11. TSETSE DISTRIBUTION IN AFRICA SEEING THE WOOD *AND* THE TREES

DAVID J. ROGERS* AND BRIAN G. WILLIAMS†
* *Department of Zoology, South Parks Road, Oxford OX1 3PS, UK and*
† *THEU, London School of Hygiene and Tropical Medicine,*
Keppel Street, London WC1E 7HT, UK

SUMMARY

After an introduction outlining the differences between the biological and statistical approaches to understanding the distribution and abundance of organisms, this chapter gives two examples of dimension-reducing statistical techniques whereby large amounts of environmental data can be processed and sifted to extract useful correlates of the distributional ranges of animal species. These techniques are illustrated using the tsetse fly, *Glossina morsitans*, as an example.

The first technique is that of linear discriminant analysis which predicts the past and present distribution of *G. morsitans* in Zimbabwe, Kenya and Tanzania with an accuracy of >80%. Conclusions from the statistical analysis coincide with previous biological interpretations of the distribution of this species in Africa. The message to emerge from the analysis is that Global Circulation Models (GCMs) will need to achieve a greater degree of accuracy than at present if they are to be useful in making predictions about changing vector distributions with global climate change.

The second technique is temporal Fourier analysis of a series of Normalized Difference Vegetation Indices (NDVIs) of Africa derived from the Advanced Very High Resolution Radiometers (AVHRR) of Earth-orbiting meteorological satellites of the National Oceanic and Atmospheric Administration (NOAA) series. The analysis captures the important characteristics (i.e. the average, amplitude and phase) of the major annual and biannual cycles of vegetation growth. Examples are given of features (the Gezira irrigation project in Sudan) and processes (the timing of the peak vegetation growth along the Nile from Uganda to the Mediterranean) which are revealed clearly by Fourier analysis. A strong association is demonstrated between the amplitude of the first term of the Fourier expansion (= the amplitude of the annual cycle of vegetation growth) and savannah woodland areas of Africa, and a similar close association is shown between the same features of the analysis and the areas infested with the tsetse *G. morsitans*.

Both discriminant analysis and Fourier analysis achieve dimension-reduction without the obfuscation of the underlying biological processes that is often associated with statistical processing of biological data (e.g. as with principal components analysis). Finer resolution of biologically important features may be possible in both space and time using satellite sensor information at a finer spatial resolution.

INTRODUCTION

The distribution and abundance of organisms can be studied in two ways. The first involves a biological approach, in which demographic rates are measured and related to obvious biotic and abiotic factors that might determine or influence them. The aim is to construct models based on the biology of the organism that describe changes in the population over time, and these models can be used to investigate the impact of proposed interventions on population size. While this approach is sound, it requires data for several generations of the study species before any sensible analysis of species' dynamics can be carried out. Resources are often limited and, as a result, such studies cover restricted areas. If the resulting biological models are applied to much larger areas, it must then be assumed that the relationships of the organisms to their abiotic and biotic environments are the same throughout their range, an assumption that is rarely tested and seldom justified.

The second approach to studying the distribution and abundance of organisms is based on a statistical analysis of the relationship between population data (presence or absence of a species, or records of its abundance) and environmental factors that are often measured for other purposes. These factors are generally abiotic (meteorological variables) or environmental (vegetation), and rarely include any measure of the distribution and abundance of natural enemies, parasites or predators. The statistical approach requires extensive data sets that encompass a wide range of environmental conditions, some suitable and some unsuitable for the study species. In the analysis one tries to define a set of statistically sound rules for predicting the presence or absence of species and their spatial abundance. Problems with the statistical approach often hinge upon the non-linear response of the predicted variable to the set of predictors (Ter Braak & Prentice 1988; Hill 1991). More sophisticated mathematical techniques can be used to overcome these problems, but are generally very difficult to interpret biologically.

