Bulletin of Entomological Research (2001) **91**, 333–346 DOI: 10.1079/BER2001103

A contribution towards simplifying area-wide tsetse surveys using medium resolution meteorological satellite data

G. Hendrickx[1]*, A. Napala[2], J.H.W. Slingenbergh[3], R. De Deken[4] and D.J. Rogers[5]

[1]FAO Trypanosomiasis project GCP-RAF-347-BEL, BP 2034 Bobo Dioulasso, Burkina Faso: [2]FAO Trypanosomiasis project GCP-RAF-347-BEL, BP 114 Sokodé, Togo: [3]FAO, AGAH, Viale delle Terme di Caracalla, 00100 Rome, Italy: [4]ITG/IMT, Nationale Straat 155, 2000 Antwerp, Belgium: [5]Department of Zoology, Oxford University, South Parks Road, Oxford, OX1 3PS, UK

Abstract

A raster or grid-based Geographic Information System with data on tsetse, trypanosomiasis, animal production, agriculture and land use has recently been developed in Togo. The area-wide sampling of tsetse fly, aided by satellite imagery, is the subject of two separate papers. This paper follows on a first paper, published in this journal, describing the generation of digital tsetse distribution and abundance maps and how these accord with the local climatic and agro-ecological setting. Such maps when combined with data on the disease, the hosts and their owners, should contribute to the knowledge of the spatial epidemiology of trypanosomiasis and assist planning of integrated control operations. Here we address the problem of generating tsetse distribution and abundance maps from remotely sensed data, using a restricted amount of field data. Different discriminant analysis models have been applied using contemporary tsetse data and remotely sensed, low resolution data acquired from the National Oceanographic and Atmospheric Administration (NOAA) and Meteosat platforms. The results confirm the potential of satellite data application and multivariate analysis for the prediction of the tsetse distribution and abundance. This opens up new avenues because satellite predictions and field data may be combined to strengthen and/or substitute one another. The analysis shows how the strategic incorporation of satellite imagery may minimize field collection of data. Field surveys may be modified and conducted in two stages, first concentrating on the expected fly distribution limits and thereafter on fly abundance. The study also shows that when applying satellite data, care should be taken in selecting the optimal number of predictor variables because this number varies with the amount of training data for predicting abundance and on the homogeneity of the distribution limits for predicting fly presence. Finally, it is suggested that in addition to the use of contemporary training data and predictor variables, training and predicted data sets should refer to the same eco-geographic zone.

*Permanent address: Avia-GIS, Elsbos 24, 2650 Edegem, Belgium
Fax: +32 3 458 2979
E-mail: ghendrickx@pandora.be

Introduction

Previous papers (Hendrickx *et al*., 1999a,b) introduced an area-wide approach to trypanosomiasis management. It was argued that spatial information on tsetse challenge, trypanosomiasis risk, animal production, farming systems and land use are essential to our understanding of the disease problem and, therefore, for rational disease control. A grid-based data capture frame was introduced to facilitate analysis in a Geographical Information System (GIS).

Field-collected data on the distribution of tsetse species (Diptera: Glossinidae) in Togo were compiled and digital maps produced, depicting fly presence or absence and (for riverine species only) fly abundance. As a next step, the fly data were related to the prevailing ecological conditions. For this, a process of cluster analysis of environmental variables, mostly obtained through satellite imagery, was first carried out in order to demarcate the main, ecologically distinct areas; when compared with the observed tsetse distribution and abundance the more essential fly habitat characteristics were revealed.

The present article takes this study one step further and explores different avenues in order to establish how the tsetse habitat may be discerned from satellite-derived environmental data while applying only minimal amounts of field-collected entomological data.

Two distinctly different types of satellite data may be used in tsetse studies:

1. High resolution data are obtained from two major sources, SPOT (Satellite Pour l'Observation de la Terre) or Landsat. Both products have a high spatial (respectively 10 and 30 m) but low temporal (26 and 16 days) resolution.
2. Low resolution data are mainly acquired from the National Oceanographic and Atmospheric Administration (NOAA) and Meteosat platforms. This imagery has a lower spatial resolution of 7.5 km pixels for NOAA data and 5 km for Meteosat, but a much higher temporal resolution, with about two NOAA images per 24 h and one image every 30 min for Meteosat.

De Wispelaere (1994) integrated SPOT derived data on vegetation and land use to discern the habitat of *Glossina morsitans submorsitans* Newstead (Diptera: Glossinidae) on the Adamawa plateau in Cameroon. Kitron *et al*. (1996) used Landsat imagery in the remote Lambwe Valley (Kenya) to predict favourable fly habitat. De La Rocque (1977) applied SPOT imagery, in combination with field survey data, in an attempt to identify major discriminating factors of tsetse presence at a 3–10 m resolution, in Sideradougou, Burkina Faso. The high resolution imagery allows in-depth studies of tsetse habitats but in relatively small areas only.

Rogers & Randolph (1993) applied NOAA derived NDVI (Normalized Difference Vegetation Index, a measurement for the greenness of vegetation) data plus ground measured temperature and elevation data, to predict the distributions of *Glossina morsitans* Westwood and *G. pallidipes* Austeni (Diptera: Glossinidae) in Kenya and Tanzania. Taking the historical fly distribution (Ford & Katondo, 1977) as a reference, satellite-derived predictor variables were selected and an accuracy of respectively 84 and 79% correct predictions was obtained when predicting the presence of *G. morsitans* and *G. pallidipes*.

For West Africa, Rogers *et al*. (1996) carried out a similar exercise and produced distribution limits of eight different tsetse species encountered in Burkina Faso and Ivory Coast, at a 0.167 degree resolution. The satellite data in this study comprised Fourier-processed NDVI, Channel 4 (linked to ground temperature) and CCD values (Cold Cloud Duration, linked to rainfall). As before, historical records served as the reference fly distribution (Laveissière & Challier, 1977, 1981). Selecting the ten best predictor variables, a 74%, 87%, 67% and 71% correct description, including all the data in the training set, was obtained when predicting the abundance, respectively, of *Glossina tachinoides* Westwood, *G. palpalis* Robineau Desvoidy, *G. m. submorsitans* and *G. longipalpis* Wiedemann.

More recently still, Robinson *et al*. (1997a,b) analysed the historical distribution of *G. m. centralis* Machado, *G. m. morsitans* Westwood and *G. pallidipes* in the common fly belt of Malawi, Mozambique, Zambia and Zimbabwe (Ford & Katondo, 1977) using NDVI, ground measured temperatures, rainfall and elevation. Multivariate techniques included were linear discriminant analysis, maximum likelihood classification and principal component analysis. For each species, the best predictor variables were selected and the discriminant functions applied to produce 84–92% correct predictions. Interestingly, the analysis successfully identified the geographical limits of both subspecies of *G. morsitans* present.

Hendrickx *et al*. (1993, 1996) and Rogers *et al*. (1994) introduced discriminant analysis of satellite data to identify tsetse habitat in Togo. The present paper follows up on these initial studies. The rationale is to seek to minimize the use of ground collected data and to optimize satellite imagery application. Results obtained using similar approaches to predict the spatial distribution of trypanosomiasis prevalence are discussed elsewhere (Hendrickx *et al*., 2000) Such results should contribute significantly towards better decision making and integrated control programmes (Hendrickx *et al*., 1997, 2001; Hendrickx, 1999).

Materials and methods

The production of comprehensive, 0.125 degree resolution tsetse distribution and abundance maps covering the whole of Togo is detailed in a previous paper (Hendrickx *et al*., 1999a). The digital map layers used in this paper include: (i) distributions of *Glossina (Nemorhina) tachinoides*, *G. (Nemorhina) p. palpalis*, *G. (Glossina) m. submorsitans* and *G. (Glossina) longipalpis*; and (ii) abundance maps of both riverine species *G. tachinoides* and *G. p. palpalis*. For these species, $\log_{10}$ transformed densities were first adjusted for seasonality and smoothed spatially. This was introduced since surveys were carried out during different times of the year in different places. Seasonal adjustments were first calculated on the basis of monthly indices, depicted from measured seasonal curves. Fly density values were thus multiplied with the index of the respective capture month. Fly catches of different survey sites were pooled to obtain the apparent fly density per species per grid. Seasonal curves at the individual fly species level were calculated for the two riverine species (*Nemorhina*), *G. tachinoides* and *G. p. palpalis*. Data available for the savannah flies (*Glossina*) *G. m. submorsitans* and *G. longipalpis* were insufficient to obtain these seasonal figures. Spatial smoothing was carried out through averaging the grid value with that of the eight, or less, adjacent grids. Missing grid-values were also replaced this way.

