Annals of Tropical Medicine and Parasitology, Vol. 90, No. 3, 225–241 (1996)

REVIEW

Predicting the distribution of tsetse flies in West Africa using temporal Fourier processed meteorological satellite data

BY D. J. ROGERS, S. I. HAY AND M. J. PACKER

Trypanosomiasis and Land Use in Africa (TALA) Research Group, Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, U.K.

Received and accepted 10 April 1996

An example is given of the application of remotely-sensed, satellite data to the problems of predicting the distribution and abundance of tsetse flies in West Africa. The distributions of eight species of tsetse, *Glossina morsitans*, *G. longipalpis*, *G. palpalis*, *G. tachinoides*, *G. pallicera*, *G. fusca*, *G. nigrofusca* and *G. medicorum* in Côte d'Ivoire and Burkina Faso, were analysed using discriminant analysis applied to temporal Fourier-processed surrogates for vegetation, temperature and rainfall derived from meteorological satellites. The vegetation and temperature surrogates were the normalized difference vegetation index and channel-4-brightness temperature, respectively, from the advanced, very-high-resolution radiometers on board the National Oceanic and Atmospheric Administration's polar-orbiting, meteorological satellites. For rainfall the surrogate was the Cold-Cloud-Duration (CCD) index derived from the geostationary, Meteosat satellite series. The presence or absence of tsetse was predicted with accuracies ranging from 67%–100% (mean=82·3%). A further data-set, for the abundance of five tsetse species across the northern part of Côte d'Ivoire (an area of about 140 000 km^2), was analysed in the same way, and fly-abundance categories predicted with accuracies of 30%–100% (mean=73·0%). The thermal data appeared to be the most useful of the predictor variables, followed by vegetation and rainfall indices. Refinements of the analytical technique and the problems of extending the predictions through space and time are discussed.

The sensitivity to climate of arthropods in general, and insect vectors in particular, has already been stressed in relation to the transmission of vector-borne disease transmission. Hay *et al.* (1996) explained how remote-sensing, satellite platforms can provide data that are suitable surrogates for the traditional meteorological data that, in the past, have been correlated with both vector abundance and vector-mortality rates. They also explained the steps in image processing that lead to a variety of vegetation, thermal and rainfall indices, with emphasis on those platforms such as the National Oceanic and Atmospheric Administration's (NOAA), polar-orbiting meteorological satellites and Meteosat satellites which provide frequent coverages from which relatively cloud-free views of the Earth's surface, or of cloud-top temperatures, can be produced.

Such multi-temporal data may be used to give a realistic picture of average monthly and annual values of vegetation and climate and have recently been used to describe the distributions of several species of tsetse fly: *Glossina morsitans* Westwood in Zimbabwe (Rogers and Williams, 1993); *G. morsitans* and *G. pallidipes* Austen in Kenya and Tanzania (Rogers and Randolph, 1993); and *G. palpalis palpalis* (Robineau-Desvoidy), *G. tachinoides* Westwood, *G. morsitans submorsitans*

0003–4983/96/030225+17 $12.00/0

Newstead, *G. longipalpis* Wiedemann and *G. fusca* Walker in Togo (Rogers *et al.*, 1994). These studies extended previous research in which tsetse-fly mortality rates (Rogers and Randolph, 1986, 1991), distribution (Gaschen, 1945), abundance (Fairbairn and Culwick, 1950) and infection rates (Ford and Leggate, 1961) and prevalence of human sleeping sickness (Rogers and Williams, 1993) were related to ground-based measures of climate based on synoptic or contemporary meteorological records.

The problems of processing large amounts of satellite data have led to the development of a variety of data-reduction methods. In the case of multi-temporal data, principal-components analysis of monthly, normalized difference vegetation indices (NDVI), derived from the advanced, very-high-resolution radiometers (AVHRR) on board the NOAA series of satellites, usually gives a first component obviously correlated with the mean vegetation index for the year. The second and third components are related to seasonality, which is especially pronounced in the savannah regions of Africa (Townshend and Justice, 1986; Eastman and Fulk, 1993). An alternative approach to data reduction, using temporal Fourier processing, gives results that have recently been related to regional- and continental-scale, biological processes (Andres *et al.*, 1994; Olsson and Eklundh, 1994; Rogers and Williams, 1994; Verhoef *et al.*, 1996). When Fourier analysis was applied to monthly NDVI data for Africa, it was found that the annual, bi-annual and tri-annual cycles (called 'components' in the analysis) explained a large part of the variability of the annual NDVI signal. Features of the Fourier analysis of the whole-Africa NDVI were related both to ecological patterns, such as the savannah regions of Africa or the Gezira irrigation scheme in southern Sudan, and to ecological processes, such as the seasonal growth of vegetation along the River Nile (Rogers and Williams, 1994). The present review describes the application of these techniques to the description of the distribution and abundance of eight species of tsetse in Côte d'Ivoire and Burkina Faso, West Africa.

MATERIALS AND METHODS

Tsetse-fly Distributions

The distributions of tsetse in Côte d'Ivoire and Burkina Faso were taken from maps published by the Office de la Recherche Scientific et Technique d'Outre-Mer (ORSTOM) (Laveissiere and Challier, 1977, 1981). These maps are compendia of information gathered over the preceding decades and record species' presence at a spatial resolution of 0·167°. The original data sources do not give complete spatial coverage, and the maps do not necessarily record fly presence in areas where a species was thought to be ubiquitous by the compiling authors (e.g. *G. palpalis* in the southern part of Côte d'Ivoire). Thus, whilst records of fly presence on these maps are historically accurate, records of absence are occasionally misleading. The distributions of eight species of tsetse were used in the present analysis: *G. morsitans submorsitans*; *G. longipalpis*; *G. palpalis* s.l.; *G. tachinoides*; *G. pallicera* Bigot; *G. fusca*; *G. nigrofusca* Newstead; and *G. medicorum* Austen.

Tsetse-fly Abundance

The abundance of flies in the northern part of Côte d'Ivoire was monitored by a joint Food and Agriculture Organization/German Technical Assistance (FAO/GTZ) project that ran from 1979 to 1980 and produced detailed maps of fly distributions at scales of 1:1 000 000 and 1:200 000 (Anon., 1982). The data at a spatial resolution of 0·250° are used in the present analysis. Flies were sampled using Challier/Laveissiere traps (Challier and Laveissiere, 1973) placed in suitable habitats by the survey teams and left for short periods before collection and removal. Given the very large area sampled and the short sampling time in each habitat, these data are likely to be affected by a number of confounding effects such as sampling errors, poor weather at the time of sampling and seasonality, so only the mean values (flies/trap per nominal 6-h trapping session) were analysed.

Some of the species present in the region were inadequately sampled by the traps used in the surveys, and supplementary catches using

hand nets were recorded separately on the maps. The coverage using this method was relatively poor, however, and these results have not been included in the present analysis, which is based entirely upon the trap catches of *G. morsitans*, *G. longipalpis*, *G. palpalis*, *G. tachinoides* and *G. fusca*.

Satellite Data

NDVIs are derived from readings in channels 1 and 2 (Ch_1 and Ch_2, respectively) of the AVHRR on board the NOAA series of meteorological satellites, being calculated as $(Ch_2 - Ch_1)/(Ch_2 + Ch_1)$.