Whilst the biological approach can investigate experimentally the links between demographic variables and the species' environment, the statistical approach relies on correlations only. Nevertheless, in an ideal world, the two approaches should lead to the same conclusions and should identify the same limiting factors. Though past attempts to unite these two approaches have had limited success, there are several reasons to be optimistic about the future. First, the increased interest in global change has highlighted the biologists' lack of knowledge of the critical determinants of the distribution and abundance of many key species (either those threatened with extinction, or those, such as pests or vectors, whose increasing ranges might increase the spread of crop failures or diseases). Second, an increasing number of ground-based environmental data sets that can be used for statistical analyses are becoming available. Third, there is now a substantial archive of satellite sensor data, a major advantage of which is its extensive and uniform coverage of large areas of the tropics where other, ground-based information is often patchily or erratically recorded. Finally, the development of Geographical Information Systems provides a means of storing and processing spatial data in ways hitherto unavailable to the general biological community.

Unfortunately many of these opportunities come with costs, in particular those of storing the large amounts of data collected, and of detecting important patterns within the data sets, the latter becoming more difficult as the volume of data increases. The risk is that we will soon be overwhelmed with data which, in their raw, unprocessed form, are worthless. The immediate need, therefore, is to develop analytical techniques that can be used as a filter to find the useful information in the available data sets and to present them in a way that makes biological sense. What we require is mathematical simplification without biological corruption. In this chapter we describe two techniques for extracting significant information from large data sets by finding a small number of linear combinations of the original variables that contain most of the relevant information. Such 'dimension reduction' is used to extract biologically useful information from environmental and satellite data sets relevant to tsetse flies, *Glossina* spp., in Africa.

BIOLOGICAL BACKGROUND

There are 22 species of tsetse, a genus currently restricted to Africa. Both sexes of all species live only on vertebrate blood. Three ecological groups of flies are recognized; the forest-dwelling *fusca* group, the forest and riverine *palpalis* group, and the savannah *morsitans* group. The life cycle of the fly is relatively straightforward. Approximately every 9 or 10 days mature female flies produce fully grown larvae viviparously and these burrow into the soil and pupate within a few minutes of larviposition. Three or more weeks later the teneral adults emerge from the soil and the life cycle continues (Buxton 1955).

The diseases transmitted by tsetse are caused by trypanosomes, flagellate Protozoa in the genera *Trypanosoma*, *Nannomonas* and *Duttonella* (Hoare 1972). The host-range of most trypanosome species is quite wide (Molyneux & Ashford 1983). Species in all three trypanosome groups infect domestic animals, with variable consequences for the hosts, but only two subspecies of *T. brucei*, *T.b. gambiense* and *T.b. rhodesiense*, affect humans, causing the generally endemic and milder West African ('Gambian') and the generally epidemic and more acute East African ('Rhodesian') sleeping sickness respectively. Animal trypanosomiasis is more widespread than human trypanosomiasis in Africa and, because most large vertebrates (including man) are more abundant in the savannah areas of Africa than elsewhere, the *morsitans* and *palpalis* groups of flies are of greater economic importance than is the *fusca* group.

ANALYSING THE DISTRIBUTION OF TSETSE

Tsetse survival in the laboratory is related to the humidity and temperature at which the flies are kept (Buxton & Lewis 1934) and fly mortality rates in the field are correlated with saturation deficit and temperature (Rogers & Randolph 1985). The biological studies leading to these conclusions unfortunately have not been carried out over large areas so that the predictions of the pan-African distributional range of tsetse arising from them (Rogers 1979; Rogers & Randolph 1986) must be treated with caution. Arising from such studies are data sets for fly mortality rates (or correlates of these rates, such as the physical size of flies) and relative population estimates in both space and time and these data sets have recently been correlated with satellite data derived from the National Oceanic and Atmospheric Administration (NOAA) series of meteorological satellites (Rogers & Randolph 1991). The NOAA data were processed to give Normalized Difference Vegetation Indices (NDVIs), calculated from two of the five channels of the satellite's Advanced Very High Resolution Radiometer (AVHRR),

$$\mathrm{NDVI} = (\mathrm{Ch2} - \mathrm{Ch1})/(\mathrm{Ch2} + \mathrm{Ch1}),$$

where Ch1 = AVHRR Channel 1 reading (= radiance at 580–680 nm wavelength, visible red) and Ch2 = AVHRR Channel 2 reading (=