Fly abundance data for *G. tachinoides* and *G. p. palpalis* were reclassed into four categories: (i) flies absent; and (ii) low, (iii) medium and (iv) high fly abundance. Class limits were set using an equal frequency distribution. The forest group flies (*Austenina*), *Glossina fusca fusca* Walker and *G. medicorum* Austen, are known vectors of the disease where contact with cattle exists (e.g. Baldry & Molyneux, 1980), but of marginal economic importance in Togo (Hendrickx *et al.*, 1999a) and are not included here.

Two types of environmental predictor variables were used; (i) ground based climatic and agricultural land use data; and (ii) remotely sensed data from two different satellite platforms, NOAA and Meteosat.

Data from the NOAA series included: the Normalized Vegetation Index (NDVI), a measure for actively photosynthesizing biomass (Tucker *et al.*, 1983), Channel 3 (CH3) data from the middle infrared band (MIR), identified as predictors of land cover categories in Nigeria (Rogers *et al.*, 1997), and, finally, the Price thermal brightness index (Price, 1984), a measure of the variability of ground temperature. In this study the decadal (= 10 days) maximum value composite images derived from the decadal 7.5 km resolution PATHFINDER data set were used, provided by the TALA research group (Zoology Department, Oxford University), for the period 1988 to 1990. The original temporal resolution is of two images per 24 h period. Mean decadal average, maximum and minimum values were extracted using the 2×2 pixel array centred as closely as possible on each 0.125 degree grid square to fit the database.

Meteosat data, provided by the ARTEMIS project (FAO, Rome), comprise Cold Cloud Duration (CCD) values, derived from measured cloud-top temperatures. In West Africa cloud-top temperatures between –30 and –60°C have been related to ground measures of rainfall (Snijders, 1991). Cold cloud duration values are expressed as the number of hours that cloud-top temperatures were below the applicable threshold, depending on the area covered, for a given time period. Here the original spatial resolution is 5 km and a new image is taken every 30 min. Data were adapted in a similar way as for NOAA data.

Three consecutive years of decadal, i.e. 10 day period, maximum-value composited images were selected; for NDVI, CH3, Price: 1988–1990 and for CCD 1989–1991. Seasonal information was extracted through a Fourier time series analysis (Rogers *et al.*, 1996). Such a data reduction technique is essential in order to reduce the size of the data set (36 decades $\times$ 3 years $\times$ 4 variables) whilst keeping a maximum of useful information. The extracted information was stored as the mean variable value as well as the amplitudes (i.e. peaks) and phases (i.e. seasonal timing) of the annual, bi-annual and tri-annual cycles for each variable. Table 1 summarizes the derived variables included in this study.

Multiple discriminant analysis is traditionally the most appropriate technique of multivariate analysis in situations where the dependent variable counts over two categories and the predictor variables are a series of metrical independent variables. An earlier version of the model adopted in this paper has been described in detail by Rogers *et al*. (1996). Predictor variables were ranked individually and the best variables selected to compute discriminant functions and produce predictive maps of fly abundance.

The selection of discriminant variables was based on the assessment of the Mahalanobis distance, D^2. Before determining D^2, outliers, defined as observations or values that differ from their respective class means by more than six standard deviations, were removed from the analysis. The variable selection rule aimed at maximizing D^2 between group centroids, first computing the best single-variable model and subsequently adding variables until the selection rule is not met anymore. A model using separate covariance matrices was adopted in this paper.

Four different ways, later referred to as analysis options one to four, were then explored in order to arrive at a ranking of the best predictor variables:

1. The best predictor variables were ranked on the basis of the complete set of abundance data, including data concerning areas with no flies, and selected variables were

Table 1. List of variables used.

Ground data	Remotely sensed data			
%Agriculture	NDVI av	Ch3 av	Price av	CCD av
Rainfall	NDVI p1	Ch3 p1	Price p1	CCD p1
Dry months	NDVI a1	Ch3 a1	Price a1	CCD a1
Maximum temp.	NDVI p2	Ch3 p2	Price p2	CCD p2
Digital elevation model	NDVI a2	Ch3 a2	Price a2	CCD a2
	NDVI p3	Ch3 p3	Price p3	CCD p3
	NDVI a3	Ch3 a3	Price a3	CCD a3
	NDVI max	Ch3 max	Price max	CCD max
	NDVI min	Ch3 min	Price min	CCD min
	NDVI range	Ch3 range	Price range	CCD range

%Agriculture, percentage land in the cultivation cycle; Rainfall, annual rainfall (mm); Dry months, number of dry months (< 30 mm rain); Maximum temp., average monthly mean maximum temperature (°C); DEM, Digital elevation model (m); Ch3, Channel 3 value; Price, Price thermal brightness index; NDVI, Normalized Difference Vegetation Index (see also text); CCD, Cold Cloud Duration (h, see also text); av, average value; p1 and a1, phase and corresponding amplitude of the annual Fourier cycle; p2 and a2, phase and corresponding amplitude of the bi-annual Fourier cycle; p3 and a3, phase and corresponding amplitude of the tri-annual Fourier cycle; max, maximum value; min, minimum value; range, maximum–minimum value.

then used to compute discriminant functions and produce predictive maps;

2. Variables were ranked on the basis of only the low, medium and high density fly abundance data, excluding data pertaining to fly absent areas. Selected variables were then used to compute discriminant functions and produce predictive maps, also for fly absent areas;

3. As above, but with predictions produced solely for areas where flies were considered present and excluding fly absent areas, thus assuming the distribution limits to be known;

4. A simple discriminant analysis was applied on the basis of fly presence and absence. This was the sole approach adopted for the savannah flies, for which no fly abundance data were available.

An essential aspect of this type of analysis is the validation of the discriminant functions obtained. To do this, the data were divided into two subsamples, bearing in mind that the total training sample should exceed 100 grids (Hair *et al*., 1995). The first subsample, the analysis sample or more commonly called the training set, was used to rank variables, to define prior probabilities of class membership and to compute the discriminant functions. The second subsample, the hold-out sample or predicted set, was used to validate the computed functions, i.e. to produce 'true predictions' based on the information contained in the training set. This approach *de facto* offers the opportunity to test the results obtained, combining field surveys (training set) and extrapolations (predicted set), the purpose of this study. An additional aim was to explore ways to economize field surveys and yet produce accurate or even improved spatial distribution layers.

Two different ways of partial incorporation of training set data were examined, applying two different types of training set data (S1 and S2):

S1: A first approach assumes no prior knowledge of fly distribution patterns (fly presence or absence) and the training set data were randomly generated from within the seasonal cluster areas earlier defined by Hendrickx *et al*. (1999a). To test the effect of a decreasing amount of training data, and thus an increase in the size of the predicted set, the exercise was repeated incorporating 75%, 66%, 50% and 33% respectively from each seasonal cluster area (fig. 1a).