Prince *et al.* (1990) have described the application of NDVI data to a range of biological problems and Hay *et al.* (1996) not only described their application to arthropod vectors of disease specifically but also reviewed alternative vegetation indices. 1982–90 ten-day 'dekadal' maximum-value-composite (MVC) NDVI data (Holben, 1986) were obtained from the Food and Agriculture Organization's (FAO) African Real Time Environmental Monitoring using Meteorological Satellites (ARTEMIS) program at 7·6 × 7·6 km resolution. The registration of these images was checked against a geo-referenced 'master' image, and corrections made where necessary. This involved shifting images by 0–3 pixels in an east–west or north–south direction, depending on the scene. The raw imagery was then corrected for satellite-sensor drift in channel 1 using calibration coefficients derived by Los (1993), and then maximum-value composited by selecting the highest value of the dekadal pixels for each site within each month, to produce a set of monthly images for further analysis.

AVHRR-channel-4-brightness temperature correlates with air temperature at the Earth's surface (Hay *et al.*, 1996). Dekadal data at 7·6 × 7·6 km spatial resolution from the archives of the Global Inventory Monitoring and Modelling Systems (GIMMS) group at the NASA Goddard Space Flight Center were maximum-value composited for the period 1987–1992. Monthly imagery was later produced, again by MVC, and used in the present analysis.

Cold-cloud-duration (CCD) imagery was obtained from the FAO-ARTEMIS program as 5-year, monthly means for the period 1988–1992. The CCD imagery has been correlated with surface rainfall measurements as part of the Tropical Applications in Meteorology of Satellite and other data (TAMSAT) program within the area covered by the present tsetse surveys (Snijders, 1991).

Digital-elevation-model (DEM) data were obtained from a 0·083°-resolution elevation surface for Africa, produced by the Global Land Information System (GLIS) of the United States Geological Survey, Earth Resources Observation Systems (USGS, EROS) data centre. The original files were resampled to a 7·6 × 7·6 km resolution image to ensure compatibility with the other data layers.

Satellite-data Processing and Data Reduction

Multi-temporal satellite data produce multivariate data-sets for each unit area (pixel) within an image. Each of the 12, monthly MVC of each of the image types forms a single axis in a multi-variate space defining the environment of the vector. Many of these axes are strongly correlated with each other because, for example, a pixel with a high NDVI in one month is likely to have a high NDVI in other months. This indicates that data reduction (i.e. ordination) could be achieved, without loss of information, by replacing the raw imagery with some combined signal derived from these highly correlated values.

The simplest combination is obviously the arithmetic mean, and seasonal variability may be captured by the variance or standard deviation of the mean. More complex ordination techniques generally involve projecting the data onto a rotated (usually orthogonal) set of axes (called 'principal components') such that the first new axis captures the largest proportion of data variance, the second captures the largest proportion of the remaining variance, and so on. Principal-components analysis (PCA) retains the same number of axes as the original data-set, but the sequential partitioning of the variance often means that many of

the axes in principal-component space are effectively redundant because they explain only a very small proportion of the variance in the original data-set (Green, 1978). Projection of the original data-set onto the principal component axes involves applying a series of coefficients or weights to the raw data, effectively to achieve the desired axis rotation (the weights are the cosines of the angles between the original and rotated co-ordinate axes). Data values in the original co-ordinate system all contribute (via their weighted values) to each principal-component-axis score. Thus, for example, a series of 12, monthly images, from January–December, may be subjected to PCA and every month would then contribute to each of the 12 resulting principal-component axes. If, however, only the first few principal-component axes explain the great majority of the variance in the original data, only these need be used in further analysis. PCA is not independent of the scale of the original axes and it is generally necessary to standardize (or transform) the raw variables to roughly similar variances before analysis, or else to use the correlation matrix in the analysis (Marriott, 1974). This makes it difficult to extend the results of PCA to other times and places because principal-component-axis rotation is uniquely determined by the original set of observational data (the 'training set').

An entirely different approach to the same problem of data reduction was suggested by the literature on time-series analysis (e.g. Chatfield, 1980). The time-series x_t may be described by a Fourier series representation where

$$x_t = a_0 + \sum_{p=1}^{N/2-1} [a_p \cos(2\pi pt/N) + b_p \sin(2\pi pt/N)] + a_{N/2}\cos\pi t \qquad (1)$$

$t = 1, 2, \ldots N$, and the coefficients a_p and b_p are defined as follows:

$$\begin{aligned} a_0 &= \bar{x} \\ a_{N/2} &= \Sigma(-1)^t x_t/N \\ a_p &= 2[\Sigma x_t \cos(2\pi pt/N)]/N \\ b_p &= 2[\Sigma x_t \sin(2\pi pt/N)]/N \end{aligned} \qquad (2)$$

where $p = 1, \ldots [(N/2) - 1]$. The component at a frequency $\omega p = 2\pi p/N$ is called the pth harmonic and, for all $P \neq N/2$, these harmonics may be written in the equivalent form

$$a_p \cos\omega_p t + b_p \sin\omega_p t = R_p \cos(\omega_p t + \Phi_p) \qquad 3)$$

where R_p is the amplitude of the pth harmonic

$$= \sqrt{(a_p^2 + b_p^2)}$$

and Φp is the phase of the pth harmonic (Chatfield, 1980)

$$= \tan^{-1}(-b_p/a_p).$$

The effect of Fourier analysis is to partition the variability of the time-series into (orthogonal and thus uncorrelated) components at frequencies of $2\pi/N$, $4\pi/N$, $6\pi/N \ldots$, π, or periods equal to 1, 1/2, 1/3, ... $2/N$ times the duration of the observations, N. If monthly observations are taken, Fourier analysis can partition the time-series into frequencies equivalent to periods ranging from as long as the whole time-series, down to 2 months (higher frequencies, i.e. shorter period cycles, cannot be distinguished by monthly data). Full Fourier analysis exactly describes the original data set [since the Fourier series in eqn (1) contains N parameters to describe N observations], but not all harmonics may be contributing equally to this description. The following relationship, known as Parseval's theorem, applies to the Fourier representation of x_t

$$\Sigma(x_t - \bar{x})^2/N = \sum_{p=1}^{N/2-1} R_p^2/2 + a^2_{N/2}. \qquad (4)$$

This equation states that a quantity very similar to the variance of the original observations [the left-hand side of the equation, but with the divisor N rather than $(N-1)$] is the sum of the contributions of each of the harmonics for values of p from 1 to $N/2$, where $R_p^2/2$ is the contribution of the pth harmonic.

The combination of the orthogonality of the harmonics in the Fourier-series representation

of multi-temporal satellite data and the (perhaps illusory) biological transparency of the interpretation of these harmonics makes this approach to data reduction especially attractive to biologists (Rogers and Williams, 1994). In effect, it may be possible to reduce a monthly or dekadal data stream covering 10 or more years to just seven variables (the mean of the whole series and the amplitude and phases of the first three Fourier components) without a significant loss of information.

For the present study, each of the NDVI, channel-4-brightness-temperature and CCD, monthly-image data-sets were subjected to temporal Fourier processing and the means, amplitudes and phases of the annual, bi-annual and tri-annual cycles calculated. These variables were stored as new image layers for analysis, at the same spatial resolution as the original imagery. The combined (i.e. annual plus bi-annual plus tri-annual cycle) Fourier description of the original signal was also calculated (a summation that essentially smooths the original data-set) and its minimum, maximum and range were recorded for use in the analysis. In addition, certain combinations of the Fourier-processed signals were calculated, such as the ratio of NDVI to mean values of channel-4-brightness temperatures, which has been shown to be a more stable indicator of vegetation type than either variable alone (Lambin and Ehrlich, 1995, 1996). A full list of predictor variables used in this study is given in Table 1.

All satellite imagery was further processed by selecting a block of 2 × 2 pixels positioned at the centre of each grid square on the fly-distribution, and the mean values for each square were used in further analyses.

