease is caused by several species of cassava mosaic begomoviruses (CMB) (family *Geminiviridae*, genus *Begomovirus*) (Bock and Woods 1983), which are propagated through planting infected stem cuttings and transmitted persistently by the whitefly *Bemisia tabaci* (Genn.) (Dubern 1994). In the late 1980s and early 1990s, an epidemic of unusually severe CMD emerged in Uganda (Otim-Nape et al. 1997) and subsequently spread to affect a large area of East and Central Africa (Legg 1999; Legg et al. 2006; Otim-Nape et al. 1997). A novel virus recombinant, *East African cassava mosaic virus-Uganda* (EACMV-UG) was shown to be associated with this "pandemic" (Zhou et al. 1997) and gave rise to unusually severe disease symptoms through a synergistic interaction with a second CMB, *African cassava mosaic virus* (ACMV). Severe CMD spread rapidly at estimated speeds of 20 to 30 km/year (Otim-Nape et al. 1997), driven by superabundant populations of *B. tabaci* (Legg and Ogwal 1998). Almost 30 years after the first reports of severe CMD from Uganda, the severe CMD pandemic continues to spread, currently advancing southward through eastern Democratic Republic of Congo and westward through central Cameroon. The pandemic of severe CMD continues to pose a threat to the world's largest producer of cassava, Nigeria, lying immediately to the west of Cameroon.

Monitoring and surveillance activities have played a vital role in the CMD management effort as newly affected areas have been identified, facilitating the targeting of control interventions. An important aspect of this monitoring work has been the development and publication of maps illustrating patterns of disease spread and attempting to predict areas likely to be affected in the near future.

One of the earliest maps of CMD was the illustration by Swanson and Harrison (1994) of the distribution of viruses causing CMD in Africa. This was a continental-level map showing countries from which EACMV and ACMV had been detected using monoclonal antibodies in an enzyme-linked immunosorbent assay protocol. More recently, country-level maps have been produced to illustrate the distribution of CMB species (Bull et al. 2006; Ndunguru et al. 2005; Ogbe et al. 2003) or CMD incidence in regions or districts within countries (IITA 2009; Otim-Nape et al. 2001). Mapping was used to illustrate the progress of the CMD pandemic during the peak of its spread in the 1990s and early 2000s. Spread was tracked through Uganda (Otim-Nape et al. 1997) and into neighboring countries (Legg 1999; Legg and Thresh 2000) and estimates of the rate of movement of the pandemic "front" were used to provide forecasts for likely future patterns of spread. In northwestern Tanzania, the first "arrival" of the CMD pandemic from neighboring Uganda was reported in 1998 (Legg 1999). By 2005, the pandemic-associated EACMV-UG was recorded from much of the Lake (Victoria) Zone of northwestern Tanzania (Ndunguru et al. 2005) and the "zone of epidemic expansion" was described as covering Kigoma, Mara, and Shinyanga Regions at approximately the same time (Legg et al. 2006). Most recently, maps have been used to describe the systematic pattern of severe CMD spread through southern Uganda, western Kenya, and northwestern Tanzania over a 16-year period from 1992 to 2007 (Legg 2010). None of the mapping presented in these reports, however, has used quantitative spatial analysis techniques, with the result that many of the statements made about rates of disease spread are subjective. Similarly, predictions for likely future patterns of epidemic spread were not based on quantitative models that considered environmental variables and spatial patterns of cassava cultivation. Consequently, much potential remains for the design of mechanistic spatial models which, in addition to describing CMD epidemic development in the geographical environment from which they are derived, also offer the possibility of predicting likely outcomes in different locations, countries, or regions.

†Corresponding author: J. P. Legg; E-mail: j.legg@cgiar.org

\*The e-Xtra logo stands for "electronic extra" and indicates that two supplementary figures and two supplementary tables are published online.

Geostatistics were first applied to analyzing patterns of CMD spread in the early 1990s in Ivory Coast (Lecoustre et al. 1989), although this comprised a field-level study and did not involve mapping. Kriging and map-based spatial analysis were first used for CMD in a study that compared the results of three different types of survey in Rwanda and Burundi conducted between 2006 and 2008 (Bouwmeester et al. 2012). In this approach, 10 environmental variable predictor maps were combined with the CMD data to develop a regression model. This model was subsequently used with simple kriging to derive predicted CMD maps for the entire land surface areas of Burundi and Rwanda. An important result of this study was the identification of zones where the survey approaches gave divergent results. Two key conclusions were that direct data are more reliable than proxy variables (e.g., yield loss as a proxy for CMD occurrence) and that irregular coverage of survey areas reduced the power of kriging-based geostatistics.

In the study described here, we aimed to address some of these deficiencies and to provide novel insights into the region-wide epidemiology of CMD by making the first attempt to monitor the spread of CMD through a predefined geographical zone within a country still being affected by the severe CMD pandemic. We planned to achieve this by carrying out three sequential annual surveys of CMD in a square of 2 by 2° (approximately 222 by 222 km) of northwestern Tanzania. Our aim was to improve structure of the data in comparison with the more usual sampling approaches that have used a structure based on administrative regions, transects limited to major roads, or areas with particular defined disease characteristics (Jeremiah et al. 2015; Sseruwagi et al. 2004b). Two key targets of the study were the application of geostatistical techniques to measure the rate and direction of spread

of the severe CMD pandemic, and an examination of the properties of the pandemic front on a relatively small scale, including host, disease, and vector interactions.

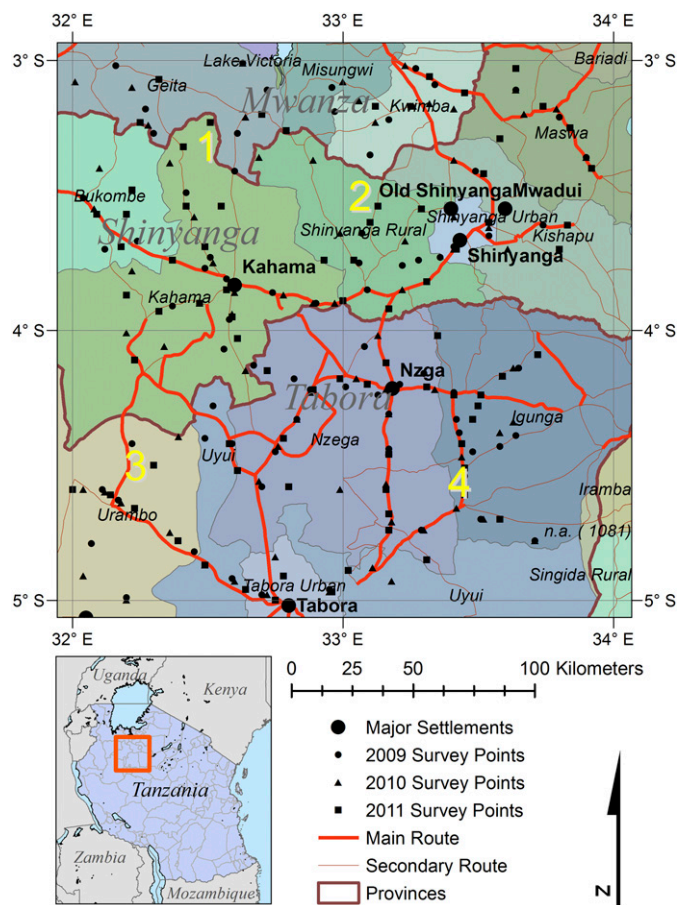
## MATERIALS AND METHODS

**Location.** Surveys were conducted in 2009, 2010, and 2011 in an area of northwestern Tanzania bounded by the 3°S and 5°S lines of latitude and the 32°E and 34°E lines of longitude shared between the regions of Mwanza in the north, Shinyanga in the center, and Tabora in the south (Fig. 1). This midaltitude area of Tanzania lies between 1,034 and 1,506 m above sea level. Climate in this area is classified as tropical, with 861 mm of annual average precipitation. A dry period is observed between June and August (in the range of 0 to 3 mm on average) and December is the wettest month of the year, with an average 151 mm of rain. Average temperature ranges from 21.6°C in June to 25.2°C in October (Hijmans et al. 2005). The vegetation is largely dry open woodland and savanna. Population density is greater in the north compared with the south (Tatem 2017) and the density of cassava is greater in the west than it is in the east (You et al. 2014) (Supplementary Fig. S1).

**Sampling and field data collection.** In each of the three study years, 80 fields were sampled in the study area, 20 in each of the four 1-by-1° quadrants. Fields were sampled at approximately 10-km intervals along motorable roads, resulting in an evenly dispersed sampling pattern. Quadrant designations were quadrant 1, located in the northwest (Q1), quadrant 2, located in the northeast (Q2), quadrant 3, located in the southwest (Q3), and quadrant 4, located in the southeast (Q4). Global Positioning System (GPS) handsets (Garmin eTrex 20 Handheld GPS Navigator; Garmin) were used to ensure that 20 fields were within the intended quadrant as well as to mark and record coordinates of each sampled field. Sampling was done in the middle months of the year when it was easy to find cassava plantings of the most suitable age (3 to 8 months after planting). Sampling periods were 20 July to 7 August 2009, 2 to 20 May 2010, and 24 May to 3 June 2011. In each field, 30 plants were sampled along two diagonals running between the corners of the field. Mean or incidence values for the data recorded were calculated from results obtained for each of the 30 sampled plants. In 2009 and 2010, the owners of the selected farms were identified and asked questions about their crop. Fields and field owners were different for each of the 3 years in which sampling was carried out. In all 3 years, records were taken of the following parameters, using the protocol described by Sseruwagi et al. (2004b).

- CMD incidence (percentage of plants expressing CMD symptoms)
- CMD severity (1-to-5 scale, where 1 represents no symptoms, 2 the mildest symptoms, and 5 the most severe foliar symptoms of CMD infection)
- CMD cutting infection (percentage of plants infected through the cutting; plants infected through the cutting are distinguished by observing CMD symptoms on the lowest first-formed leaves)
- CMD whitefly-borne infection (percentage of plants infected through current season whitefly-borne infection; plants with current season whitefly-borne infection are distinguished by noting the absence of CMD symptoms on the lowest first-formed leaves)
- Whitefly abundance (number of *B. tabaci* adult whiteflies on the top five leaves of the tallest shoot per plant)

Two questions were posed to farmers owning the sampled fields in 2009 and 2010. The first question was, “Do you recognize CMD?” When posing this question, an image of plants expressing CMD symptoms was shown to the farmer. Responses were either yes or no. The second question was: “What yield loss did you experience in the previous year caused by CMD?” Responses were percentage estimates.



**Fig. 1.** Sampled zone of northwestern Tanzania, illustrating major settlements and major and minor routes that were used for the survey activities. Survey points are represented and quadrant numbers highlighted in yellow, demonstrating the even distribution of sampling locations.

**Laboratory sample collection.** In 2010, in each surveyed field, up to 100 *B. tabaci* adults were collected with an aspirator from young leaves of cassava plants following the completion of the field data assessments. Whitefly adults were immediately preserved in 90% ethanol and stored in microfuge tubes prior to laboratory analysis. In 2009 and 2010, in each surveyed field, a single leaflet was collected from the first fully opened leaf of a plant showing CMD symptoms. The leaflet was dried by placing it between sheets of blank newsprint, covered with corrugated cardboard, and held in a herbarium press prior to being stored at room temperature in the laboratory.