S2: A second approach presumes a minimal prior knowledge of fly distributions. For this, Togo was divided into two parts from a field survey point of view: (i) high preference areas comprising areas with known or presumed riverine fly distribution limits and/or savannah fly presence; and (ii) low preference areas. As with the seasonal cluster areas, a sequential random set of subsamples was then generated, each time incorporating a fixed percentage of the training set. The rationale for the incorporation of S2 is that there may be a premium on concentrating field survey activities in

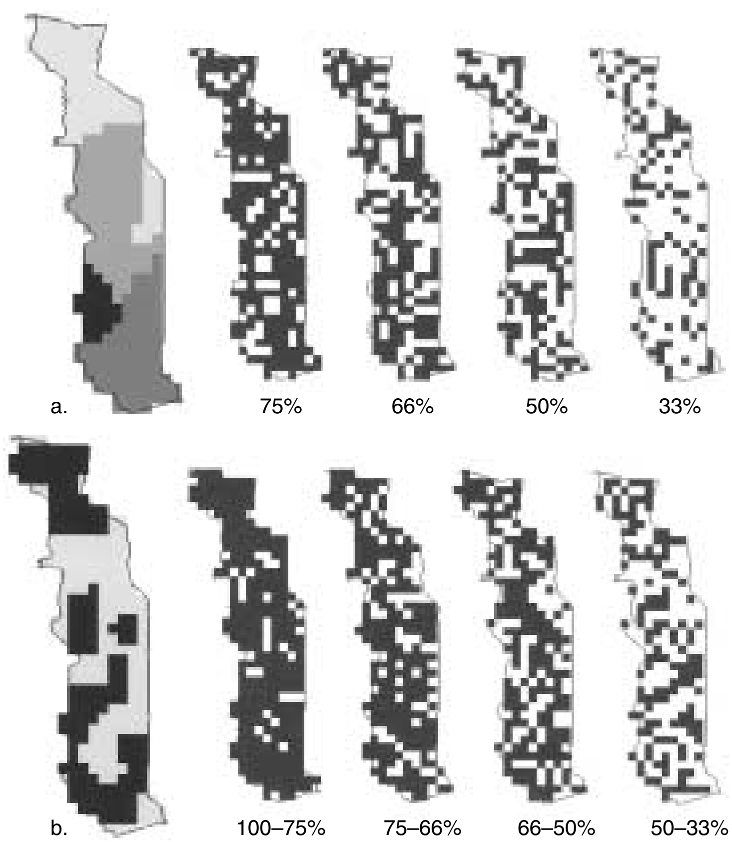

Fig. 1. Maps showing randomly selected grid-squares per sample: (a) sample 1 (S1), selected in satellite derived seasonal clusters (cf. first map of series a; four seasonal clusters defined by Hendrickx *et al*. (1999a)); (b) sample 2 (S2), selected according to preferential sampling areas (cf. first map of series b; ■, high preference; □, low preference). See text for more details.

the area of expected fly limits. To maintain the analogy with S1, the following tests were run: taking 100% of the preferential area as part of the training set plus 75% of the lower preference area, 75–66%, 66–50% and 50–33%. The training sets thus cover respectively 89%, 74%, 62% and 45% of the country surface area (fig. 1b).

Specific discriminant functions were computed per fly species for the different scenarios, with the partial incorporation of two types of training set data, and four different ways to arrive at the ranking of predictor variables, all run firstly to reproduce the original training set in its entirety and, secondly to approximate the observed 'true prediction'.

Map displays of predictions were compared with observed distributions and abundances and with prediction outputs using a 100% training set. For each predicted map the 'hit ratio' was computed, i.e. the percentage of predicted grid-square values corresponding with the observed values or percentage 'correct' answers, per abundance class, for the part of the area where flies are present and for the total area. It is a relatively simple and intuitive measure of agreement.

Classification matrices to assess the hit ratio were prepared separately for the training and the predicted set. The hit ratios given for the predicted sample are a measure of the predictive value of the model, whilst those given for the training set reflect the ability of the model to describe the part of the data set 'known to the analysis'.

The computed classification matrices also allowed a study of the error distribution. Given the aim of our study, emphasis was placed on the occurrence of major errors, i.e. of more than one class, since these could generate major decision-taking errors during application of the methodology for surveying. Errors were considered to be over- or under-estimations of abundance.

The predictive accuracy obtained with the predicted set was further tested using the Kappa Coefficient, or Index of Agreement (Cohen, 1960), widely used for this purpose (Congalton, 1991):

$$K = \frac{p_o - p_c}{1 - p_c}$$

where p_o = the sum of the proportions of predictions in concordance with observed values and p_c = the sum of proportions of concordant predictions expected by chance alone. Significance of K is best assessed using the ranges of agreement set by Landis & Koch (1977): poor $K < 0.4$; good $0.4 < K < 0.75$; excellent $K > 0.75$.

A Poisson distribution was used to test differences between used prediction models. However, since data overdispersion was detected, a negative binomial regression model was used to account for this dispersion.

Tables are given describing only the most frequently selected variables per fly species and per type of analysis approach.

Discriminant functions can be validated in two ways: (i) internally, i.e. part of the data set is used to predict the other part as described above; and (ii) externally, i.e. the data set is used to predict distribution and or abundance elsewhere. To test such results, updated distribution or abundance maps of the new predicted area are required. Since no such data were available for neighbouring Ghana and Benin, the areas of choice for such an exercise, the potential for predicting the fly distribution elsewhere was verified by comparing previous fly predictions for Togo fly distributions, prepared on the basis of discriminant functions obtained from studies in Ivory Coast and Burkina Faso (Rogers *et al*., 1996), to the actual fly distributions now reported.

Results

Abundance and distribution of riverine tsetse (subgenus Nemorhina)

Glossina (Nemorhina) tachinoides

The results of the *G. tachinoides* predictions are given in tables 2 and 3 and fig. 2. To avoid data overload, only part of the obtained results are shown in fig. 2 whilst a full summary is given in table 1. Mapped outputs suggest a satisfactory overall fit of predicted and observed results for *G. tachinoides*: the analysis produces sound geo-climatic areas with fly absence and low, medium or high abundance.

When predicting all four fly density classes (i.e. absent, low, medium, high) results improved for both types of training sets when the second analysis option was applied (calculation of Mahanolobis distance without incorporation of fly-free areas). In particular, the prediction of fly absence improves. However, the different ways in which the variable ranking was determined had little effect on the accuracy of the fly abundance predictions.

When comparing the two training set types, i.e. the 75% training-set of sample 1 with the 75–66% (*de facto* 74%) training set of sample 2 and also respectively the 66% and 66–50% (*de facto* 62%) training sets, best overall hit ratios were obtained using the first sample technique, assuming no prior knowledge of the tsetse distribution.

As expected, and shown in table 2, a diminished amount of training set data in general translated into diminished predictive power. However, the reduction in the amount of training set data was generally not matched by commensurate decreases in predictive power. In fact, the results suggested that with a reduction of the amount of training set data by two thirds, the predictive power became reduced by only about 10%.

Table 2 also illustrates that fly absence was best predicted with a simple two class presence/absence model (option 4), whilst the best results with fly abundance predictions (LMH) were obtained when predicting abundance within fly limits (option 3). However, these results were only indicative since no statistical significance could be shown. A comparison with results obtained using different sets of randomly selected samples might be needed to further clarify this issue.

The total hit ratio of the predicted set was on average 22% (± 2) lower than for the training set (see example in fig. 3). Adding more variables to the model added to its predictive power until a maximum was reached (13 predictor variables in this case) beyond which predictions even deteriorated.

The most frequently selected predictor variables are given in table 3. Two distinct types of variables occurred. The variables ranking highest in the predictions were characterized by a low-to-high or high-to-low gradient in variable value from fly absent to fly abundant situations; clearly these variables contributed towards the determination of the 'broad brush' overall abundance class. Lower ranking variables may not show such a gradient (bold print, table 3) but instead contributed to fine-tune local variability in abundance levels.