Data Analysis

The reduced-dimension data-set produced by the methods outlined above form the set of predictor variables for describing the species' distributions and abundance. Of the methods available, such as correspondence analysis (Ter Braak, 1986; Hill, 1991), projection pursuit, nearest neighbour and neural network analysis (Williams *et al.*, 1992), this review concentrates on the use of various forms and modifications of discriminant analysis; these are relatively easy to apply and provide biological insight into the nature of the limits to the distribution and abundance of vector species. The simple problem of describing the distribution of a vector is taken to illustrate the techniques.

In its simplest form, discriminant analysis assumes a multi-variate normal distribution of the predictor variables and a common within-group co-variance of the variables for all points defining vector presence and vector absence. The mean values of the predictor variables in sites of vector presence and absence, and the within-group co-variance matrix, are estimated from representative samples from reliable distribution maps (the 'training sets'). Means of multi-variate distributions are referred to as centroids and are defined by mathematical vectors ($\bar{x}_n$) where n is the number of dimensions (i.e. variables). The Mahalanobis distance, D^2, is the distance between two multi-variate distribution centroids, or between a sample point and a centroid, and is a generalization of the traditional squared Euclidean distance, d^2:

$$d_{12}^2 = (\overline{\mathbf{x}_1} - \overline{\mathbf{x}_2})'(\overline{\mathbf{x}_1} - \overline{\mathbf{x}_2}) = \mathbf{d}'\mathbf{d}$$

and

$$D_{12}^2 = (\overline{\mathbf{x}_1} - \overline{\mathbf{x}_2})'\mathbf{C}_w^{-1}(\overline{\mathbf{x}_1} - \overline{\mathbf{x}_2}) = \mathbf{d}'\mathbf{C}_w^{-1}\mathbf{d} \quad (5)$$

where d_{12}^2 and D_{12}^2 are, respectively, the Euclidean and Mahalanobis distances between group 1 (e.g. for vector absence) and group 2 (e.g. for vector presence), d is $(\bar{x}_1 - \bar{x}_2)$, with the subscripts again referring to the two groups (or, alternatively, 1 and 2 might refer to a point and a centroid), and $\mathbf{C}_w^{-1}$ is the inverse of the within-groups co-variance (=dispersion) matrix (Green, 1978). [In eqn (5), the subscript n for the number of variables has been dropped for clarity.] Thus, the Mahalanobis distance is the distance between the sample centroids adjusted for their common co-variance. In the case of the Euclidean distance, d_{12}^2, the co-variances are zero, so that the co-variance matrix $\mathbf{C}_w$ equals $\mathbf{C}_w^{-1}$ or $\mathbf{I}$, the identity matrix (with values of 1 along the diagonal and 0 elsewhere). This reduces the equation for D_{12}^2 to that for d_{12}^2. If the

TABLE 1

Predictor variables used in the analyses of tsetse fly distributions in Côte d'Ivoire and Burkina Faso, and their observed maximum and minimum values in the training-set data

Abbreviation	*Name*	*Maximum*	*Minimum*
ALTITUDE			
Elev	Elevation (m)	829·8	0
VEGETATION			
NDmean	Normalized difference vegetation index (NDVI) mean	0·492	− 0·082
NDp1	NDVI phase 1	11·23	3·13
NDa1	NDVI amplitude 1	0·19	0·012
NDp2	NDVI phase 2	4·78*	1·6
NDa2	NDVI amplitude 2	0·083	0·009
NDp3	NDVI phase 3	3·925*	0·05
NDa3	NDVI amplitude 3	0·033	0·002
NDmax	NDVI maximum	0·597	− 0·061
NDmin	NDVI minimum	0·4	− 0·106
NDrange	NDVI range	0·408	0·04
RAINFALL			
CCDmean	Cold-cloud duration (CCD) mean (h)	111	27
CCDp1	CCD phase 1	6·9	2·65
CCDa1	CCD amplitude 1 (h)	66·75	23
CCDp2	CCD phase 2	4·2	0·5
CCDa2	CCD amplitude 2 (h)	72·25	2·75
CCDp3	CCD phase 3	3·975	0·025
CCDa3	CCD amplitude 3 (h)	27·8	1
CCDmax	CCD maximum (h)	223·6	81·9
CCDmin	CCD minimum (h)	52·2	− 30·6†
CCDrange	CCD range (h)	192·4	85·6
TEMPERATURE			
CH4mean	Thermal (AVHRR-channel-4) radiance mean (°C)	41	21
CH4p1	Thermal phase 1	2·7	0·8
CH4a1	Thermal amplitude 1 (°C)	11·5	1·15
CH4p2	Thermal phase 2	5·2	1·025
CH4a2	Thermal amplitude 2 (°C)	3·8	0·125
CH4p3	Thermal phase 3	3·8	0·225
CH4a3	Thermal amplitude 3 (°C)	1·8	0·1
CH4max	Thermal maximum (°C)	50·5	21·9
CH4min	Thermal minimum (°C)	34·8	18·6
CH4range	Thermal range (°C)	26·6	2·5
MIXED			
NDp-Cdp	NDVI phase-CCD phase	6·93	0·10
NDm/CDm	100 × NDVI mean/CCD mean	0·742	− 0·130
NDm/CH4m	100 × NDVI mean/thermal mean	2·097	− 0·369
NDp1-CH4p1	NDVI phase 1-thermal phase 1	10·025	2·175
NDa1/CH4a1	100 × NDVI amplitude 1/thermal amplitude 1	3·674	0·065

*Values are the timing of the maxima of the first of the bi-annual or tri-annual cycles.
†Negative values are possible in Fourier harmonics.

problem is to predict only to which of the groups of 'presence' or 'absence' a new point belongs, it is simply necessary to calculate the two values of D^2 between the point and each of the two centroids. The point is then assigned to the group to which it is closest in multi-variate space (i.e. the one which gives the smallest D^2). This assignment rule is obviously an over-simplification since the values of D^2 may differ by only a little, or by a very large amount. There is always a probability, however slight, that the observation in fact belongs to the group to which it was not assigned.

The 'posterior probability' replaces the simple prediction of group membership by calculating the probability with which any observation belongs to each group as follows

$$P(1|x) = \frac{p_1 e^{-D_1^2/2}}{\sum_{g=1}^{2} p_g e^{-D_g^2/2}}$$

and

$$P(2|x) = \frac{p_2 e^{-D_2^2/2}}{\sum_{g=1}^{2} p_g e^{-D_g^2/2}} \qquad (6)$$

where $P(1|x)$ is the posterior probability that observation x belongs to group 1 and $P(2|x)$ the posterior probability that it belongs to group 2 (Green, 1978). p_1 and p_2 are the prior probabilities of belonging to the two groups, defined as the probabilities with which any observation might belong to either group, given prior knowledge or experience of the situation (often, when applied, based on the training-set data). In the absence of any prior experience, it is usual to assume equal prior probability of belonging to any of the groups. Where there are only two groups, for absence and presence, p_1 and p_2 are both 0·5. Equation (6) assumes that observation x must come from either group 1 or group 2; the possibility it belongs to neither is discounted. Once again, the assumption in eqn (6) is of multi-variate normality, the other terms of the multi-variate normal equation cancelling out (Tatsuoka, 1971).