**Whitefly genotyping.** Subsamples from each of the field-collected whitefly samples were used for genotyping using the standardized mitochondrial DNA cytochrome oxidase I marker. DNA extraction and polymerase chain reaction (PCR) were performed using the procedure of Frohlich et al. (1999) with primers C1-J-2195 and L2-N-3014 (Simon et al. 1994). PCR amplicons, approximately 800 bp in length, were sent to MacroGen Inc. for sequencing. Sequences were trimmed, edited, and aligned with the ClustalW facility of MEGA6.0 (Tamura et al. 2011), and genetic relationships between sequences were examined using the maximum-likelihood phylogenetic analysis procedure of MEGA 6.0 with default settings and 1,000 bootstraps. Phylogenetic trees, incorporating known standard sequences obtained from GenBank, were used to identify the genotype groups of the sequences obtained from the Tanzanian samples.

**Virus testing.** Total nucleic acids were extracted from leaf samples using a modified cetyltrimethylammonium bromide protocol (Lodhi et al. 1994). Standard PCR was then carried out using specific primers for EACMV-UG. These were UV-ALI/F and ACMV-CP/R3 (Zhou et al. 1997). PCR conditions were as described by Sseruwagi et al. (2004a). PCR products were run on a 1.8% agarose gel and PCR amplicons were visualized under UV light using a DigiDoc-It (UVP/LLC) gel documentation system. Based on these results, samples were identified as positive or negative for the presence of EACMV-UG.

**Hotspot index calculation.** A hotspot index was calculated for each sampled location. This aimed to provide a measure of the overall potency of the CMD epidemic conditions at the field by combining index values for CMD incidence, CMD severity, whitefly abundance, and the relative proportion of whitefly-borne versus cutting CMD infection (equal to the former divided by the latter). Each of the four variables was categorized by the range of values corresponding to four index scores (between 1 and 4) and for each field. The index scores for each variable were summed to produce the hotspot index value for the field, with a minimum value of 4 and a maximum of 16 (Table 1).

**Two-way comparisons and correlation analyses for tabular data.** CMD incidence, logarithmically transformed  $\ln(x + 1)$  whitefly abundance, hotspot index, and farmer's previous year yield loss estimate were analyzed with a two-way analysis of variance (ANOVA) procedure in the SigmaStat component of SigmaPlot 11.0 (Systat Software Inc.). The Holm-Sidak method was used for pairwise multiple comparisons and, in all cases, treatments were Year and Quadrant. CMD mean severity scores from only those fields and plants where disease was detected were analyzed in R with *aov* function and Tukey's method for pairwise multiple comparisons between factors Year and Quadrant. Binary response data (farmer's CMD symptom recognition and presence of EACMV-UG) were analyzed with binomial generalized linear model (GLM) with logit link in R (*glm* function). Pairwise correlations between variables and *P* values were examined using the Spearman's rank correlation coefficient using *rcorr* function in the 'Hmisc' package in R (Harnell 2017). Whitefly genotypes occurring in each of the quadrants were compared using Kruskal-Wallis one-way ANOVA on ranks. For this analysis, two genotype groupings were designated: sub-Saharan Africa 1–subgroup 1 (SSA1-SG1) (one group) and all other genotypes (second group).

**Spatial prediction.** Ordinary kriging was used to predict the likely values for the variables of interest at unsampled locations.

Kriging is a geostatistical interpolation method that uses weighted averages of the neighboring observations for unbiased spatial prediction with minimized errors. The weights are derived from the experimental variogram, with semivariances calculated for all neighbors at a chosen discrete number and length of lags, which differed for each year and variable of interest (Oliver and Webster 2014). A smoothing function is applied to variogram values to represent the changes in the spatial correlation of observations over geographical distance. The shape of the function is characterized by the nugget (small-scale unresolved variation at short distance or due to measurement error), sill (variance of the random process), and range (distance beyond which no spatial autocorrelation is observed). For each year and for each variable, spherical, exponential, mathematical, and power variogram functions were calculated and the chosen best fit function was applied to fit the variogram model using the weighted least squares fitting method. The R package 'gstat' was used to calculate the function parameters and calculate kriging surface and kriging prediction error (Bivand et al. 2008; Pebesma 2004). The kriging function was used to predict the response variable for every cell of the predefined grid in the area of interest. It also returns the kriging variance at each location, which can be interpreted as the prediction error. The kriging variance tends to be higher in unsampled locations and locations with a limited number of neighbors and, therefore, is influenced by the sampling density, the sampling pattern, and the choice of variogram function (Goovaerts 1997). Our sampling design, in which the target area was predefined and the sampling density fixed at 20 evenly dispersed fields per 1-by-1° quadrant, helped to reduce errors in kriging variance. In some instances, values below or above the possible range of values were predicted. Because these values were not realistic, maximum and minimum values were set.

**Defining the front and affected area of the CMD pandemic.** Kriging prediction surfaces were examined to detect the steepest gradient of change between high and low incidence and vector abundance regions in each survey year. Six histograms with the number of pixels per range of values category (0 to 1 with a 0.05 increment for CMD incidence and 0 to 1.2 with a 0.5 increment for whitefly log mean values) were constructed to guide the choice of the steepest gradient of change and identify the disease pandemic front. This selection was based on the assumption that the category with the smallest number of pixels (smallest geographical area) corresponded to the greatest gradient in incidence/whitefly abundance values. Once the CMD incidence or whitefly log mean values of the front had been determined, all areas with values equal to or greater than these were considered to be pandemic-affected and all areas with values less than these were considered to be nonpandemic. To define the predominant trend and angle of the front, we interpolated disease incidence and logarithmically transformed mean whitefly values with a first-order polynomial function using the Spatial Analyst tool in ArcGIS 10.2.1 (ESRI) and the angle of the predominant direction of the trend against the x axis was recorded. Net movement of the front in latitudinal and longitudinal directions was estimated using the net difference in the extent of the pandemic and nonpandemic area extent between 2009 and 2010 and between 2010 and 2011. The distance of change was calculated for every row of the prediction surface grid with the edge size of 1.11 km, and the mean and maximum net changes in the

TABLE 1. Classification table for scoring hotspot index values at each sampled location

Variable <sup>a</sup>	Hotspot index score			
	1	2	3	4
CMD incidence	0	0–0.33	0.33–0.67	0.67–1
CMD severity	0	2–2.5	2.5–3	3–5
Whitefly abundance	0	0–1	1–3	3–100
Whitefly infection	0	0–0.1	0.1–0.3	0.3–1

<sup>a</sup> CMD = cassava mosaic disease.

distance “moved” by the front were recorded. Trigonometric sine and cosine functions were used to calculate the net distance in the predominant direction using directional net change in distance values and angle values derived from the first polynomial trend surfaces.

**Correlations between survey outcomes, cassava density, monthly weather averages, and population distribution.** To estimate population density in the region and at sampled locations, fine-scale gridded global population data (1-km resolution) were obtained from the WorldPop project (<http://www.worldpop.org.uk/>)

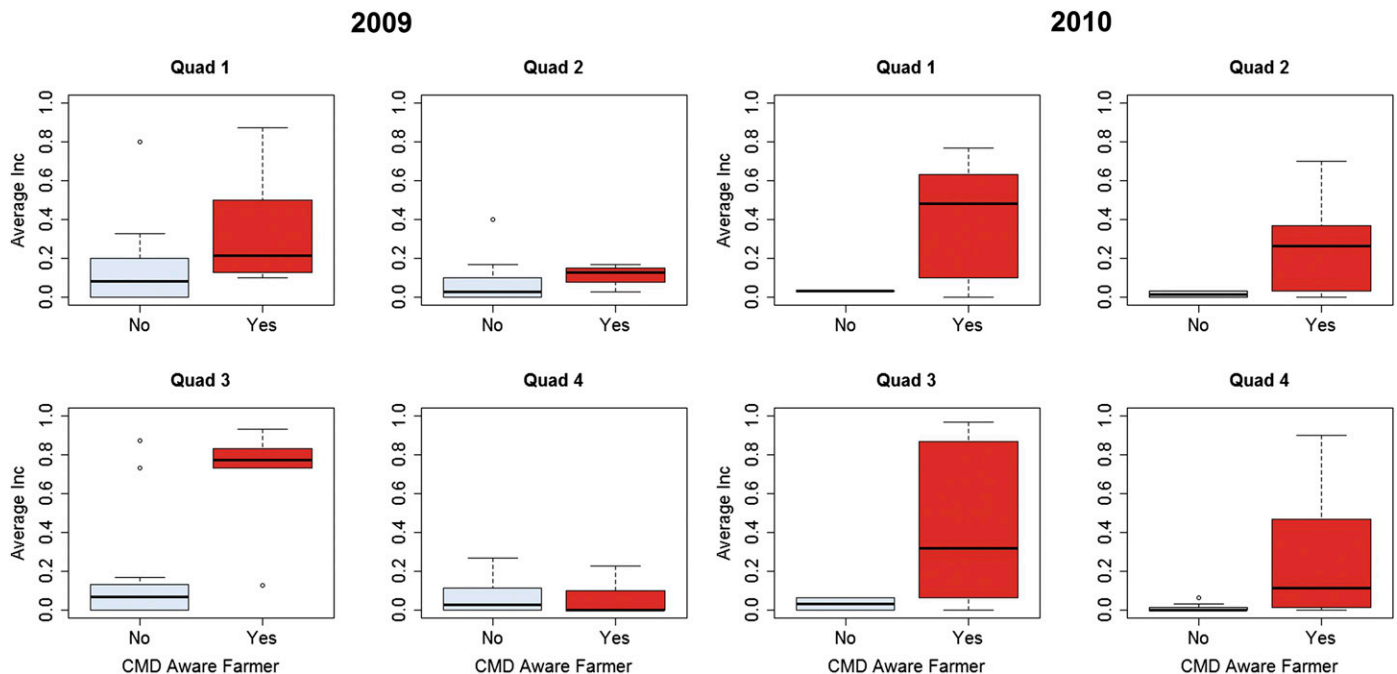
(Linard et al. 2012; Tatem 2017) (Supplementary Table S1). Spatial Production Allocation Model data (MapSPAM2005v2.0; <http://mapspam.info/>) were used to represent the heterogeneity of the spatial distribution and density of the area under cassava cultivation at 5-min resolution, corresponding to approximately 10 by 10 km at the equator (You et al. 2014). Due to the scarcity of fine spatial and temporal resolution weather station data in the area, satellite-derived datasets for the estimates of the climatic conditions were obtained. Tropical Rainfall Measuring Mission (TRMM 3B43; <https://trmm.gsfc>

TABLE 2. Average values of sampled locations for each quadrant (Q) of the study area for each of the three survey years<sup>a</sup>

Variable <sup>b</sup>	Year	Q1	Q2	Q3	Q4
CMD Incidence	2009	23.8 (5.6)	7.8 (2.1)	29.2 (7.9)	6.7 (2.0)
	2010	39.5 (6.3)	22.8 (4.8)	38.2 (8.3)	16.5 (6.1)
	2011	39.7 (4.8)	10.3 (2.6)	41.8 (8.0)	13.7 (3.3)
CMD severity	2009	2.37 (0.08)	2.06 (0.03)	2.37 (0.07)	2.24 (0.12)
	2010	3.17 (0.07)	3.21 (0.09)	3.13 (0.12)	2.92 (0.03)
	2011	3.04 (0.09)	2.65 (0.07)	2.90 (0.11)	2.42 (0.08)
Whitefly-borne infection	2009	0 (0)	0 (0)	0 (0)	0 (0)
	2010	5.8 (1.5)	2.5 (1.0)	5.7 (1.5)	0.33 (0.23)
	2011	18.2 (3.7)	0.33 (0.23)	10.5 (3.3)	0.17 (0.17)
Cutting-borne infection	2009	23.8 (5.6)	7.8 (2.1)	29.2 (7.9)	6.7 (2.0)
	2010	33.7 (5.7)	20.3 (4.4)	32.5 (7.8)	16.2 (6.1)
	2011	21.5 (2.7)	10.0 (2.5)	31.3 (6.6)	13.5 (3.3)
Whitefly abundance (log mean)	2009	0.63 (0.18)	0.40 (0.087)	0.33 (0.22)	0.008 (0.005)
	2010	7.7 (4.7)	0.42 (0.11)	2.3 (0.51)	0.23 (0.11)
	2011	2.8 (0.97)	0.2 (0.13)	3.9 (2.2)	0.1 (0.03)
Prevalence of EACMV-UG	2009	0.20 (0.09)	0.25 (0.10)	0.15 (0.08)	0 (0)
	2010	0.40 (0.11)	0.40 (0.11)	0.45 (0.11)	0.05 (0.05)
	2011	NA	NA	NA	NA
Farmer recognition of CMD	2009	0.50 (0.11)	0.15 (0.08)	0.25 (0.10)	0.25 (0.10)
	2010	0.9 (0.07)	0.9 (0.07)	0.9 (0.07)	0.6 (0.11)
	2011	NA	NA	NA	NA
Farmer losses last year	2009	21.5 (6.04)	4.5 (3.12)	9.5 (4.1)	2.5 (2.5)
	2010	49.6 (5.7)	26.4 (4.1)	64.0 (7.0)	17.7 (6.2)
	2011	NA	NA	NA	NA
Hotspot index value	2009	7.7 (0.39)	6.6 (0.22)	6.85 (0.52)	5.3 (0.32)
	2010	10.85 (0.55)	8.8 (0.58)	10.7 (0.74)	7.3 (0.54)
	2011	11.35 (0.55)	6.4 (0.44)	10.8 (0.78)	6.15 (0.37)

<sup>a</sup> Values in parentheses represent standard errors and NA = not applicable.