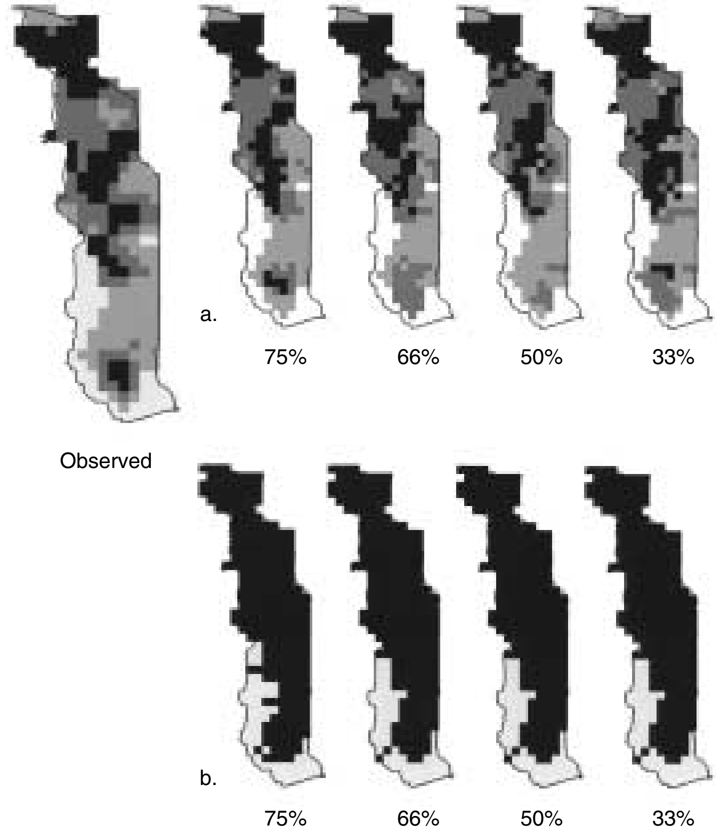

Fig. 2. Maps showing abundance and distribution predictions for *Glossina tachinoides* as compared to observed values. Results are shown for subsequent training sets including respectively 75%, 66%, 50% and 33% of the grid cells: (a) predictions of abundance within fly limits (□, absent; □, low; ■, medium; ■, high); (b) predictions of fly distribution limits. See text and tables 2 and 3 for more details. The upper left map shows the observed abundance.

When analysing the occurrence of the most frequently selected variables compared to their proportional occurrence in the data base, it appeared that from the included satellite data the derived rainfall data (CCD) scored best and the ground temperature variability (Price index) and vegetation index (NDVI) least. It was interesting to note that though the Price index did not score well as a whole, the Price amplitude 1 defined alone the limits of the fly distribution with great accuracy irrespective of the size of the training set used as seen in fig. 2b, the areas of fly absence being areas with a low Price amplitude 1 value.

Glossina (Nemorhina) palpalis palpalis

Selected prediction results are given in tables 4 and 5 and fig. 4. While the north–south gradient of fly abundance was reproduced more accurately, the lower density areas in the southern part of the country tended to disappear. This influenced overall prediction results which were less accurate as for *G. tachinoides*.

As for *G. tachinoides*, no statistically significant differences could be shown between the different methods used and abundance prediction trends were similar: fly abundance was best predicted within fly limits and fly absence with a two class presence/absence model except for the linear fly presence on the Oti river at the northern edge of its distribution (fig. 4b, see arrow).

Again, as with *G. tachinoides*, no proportional drop in overall hit ratio was observed when the amount of training set data decreases, but the results of the smallest training set were consistently inferior to those obtained with the largest. Still, comparatively little predictive power was lost even when using only one-third of the data.

No single variable was capable of predicting both distribution limits, i.e. in the dry north and in the humid south. This may be attributed to the fact that both these areas of absence are eco-climatically most distinct and thus not defined by the same predictor variables; the timing of vegetation growth (NDVI phase 1) demarcates the northern limit and a variable related to the amount of rainfall (CCD amplitude 3) the south western plateau absence area. The best total absence predictions were obtained with four to seven variables.

Figure 5 shows the regression obtained when plotting the amount of training set data against the number of variables required for optimal predictions, pooling all data for *G. tachinoides* and *G. p. palpalis*. When the amount of training set data decreased, the addition of more predictive variables did not further improve results.

The most frequently selected variables are given in table 5.

Table 2. Predictions for *Glossina tachinoides*.

Analysis option		# var.	Hit ratio (%)						Major error (%)		*Kappa*
			Absent	Low	Medium	High	LMH	Total	+	−	
D^2 calculated using all classes	$S1_{75}$	16	62	81	57	81	72	71	3	3	0.601
	$S1_{66}$	13	73	63	65	71	66	67	3	3	0.550
	$S1_{50}$	7	74	59	44	65	55	58	2	5	0.433
	$S1_{33}$	10	58	71	54	68	64	63	2	1	0.499
	Average values		**67**	**69**	**55**	**71**	**64**	**65**	**2.6**	**2.9**	
D^2 calculated omitting the absent class	$S1_{75}$	13	85	90	25	76	63	67	0	6	0.553
	$S1_{66}$	14	80	71	65	81	72	73	3	3	0.582
	$S1_{50}$	7	87	64	38	70	56	61	3	5	0.470
	$S1_{33}$	8	70	62	57	68	63	64	3	7	0.514
	Average values		**81**	**72**	**46**	**74**	**64**	**66**	**2.1**	**5.3**	
Predictions within fly limits	$S1_{75}$	13	–	95	26	76	65	–	0	6	0.474
	$S1_{66}$	13	–	71	71	81	74	–	0	2	0.611
	$S1_{50}$	11	–	76	59	54	63	–	2	3	0.409
	$S1_{33}$	8	–	64	54	75	64	–	2	1	0.527
	Average values		–	**77**	**53**	**72**	**67**	–	**1.0**	**3.0**	
Present/absent	$S1_{75}$	4	92	–	–	–	91	91	–	–	0.687
	$S1_{66}$	1	80	–	–	–	92	91	–	–	0.631
	$S1_{50}$	1	87	–	–	–	94	93	–	–	0.742
	$S1_{33}$	1	79	–	–	–	94	91	–	–	0.691
	Average values		**85**	–	–	–	**93**	**92**	–	–	
D^2 calculated using all classes	$S2_{100}$	22	60	85	20	100	67	66	0	0	0.528
	$S2_{75}$	14	85	71	57	57	60	64	0	5	0.521
	$S2_{66}$	18	30	49	71	57	58	53	6	1	0.364
	$S2_{50}$	15	46	60	62	70	63	61	9	1	0.463
	Average values		**55**	**66**	**53**	**71**	**62**	**61**	**3.8**	**1.6**	
D^2 calculated omitting the absent class	$S2_{100}$	18	60	77	80	43	70	69	0	3	0.559
	$S2_{75}$	10	85	76	52	60	62	65	1	5	0.530
	$S2_{66}$	6	75	72	54	70	66	68	3	4	0.561
	$S2_{50}$	6	82	54	49	72	58	62	3	5	0.487
	Average values		**76**	**70**	**59**	**61**	**64**	**66**	**1.7**	**4.2**	

SX_Y, chosen sample method (cf. fig. 1) where X refers to sample option 1 or 2 and Y to the percentage grids included in the trainingset; # var., the number of predictor variables best predicting the predicted set; hit ratio (%), percentage of predicted grid-squares corresponding with observed results; LMH, combined results for the three presence classes (low–medium and high); major err. (%), errors of more than one class; +, overestimation; −, underestimation; *Kappa*, *Kappa* coefficient of agreement.

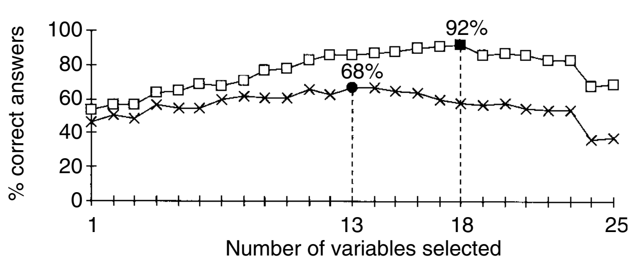

Fig. 3. Relationship between the number of variables selected (X) and the percentage of correct answers (Y). Example given here: prediction of *Glossina tachinoides* abundance, Mahanalobis distance computed including all classes, using 66% of the grid cells (training set, □) to predict the remaining 33% of the grid cells (predicted set, ×).

Here also trends were similar to *G. tachinoides*. Tables 3 and 5 reveal that about half the variables selected appeared in both cases: 8/17 for *G. tachinoides* and 8/15 for *G. p. palpalis*. Both types of variables, those that show a low–high gradient according to fly abundance and those that do not, were included. As for *G. tachinoides*, CCD or rainfall derived predictor variables scored best.

Distribution of savannah tsetse (subgenus Glossina*)*

Glossina (Glossina) morsitans submorsitans

The results of the predictions are given in table 6 and fig. 6. In general, outputs obtained with the second training set were somewhat better. Though the overall hit ratio was similar for both data sets, the prediction of fly presence improved using the second sample. As for *G. p. palpalis*, a combination of variables was required to predict fly distribution limits.