The above formulae apply only to those situations in which a common co-variance matrix can be assumed. In many cases of distribution data, however, this does not apply because animals do not live within a random subset of environmental space, but within a rather unusual subset, with specific environmental conditions which cannot be described by general environmental conditions. The result is that the co-variances of the variables within a distributional range are often different from those of the same variables outside the distributional limits. This requires a modification of eqns (5) and (6), to allow for different, within-group co-variance matrices. Equation (6) is then modified as follows

$$P(1|x) = \frac{p_1 |C_1|^{-1/2} e^{-D_1^2/2}}{\sum_{g=1}^{2} p_g |C_g|^{-1/2} e^{-D_g^2/2}}$$

and

$$P(2|x) = \frac{p_2 |C_2|^{-1/2} e^{-D_2^2/2}}{\sum_{g=1}^{2} p_g |C_g|^{-1/2} e^{-D_g^2/2}} \qquad (7)$$

where $|C_1|$ and $|C_2|$ are the determinants of the co-variance matrices for groups $g=1$ and $g=2$, respectively [the Mahalanobis distances in eqn (7), calculated from eqn (5), are now evaluated using the separate within-group co-variance matrices] (Tatsuoka, 1971). With unequal co-variance matrices, the discriminant axis (strictly speaking a plane) that separates the two groups in multi-variate space is no longer linear.

It is relatively straightforward to extend eqns (5) to (7) to situations in which more than two groups (absence/presence) are encountered. The most obvious example is when vector abundance data are 'binned' into more than two groups, with each bin defining a range of vector densities. Examples are given here of binning the abundance data for the five species of tsetse in northern Côte d'Ivoire into three or five abundance classes, with approximately equal sample sizes (although using the

rule that no abundance level appeared in more than one class).

A more subtle application of eqns (5) to (7) is when distributional data of a single species are drawn from more than one data source, extending across a wide geographical region. Here there may be regional variation in areas of vector absence that gives different co-variance matrices in different areas. There may also be different sub-specific- or strain-variation responses of the vectors to environmental conditions in the different regions, again requiring different co-variance matrices defining fly presence in the different areas. The statistical significance of any differences found may be tested using Bartlett's χ^2 approximation (B) for testing co-variance matrix equality (Green, 1978), defined as follows

$$B=(m-G)\ln|C_w|-\sum_{g=1}^{G}(m_g-1)\ln|C_g| \quad (8)$$

where m is the total number of observations of all groups ($m=m_1+m_2+\ldots m_G$) and G is the total number of groups (i.e. two in the simple case of presence/absence). B is approximately distributed as χ^2 with $\frac{1}{2}[(G-1)(n)(n+1)]$ degrees of freedom, where n is the number of variables contributing to the co-variance matrices. $|C_w|$ and $|C_g|$ respectively refer to the determinants of the within-groups co-variance matrix of all groups combined or of each group, g, separately. *A priori*, the best approach to analysing multiple data-sets from large areas is to keep them separate initially and then to combine co-variance matrices appropriately only when they can be shown not to differ significantly. In practice, however, this may result in rather small sample sizes giving unreliable co-variance matrices (Lark, 1994). In the present study, both approaches were tested. Whilst there was an improvement in the predicted distribution of some species when the data for Côte d'Ivoire and Burkina Faso were kept separate, for others the overall fit was worse. Only the overall fits are shown, with comments on the alternative approach.

In the analyses, all of the map data were used as the training sets for each tsetse species, and the predictor satellite variables were selected in a forward, step-wise manner, the criterion for inclusion being that the addition of the selected variable caused the greatest increase in the Mahalanobis distance [eqn (5)] compared with all other variables during that round (since unequal co-variance matrices were assumed in the analysis, the Mahalanobis distance calculated for each comparison was the sum of the distance between the presence and absence category and between the absence and presence category). Variables were selected in order of their ability to separate the different groups either of presence/absence or of density classes. A total of 10 variables (out of 36) was selected and those chosen were later used to produce maps of posterior probabilities [eqn (7)] which represent the probabilities with which each grid-square falls into the category of fly presence or absence. These predicted maps cover the region from Côte d'Ivoire in the west, to Togo in the east and are based on sampling the satellite and other data files at a spatial resolution of 0·125°.

No transformation of the raw variables was undertaken before analysis, to make biological interpretation of the results more straightforward. The method of variable selection, using Mahalanobis distances, overcomes the potential effect of unequal co-variances arbitrarily determining the importance of the predictor variables.

The ability of the technique to describe the observed distribution and abundance data was measured in several ways. The overall percentage correct predictions (of presence/absence, or of abundance class) were calculated together with the percentages of false-positive and false-negative predictions (i.e. false predictions of presence or absence, respectively). Finally, the sensitivity (ability to predict presence correctly) and specificity (ability to predict absence correctly) were also calculated. In the case of the abundance data, the percentage correct assignment to each density class was recorded.

RESULTS

Table 1 lists the 36 predictor variables available to the analysis, and Fig. 1 (a–h) shows the