<sup>b</sup> CMD = cassava mosaic disease and EACMV-UG = *East African cassava mosaic virus-Uganda*.



**Fig. 2.** Farmer awareness of the cassava mosaic disease (CMD) symptoms per year and quadrant compared with the respective field disease incidence rates in 2009 and 2010 surveys.



nasa.gov/) monthly data for 2009 to 2011 at 0.25-by-0.25° spatial resolution were obtained (Huffman et al. 2007). The MODIS land surface temperature (LST) product (<https://modis-land.gsfc.nasa.gov/temp.html>) represents estimates of the land temperature variability in the region based on thermal bands of reflectance from the earth's surface (Justice et al. 2002). Default Kelvin degree units were converted to degree Celsius and minimum, maximum, and mean LST values were calculated for each month of the 3 years of interest. To represent the relative climatic variability in the study zone over the 3 years of interest, 3-year means and departures from the mean were illustrated in a set of three maps (Supplementary Fig. S2). Population, cassava cropping distribution, population density, and altitude information obtained in the survey were used to construct a Spearman rank correlation matrix against those values and CMD incidence, CMD severity, and whitefly abundance using the 'Hmisc' and 'stats' packages in R (Harnell 2017; R Core Team 2017).

**Estimating the losses due to CMD in pandemic and non-pandemic regions.** Cassava production losses in the pandemic (behind the pandemic front) and nonpandemic (ahead of the pandemic front) regions of the study area were assessed using FAOSTAT estimates for cassava yield in Tanzania for each year of the study ([www.fao.org/faostat/](http://www.fao.org/faostat/)), area under cassava cultivation obtained from MapSPAM, predicted incidence values based on survey data, and constant values of average losses due to CMD in pandemic and nonpandemic regions based on previously published CMD yield loss ( $L$ ) estimates (Legg et al. 2006). Using this set of variables, we derived the value of the potential yield per hectare ( $Y$ ) in the study region (equation 1, below). Actual production in the CMD pandemic ( $Prod_p$ ) and nonpandemic ( $Prod_n$ ) areas is equal to the sum of the average disease incidence in the respective zone (pandemic =  $i_p$  and nonpandemic =  $i_n$ ) multiplied by potential yield ( $Y$ ), area under cultivation ( $H_p$  and  $H_n$ ), and  $1 -$  average losses ( $L_p$  and  $L_n$ ) plus the product of the ratio of healthy crop ( $1 - i$ ), the yield potential value ( $Y$ ), and harvested area  $H_p$  or  $H_n$ , depending on the zone (equations 2 and 3). Production loss ( $Loss$ ) is equal to the difference between the production potential (harvested area multiplied by the yield potential

value for each year) and the sum of actual production in the pandemic region ( $Prod_p$ ) and nonpandemic region ( $Prod_n$ ) (equation 4). The economic loss was then calculated by multiplying the production loss by the estimated market value for fresh tuberous roots of cassava of US \$60/t (Ndyetabula et al. 2016; Sewando 2012).

$$Y = [Y_a \times (H_p + H_n)] / [H_p(i_p \times \{1 - L_p\} + \{1 - i_p\}) + H_n(i_n \times \{1 - L_n\} + \{1 - i_n\})] \quad (1)$$

$$Prod_p = H_p \times Y \times i_p \times (1 - L_p) + H_p \times Y \times (1 - i_p) \quad (2)$$

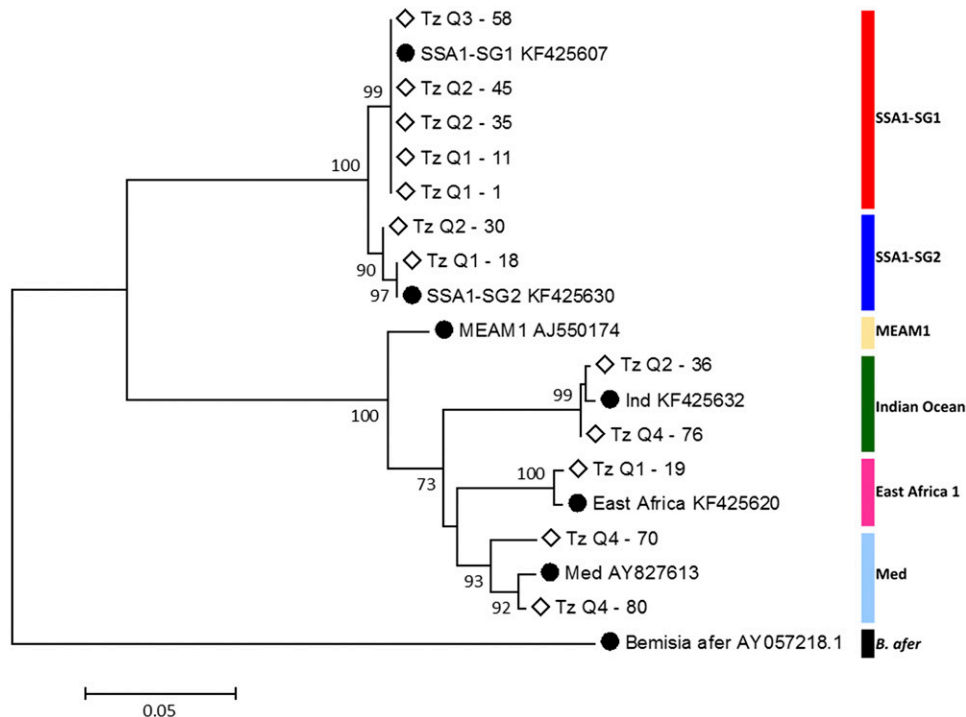
$$Prod_n = H_n \times Y \times i_n \times (1 - L_n) + H_n \times Y \times (1 - i_n) \quad (3)$$

$$Loss = (H_p + H_n) \times Y - Prod_p + Prod_n \quad (4)$$

## RESULTS

**Changes in CMD associated variables.** *CMD incidence and EACMV-UG presence.* Disease incidence varied significantly between years and quadrants (Table 2; Supplementary Table S2). Incidence in both 2010 and 2011 was greater than in 2009 but there was no change from 2010 to 2011. Quadrants Q1 (northwest) and Q3 (southwest) had higher incidences than Q2 (northeast) and Q4 (southeast). There was no significant interaction between year and quadrant; therefore, the pattern of incidence remained stable over time in all quadrants. Significant variation in EACMV-UG presence was observed between years and quadrants, with no interactions between factors (Table 2). EACMV-UG presence was significantly higher in 2010 compared with 2009 and significantly lower in Q4 compared with other factors across the study period.

*CMD severity.* Mean disease severity scores from diseased plants varied between 2009 and 2010, between 2010 and 2011, and between quadrants (Table 2). The northwest quadrant (Q1) had a significantly



**Fig. 3.** Phylogenetic tree of mitochondrial DNA cytochrome oxidase I sequences for selected *Bemisia tabaci* (Genn.) adult whiteflies collected from cassava fields sampled in the 2010 quadrant survey of northwestern Tanzania. The tree was generated in MEGA 6.0 using the maximum-likelihood procedure with 1,000 bootstraps. Open diamonds represent sequences from this study and filled circles are control sequences from the GenBank database. SSA1-SG1 and SG2 = sub-Saharan Africa 1-subgroup 1 and 2, respectively, and Med = Mediterranean.

higher average severity score compared with the eastern quadrants (Q2 and Q4), and Q3 (southwest) had higher severity scores compared with Q4 (southeast). There were no significant interactions between years and quadrants.

**Vector abundance.** Whitefly abundance varied between years and quadrants and there was a significant interaction between the two factors (Table 2). For both Q1 and Q3, there was a significant increase in whitefly abundance from 2009 to 2010 but no further change after that. There were no significant differences over time in Q2 and Q4. In common with the disease variables, the most significant differences through the study period were recorded in the two western quadrants.

**Farmers' responses.** Farmers' awareness of disease symptoms varied significantly between years and quadrants, with significantly higher overall awareness observed in 2010 (Fig. 2; Table 2). Q4 (southeast) had the least awareness of disease symptoms in the study period. There were significant differences in the level of farmer-recalled losses due to CMD between years and quadrants, and there were significant interactions between the two factors ( $df = 3$ ,  $F = 5.8$ ,  $P < 0.001$ ) (Table 2). In 2009, there were no perceived differences in losses between quadrants. In 2010, however, greater losses were perceived in Q1 and Q3 compared with Q2 and Q4 and significant changes were observed between eastern and western quadrants.

**Hotspot value.** There were significant differences in hotspot index value between years and quadrants and there was a significant interaction between both factors (Table 2). Average hotspot index value increased from 6.5 (in 2009) to 9.3 (in 2010), and was high in Q1 (10.0) and Q3 (9.6) but low in Q2 (7.2) and Q4 (6.0). There was no significant increase in hotspot index value from 2010 to 2011.

The greatest contrasts in hotspot index value between quadrants occurred between Q1 and Q4 and between Q1 and Q2 in 2011 and the greatest increase over time was recorded for Q3, where the mean hotspot value increased from 6.9 in 2009 to 10.8 in 2011.

**Whitefly genotypes.** Five *B. tabaci* genotypes were identified through 46 sequences obtained from samples collected from cassava plants in the study area. These included Indian Ocean (Ind), Mediterranean (Med), and East Africa 1 (EA1), all of which are polyphagous, and two genotypes known to primarily colonize cassava: SSA1-SG1 and sub-Saharan Africa 1–subgroup 2 (SSA1-SG2). Genetic relationships between the groups identified were described through the maximum-likelihood phylogenetic analysis (Fig. 3). The largest proportion of sequences were of SSA1-SG1 (65.2%), followed by SSA1-SG2 (17.4%), Med (8.7%), Ind (4.3%), and EA1 (4.3%). Whitefly genotypes differed between quadrants and, using Dunn's method for multiple comparison, Q3 had a significantly higher proportion of SSA1-SG1 when compared with Q4. Overall, the data showed that the putative superabundant genotype of *B. tabaci* occurred most frequently across the sampled area. Although SSA1-SG1 was much more common than other genotypes in the western quadrants of the study area (Q1 = 76.5% and Q3 to 100%), it was less frequent than other genotypes occurring in the eastern quadrants (Q2 = 45.5% and Q4 = 14.3%).