As with the riverine species, rainfall linked CCD data were more frequently selected over the other satellite data and ground measured variables also significantly contributed to the analysis.

Glossina (Glossina) longipalpis

Prediction results are given in table 7 and fig. 7. As for *G. m. submorsitans*, presence prediction results improved with

Table 3. Most frequently selected variables for predicting *Glossina tachinoides*.

Variable	Av. rank	Average values (Standard Errors)			
		Absent	Low	Medium	High
Price amp 1	1.9	3.96 (0.16)	6.67 (0.19)	8.34 (0.22)	8.72 (0.17)
CCD amp 3	2	66.8 (2.78)	62.7 (1.23)	65.2 (1.04)	72.7 (0.87)
Ch3 range	2.8	11.7 (0.26)	15.2 (0.22)	17.2 (0.18)	17.3 (0.16)
Price range	3	9.10 (0.34)	14.0 (0.41)	17.0 (0.42)	18.6 (0.40)
Ch3 phase 2	3.7	2.29 (0.07)	2.18 (0.13)	2.99 (0.16)	3.51 (0.16)
Ch3 phase 1	5.8	1.83 (0.02)	1.69 (0.01)	1.73 (0.01)	1.79 (0.01)
Agriculture	6	41.3 (3.6)	40.8 (2.1)	37.8 (2.1)	32.0 (2.4)
NDVI phase 2	7	4.40 (0.06)	3.95 (0.06)	3.78 (0.07)	3.15 (0.12)
CCD phase 1	7.1	4.46 (0.08)	5.07 (0.08)	5.71 (0.08)	5.96 (0.06)
CCD amp 2	7.2	120.8 (3.83)	77.5 (3.26)	80.4 (3.01)	65.4 (3.40)
CCD phase 3	7.5	2.14 (0.07)	2.35 (0.09)	1.98 (0.15)	1.65 (0.17)
CCD av	7.9	99.8 (9.85)	78.6 (1.12)	77.8 (1.00)	71.2 (1.33)
NDVI min	8	0.191(0.008)	0.102 (0.006)	0.098 (0.004)	0.081 (0.005)
Ch3 phase 3	8.4	2.11 (0.08)	2.03 (0.07)	2.39 (0.09)	2.27 (0.08)
Rainfall	9.8	1250 (31)	1210 (16)	1240 (12)	1200 (19)
CCD phase 2	10.3	3.41(0.07)	2.99 (0.05)	2.88 (0.04)	2.94 (0.03)
Dry months	11.2	2.3 (0.1)	3.3 (0.1)	3.8 (0.1)	4.1 (0.1)

Av. rank, average rank of selected variables. See table 1 for the definition of the different variables.

Table 4. Predictions for *Glossina p. palpalis*.

Analysis option		# var.	Hit ratio (%)						Major error (%)		*Kappa*
			Absent	Low	Medium	High	LMH	Total	+	−	
Predictions within fly limits	$S1_{75}$	15	–	41	48	79	58	–	6	6	0.359
	$S1_{66}$	21	–	57	62	68	62	–	2	1	0.421
	$S1_{50}$	17	–	70	61	56	62	–	3	2	0.435
	$S1_{33}$	3	–	46	40	80	56	–	6	2	0.334
	Average values	–	**54**	**53**	**71**	**60**	–	**4.2**	**2.6**		
Present/absent	$S1_{75}$	7	93	–	–	–	97	96	–	–	0.873
	$S1_{66}$	5	92	–	–	–	95	95	–	–	0.771
	$S1_{50}$	4	87	–	–	–	98	96	–	–	0.847
	$S1_{33}$	4	86	–	–	–	94	93	–	–	0.742
	Average values		**90**	–	–	–	**96**	**95**	–	–	

$S1_Y$, chosen sample method (cf. fig. 1) where 1 refers to sample option 1 and Y to the percentage grids included in the training set; # var., the number of predictor variables best predicting the predicted set; hit ratio (%), percentage of predicted grid-squares corresponding with observed results; LMH, combined results for the three presence classes (low–medium and high); major err. (%), errors of more than one class; +, overestimation; −, underestimation; *Kappa*, *Kappa* coefficient of agreement.

the second data set. Also here rain-related CCD predictor variables were predominant and ground measured data contributed significantly to the predictions.

Comparison with predictions based on data from Ivory Coast and Burkina Faso

The dataset used in this paper offered a unique opportunity to compare obtained results with previously made predictions of fly presence in Togo using training data from other West African countries. This could help us gain insight into how training sets have to be selected in order to obtain improved results. Figure 8 was redrawn from the prediction results published earlier by Rogers *et al*. (1996), using historical, and thus outdated, fly distributions from Ivory Coast and Burkina Faso as training set data to predict the fly distribution in Togo, applying a similar set of remotely sensed predictor variables as described here.

When these tentative predictions were compared with the observed fly distributions in Togo, it showed that generally there was no good match between the two. Though some eco-geographical logic patterns appeared when depicting fly-presence areas, the satellite-derived environmental correlates of fly distribution failed to fine-tune actual distribution limits in unknown territories.

Discussion

From a user's point of view, the usefulness of the results described here are likely to vary with the amount and type of errors produced. The level of accuracy, in turn, is a function of the amount of resources allocated. Pest

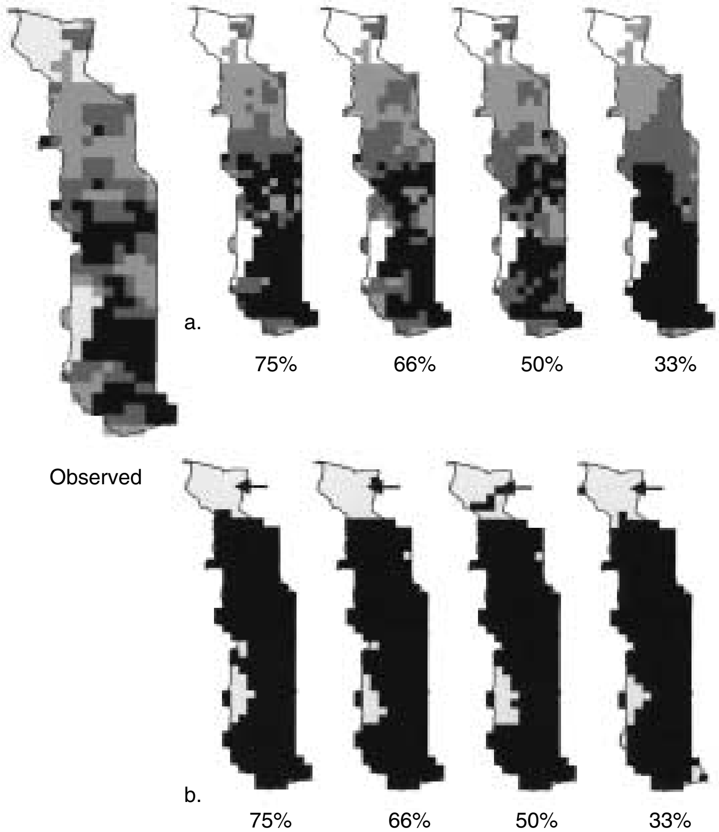

Fig. 4. Maps showing abundance and distribution predictions for *Glossina p. palpalis* as compared to observed values. Results are shown for subsequent training sets including respectively 75%, 66%, 50% and 33% of the grid cells: (a) predictions of abundance within fly limits (□, absent; □, low; ■, medium; ■, high); (b) predictions of fly distribution limits. See text and tables 4 and 5 for more details. Arrow indicates false absence prediction at northern fly limit. The upper left map shows the observed abundance.