radiance at 725–1100 nm wavelength, near infra-red). AVHRR data, which have many environmental applications (Prince *et al.* 1990), are now routinely processed to produce NDVIs, which are the most widely available of a number of vegetation indices (Jackson & Huete 1991). The NDVI is directly related to the photosynthetic activity of plants (Tucker & Sellers 1986) and has also been correlated with vegetation type on a continental scale (Tucker *et al.* 1985), vegetation biomass (Huete & Jackson 1987) and seasonal crop production (Bartholomé 1988). The spatial resolution of the AVHRR sensors is at best 1.1 km, but when the data are stored on-board the satellites the limited tape storage facilities reduce this to *c.* 4 km. Later image registration to a map co-ordinate system results in NDVI picture elements ('pixels') with a spatial resolution of *c.* 8 km, considerably poorer than the 10 m spatial resolution of the SPOT satellite's panchromatic sensor or the 30 m spatial resolution of the LANDSAT's Thematic Mapper (TM). However, AVHRR data are available on a daily basis from each of two satellites (Huh 1991), whilst the higher spatial resolution data are available only every 16 days from LANDSAT or nominally every 26 days from SPOT (the receptors of which can be directed from the ground to review particular sites up to seven times during this interval) (Hugh-Jones 1989; Cracknell & Hayes 1991). The high temporal frequency of AVHRR data allows several images taken in a short period of time to be combined in order to eliminate clouds from the final image product. Clouds have very low NDVI values and automated cloud removal is usually achieved by selecting the maximum NDVI value for each site in each 10-day period ('maximum value compositing').

Our current view is that NDVIs effectively integrate a variety of environmental factors that are important for tsetse fly survival. For example, vegetation growth is related to both rainfall and temperature, and active vegetation provides conditions of relatively high humidity in which flies thrive. Satellite imagery therefore provides a link between the intensive biological approach and the extensive statistical approach, providing both an interpretation of the biological result and a data layer for use in the statistical analyses.

In the absence of extensive biological information on tsetse, alternative statistical techniques have been used to 'fit' tsetse distributions to environmental data sets derived from ground-based meteorological recordings and from satellite sensors (Rogers & Randolph 1993; Rogers & Williams 1993; Gareth Staton, personal communication). The meteorological records are interpolated to calculate values of meteorological variables within cells of a geographical data base and the satellite sensors