**Correlation analyses. Surveyed variables.** Considering datasets for all 3 years combined, strong positive correlations were demonstrated between CMD incidence and mean severity score (0.73) and moderately positive correlations with farmers' disease recognition (0.47) and farmers' estimate of loss in the previous year (0.55). Mean severity scores had the strongest correlation with

TABLE 3. Pairwise Spearman rank correlation coefficient table representing  $\rho$  values and  $P$  values (in italics)<sup>a</sup>

Parameters	Inc	Sev	Cut inf	Wf inf	Wf abund	Farmer recog	Farmer loss	EACMV-UG
Inc	1	0.7308	0.9763	0.6083	0.5192	0.4684	0.5473	0.2509
	...	<i>0.0000</i>	<i>0.0000</i>	<i>0.0000</i>	<i>0.0000</i>	<i>0.0000</i>	<i>0.0000</i>	<i>0.0014</i>
Sev	...	1	0.7060	0.5541	0.4209	0.4805	0.5666	0.3302
	...	...	<i>0.0000</i>	<i>0.0000</i>	<i>0.0000</i>	<i>0.0000</i>	<i>0.0000</i>	<i>0.0000</i>
Cut inf	...	...	1	0.7060	0.4489	0.4451	0.5195	0.2270
	...	...	...	0	<i>0.0000</i>	<i>0.0000</i>	<i>0.0000</i>	<i>0.0039</i>
Wf inf	...	...	...	1	0.5847	0.4507	0.5839	0.2686
	...	...	...	...	<i>0.0000</i>	<i>0.0000</i>	<i>0.0000</i>	<i>0.0006</i>
Wf abund	...	...	...	...	1	0.4340	0.5242	0.1762
	...	...	...	...	...	<i>0.0000</i>	<i>0.0000</i>	<i>0.0258</i>
Farmer recog	...	...	...	...	...	1	0.8323	0.2029
	...	...	...	...	...	...	<i>0.0000</i>	<i>0.0101</i>
Farmer loss	...	...	...	...	...	...	1	0.2557
	...	...	...	...	...	...	...	<i>0.0011</i>
EACMV-UG	...	...	...	...	...	...	...	1

<sup>a</sup> Inc = cassava mosaic disease (CMD) incidence, Sev = CMD severity, Cut inf = proportion of cutting-borne infections, Wf inf = proportion of whitefly-borne infections, Wf abund = mean whitefly number per field, Farmer Recog = farmers awareness of CMD, Farmer Loss = estimated losses in yield due to CMD, and EACMV-UG = East African cassava mosaic virus–Uganda.

TABLE 4. Pairwise Spearman rank correlation coefficient table representing  $\rho$  and  $P$  values (in italics), including source data in parentheses<sup>a</sup>

Variables	CMD incidence		CMD severity		Whitefly abundance	
	$\rho$	$P$	$\rho$	$P$	$\rho$	$P$
Harvested area (MAPSPAM) <sup>b</sup>	0.154	<i>0.017*</i>	0.147	<i>0.023*</i>	0.128	<i>0.048*</i>
Production (MAPSPAM)	0.214	<i>0.001*</i>	0.116	<i>0.074</i>	0.044	<i>0.497</i>
Mean LST (MODIS) <sup>c</sup>	−0.230	<i>0.000*</i>	−0.121	<i>0.062</i>	−0.113	<i>0.079</i>
Min LST (MODIS)	−0.161	<i>0.013*</i>	−0.054	<i>0.403</i>	−0.025	<i>0.702</i>
Max LST (MODIS)	−0.291	<i>0.000*</i>	−0.123	<i>0.058</i>	−0.201	<i>0.002*</i>
Annual precipitation (TRMM) <sup>d</sup>	0.145	<i>0.024*</i>	−0.070	<i>0.280</i>	0.047	<i>0.469</i>
Population (WorldPop) <sup>e</sup>	−0.075	<i>0.244</i>	−0.044	<i>0.495</i>	0.020	<i>0.763</i>
Altitude (survey data)	−0.313	<i>0.000*</i>	−0.226	<i>0.000*</i>	−0.275	<i>0.000*</i>
Longitude (survey data)	−0.484	<i>0.000*</i>	−0.429	<i>0.000*</i>	−0.471	<i>0.000*</i>
Latitude (survey data)	0.123	<i>0.058</i>	0.110	<i>0.090</i>	0.246	<i>0.000*</i>

<sup>a</sup> CMD = cassava mosaic disease and \* indicates significance at  $\alpha = 0.05$ .

<sup>b</sup> Cassava harvested area and production estimates were obtained from the MapSPAM2005v2.0 production allocation model (You et al. 2014) (<http://mapspam.info>).

<sup>c</sup> Mean, minimum, and maximum land surface temperatures (LST) were obtained from MODIS (<https://modis-land.gsfc.nasa.gov/temp.html>).

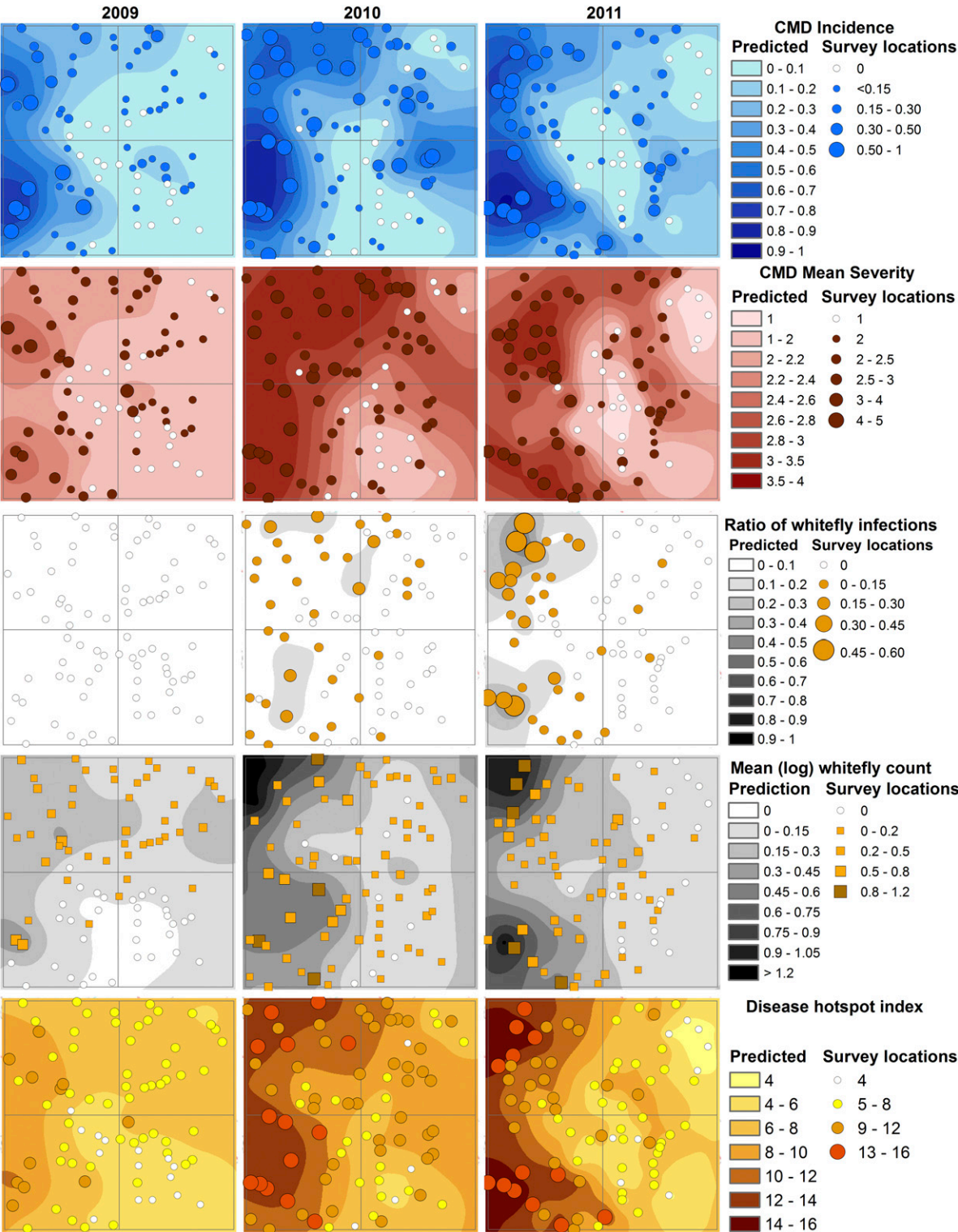
<sup>d</sup> Annual precipitation in survey locations obtained from Tropical Rainfall Measuring Mission (TRMM 3B43; <https://trmm.gsfc.nasa.gov/>).

<sup>e</sup> Population density in survey locations obtained from WorldPop (<http://www.worldpop.org.uk/>).

cutting infection (0.71), farmer recognition of CMD (0.48), and their estimate of CMD-associated loss (0.57) (Table 3). Whitefly abundance was correlated with overall CMD incidence (0.52) and previous year loss estimates (0.52), and it was more strongly correlated with whitefly-borne compared with cutting-borne CMD infection (0.58 versus 0.45, respectively). The presence of EACMV-UG had a slightly stronger correlation with whitefly-borne CMD infection compared with cutting-borne CMD (0.27 versus 0.23). The strongest correlation (0.71) was observed between farmers'

recognition of disease symptoms and the estimate of losses due to CMD in the previous year (Table 3).

*Incidence, severity, and whitefly abundance versus environmental variables.* Weak linear relationships were observed between surveyed variables and monthly weather averages, cassava distribution, and altitude at surveyed points (Table 4). No significant correlation was observed between human population density and the surveyed variables. Cassava production and area of cultivation had very weak, albeit significant positive correlations with