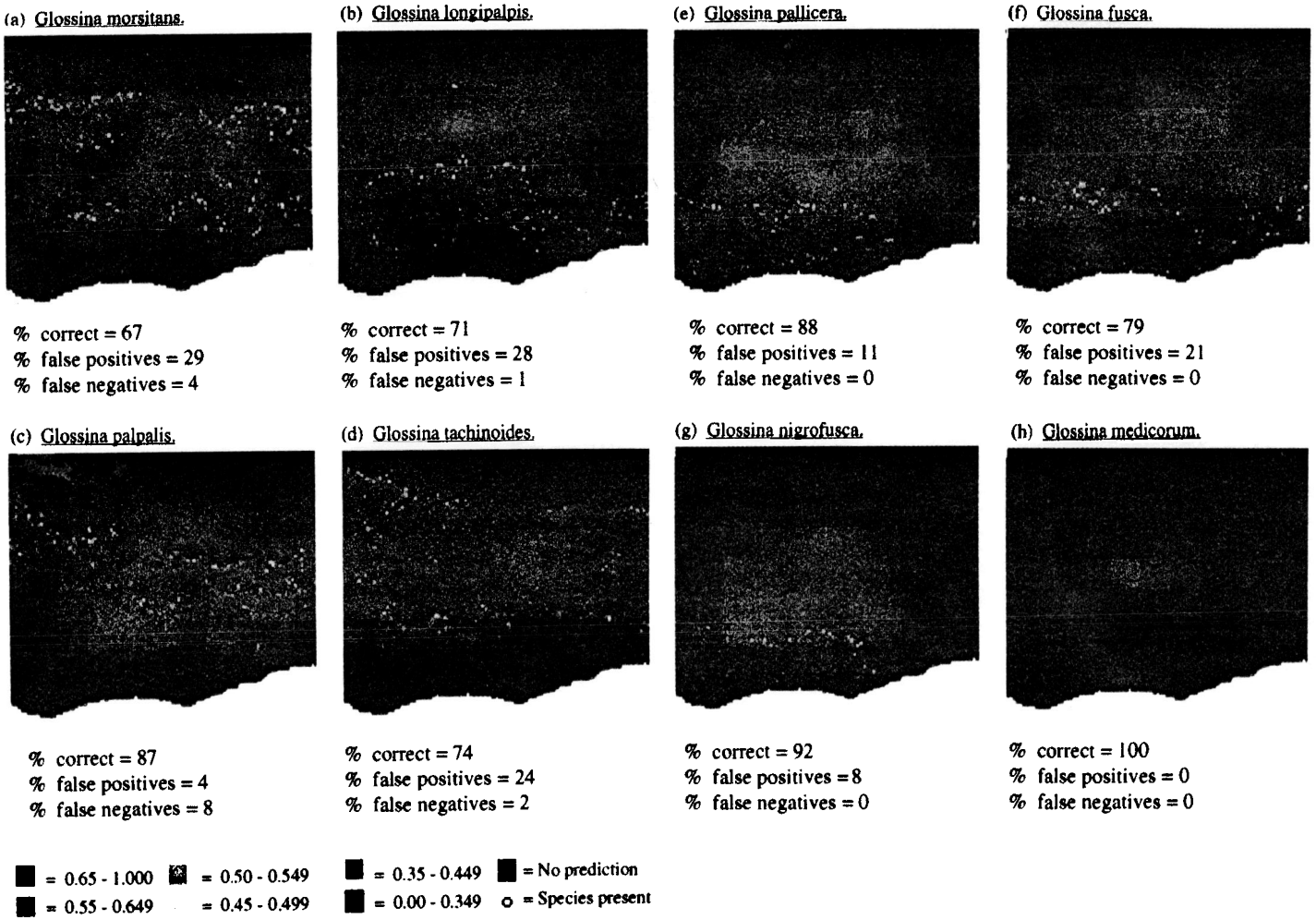

Fig. 1. Discriminant analysis predictions of areas of suitability for *Glossina morsitans* (a), *G. longipalpis* (b), *G. palpalis* (c), *G. tachinoides* (d), *G. pallicera* (e), *G. fusca* (f), *G. nigrofusca* (g) and *G. medicorum* (h) in Côte d'Ivoire and Burkina Faso, based on the satellite predictor variables listed in Table 2. Suitability is on a colour scale from red (low suitability) to green (high suitability). The recorded distribution of each species is also indicated. 'No prediction' refers to areas where satellite data were more extreme than any in the training set. The accuracies of predictions, shown below each figure, refer to the training-set data (see text for further details).

observed and predicted distributions for each of the eight species present, and the accuracy of each, using the 10 most important predictor variables. Table 2 lists the predictor variables in order of importance for each species and lists the accuracy of predictions when one, five or all 10 predictor variables were used. When the three data-sets for each species (two from Côte d'Ivoire and one from Burkina Faso) were kept separate in the analysis, and when variables were chosen during training on their ability to distinguish presence and absence within each country only (between-country comparisons were confused by the large variation in some of the predictors across the full geographical range), the accuracies of the final maps increased considerably for some species (e.g. *G. morsitans*, 83% correct; *G. longipalpis*, 91% correct; *G. tachinoides*, 85% correct) but decreased for others (e.g. *G. palpalis*, 50% correct). This was possibly because, for *G. palpalis*, areas of absence in one of the samples were rather similar in satellite characteristics to areas of presence in another. This indicates a certain degree of adaptation of the tsetse species to local conditions (the geographical area covered includes the two sub-species of *G. palpalis*).

Figure 2 (a–e) shows the observed and predicted abundance classes of the five species sampled in the north of Côte d'Ivoire and Table 3 lists the predictor variables used and the accuracy of the predictions. Given the relatively small ranges of density in each class, the results are surprisingly good. Table 4 lists the mean values of the predictor variables for the five density classes for each of the three most widespread species, *G. morsitans*, *G. tachinoides* and *G. palpalis*. In many cases there is a gradual, monotonic increase or decrease in the mean values of the predictor variables across the density classes. Only very slight differences between the predictor variables are found in areas of absence or presence at the different densities.

DISCUSSION

Advances in micro-computing technology, the increasing availability of remotely-sensed, satellite imagery and the development of novel image-processing and analytical techniques all contribute to an increasing ability to predict the distribution and abundance of natural resources and disease vectors using remotely-sensed, satellite imagery. The success of the approach described in this review is all the more remarkable when the relatively poor quality of the distribution data is remembered, together with the long periods of time over which they were recorded. The satellite data were gathered during the 1980s, when particularly severe droughts affected much of the study area, resulting in rapid changes of the distributional limits of some of the species considered. *Glossina tachinoides*, for example, extended its range south, and apparently replaced *G. palpalis* over large areas of central Côte d'Ivoire (Clair, 1987; DJR, unpubl. obs.). Curiously, and perhaps significantly, when each data-set is kept separate within the analysis, the predicted area of suitability for *G. tachinoides* shifts southwards into the same central and southern regions of Côte d'Ivoire that the species invaded in the 1980s. These fly advances may be facilitated by only slight changes in the average environmental conditions in the newly invaded areas. For example, the present analysis indicates that the difference between the mean NDVI in sites of presence or absence of *G. tachinoides* varies between $-0{\cdot}08$ and $+0{\cdot}04$ (in Côte d'Ivoire and Burkina Faso, respectively), each a small fraction of the total range shown by this variable across these two countries (0·53; Table 1). As some areas become more suitable for flies, however, other areas become less suitable, so that the distributional limits shift slowly with time. The expansion and subsequent contraction of tsetse-fly distributions have been previously recorded, but global environmental change will bring about permanent range shifts in addition to these relatively short-term variations.

As more predictor variables are added into the analysis, the predictions tend to become more accurate (Table 2) and more well-defined, so that with 10 predictor variables the analysis identifies areas of suitability with either a relatively high or relatively low

TABLE 2

*The 10 most important predictor variables used to describe the distribution of tsetse in Côte d'Ivoire and Burkina Faso, and the accuracy of predictions when using one, five or 10 variables**

	Species of tsetse fly							
	G. morsitans	*G. longipalpis*	*G. palpalis*	*G. tachinoides*	*G. pallicera*	*G. fusca*	*G. nigrofusca*	*G. medicorum*
RANK								
1	CH4range	CCDp2	NDmax	CH4range	CH4mean	CH4min	CH4mean	CCDp2
2	CCDmin	NDmax	CH4a2	NDm/CH4m	NDrange	CH4mean	CH4p3	CH4a1
3	NDp-CDp	CH4a1	CH4a1	CH4p2	CH4range	NDrange	NDrange	NDmin
4	NDa2	NDp1	CH4mean	NDmean	CCDa2	CH4p2	NDa1/CH4a1	NDa1/CH4a1
5	NDa1/CH4a1	NDa1	CH4min	CH4mean	CCDp2	CH4p3	CH4p2	CCDmax
6	CCDa2	CCDmean	NDm/CH4m	CH4a2	CCDa1	CCDp2	CH4a1	CCDmean
7	NDa3	CCDa1	CH4a3	NDa2	CCDp1	CH4a3	NDa1	NDa1
8	NDmean	NDp3	CCDp2	Elev	CCDmax	CCDa1	CCDp1	NDmax
9	CH4a2	CH4mean	CCDa3	CH4a1	NDp2	CCDmean	CCDmin	NDp1
10	NDm/CH4m	CH4a3	CH4p2	CCDmin	NDa3	NDm/CCDm	CCDrange	CH4range

	Number of variables used																							
	1	*5*	*10*	*1*	*5*	*10*	*1*	*5*	*10*	*1*	*5*	*10*	*1*	*5*	*10*	*1*	*5*	*10*	*1*	*5*	*10*	*1*	*5*	*10*
ACCURACY																								
Correct (%)	64	65	67	59	67	71	87	88	87	66	70	74	88	83	88	52	77	79	88	89	92	32	64	100
False positive (%)	31	31	29	40	33	28	7	6	4	30	28	24	12	17	11	47	22	21	12	11	8	68	36	0
False negative (%)	5	4	4	1	1	1	6	6	8	4	2	2	1	0	0	1	1	0	0	0	0	0	0	0
Sensitivity	0·89	0·91	0·9	0·9	0·97	0·98	0·93	0·93	0·90	0·90	0·95	0·95	0·93	0·96	0·96	0·95	0·93	0·97	0·93	0·98	1·0	0·91	1·0	0·91
Specificity	0·45	0·46	0·4	0·4	0·59	0·64	0·59	0·64	0·74	0·53	0·56	0·63	0·87	0·82	0·88	0·47	0·75	0·77	0·88	0·89	0·91	0·32	0·64	1·0

*See Table 1 for an explanation of the abbreviations.