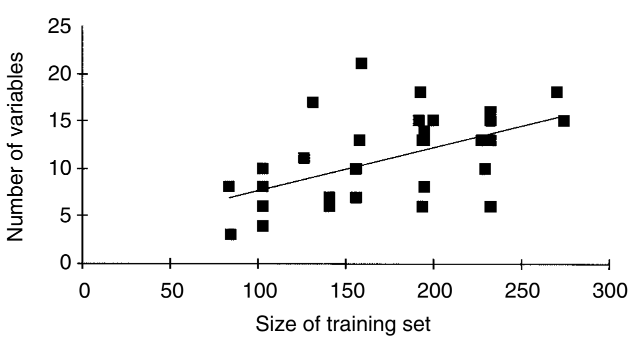

Fig. 5. Relationship between the size of the training set (X) and the number of variables needed for best prediction of the predicted set (Y): r = 0.527, n = 32, $P < 0.005$.

management economics will in practice tell us how much accuracy and detail is required; an over-estimation of the tsetse problem might induce unnecessary costs whilst underestimating it might result in priority areas not being identified.

Where abundance data were available, as is the case of *G. tachinoides* and *G. p. palpalis*, the least errors were obtained when first predicting fly presence/absence, with a two class model, and next the abundance within the fly distribution limits established. Nevertheless, no significant statistical difference could be shown between the used models and therefore pest management economics may guide sample design. It should be more economic to concentrate sampling in a first phase at the expected fly limits in order to move on to define fly abundance in the expected/identified fly presence areas in a second phase.

When only distribution data were available, i.e. for *G. m. submorsitans* and *G. longipalpis*, both types of training set data yielded comparable overall results except for the prediction of fly presence which improved with the second type of training set. This should be of no surprise because these preferential areas, as they were defined, *de facto* provided for a superior training set.

In previous studies (Hendrickx *et al.*, 1993, 1996; Rogers *et al.*, 1994, 1996) the number of predictor variables was arbitrarily set to 10. Here it is conclusively shown that the size of the training set influences the optimal number of variables; more training data generally allow incorporation of a higher number of variables (fig. 5), so as to achieve optimal prediction results, but as fig. 3 has shown, there is clearly a limit beyond which the incorporation of additional

Table 5. Most frequently selected variables for predicting *Glossina p. palpalis*.

Variable	Av. rank	Average values (Standard Errors)			
		Absent	Low	Medium	High
Ch3 phase 2	2.3	3.96 (0.18)	3.27 (0.16)	2.58 (0.14)	1.92 (0.06)
CCD max	2.8	362.2 (10.33)	328.6 (7.31)	346.5 (5.00)	382.3 (3.87)
Price amp 2	3.5	3.28 (0.23)	1.81 (0.09)	1.51 (0.09)	1.20 (0.04)
CCD amp 3	3.7	79.2 (1.41)	62.5 (1.24)	64.8 (1.23)	66.7 (1.30)
NDVI phase 1	4.2	7.99 (0.08)	7.39 (0.04)	7.26 (0.03)	7.10 (0.03)
CCD phase 3	4.2	0.91 (0.16)	1.77 (0.15)	2.47 (0.12)	2.40 (0.06)
CCD av	4.9	69.2 (2.95)	83.6 (6.07)	79.0 (1.06)	82.7 (0.73)
NDVI phase 2	5.5	2.68 (0.15)	3.72 (0.09)	3.95 (0.07)	4.18 (0.04)
CCD phase1	6.5	5.70 (0.11)	5.70 (0.10)	5.50 (0.09)	4.83 (0.07)
Price min	6.7	309.3 (0.58)	306.9 (0.19)	306.2 (0.18)	305.6 (0.14)
g-DEM	6.9	300.0 (24.9)	244.7 (13.7)	253.9 (13.8)	195.5 (13.6)
Ch3 amp 1	7.2	6.41 (0.16)	8.00 (0.15)	7.71 (0.15)	6.92 (0.14)
CCD amp 1	8.3	216.1 (1.79)	214.0 (2.74)	215.6 (2.44)	203.0 (2.35)
CCD phase 2	9.4	3.12 (0.02)	2.90 (0.04)	2.94 (0.05)	3.14 (0.06)
CCD amp 2	10.3	56.0 (6.18)	78.6 (3.39)	85.0 (3.39)	96.9 (2.84)

Av. rank, average rank of selected variables. See table 1 for the definition of the different variables.

Table 6. Predictions for *Glossina m. submorsitans*.

# var.	TS	Hit ratio (%)			*Kappa*
		Absent	Present	Total	
9	$S1_{75}$	96	43	91	0.413
6	$S1_{66}$	88	85	88	0.544
7	$S1_{50}$	98	37	91	0.456
3	$S1_{33}$	85	64	83	0.361
Average values		**92**	**57**	**88**	
5	$S2_{100}$	–	–	–	–
4	$S2_{75}$	94	60	90	0.543
2	$S2_{66}$	89	73	88	0.468
4	$S2_{50}$	86	79	86	0.409
Average values		**90**	**71**	**88**	

TS, training set; SX_Y, chosen sample method (cf. fig. 1) where X refers to sample option 1 or 2 and Y to the percentage grids included in the training set; # var., the number of predictor variables best predicting the predicted set; hit ratio (%), percentage of predicted grid-squares corresponding with observed results; *Kappa*, *Kappa* coefficient of agreement.

variables may be counter-productive. In any case, the rule suggested by Hair *et al*. (1995) of one predictor variable per minimum 20 grid-squares does not apply here; best predictions were obtained with on average 13.2 to 17 grid-squares per predictor variable, for a total sample of respectively 100 to 250 grid-squares. With an increased number of predictors an 'over-fit' of the discriminant function, i.e. bias towards the training set, may readily appear. Thus, caution is required when deciding on the number of variables to include, especially when smaller amounts of training set data are involved, as is the case in this study. In the case of a two class model, the required number of variables mainly depends on how homogeneous the distribution pattern is.

The restricted number of 311 grid-squares in the data base did not allow the testing of training sets containing less than 33% of the data because the average class size then became too small (see Lark, 1994). It is believed, however, that with larger data bases proportionally much smaller proportions of training sets might be sufficient to reproduce the original set and to make a prediction provided that enough variability is preserved per class, to avoid an over-fit of the discriminant function. In fact, it has been shown that in two class models for the prediction of vector distribution over large areas, as little as 1–5% of the training set data still generated satisfactory results (e.g. Rogers & Randolph, 1993; Robinson *et al*., 1997). A similar trend has been suggested experimentally by Lark (1994) using an artificially built 10,000 grid-square data base divided into four classes, where predictions did not tend to improve significantly using a stratified randomly selected training set of 25 grid-squares per class or a total set of one percent of the original data.

The poor results obtained with the prediction of fly distributions in Togo on the basis of data from Ivory Coast and Burkina Faso may be ascribed to two main reasons (Rogers *et al*., 1996): (i) problems with the data itself, in that no contemporary data sets were used; this applies to compatibility between the training set data, the satellite imagery and the observed fly distributions in Togo; a mismatch may produce faulty results; and (ii) species truly adapt to local conditions making fly ecology location specific. It is also possible of course that both (i) and (ii)

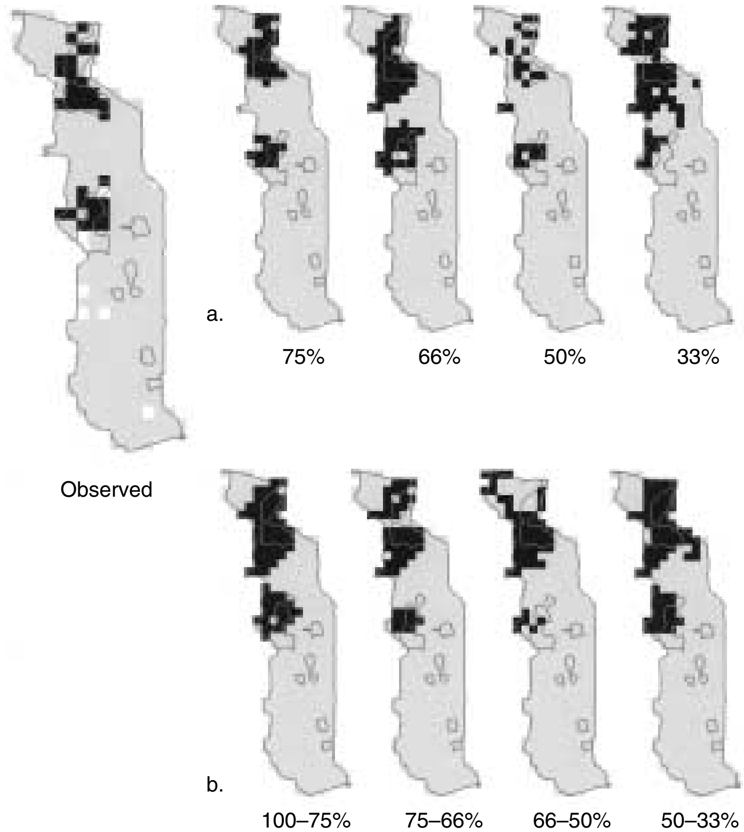

Fig. 6. Maps showing predictions of the distribution of *Glossina m. submorsitans* using (a) sample 1 and (b) sample 2 (■, absent; ■, present). Outlines depict protected areas.