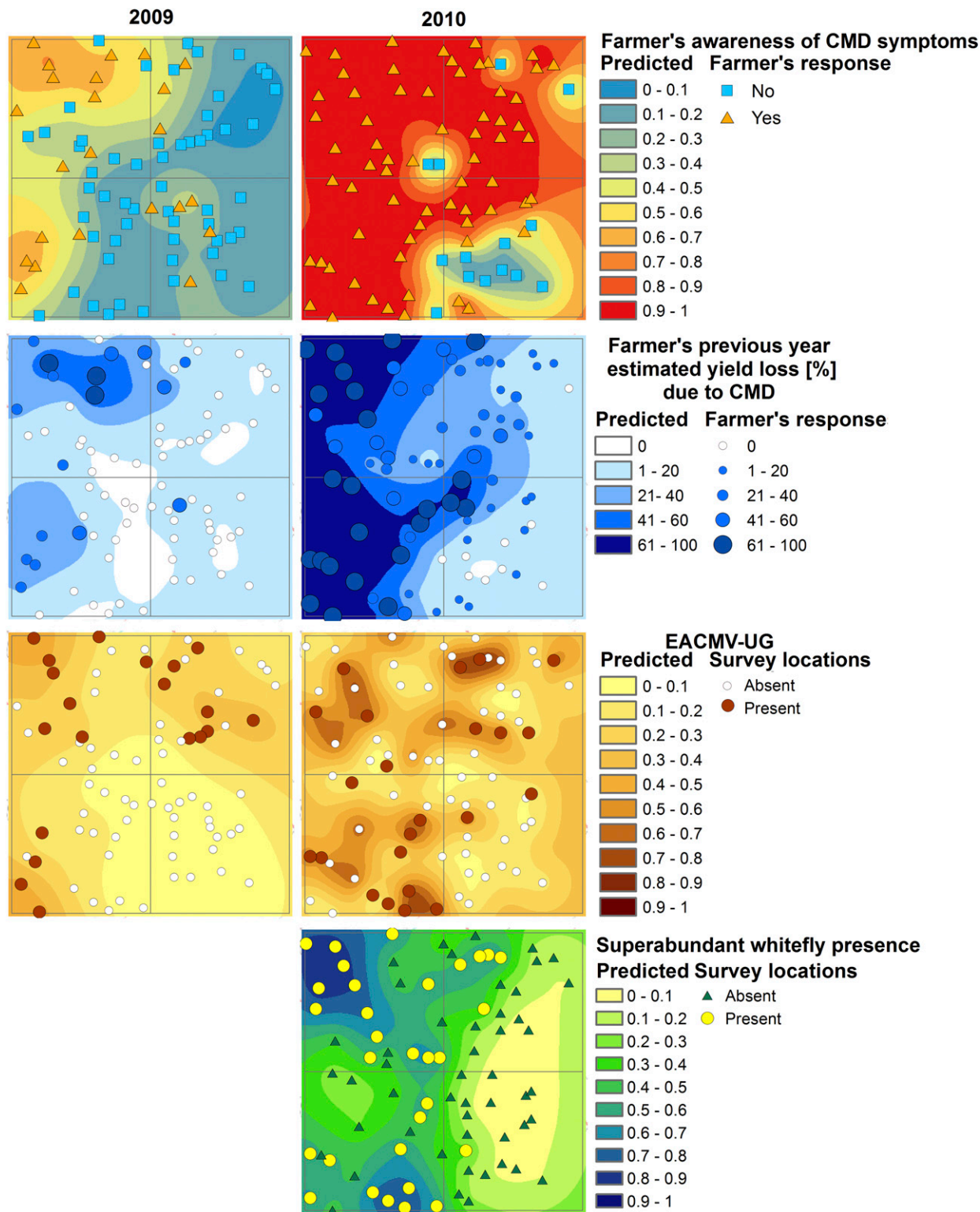


**Fig. 4.** Field survey results and kriging predictions at unsampled locations for cassava mosaic disease (CMD) incidence, CMD severity, ratio of plants infected by whitefly, vector abundance (natural log transformed mean whitefly count), and disease hotspot index.



CMD incidence. Mean, minimum, and maximum LST had very weak but significant negative correlations with CMD incidence and severity and whitefly abundance in a range of  $-0.11$  to  $-0.25$ . Altitude had significant weak correlations with CMD incidence and severity and vector abundance ( $-0.31$ ,  $-0.23$ , and  $-0.27$ , respectively). Longitude had moderate correlations with the same three surveyed variables ( $-0.48$ ,  $-0.43$ , and  $-0.47$ ), whereas latitude was weakly positively correlated with whitefly abundance.

**Spatial prediction of surveyed variables.** Large differences were observed in the distribution of all the variables of interest, mainly between 2009 and 2010 but to a lesser degree between 2010 and 2011 (Figs. 4 and 5). A pronounced gradient was apparent between the western zone of the study area, in which there were higher CMD incidence scores in 2009, and the largely uninfected central and southeastern portion of the study area (Fig. 4). A similar trend was observed for CMD severity scores, where the most severe



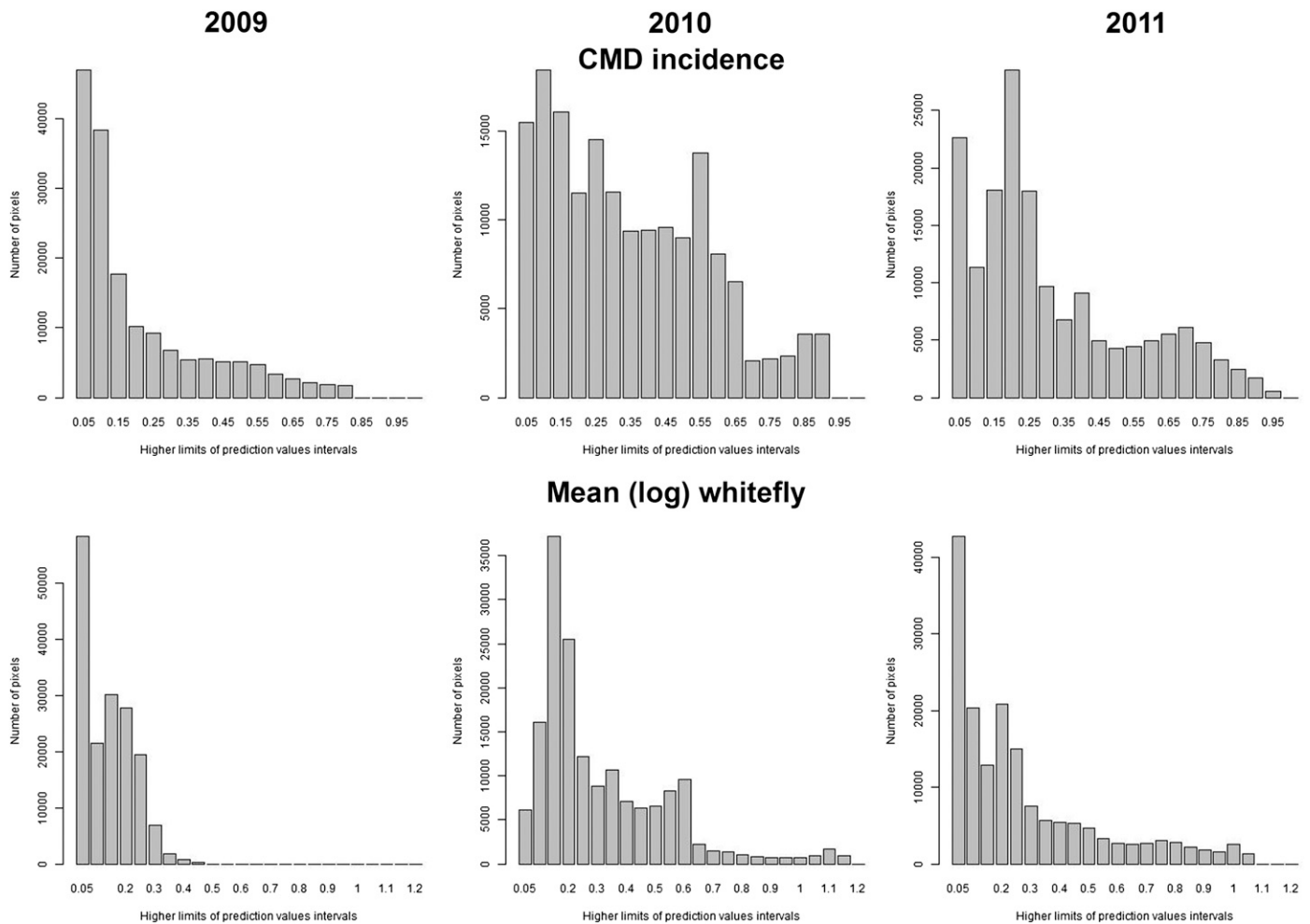
**Fig. 5.** Field survey results and kriging predictions at unsampled locations representing farmers' awareness of cassava mosaic disease (CMD) symptoms, estimated loss in yield the previous year, *East African cassava mosaic virus-Uganda* (EACMV-UG) presence, and the distribution of superabundant whitefly.



symptoms were predicted for the western part of the study area in 2009 and for western and northern areas in 2010. Little change was apparent in CMD severity between 2010 and 2011, although the distribution of areas with relatively severe CMD shifted to the south. No whitefly-borne infections were observed in the 2009 survey whereas, in the two subsequent years, a steady pattern of increase was apparent. Vector densities in the study area increased from 2009 to 2010, with the highest densities observed in the northwestern corner (Q1). The most striking change in the spatial distribution of whiteflies from 2010 to 2011 was the increase in abundance in the southwestern corner (Q3).

In common with most of the component variables, the greatest spatial changes observed comprised the increase from 2009 to 2010, combined with a shift in the primary zone of increase from the northwest (Q1) to the southwest (Q3) from 2010 to 2011. The greatest of all spatial changes in the variables for which data were collected

was recorded for the level of farmer awareness of CMD symptoms. Small numbers of farmers, concentrated in the northwestern corner, were aware of the disease in 2009 but, by 2010, most farmers throughout the study area were aware, with the exception of those in the southeastern quadrant (Q4). Similar patterns of spatial change were observed for farmer perceptions of CMD-associated yield loss, although the greatest area of change for this variable was in the western part of the study zone (Q1 and Q3), which coincided with the zone in which there were greatest increases in the CMD-associated variables. EACMV-UG became more prevalent in the surveyed area over the period 2009 to 2010 as its frequency of detection doubled. There was a clear pattern of EACMV-UG spread into the northeastern and southwestern quadrants, although this pattern of change was strongest in the southwest. The spatial distribution of the *B. tabaci* genotype that has been associated with superabundance (SSA1-SG1) had a strong longitudinal gradient, with a high



**Fig. 6.** Histograms representing the number of pixels (square units of area) containing values in a certain range of cassava mosaic disease (CMD) incidence (above) and log mean whitefly count (below) derived from kriging prediction surfaces.

**TABLE 5.** Total area affected by defined cassava mosaic disease (CMD) and whitefly pandemic zones for each year and in each quadrant of the study area

Pandemic area affected	2009		2010		2011	
	CMD	Whitefly	CMD	Whitefly	CMD	Whitefly
Quadrant 1	789.8	0.0	2,809.8	2,152.8	1,604.5	1,632.5
Quadrant 2	0.0	0.0	149.4	0.0	0.0	0.0
Quadrant 3	1,426.2	29.9	1,853.7	2,396.7	2,343.1	2,062.5
Quadrant 4	0.0	0.0	550.7	0.0	41.9	0.3
Sum (km <sup>2</sup> )	2,215.9	29.9	5,363.6	4,549.5	3,989.5	3,695.4
Total area of harvest (km <sup>2</sup> )	200.1	2.4	571.3	471.2	372.9	436.1

frequency of occurrence in the west and little presence in the east. This genotype was recorded from only 1 of the 20 sampled sites in the southeastern quadrant (Q4).

**Characterization and movement of the CMD pandemic front.** Histograms were generated which illustrate the number of pixels (units of area) present in each of a series of prediction value categories (ranges) for each of the 2010 and 2011 datasets (Fig. 6). For CMD incidence, the range of predicted values in the 0.45 to 0.50 probability of disease presence represented a small fraction of the prediction surface and, therefore, was identified as representing the disease front. A similar approach for the whitefly abundance data led to the choice of the 0.4 log mean whitefly abundance value to represent the whitefly front. Although the CMD front was clearly present in the far western part of the surveyed area in 2009, the whitefly front was represented by only a small patch in the southwestern zone (Q3). In 2010, the disease front made a clear eastward movement, particularly in the northwestern quadrant (Q1), and the whitefly front similarly advanced to the east, exceeding the range of the CMD front in the southwestern quadrant (Q3). By 2011, the fronts for both disease and whiteflies had stabilized. The position of the whitefly front was the same from 2010 to 2011 but the disease front continued to advance eastward in Q3, while retracting slightly in Q1.

The CMD pandemic area was 2,216 km<sup>2</sup> in 2009 and increased to 5,364 km<sup>2</sup> in 2010, before shrinking slightly to 3,939 km<sup>2</sup> in 2011 (Table 5). When using whitefly abundance to classify the pandemic

zone, the area affected increased from 30 km<sup>2</sup> in 2009 to 4,549 km<sup>2</sup> in 2010 before declining to 3,695 km<sup>2</sup> in 2011. Q2 remained entirely outside the pandemic-affected zones in 2009 and 2011 and Q4 was unaffected in 2009. The harvested area in the pandemic zone was also highest in 2010 (571 and 471.2 km<sup>2</sup> under CMD incidence and whitefly pandemic zones, respectively).