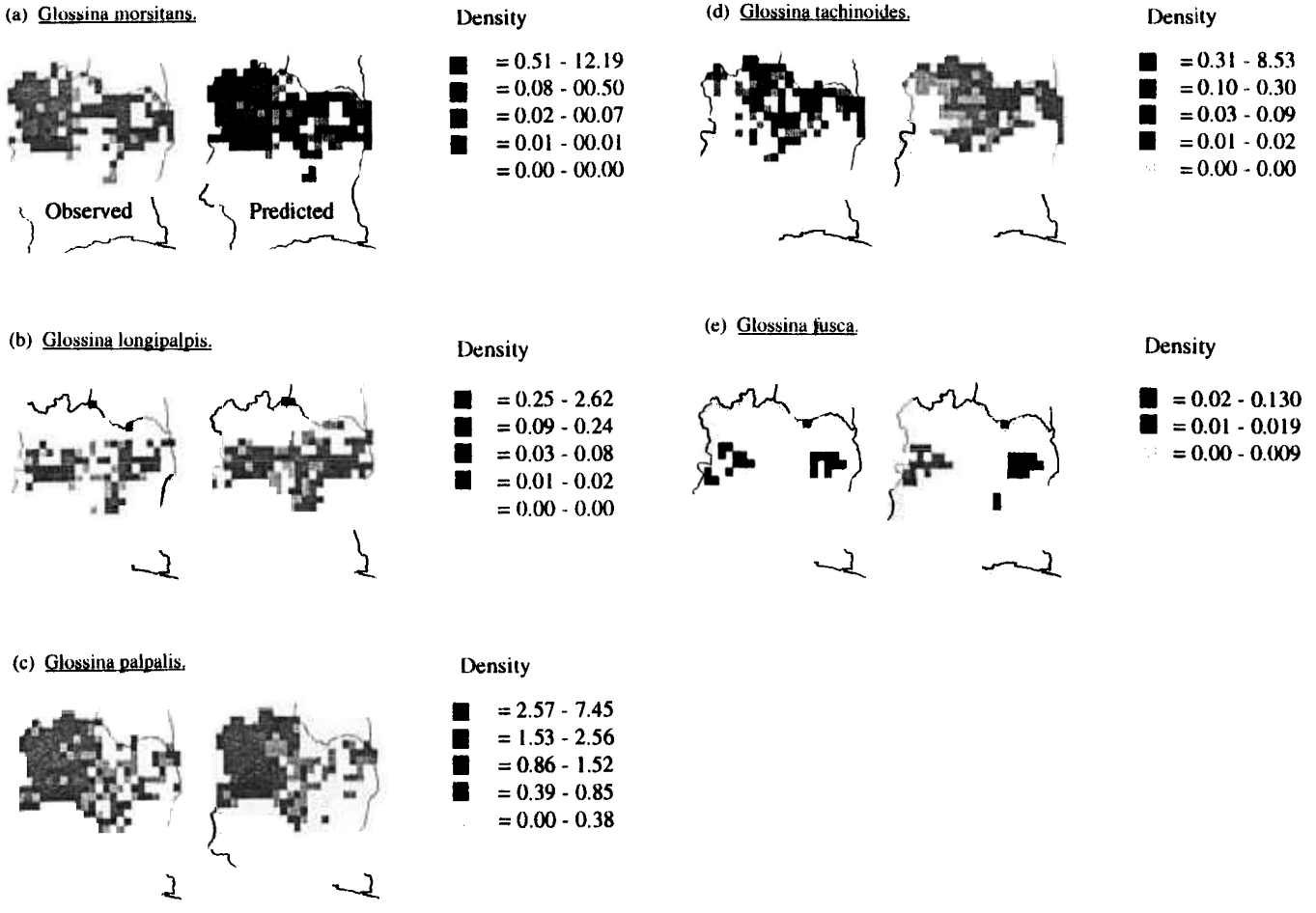

Fig. 2. Discriminant analysis predictions of the apparent density (flies/trap.day) of *Glossina morsitans* (a), *G. longipalpis* (b), *G. palpalis* (c), *G. tachinoides* (d) and *G. fusca* (e) in northern Côte d'Ivoire, based on the satellite predictor variables listed in Table 3. Apparent densities were divided into three classes (*G. fusca*) or five classes (other species) by the analysis to give approximately equal sample sizes.

TABLE 3

The 10 most important predictor variables used to describe the apparent density of tsetse (flies/trap.day) in northern Côte d'Ivoire, and the accuracy of the predictions of the abundance classes (five for all species except G. fusca*)**

	Species of tsetse fly				
	G. morsitans	*G. longipalpis*		*G. tachinoides*	*G. fusca*
RANK					
1	NDp-CDp	CCDa2	CH4p3	NDm/CH4m	CCDa2
2	NDm/CCDm	CCDa1	CH4a2	Elev	CH4a2
3	CH4a1	NDp-CDp	NDmax	CH4a2	NDp1
4	CCDp2	CH4a2	CCDa3	NDp2	CH4a1
	CH4p2	CCDp2	CH4min	CCDmin	Elev
	CH4range	NDa1/CH4a1	CH4max	NDm/CCDm	CCDp2
	NDa2	CH4min	CCDp1	CCDrange	CCDmin
8	CCDmax	CCDa3	CCDa1	NDrange	CCDrange
9	CCDa2	CH4mean	CH4p2	NDa3	NDp1-CH4p1
10	NDp2	CCDp3	NDa3	CH4range	NDmax
	Density and [accuracy of prediction (%)]				
	0·00–0·00 (58)	0·00–0·00 (81)	0·00–0·38 (78)	0·00–0·00 (65)	0·00–0·00 (98)
	0·01–0·01 (100)	0·01–0·02 (80)	0·39–0·85 (50)	0·01–0·02 (78)	0·01–0·01 (100)
	0·02–0·07 (72)	0·03–0·08 (76)	0·86–1·49 (30)	0·03–0·09 (78)	0·02–0·13 (100)
	0·08–0·50 (51)	0·09–0·20 (94)	1·53–2·53 (59)	0·10–0·30 (62)	
	0·51–12·2 (63)	0·25–2·62 (83)	2·57–7·45 (58)	0·31–8·53 (65)	

*See Table 1 for an explanation of the abbreviations.

probability; and intermediate probabilities (i.e. within the range 0·35–0·65) are very scarce (Fig. 1). With the exception of *G. palpalis*, the percentage of false positives (i.e. false predictions of presence) always outnumbers the percentage of false negatives (false predictions of absence), often considerably (Table 2 and Fig. 1). The former indicate areas of apparent suitability for flies that are not occupied by them, or in which flies were not recorded by the surveys. This is a feature of many distribution maps (i.e. species do not always occupy all areas that are suitable for them, or are not always found, even when they occur there). False-negative predictions, however, indicate a more serious situation where the technique, for one reason or another, has failed to define the full range of conditions in which the flies can survive. The percentage of false negatives is highest (8%) for *G. palpalis*, the most widespread species, and the errors arise because the analysis fails to identify its recorded northern limits in Burkina Faso [Fig. 1(c)]. The same environmental changes that caused a southward extension of *G. tachinoides* may, however, have caused a retreat of *G. palpalis* (a species adapted to moister conditions) from the northern part of its range, thus explaining the high error rate.