Table 7. Best predictions for *Glossina longipalpis*.

# var.	TS	Hit ratio (%)			*Kappa*
		Absent	Present	Total	
9	$S1_{75}$	89	80	87	0.625
6	$S1_{66}$	82	100	85	0.594
7	$S1_{50}$	90	65	86	0.506
3	$S1_{33}$	82	51	77	0.291
Average values		**86**	**74**	**84**	
5	$S2_{100}$	–	–	–	–
4	$S2_{75}$	86	89	86	0.521
2	$S2_{66}$	85	90	85	0.445
4	$S2_{50}$	86	90	86	0.374
Average values		**86**	**90**	**86**	

TS, training set; SX_Y, chosen sample method (cf. fig. 1) where X refers to sample option 1 or 2 and Y to the percentage grids included in the training set; # var., the number of predictor variables best predicting the predicted set; hit ratio (%), percentage of predicted grid-squares corresponding with observed results; *Kappa*, *Kappa* coefficient of agreement.

above apply. Several observations are in favour of the first argument. Fly distribution data collated at different points in time (Mawuena & Itard, 1981; Clair, 1987) suggest a southwards expansion of the distribution range of *G. tachinoides* in Ivory Coast and Togo. More recently Hendrickx *et al*. (1999a) observed a change in the northern distribution range of *G. palpalis* in Togo. The fact that these fly distributions are dynamic may possibly relate to a more general climate shift in coastal West Africa as a whole; the satellite derived environmental correlates may be applied to predict these dynamics provided that data are available over a wide enough area.

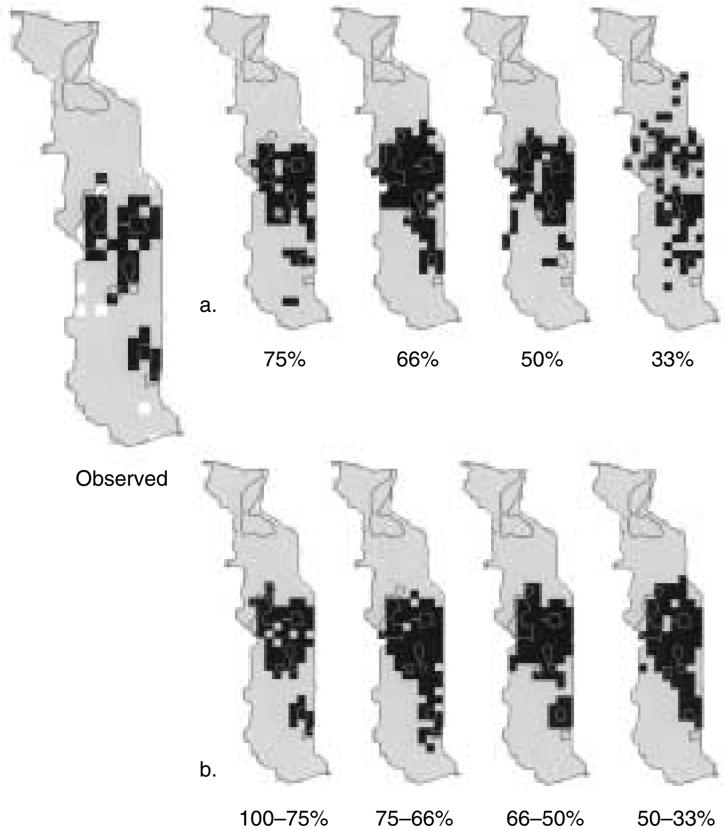

Fig. 7. Maps showing predictions of the distribution of *Glossina longipalpis* using (a) sample 1 and (b) sample 2 (■, absent; ■, present). Outlines depict protected areas.

Factors such as those described above which locally influence distribution patterns will also affect fly abundance. *Glossina tachinoides* and *G. p. palpalis* in Togo show locally higher densities at the fringe of their respective distribution limits (Hendrickx *et al*., 1999a). Also, it was shown that agriculture has much local influence on the distribution and abundance of *G. tachinoides*, *G. p. palpalis*, *G. m. submorsitans* and *G. longipalpis*. Such changes in vegetation patterns might not always show up in medium resolution imagery. In addition, the relatively coarse resolution of 0.125 degree pixels, used in this study, contributed to further blur the picture. However (e.g. Rogers *et al*., 1997), NOAA data may assist in the demarcation of land cover patterns suggesting that higher resolution data (1 km pixels), such as becoming now more widely available, should contribute to improve results.

This further stresses the fact that particular care should be taken with regard to the selection of the training set which should not only be contemporary to the predictor variables but also reflect as much eco-geographical variability as affordable in order to be representative for the entire area covered. This will become increasingly complex when higher resolution data sets will be involved. In practice, sampling within a target area (e.g. identified fly presence area) should be stratified based on a set of known land cover categories.

In conclusion, the results presented here confirm the potential for application of satellite imagery and multivariate analysis to predict tsetse distribution and abundance, and thus to integrate field and satellite data. To obtain optimal results, field surveys may be conducted in two stages, first concentrating on expected fly limits and thereafter on fly abundance, within the limits first established, adopting a split sample approach. Special care should be taken in selecting the optimal number of predictor variables. Ideally, training data should be contemporary to the predictor variable set and also originate from the same area as the predicted set.

Acknowledgements

Normalized Difference Vegetation Index and Cold Cloud Duration data were kindly provided by the Africa Real Time Environmental Monitoring Information System Project (ARTEMIS, FAO, Rome). Channel 3, 4 and 5 were provided by the Global Inventory Monitoring and Modelling Systems group (GIMMS, NASA). All satellite data transformation and Fourier analysis were performed by the Trypanosomiasis and Land use in Africa (TALA) research group, Zoology Department, Oxford, UK. Computer programs were developed by D.J. Rogers. The contribution of Niko Speybroeck, Tropical Institute, Antwerp, Belgium, is acknowledged for assistance with testing statistical significance between the models used. All field activities were conducted within the framework of the Belgian funded FAO project, GCP-TOG-013-BEL, of the Ministère Togolais de l'Agriculture, de l'Elevage et des Pêches. The authors are particularly grateful to K. Adoméfa, M. Denke, N.

	a. *G. tachinoides*	b. *G. palpalis*	c. *G. m. submorsitans*	d. *G. longipalpis*
Hit ratio	63%	81%	48%	82%
Absence	98%	15%	46%	95%
Presence	55%	93%	68%	2%
Kappa	0.291	0.100	0.047	–0.03
z	4.56	(0.86)	(0.87)	(–0.28)

Fig. 8. Maps showing predictions for Togo of the distribution of *Glossina tachinoides*, *G. palpalis*, *G. m. submorsitans* and *G. longipalpis* by Rogers *et al*. (1996) using historical distribution limits in Ivory Coast and Burkina Faso (■, absent; ■, present). A comparison is made (hit ratios) with present distribution limits (Hendrickx *et al*., 1999a).

Doumandji, J. Mortelmans and J. Vercruysse for their permanent help and support. This publication arises in part from funding provided by the Department for International Development (DfID), UK (Livestock Production Programme ZC0012, to DJR). However, the DfID can accept no responsibility for information provided or views expressed.