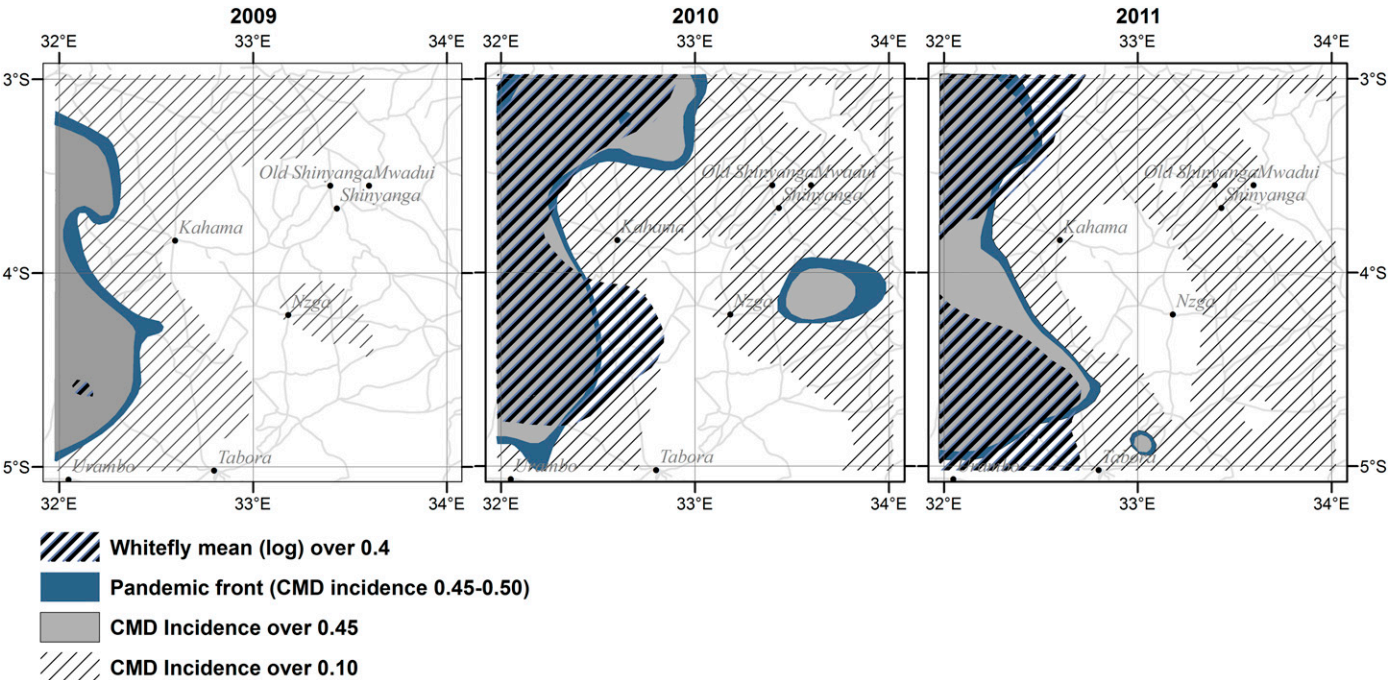
The predominant direction of the pandemic front movement varied between pandemic front type (CMD versus whitefly) and year (Table 6). The CMD front followed a longitudinal trend from east to west in 2009. The pattern changed in 2010, with a much more pronounced 19° tilt in the same direction, while, in 2011, the angle switched to a northeastern tilt of 6.5° (Table 6; Fig. 7). Similar trends were observed for the direction of movement of the whitefly front. The net distance of the pandemic front movement between 2009 and 2010 was 22.9 km in the southeastern direction on average (Table 7). Between 2010 and 2011, the CMD front progressed by an average 7.5 km to the northeast but retracted, on average, by 17.3 km to the southwest. The pandemic trend based on whitefly abundance data had a very pronounced southeastern tilt in 2009 and 2010 (41 and 29°, respectively), which became weaker in 2011 (4.5°). The average distance of southeastern movement of the whitefly pandemic zones was larger between 2009 and 2010 (43.6 km) compared with 2010 and 2011 (9.6 km).

**Cassava mosaic disease (CMD) pandemic front estimates.** The harvested area of cassava within the CMD pandemic zone increased from 20,009 ha in 2009 to over 57,127 ha in 2010, although it decreased by almost 20,000 ha in 2011 (Table 8; Fig. 8). Based on FAOSTAT yield estimates for Tanzania of 5.5 t/ha (2009), 5.2 t/ha (2010), and 6.3 t/ha (2011), we derived potential yield values for the study area of 6.3 t/ha in 2009, 7.5 t/ha in 2010, and 8.2 t/ha in 2011 (Table 8). In 2009, the estimated loss due to CMD was 12.6% across the study area. The CMD pandemic area suffered losses estimated at 41.7% compared with nonpandemic area losses of 4.5%. Both production and estimated losses in the pandemic region were greatest in 2010 (228,083 and 186,212 t, respectively). The percent loss of total cassava production due to CMD was highest in 2010 (30.2% across the study area), although the losses in the pandemic zone were highest in 2011 (46.8%). The spatial distribution of losses per unit of area is illustrated in Figure 9. Total losses of cassava production in absolute

TABLE 6. Predominant angle of the first polynomial trend function for cassava mosaic disease (CMD) incidence and whitefly densities as measured from the y axis<sup>a</sup>

Year	Front type	Angle
2009	CMD incidence	-0.5°
	Whiteflies	41°
2010	CMD incidence	20°
	Whiteflies	29°
2011	CMD incidence	-6.5°
	Whiteflies	4.5°

<sup>a</sup> Positive values indicate tilt to the southeast and negative values indicate tilt to the northwest.



**Fig. 7.** Cassava mosaic disease (CMD) pandemic front in 2009 (left), 2010 (center), and 2011 (right). Pandemic zones were defined as areas with predicted values of disease incidence of 0.45 or higher and whitefly log-transformed means over 0.4.

terms were 71,104 t in 2009, 203,030 t in 2010, and 168,733 t in 2011. The total combined value of lost cassava production is estimated to be approximately US\$26.5 million over the 3-year period of the study.

## DISCUSSION

We used a novel approach to sample variables associated with CMD, in which a predetermined study zone was delineated that covered an area of 2 by 2° (approximately 222 by 222 km) in northwestern Tanzania. This was selected based on the prior knowledge that the CMD pandemic was moving into this zone from the northwest. Our approach differed from all previous studies of CMD regional epidemiology in not using a province- or country-level spatial scale. Otim-Nape et al. (1997) produced the first estimates for the rate of CMD pandemic spread (approximately 20 to 30 km/year) based on field studies in Uganda. Subsequently, Legg (2010) estimated rates of pandemic spread of up to 100 km/year in East Africa around Lake Victoria. In both of these cases, however, distance values were estimates. In the current study, the sampling design made it possible to take a more mechanistic approach to mapping the CMD pandemic front and using the maps to determine the rate, direction, and economic consequences of movements in the CMD pandemic front.

TABLE 7. Net change in distance and predominant direction of the pandemic front<sup>a</sup>

Movement	Net movement (km) for pandemic front type, years, and predominant direction				
	Incidence			Whitefly	
	2009–2010	2010–2011		2009–2010	2010–2011
	SE	SW	NE	SE	SE
Mean	22.9	17.3	7.5	43.6	9.6
Maximum	85.1	75.3	37.6	74.0	65.9

<sup>a</sup> Latitudinal and longitudinal net values are obtained by measuring the mean and maximum value of the net change in the latitudinal and longitudinal extent of the pandemic front. Southeastern (SE) and northwestern (NW) values are calculated using the angle value, spread distance, and sine function.

Mapping the pattern of CMD spread through the target zone over a 3-year period (from 2009 to 2011) revealed a strong spatial association between the CMD-associated variables that were recorded, which included incidence and severity of CMD, whitefly-borne CMD infection, and whitefly abundance. All variables increased in the western parts of the study area between 2009 and 2010. Several studies have described the temporal links between such factors (Bigirimana et al. 2004; Legg 1999; Otim-Nape et al. 1997) but the current study is the first to present the spatial features of these factors. Data presented here also demonstrate that there was little significant change in the CMD-associated variables from 2010 to 2011, and CMD severity actually declined in 2011 in comparison with the previous year. Whitefly abundance increased from 2009 to 2010 but was generally lower in 2011 than in 2010, notably in the northwestern quadrant of the study area, which is a probable reason for the lack of change in CMD incidence there, although the reduction in severity might be due to environmental conditions. The weather was generally cooler and wetter in 2011 compared with the hotter and drier weather in 2010. Data from the current study, however, did not allow us to prove that weather factors caused this change. Another factor contributing to the decrease in CMD severity in 2011 could have been the increased frequency of whitefly-borne relative to cutting-borne infections, because it has been previously shown that whitefly-borne infections are associated with milder disease symptoms (Thresh et al. 1994). Changes in whitefly abundance from year to year might have been partly due to differences in weather conditions during the periods when surveys were conducted. However, the most significant pattern of variation in whitefly abundance recorded during this study was between quadrants during the same season. This gives us confidence in the significance of the spatial associations demonstrated through the study and the overall conclusions drawn about the pattern of CMD pandemic spread. These findings are further corroborated by similar findings from several studies that have confirmed spatial associations between whitefly abundance changes and patterns of CMD pandemic spread (Legg 2010; Legg and Ogwal 1998; Otim-Nape et al. 1996).

It was notable that SSA1-SG1 *B. tabaci* individuals and EACMV-UG were both detected significantly “ahead” of areas affected by the pandemic zone. This suggests that these two components, which have been described as the drivers of the

TABLE 8. Assessment of the cassava actual and potential production and losses attributed to cassava mosaic disease (CMD) in pandemic and nonpandemic regions

Cassava production and loss assessment	2009	2010	2011
Harvested area, pandemic region (ha) <sup>a</sup>	20,009.9	57,126.8	37,289.4
Harvested area, nonpandemic region (ha) <sup>a</sup>	69,912.8	32,795.9	52,633.3
Total (ha) <sup>a</sup>	...	89,922.7	...
Yield (t/ha) <sup>b</sup>	5.4712	5.2096	6.2808
Yield potential in the area (t/ha) <sup>c</sup>	6.26	7.47	8.16
Production, pandemic region (t)	73,032.8	240,378.3	161,775.6
Production, nonpandemic region (t)	418,952.2	228,083.0	403,010.9
Estimated production without CMD loss, pandemic region (t)	125,300.5	426,590.7	304,178.2
Estimated production without CMD loss, nonpandemic region (t)	437,788.7	244,901.2	429,341.9
Estimated total production without CMD loss (t)	563,089.2	671,491.9	733,520.1
Pandemic region losses (t) <sup>c</sup>	52,267.6	186,212.4	142,402.6
Pandemic region losses (%)	41.71	43.65	46.82
Nonpandemic region losses (t) <sup>c</sup>	18,836.5	16,818.2	26,331.0
Nonpandemic region losses (%)	4.50	7.37	6.53
Total losses (t)	71,104.1	203,030.6	168,733.6
Total losses (%)	12.63	30.24	23.00
Value of cassava (at US\$60/t) <sup>d</sup>			
Losses value in the pandemic region (USD)	3,136,058.8	11,172,742.9	8,544,155.5
Losses value in the nonpandemic region (USD)	1,130,188.1	1,009,093.5	1,579,860.6
Total estimated losses in the study area (USD)	4,266,246.9	12,181,836.4	10,124,016.1

<sup>a</sup> Cassava harvested area estimates were obtained from the MapSPAM2005v2.0 spatial production allocation model (You et al. 2014) (<http://mapspam.info>).