Whilst the ability to describe the training-set data is the first criterion for a successful statistical description of a species distribution, the technique is only of real use when it can be used to describe distributions in other places, and at other times. The maps presented in Fig. 1 show predicted distributions of the study species in Ghana and Togo. The predictions for Togo show both similarities and differences with the recently mapped tsetse distributions in this country (Rogers *et al.*, 1994; Hendrickx *et al.*, 1995). Predictions for *G. palpalis* and *G. tachinoides* are rather better

TABLE 4

Mean values of the predictor variables for each of the apparent density classes of G. morsitans, G. palpalis *and* G. tachinoides *in northern Côte d'Ivoire**

	Predictor variable										
G. morsitans	*NDp-CDp*	*NDm/CCDm*	*CH4a1*	*CCDp2*	*CH4p2*	*CH4range*	*NDa2*	*CCDmax*	*CCDa2*	*NDp2*	*Sample*
DENSITY											
0·00–0·00	1·97	0·56	7·69	3·29	1·95	15·71	0·04	136·25	32·32	3·57	65
0·01–0·01	1·88	0·57	8·10	3·32	1·96	16·42	0·04	127·43	28·49	3·50	17
0·02–0·07	1·53	0·55	8·32	3·34	2·00	16·9	0·04	132·49	27·13	3·43	39
0·08–0·50	1·38	0·55	8·44	3·15	2·03	17·29	0·04	133·15	26·77	3·49	41
0·51–12·18	1·4	0·53	8·81	3·09	1·96	17·89	0·04	135·43	27·80	3·51	43
G. palpalis	*CH4p3*	*CH4a2*	*NDmax*	*CCDmean*	*CH4min*	*CH4max*	*CCDp1*	*CCDa1*	*CH4p2*	*NDa3*	*Sample*
DENSITY											
0·00–0·38	1·16	1·38	0·47	12·98	21·00	38·23	5·30	42·09	1·92	0·01	41
0·39–0·85	1·23	1·42	0·48	13·96	20·93	37·79	5·33	41·62	1·93	0·01	42
0·86–1·49	1·40	1·48	0·49	15·01	21·22	38·08	5·42	42·75	2·01	0·01	43
1·53–2·53	1·55	1·71	0·51	13·37	20·59	36·98	5·62	44·77	1·95	0·01	41
2·57–7·45	1·74	1·63	0·50	14·39	21·31	37·79	5·64	46·62	2·06	0·01	45
G. tachinoides	*NDm/CH4m*	*Elev*	*CH4a2*	*NDp2*	*CCDmin*	*NDm/CCDm*	*CCDrange*	*NDrange*	*NDa3*	*CH4range*	*Sample*
DENSITY											
0·00–0·00	1·34	343·5	1·62	3·66	9·79	0·55	129·1	0·29	0·01	15·92	110
0·01–0·02	1·19	357·8	1·42	3·44	4·80	0·56	123·2	0·28	0·01	16·98	27
0·03–0·09	1·15	332·8	1·44	3·31	7·05	0·55	122·0	0·30	0·01	17·49	23
0·10–0·30	1·08	314·0	1·41	3·28	2·47	0·54	125·6	0·29	0·01	18·27	26
0·31–8·53	1·11	304·2	1·38	3·34	5·31	0·54	123·1	0·29	0·01	18·25	23

*See Table 1 for an explanation of the abbreviations.

than those for *G. morsitans* and *G. longipalpis*, although there has also been a southward extension of *G. tachinoides* in Togo that is not predicted by the present analysis, and does not appear on previous maps for tsetse in Togo (Ford and Katondo, 1977). Clearly, therefore, the distribution of tsetse in Togo, and presumably elsewhere, has changed from the historical picture that forms the basis of much of the present analysis and, in the light of recent environmental changes in Africa, it is unlikely that non-contemporary satellite data will give an entirely satisfactory fit. The ideal approach to tsetse mapping is therefore to use contemporary satellite and fly-distribution data to define the areas of suitability for each tsetse species and, from this, to make predictions for other places and times.

A further complication arises from the adaptation of each species to local conditions, about which very little is known at present. Species such as *G. morsitans*, and even its subspecies, occupy vast areas of Africa (e.g. greater than 40° of longitude for *G. m. submorsitans*) and almost certainly show biological variation across such distances. Behavioural and ecological differences within *G. pallidipes*, a widespread species in East and southern Africa, have already been shown (Rogers, 1990; Baylis and Nambiro, 1993). These effects mean that the characterization of a species' habitat in one area may not easily be extended to other, remote areas. Given contemporary satellite and distributional data, however, the approach suggested may be used to estimate the degree of difference in habitat types across wide geographical areas. Whilst satellite data may not immediately explain within-species differences across large geographical areas, they may, in the first instance, illuminate such differences and hence lead to a better biological understanding of them.

Of the three main satellite data types used in the present analysis (NDVI, AVHRR-channel-4-brightness temperature and CCD imagery), the thermal channel data appear most frequently in the predictor variables for fly distribution. On 19 occasions, one or other thermal variable is in the top five predictor variables for the eight species of tsetse considered, whilst NDVI and CCD variables appear 10 and six times, respectively (some combination of these variables making up the remainder). Considering only those species for which the abundance data are also available, the figures become 14 for thermal, six for NDVI and two for CCD data. Thermal-channel data also predominate in describing the abundance data, but less so than for the distribution data (with nine, three and seven occurrences for the thermal, NDVI and CCD data, respectively). The low importance of the CCD imagery in determining distribution and its relatively greater importance in predicting abundance indicates that fly distribution limits in this part of Africa may be more sensitive to temperature, whilst abundance within the distributional limits is some function of rainfall, which determines vegetation growth.

CONCLUSIONS

It is clear that remotely-sensed satellite imagery can be a powerful tool in our ability to investigate large-area phenomena such as the distribution and abundance of the insect vectors of disease. The application of temporal Fourier processing to multi-temporal, meteorological, satellite imagery allows characterization of habitat 'fingerprints' in the form of means and the seasonal timing and seasonal extremes of values of temperature, rainfall and vegetation surrogates. Insect-vector distributions clearly depend on habitat types and so should be amenable to statistical descriptions based on such habitat fingerprinting. In future, satellites giving data with higher spatial and spectral resolution will provide a much more fine-grained view of natural habitats (Hay *et al.*, 1996) and it is timely to prepare now for the wealth of new data these satellites will provide.

A statistical description of a tsetse-fly habitat or distribution is, however, no substitute for a full, biologically based understanding of the same phenomenon. Such an understanding comes from a study of the underlying demographic processes but, as explained elsewhere (Rogers and Randolph, 1993; Randolph, 1994),

the information on such processes is often lacking, leaving the statistical approach as the only one available at present. The long-term aim of the present work is to produce risk maps for tsetse-borne diseases based on a sound biological understanding of epidemiological processes. Satellite imagery provides a way of revealing the patterns in epidemiological processes from which such an understanding will eventually arise.

ACKNOWLEDGEMENTS. The channel-4-brightness-temperature data were supplied by Drs C. Tucker and C. Justice of the Global Inventory Monitoring and Modelling Systems (GIMMS) group at the National Aeronautics and Space Administration (NASA) Goddard Space Flight Center (GSFC) during work (by S.I.H.) towards a NASA Planetary Biology Internship. The NDVI and Meteosat data were kindly supplied by F. Snijders of the ARTEMIS project at FAO, Rome. Drs G. R. W. Wint and S. E. Randolph kindly read and commented on the manuscript. This work is funded by a grant from the Overseas Development Administration (ODA), U.K., under the Livestock Protection Programme (R5794) and administered through the Natural Resources Institute (NRI).