References

Baldry, D.A.T. & Molyneux, D.H. (1980) Observations on the ecology and trypanosome infections of a relict population of *Glossina medicorum* Austen in the Komoe Valley of Upper Volta. *Annals of Tropical Medicine and Parasitology* **74**, 79–91.

Clair, M. (1987) Données récentes sur la répartition des Glossines au Niger, au Burkina Faso et en Côte d'Ivoire. pp. 345–350 in *International Scientific Council for Trypanosomiasis Research and Control (ISCTRC), Nineteenth Meeting, Lomé, Togo, 1987, publ. 1989*. Organisation of African Unity – Scientific and Technical Research Commission (OUA-STRC), Nairobi, Kenya.

Cohen, J. (1960) A coefficient of agreement for nominal scale. *Educational and Psychological Measurement* **20**, 37–46.

Congalton, R.G. (1991) A review of assessing the accuracy of classification of remotely sensed data. *Remote Sensing of Environment* **37**, 35–46.

de La Rocque, S. (1977) *Identification des facteurs discriminants majeurs de la présence des glossines dans une zone agro-pastorale du Burkina Faso. Intérêt pour la prévision du risque trypanosomien*. 212 pp. Thèse de doctorat, Université de Montpellier II Sciences et techniques du Languedoc, France.

De Wispelaere, G. (1994) Contribution of satellite remote sensing to the mapping of land use and of potential *Glossina* biotopes. In: *A systematic approach to tsetse and trypanosomosis control. Proceedings of the FAO panels of experts, Rome, Italy, 1–3 December 1993. FAO Animal Production and Health Paper* **121**, 74–89.

Ford, J. & Katondo, K.M. (1977) The distribution of tsetse flies in Africa in 1973. Organisation of African Unity – Scientific and Technical Research commission (OUA–STRC). London, Cook, Hammond and Kell.

Hair, J.F., Anderson, R.E., Tatham, R.L. & Black, W.C. (1995) *Multivariate data analysis with readings*. New Jersey, Prentice-Hall Inc.

Hendrickx, G. (1999) *Georeferenced decision support methodology towards trypanosomosis management in West Africa*. 176 pp. PhD thesis, State University of Ghent, Belgium.

Hendrickx, G., Rogers, D.J., Napala, A. & Slingenbergh, J.H.W. (1993) Predicting the distribution of riverine tsetse and the prevalence of bovine trypanosomosis in Togo using ground-based and satellite data. pp. 218–232 in *International Scientific Council for Trypanosomiasis Research and Control (ISCTRC), Twenty Second Meeting, Kampala, Uganda, 1993, publ. 1995*. Organisation of African Unity – Scientific and Technical Research Commission (OUA–STRC), Nairobi, Kenya.

Hendrickx, G., Napala, A., Rogers, D.J. & Slingenbergh, J.H.W. (1996) Contribution des satellites à la lutte contre la trypanosomose animale africaine. pp. 441–453 *in* Demarée, G., Alexandre, J. & De Dapper, M. (*Eds*) *Proceedings of the International Conference on Tropical Climatology, Meteorology and Hydrology, Brussels, Belgium, May 22–24, 1996. publ. 1998*.

Hendrickx, G., Slingenbergh, J., Dao, B., Bastiaensen, P. & Napala, A. (1997) Systèmes d'Information Géographique (SIG), outils puissants de prise de décision. pp. ???–??? in *International Scientific Council for Trypanosomiasis Research and Control (ISCTRC), Twenty Fourth Meeting, Maputo, Mozambique, 1997*. Organisation of African Unity – Scientific and Technical Research Commission (OUA–STRC), Nairobi, Kenya.

Hendrickx, G., Napala, A., Dao, B, Batawui, D., De Deken, R., Vermeylen, A. & Slingenbergh, J.H.W. (1999a) A systematic approach to area-wide tsetse distribution and abundance maps. *Bulletin of Entomological Research* **89**, 231–244.

Hendrickx, G., Napala, A., Dao, B., Batawui, D., Bastiaensen, P., De Deken, R., Vermeylen, A., Vercruysse, J. & Slingenbergh, J.H.W. (1999b) The area-wide epidemiology of bovine trypanosomosis and its impact on mixed farming in subhumid West Africa; a case study in Togo. *Veterinary Parasitology* **84**, 13–31.

Hendrickx, G., Napala, A., Slingenbergh, J.H.W., De Deken, R., Vercruysse, J. & Rogers, D.J. (2000) The spatial patterns of trypanosomosis predicted with the aid of satellite imagery. *Parasitology* **120**, 121–134.

Hendrickx, G., Napala, A., Dao, B., De Deken, R., Bastiaensen, P., Vercruysse, J. & Slingenbergh, J.H.W. (2001) Trypanosomosis control in West Africa: why, where and how? A case study in Togo. *Acta Tropica*, in press.

Kitron, U., Otieno, L.H., Hungerford, L.L., Odulaja, A., Brigham, W.U., Okello, O.O., Joselyn, M., Mohamed-Ahmed, M.M. & Cook, E. (1996) Spatial analysis of the distribution of tsetse flies in the Lambwe Valley, Kenya, using Landsat TM satellite imagery and GIS. *Journal of Animal Ecology* **65**, 371–380.

Landis, J.R. & Koch, G.C. (1977) The measurement of observer agreement for categorical data. *Biometrics* **33**, 159–174.

Lark, R.M. (1994) Sample size and class variability in the choice of a method of discriminant analysis. *International Journal of Remote Sensing* **15**, 1551–1555.

Laveissière, C. & Challier, A. (1977) La répartition des glossines en Haute Volta, Cartes à 1/2.000.000. Notice Explicative N° 69. Paris: Office de la Recherche Scientifique et Technique d'Outre Mer.

Laveissière, C. & Challier, A. (1981) La répartition des glossines en Côte d'Ivoire, Cartes à 1/2.000.000. Notice Explicative N° 89. Paris: Office de la Recherche Scientifique et Technique d'Outre Mer.

Mawuena, K. & Itard, J. (1981) Présence de *Glossina tachinoides* Westwood, 1850 (Diptera: Glossinidae) dans le Sud du Togo. *Revue d'Elevage et de Médecine Vétérinaire des Pays Tropicaux* **34**, 47–53.

Price, J.C. (1984) Land surface temperature measurement for the split window channels of the NOAA 7 advanced very high resolution radiometer. *Journal of Geophysical Research* **89**, 7231–7237.

Robinson, T., Rogers, D.J. & Williams, B. (1997a) Univariate analysis of tsetse habitat in the common fly belt of Southern Africa using climate and remotely sensed vegetation data. *Medical and Veterinary Entomology* **11**, 223–234.

Robinson, T., Rogers, D.J. & Williams, B. (1997b) Mapping tsetse habitat suitability in the common fly belt of Southern Africa using multivariate analysis of climate and remotely sensed vegetation data. *Medical and Veterinary Entomology* **11**, 235–245.

Rogers, D.J. & Randolph, S.E. (1993) Distribution of tsetse and ticks in Africa, past, present and future. *Parasitology Today* **9**, 266–271.

Rogers, D.J., Hendrickx, G. & Slingenbergh, J.H.W. (1994) Tsetse flies and their control. *Revue Scientifique et Technique de l'Office International des Epizooties* **13**, 1075–1124.

Rogers, D.J., Hay, S.I. & Packer, M.J. (1996) Predicting the distribution of tsetse flies in West Africa using temporal, Fourier processed meteorological satellite data. *Annals of Tropical Medicine and Parasitology* **90**, 225–241.

Rogers, D.J., Hay, S.I., Wint, G.R.W. & Packer, M.J. (1997) Mapping land-cover over large areas using public domain meteorological satellite: a case study for Nigeria. *International Journal of Remote Sensing* **18**, 3297–3303.

Snijders, F.L. (1991) Rainfall monitoring based on Meteosat data – a comparison of techniques applied to the Western Sahel. *International Journal of Remote Sensing* **12**, 1331–1347.

Tucker, C.J., Vanpraet, C., Boerwinkel, E. & Gaston, A. (1983) Satellite remote sensing of total dry matter production in the Senegalese Sahel. *Remote Sensing and Environment* **13**, 461–474.

(Accepted 22 March 2001)