<sup>b</sup> Cassava yield estimates (in tons per hectare [t/ha]) in 2009 to 2011 were obtained from FAOSTAT (<http://www.fao.org/faostat/>).

<sup>c</sup> Yield potential in the area was calculated using FAOSTAT actual yield figures and 72 and 35% losses in the pandemic and nonpandemic region, respectively (equation 1) (Legg et al. 2006).

<sup>d</sup> Cassava value in U.S. dollars (USD) per ton (Ndyetabula et al. 2016; Sewando 2012).

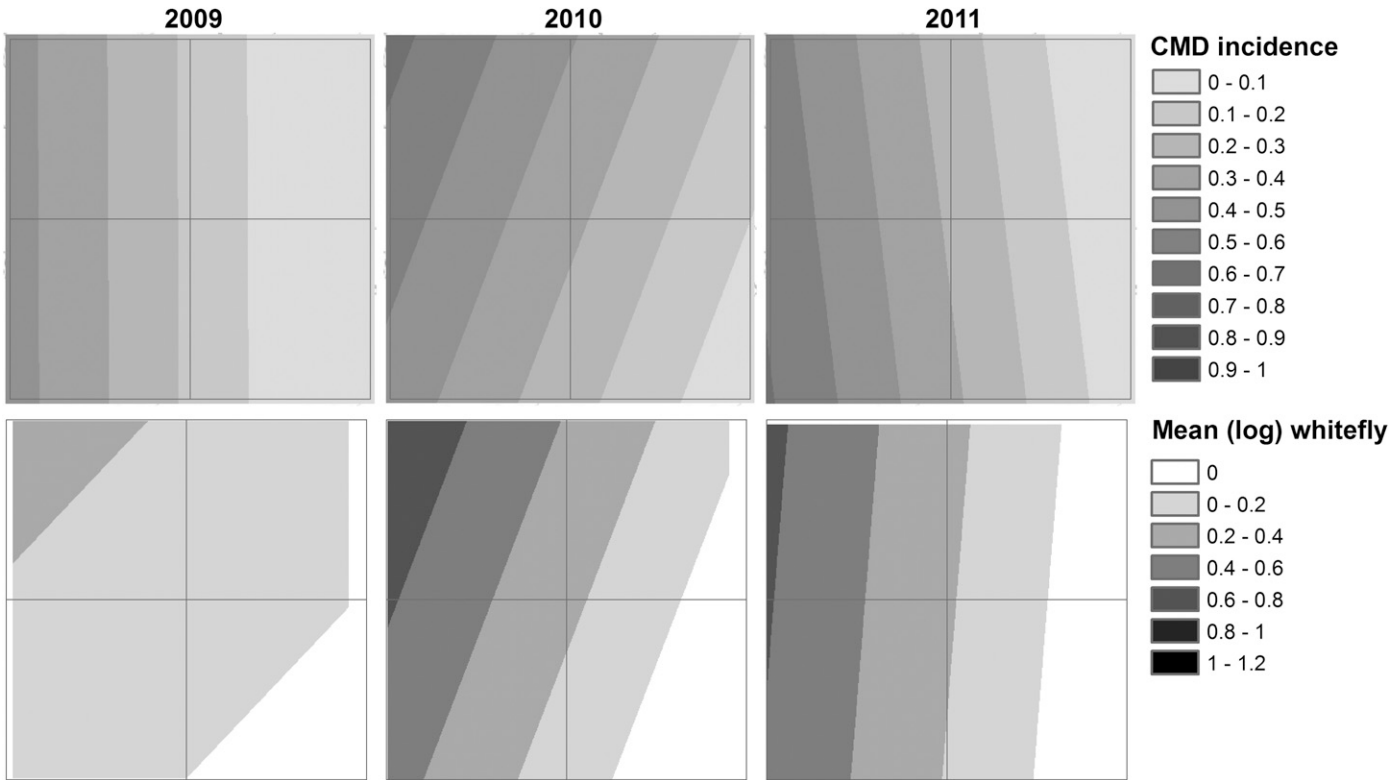


pandemic (Legg et al. 2014; Pita et al. 2001), move to areas in advance of the disease front prior to the appearance of high incidences of severe CMD.

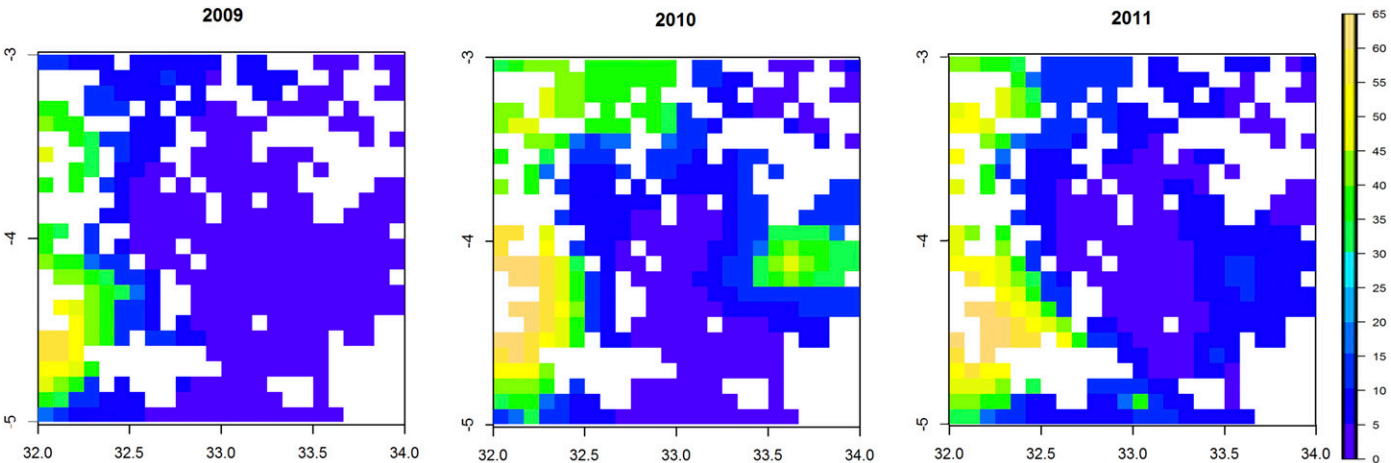
The hotspot index developed in this study aims to highlight those parts of the study area in which CMD epidemic conditions were most extreme and which, therefore, would represent the greater production risk to cassava growers. The progression over time for hotspot index matched the patterns of change of its components.

It is relatively straightforward for farmers to recognize the symptoms of CMD and, therefore, symptom recognition has been an important component of farmer training programs for the management of cassava pests and diseases (Msikita et al. 2000; Thresh and Cooter 2005). Our data demonstrated a very strong spatial association between

the presence in cassava fields of high incidences of severe CMD and the likelihood that farmers recognize the disease. This recognition was also strongly associated with a perception of production losses in the previous year. In 2009, there were clear spatial associations between farmer disease recognition and the perception of loss, as well as with the other CMD-associated variables such as CMD incidence and severity. However, in the second year in which farmer responses were assessed (2010), perceptions of loss continued to match other CMD-associated variables but farmer recognition of CMD increased to a much greater extent, with almost all farmers interviewed recognizing the disease. The only exception to this was in the extreme southeastern corner, which was least affected overall by CMD. This wider-than-anticipated spread of symptom recognition by farmers is an indication



**Fig. 8.** First-order polynomial trend based on cassava mosaic disease (CMD) incidence (above) and mean log whitefly count (below) in each year of the survey, revealing the slight variation in the predominant directions of pandemic front movements.



**Fig. 9.** Estimated losses (%) in cassava production due to cassava mosaic disease for the 3 years of the survey, based on predicted mean incidence values in the region, cassava cultivated area in hectares (source: MapSPAM2005v2.0), and derived potential yield estimates. Estimates of losses for infected plants are 72% in the pandemic region and 35% in the nonpandemic region (Legg et al. 2006).

that farmers were not relying solely on what they could see in their own fields but, instead, were receiving information about this emerging problem either from friends or relatives in affected areas or via the extension system. This is supported by anecdotal evidence from the survey teams that both extension and local research staff were active in providing training information to farmers on the effects of cassava viruses during the period of the study. This highlights the potential effectiveness of awareness-raising efforts for CMD management; although, to be most effective, these would also need to be combined with efforts to supply virus-resistant germplasm, which is the frontline approach for managing CMD (Dixon et al. 2003).

Perhaps the most unique aspect of our analysis of the CMD pandemic in northwestern Tanzania was the use of several sets of spatial data to mechanistically define the CMD pandemic front. This was achieved by identifying categories of CMD incidence and whitefly abundance at which the spatial gradient between high and low values was greatest. As far as we are aware, this is the first time that this approach has been used to define and plot the advancing front of a plant virus epidemic. After defining the pandemic front, it was possible to demonstrate that it moved in a southeasterly direction approximately 23 km from 2009 to 2010 and 14 km northwest from 2010 to 2011. The maximum distance moved during the overall period of study was 85 km southeast between 2009 and 2010. It is important to mention that the front line is approximate and is somewhat affected by sampling distribution and the choice of variogram parameters. In our studies with evenly and randomly sampled fields, this potential bias was reduced.

The study reveals that the CMD pandemic front is relatively complex, and that it may not always be readily described with a smooth line running through contiguous areas sharing common epidemiological characteristics. It will certainly be affected by the heterogeneity of the landscape, which is a feature of the part of Tanzania surveyed in this study. Factors that can influence the ease and accuracy with which a front line can be constructed include the lack of reliable data on host (cassava) distribution and variation in host density, as well as the relative resistance or susceptibility to virus infection of the varieties being grown. Correlation analyses did not reveal strong associations between host distribution and human population in the landscape, or with the surveyed CMD variables, albeit multivariate relationships could be further investigated in future studies using alternative approaches that take into account spatial and temporal correlation between sampled covariates.

Defining the front also facilitated the calculation of yield losses in areas affected by the pandemic and areas not yet affected. Legg et al. (2006) used a whole-country- or region-level approach to estimate production losses due to CMD for all of the cassava-producing countries of Africa. Continent-wide losses were estimated at 34 million t, which would be equivalent to slightly more than US\$2 billion, assuming a farm-gate fresh root production value of US\$60/t (Sewando 2012). In the current study, losses predicted for the nonpandemic region ranged from 4.5 to 7.4% during the period 2009 to 2011, whereas losses in the pandemic-affected region were predicted to be much higher, ranging from 42 to 47%. Overall economic losses predicted for the 3-year period were relatively low in 2009 (US\$4.2 million) and increased greatly in 2010 (US\$12.2 million), before declining slightly in 2011 (US\$10.1 million). Pandemic-associated losses for the seven affected countries in 2006 averaged 47% (Legg et al. 2006), which closely approximates the maximum loss value obtained for the northwestern Tanzanian study area described here. It is important to point out, however, that there are significant limitations to the FAOSTAT and MapSPAM datasets. FAOSTAT data are provided at the whole-country level, while the MapSPAM model did not predict cassava cultivation at several points where cassava fields were sampled. Nevertheless, these datasets represent the best proxy available and provide a useful basis from which to develop yield loss estimation approaches. In response to the large production losses associated with CMD infection, there have been widespread efforts

to apply management approaches, most notably in Uganda, which was the country first affected.

Kriging and techniques of spatial analysis have been used to plot the movement of the CMD pandemic in Tanzania in this study. Although this has been developed for a specific pathosystem, the approaches described should be of equal value in monitoring patterns of development of the CMD pandemic in other parts of Africa, as well as more broadly for the study of plant virus epidemics at the landscape scale. The parts of Africa currently threatened by severe CMD are Cameroon and Eastern Nigeria, as well as the southern shores of Lake Tanganyika in the Democratic Republic of Congo and Zambia. The methods presented here should be of particular value in identifying CMD-affected zones, determining the direction and rate of disease spread, quantifying production losses, and helping to target control efforts.