REFERENCES

ANDRES, L., SALAS, W. A. & SKOLE, D. (1994). Fourier-analysis of multi-temporal AVHRR data applied to land cover classification. *International Journal of Remote Sensing*, 15, 1115–1121.

ANON. (1982). *Cartographie de la Repartition des Glossines*. Maison-Valfort, France: Institut d'Elevage et de Médecine Vétérinaire des Pays Tropicaux (IEMVT).

BAYLIS, M. & NAMBIRO, C. O. (1993). The responses of *Glossina pallidipes* and *G. longipennis* (Diptera: Glossinidae) to odour-baited traps and targets at Galana Ranch, south-eastern Kenya. *Bulletin of Entomological Research*, 83, 145–151.

CHALLIER, A. & LAVEISSIERE, C. (1973). Un nouveau piège pour la capture des glossines (Glossina: Diptera, Muscidae): description et essais sur la terrain. *Cahiers ORSTOM, Série Entomologie, Médicale et Parasitologie*, 11, 315–317.

CHATFIELD, C. (1980). *The Analysis of Time-series: an Introduction*. London: Chapman & Hall.

CLAIR, M. (1987). Données récentes sur la repartition des Glossines au Niger, au Burkina Faso et en Côte d'Ivoire. In *International Scientific Council for Trypanosomiasis Research and Control (ISCTRC), 19th Meeting, Lomé, Togo, 1987*, pp. 345–350. Nairobi: Organization of African Unity.

EASTMAN, J. R. & FULK, M. (1993). Long sequence time-series evaluation using standard principal components. *Photogrammetric Engineering and Remote Sensing*, 59, 991–996.

FAIRBAIRN, H. & CULWICK, A. T. (1950). Some climatic factors influencing populations of *Glossina swynnertoni*. *Annals of Tropical Medicine and Parasitology*, 44, 27–33.

FORD, J. & KATONDO, K. M. (1977). *The Distribution of Tsetse Flies in Africa*. London: Hammond & Kell.

FORD, J. & LEGGATE, B. M. (1961). The geographical and climatic distribution of trypanosome infection rates in *G. morsitans* group of tsetse-flies. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 55, 383–397.

GASCHEN, H. (1945). Les glossines de l'Afrique Occidentale Francaise. *Acta Tropica*, 2 (Suppl.), 1–131.

GREEN, P. E. (1978). *Analyzing Multivariate Data*. Chicago: Dryden Press.

HAY, S. I., TUCKER, C. J., ROGERS, D. J. & PACKER, M. J. (1996). Remotely sensed surrogates of meteorological data for the study of the distribution and abundance of arthropod vectors of disease. *Annals of Tropical Medicine and Parasitology*, 90, 1–19.

HENDRICKX, G., ROGERS, D. J., NAPALA, A. & SLINGENBERGH, J. H. W. (1995). Predicting the distribution of riverine tsetse and the prevalence of bovine trypanosomiasis in Togo using ground-based and satellite data. In *International Scientific Council for Trypanosomiasis Research and Control (ISCTRC), 22nd Meeting, Kampala, Uganda, 1993*, pp. 218–227. Nairobi: Organization of African Unity.

HILL, M. O. (1991). Patterns of species distribution in Britain elucidated by canonical correspondence analysis. *Journal of Biogeography*, 18, 247–255.

HOLBEN, B. N. (1986). Characteristics of maximum-value composite images from temporal AVHRR data. *International Journal of Remote Sensing.* 7, 1417–1434.

LAMBIN, E. F. & EHRLICH, D. (1995). Combining vegetation indices and surface temperature for land-cover mapping at broad spatial scales. *International Journal of Remote Sensing.* 16, 573–579.

LAMBIN, E. F. & EHRLICH, D. (1996). The surface temperature-vegetation index space for land-cover change analysis. *International Journal of Remote Sensing.* 17, 1–15.

LAVEISSIERE, C. & CHALLIER, A. (1977). *La Repartition des Glossines en Haute-Volta. Cartes à 1/2 000 000.* Notice Explicative No 69. Paris: Office de la Recherche Scientific et Technique d'Outre-Mer.

LAVEISSIERE, C. & CHALLIER, A. (1981). *La Repartition des Glossines en Côte d'Ivoire. Cartes à 1/2 000 000.* Notice Explicative No 89. Paris: Office de la Recherche Scientific et Technique d'Outre-Mer.

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.

LOS, S. O. (1993). Calibration adjustment of the NOAA AVHRR normalized difference vegetation index without recourse to component channel 1 and 2 data. *International Journal of Remote Sensing.* 14, 1907–1917.

OLSSON, L. & EKLUNDH, L. (1994). Fourier series for analysis of temporal sequences of satellite sensor imagery. *International Journal of Remote Sensing.* 15, 3735–3741.

MARRIOTT, F. H. C. (1974). *The Interpretation of Multiple Observations.* London: Academic Press.

PRINCE, S. D., JUSTICE, C. O. & LOS, S. O. (1990). *Remote Sensing of the Sahelian Environment: a Review of the Current Status and Future Prospects.* Ispra, Italy: Technical Centre for Agricultural and Rural Development.

RANDOLPH, S. E. (1994). Population dynamics and density-dependent seasonal mortality indices of the tick *Rhipicephalus appendiculatus* in eastern and southern Africa. *Medical and Veterinary Entomology.* 8, 351–368.

ROGERS, D. J. (1990). A general model for tsetse populations. *Insect Science and its Applications.* 11, 331–346.

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

ROGERS, D. J. & RANDOLPH, S. E. (1986). Distribution and abundance of tsetse flies (*Glossina* spp.). *Journal of Animal Ecology.* 55, 1007–1025.

ROGERS, D. J. & RANDOLPH, S. E. (1991). Mortality rates and population density of tsetse flies correlated with satellite imagery. *Nature.* 351, 739–741.

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. & WILLIAMS, B. G. (1993). Monitoring trypanosomiasis in space and time. *Parasitology.* 106 (Suppl. 1), S77–S92.

ROGERS, D. J. & WILLIAMS, B. G. (1994). Tsetse distribution in Africa, seeing the wood and the trees. In *Large-scale Ecology and Conservation Biology*, eds Edwards P. J., May, R. M. & Webb, N. R. pp. 249–273. Oxford: Blackwell Scientific.

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.

TATSUOKA, M. M. (1971). *Multivariate Analysis: Techniques for Educational and Psychological Research.* New York: John Wiley & Sons.

TER BRAAK, C. J. F. (1986). Canonical correspondence-analysis—a new eigenvector technique for multivariate direct gradient analysis. *Ecology.* 67, 1167–1179.

TOWNSHEND, J. R. G. & JUSTICE, C. O. (1986). Analysis of the dynamics of African vegetation using the normalized difference vegetation index. *International Journal of Remote Sensing.* 7, 1435–1445.

VERHOEF, W., MENENTI, M. & AZZALI, S. (1996). A color composite of NOAA-AVHRR-NDVI based on time-series analysis (1981–1992). *International Journal of Remote Sensing,* 17, 231–235.

WILLIAMS, B., ROGERS, D. J., STATON, G., RIPLEY, B. & BOOTH, T. (1992). Statistical modelling of georeferenced data: mapping tsetse distribution Zimbabwe using climate and vegetation data. In *Modelling Vector-borne and other Parasitic Diseases*, pp. 267–280. Nairobi: International Laboratory for the Research into Animal Diseases (ILRAD).