## ACKNOWLEDGMENTS

We thank F. van den Bosch, K. Hall, and two anonymous reviewers for their valuable comments on the manuscript. A. Szyniszewska gratefully acknowledges funding from the Bill and Melinda Gates Foundation grant to the University of Cambridge. Rothamsted Research receives support from the Biotechnology and Biological Sciences Research Council, United Kingdom. J. Legg acknowledges funding support from the Roots, Tubers, and Bananas program of the Consortium of International Agricultural Research Centers. The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript. The International Institute of Tropical Agriculture is hosted in Tanzania by the Government of Tanzania, and expresses its appreciation for the local support that it receives.

## LITERATURE CITED

- Bigirimana, S., Barumbanze, P., Obonyo, R., and Legg, J. P. 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 Pathol.* 53:231.
- Bivand, R. S., Pebesma, E. J., and Gomez-Rubio, V. 2008. *Applied Spatial Data Analysis with R*. Springer, New York.
- Bock, K. R., and Woods, R. D. 1983. The etiology of African cassava mosaic disease. *Plant Dis.* 67:994-995.
- Bouwmeester, H., Heuvelink, G. B. M., Legg, J. P., and Stoorvogel, J. J. 2012. Comparison of disease patterns assessed by three independent surveys of cassava mosaic disease in Rwanda and Burundi. *Plant Pathol.* 61:399-412.
- Bull, S. E., Briddon, R. W., Sserubombwe, W. S., Ngugi, K., Markham, P. G., and Stanley, J. 2006. Genetic diversity and phylogeography of cassava mosaic viruses in Kenya. *J. Gen. Virol.* 87:3053-3065.
- Dixon, A. G. O., Bandyopadhyay, R., Coyne, D., Ferguson, M., Ferris, R. S. B., Hanna, R., Hughes, J., Ingelbrecht, I., Legg, J., Mahungu, N., Manyong, V., Mowbray, D., Neuenchwander, P., Whyte, J., Hartmann, P., and Ortiz, R. 2003. Cassava: From a poor farmer's crop to a pacesetter of African rural development. *Chron. Hortic.* 43:8-14.
- Dubern, J. 1994. Transmission of African cassava mosaic geminivirus by the whitefly (*Bemisia tabaci*). *Trop. Sci.* 34:82-91.
- Frohlich, D. R., Torres-Jerez, I., Bedford, I. D., Markham, P. G., and Brown, J. K. 1999. A phylogeographic analysis of the *Bemisia tabaci* species complex based on mitochondrial DNA markers. *Mol. Ecol.* 8:1683-1691.
- Goovaerts, P. 1997. *Geostatistics for Natural Resources Evaluation*. Oxford University Press, New York.
- Harrell, F. E., Jr. 2017. Hmisc: Harrell Miscellaneous. R package version 4.0-3. Online publication.
- Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., and Jarvis, A. 2005. Very high resolution interpolated climate surfaces for global land areas. *Int. J. Climatol.* 25:1965-1978.
- Huffman, G. J., Bolvin, D. T., Nelkin, E. J., Wolff, D. B., Adler, R. F., Gu, G., Hong, Y., Bowman, K. P., and Stocker, E. F. 2007. The TRMM multi-satellite precipitation analysis (TMPA): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales. *J. Hydrometeorol.* 8:38-55.
- IITA. 2009. Cassava disease surveillance surveys. Mapping report. Great Lakes Cassava Initiative. Online publication. International Institute of Tropical Agriculture. [http://www.iita.org/wp-content/uploads/up\\_b1/Annual\\_Survey\\_2009\\_Maps\\_Report\\_Final.pdf](http://www.iita.org/wp-content/uploads/up_b1/Annual_Survey_2009_Maps_Report_Final.pdf)
- Jeremiah, S. C., Ndyetabula, I. L., Mkamilo, G. S., Haji, S., Muhanna, M. M., Chuwa, C., Kasele, S., Bouwmeester, H., Ijumba, J. N., and Legg, J. P. 2015. The dynamics and environmental influence on interactions between cassava brown streak virus disease and the whitefly, *Bemisia tabaci*. *Phytopathology* 105:646-655.

- Justice, C. O., Townshend, J. R. G., Vermote, E. F., Masuoka, E., Wolfe, R. E., Saleous, N., Roy, D. P., and Morisette, J. T. 2002. An overview of MODIS Land data processing and product status. *Remote Sens. Environ.* 83:3-15.
- Lecoustre, R., Fargette, D., Fauquet, C., and de Reffye, P. 1989. Analysis and mapping of the spatial spread of African cassava mosaic virus using geostatistics and the kriging technique. *Phytopathology* 79:913-920.
- Legg, J. P. 1999. Emergence, spread and strategies for controlling the pandemic of cassava mosaic virus disease in east and central Africa. *Crop Prot.* 18:627-637.
- Legg, J. P. 2010. Epidemiology of a whitefly-transmitted cassava mosaic geminivirus pandemic in Africa. Pages 233-257 in: *Bemisia: Bionomics and Management of a Global Pest*. A. Stansly and S. E. Naranjo, eds. Springer, Dordrecht, Heidelberg, London, New York.
- Legg, J. P., and Ogwal, S. 1998. Changes in the incidence of African cassava mosaic geminivirus and the abundance of its whitefly vector along south-north transects in Uganda. *J. Appl. Entomol.* 122:169-178.
- Legg, J. P., Owor, B., Sseruwagi, P., and Ndunguru, J. 2006. Cassava mosaic virus disease in East and Central Africa: Epidemiology and management of a regional pandemic. *Adv. Virus Res.* 67:355-418.
- Legg, J. P., Sseruwagi, P., Boniface, S., Okao-Okuja, G., Shirima, R., Bigirimana, S., Gashaka, G., Herrmann, H.-W., Jeremiah, S. C., Obiero, H. M., Ndyetabula, I., Tata-Hangy, W., Masembe, C., and Brown, J. K. 2014. Spatio-temporal patterns of genetic change amongst populations of cassava *Bemisia tabaci* whiteflies driving virus pandemics in East and Central Africa. *Virus Res.* 186:61-75.
- Legg, J. P., and Thresh, J. M. 2000. Cassava mosaic virus disease in East Africa: A dynamic disease in a changing environment. *Virus Res.* 71:135-149.
- Linard, C., Gilbert, M., Snow, R. W., Noor, A. M., and Tatem, A. J. 2012. Population distribution, settlement patterns and accessibility across Africa in 2010. *PLoS One* 7:e31743.
- Lodhi, M. A., Ye, G. N., Weeden, N. F., and Reisch, B. I. 1994. A simple and efficient method of DNA extraction from grapevine cultivars and *Vitis* species. *Plant Mol. Biol. Rep.* 12:6-13.
- Msikita, W., James, B., Nnodu, E., and Legg, J. 2000. Disease Control in Cassava Farms. IITA, Ibadan, Nigeria.
- Ndunguru, J., Legg, J. P., Aveling, T. A. S., Thompson, G., and Fauquet, C. M. 2005. Molecular biodiversity of cassava begomoviruses in Tanzania: Evolution of cassava geminiviruses in Africa and evidence for East Africa being a center of diversity of cassava geminiviruses. *Virology* 337:21-31.
- Ndyetabula, I. L., Merumba, S. M., Jeremiah, S. C., Kasele, S., Mkamilo, G. S., Kagimbo, F. M., and Legg, J. P. 2016. Analysis of interactions between cassava brown streak disease symptom types facilitates the determination of varietal responses and yield losses. *Plant Dis.* 100:1388-1396.
- Ogbe, F. O., Thottappilly, G., Dixon, A. G. O., and Mignouna, H. D. 2003. Variants of East African cassava mosaic virus and its distribution in double infections with African cassava mosaic virus in Nigeria. *Plant Dis.* 87:229-232.
- Oliver, M. A., and Webster, R. 2014. A tutorial guide to geostatistics: Computing and modelling variograms and kriging. *Catena* 113:56-69.
- Otim-Nape, G. W., Alicai, T., and Thresh, J. M. 2001. Changes in the incidence and severity of cassava mosaic virus disease, varietal diversity and cassava production in Uganda. *Ann. Appl. Biol.* 138:313-327.
- Otim-Nape, G. W., Bua, A., Thresh, J. M., Baguma, Y., Ogwal, S., Semakula, G. N., Acola, G., Byabakama, B., and Martin, A. 1997. Cassava Mosaic Virus Disease in Uganda: The Current Pandemic and Approaches to Control. Natural Resources Institute, Chatham, UK.
- Otim-Nape, G. W., Thresh, J. M., and Fargette, D. 1996. *Bemisia tabaci* and cassava mosaic virus disease in Africa. Pages 319-350 in: *Bemisia 1995: Taxonomy, Biology, Damage, Control and Management*. D. Gerling and R. T. Meyer, eds. Intercept, Andover, UK.
- Pebesma, E. J. 2004. Multivariable geostatistics in S: The gstat package. *Comput. Geosci.* 30:683-691.
- Pita, J. S., Fondong, V. N., Sangaré, A., Otim-Nape, G. W., Ogwal, S., and Fauquet, C. M. 2001. Recombination, pseudorecombination and synergism of geminiviruses are determinant keys to the epidemic of severe cassava mosaic disease in Uganda. *J. Gen. Virol.* 82:655-665.
- R Core Team. 2017. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna. Online publication. <https://www.r-project.org/>
- Sewando, P. P. 2012. Urban markets-linked cassava value chain in Morogoro Rural Districts, Tanzania. *J. Sust. Dev. Afr.* 14:283-300.
- Simon, C., Beckenbach, A., Crespi, B., Liu, H., and Flook, P. 1994. Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. *Ann. Entomol. Soc. Am.* 87:651-701.
- Sseruwagi, P., Rey, M. E. C., Brown, J. K., and Legg, J. P. 2004a. The cassava mosaic geminiviruses occurring in Uganda following the 1990s epidemic of severe cassava mosaic disease. *Ann. Appl. Biol.* 145:113-121.
- Sseruwagi, P., Sserubombwe, W. S., Legg, J. P., Ndunguru, J., and Thresh, J. M. 2004b. Methods of surveying the incidence and severity of cassava mosaic disease and whitefly vector populations on cassava in Africa: A review. *Virus Res.* 100:129-142.
- Swanson, M. M., and Harrison, B. D. 1994. Properties, relationships and distribution of cassava mosaic geminiviruses. *Trop. Sci.* 34:15-25.
- Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S. 2011. MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* 28:2731-2739.
- Tatem, A. J. 2017. WorldPop, open data for spatial demography. *Sci. Data* 4:170004.
- Thresh, J. M., and Cooter, R. J. 2005. Strategies for controlling cassava mosaic virus disease in Africa. *Plant Pathol.* 54:587-614.
- Thresh, J. M., Fargette, D., and Otim-Nape, G. W. 1994. Effects of African cassava mosaic geminivirus on the yield of cassava. *Trop. Sci.* 34:26-42.
- Thresh, J. M., Otim-Nape, G. W., Legg, J. P., and Fargette, D. 1997. African cassava mosaic disease: The magnitude of the problem. *Afr. J. Root Tuber Crops* 2:13-19.
- You, L., Wood-Sichra, U., Fritz, S., Guo, Z., See, L., and Koo, J. 2014. Spatial production allocation model (SPAM) 2005 v2.0. Online publication. <http://mapspam.info>
- Zhou, X., Liu, Y., Calvert, L., Munoz, C., Otim-Nape, G. W., Robinson, D. J., and Harrison, B. D. 1997. Evidence that DNA-A of a geminivirus associated with severe cassava mosaic disease in Uganda has arisen by interspecific recombination. *J. Gen. Virol.* 78:2101-2111.