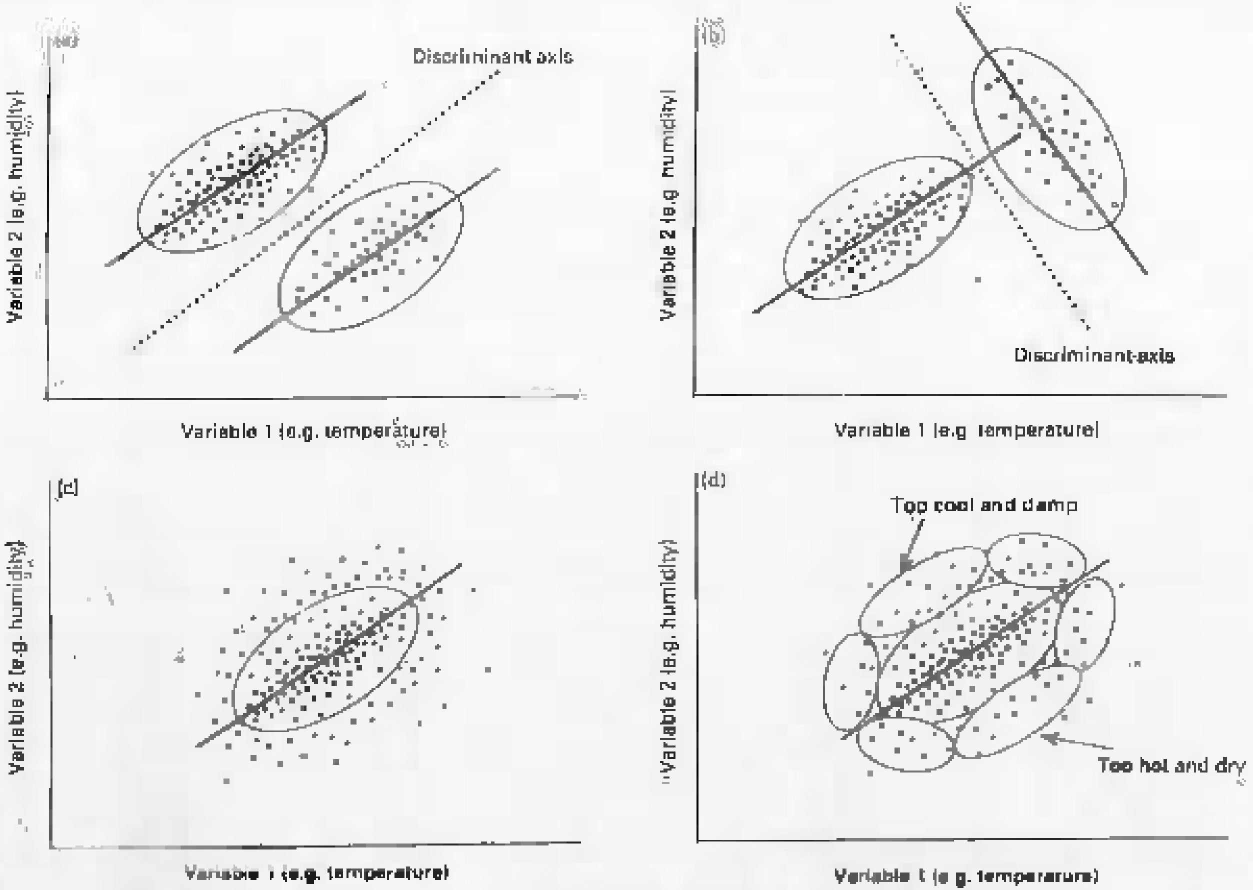

FIG. 11.1 Diagrammatic representation of the problems of applying linear discriminant analysis, using two predictor variables, to the study of animal distributions. (a) Areas of presence and absence have the same variances and co-variances and the discriminant axis easily separates them. (b) and (c) show two complications that arise when variances and covariances are not equal. It may still be possible to use the same procedures of discriminant analysis (as in (b)) or modifications of it (as in (d)). From Rogers & Randolph (1993).

provide further data layers for the analysis. Of the techniques tested we have found that linear discriminant analysis, whilst not the most accurate (Gareth Staton, personal communication), provides considerably more biological insight than many of the alternatives. The theory of this technique is shown diagramatically in Fig. 11.1. Using discriminant analysis a new observation is assigned to one of several categories that have previously been characterized in terms of means, variances and co-variances of a number of predictor variables which previous experience suggests may be important in discriminating between the categories. In the examples in the literature (e.g. Marriott 1974) the categories are classes such as species or species' groups and the discriminating variables are measurements of different parts of the organisms concerned. In the present case there are only two categories, presence and absence of tsetse (there are very few extensive data sets containing information on fly abundance), and the predictors are environmental variables such as temperature, rainfall, etc. Each category is characterized by a group mean in the *n*-dimensional space of the predictor variables and the way in which the sample points cluster around this group mean is described by the covariance matrix of the predictor variables. Given a new point in the *n*-dimensional space, it is possible to determine the probability with which it belongs to each group. This is analagous to the univariate case where both the difference from the group mean and the standard deviation around the group mean are used to determine in which percentile an observation lies. In the multivariate case co-variances between the predictor variables are also important in determining category characteristics and therefore assignment rules, since an observation may be well within the cluster of points around one group mean, but physically closer (in *n*-dimensional space) to another group mean.

One of the assumptions of discriminant analysis is that the variances and co-variances of the predictor variables are the same around each of the group means (as in Fig. 11.1a). Observations from all groups are combined to estimate the within-group co-variance matrix of the predictor variables, and this is used to determine the probability that an unclassified/new point belongs to each group in turn (Green 1978). There are reasons to believe that this assumption may not be strictly valid for distributional data (since species presumably select, or are selected by, a rather well-defined and non-random subset of environmental conditions), but discriminant analysis is relatively robust to violations of the initial assumptions (Marriott 1974). Alternatively assignment can be made on the basis of the co-variance matrix of each category separately (Tatsuoka 1971).

The output of discriminant analysis for a single observation is a set of simple probabilities of belonging to each group in turn, and all probabilities within the set sum to 1.0 (i.e. it is assumed that the observation is drawn from the populations that gave the co-variance matrix or matrices used to make the prediction). Probabilities of presence are worked out for each site in turn (i.e. each cell within the geographical data base), using that site's set of predictor variables, and a map of the result is produced. (Two-group discrimination can be treated as a linear regression between the distributional data, coded 0 for absence and 1 for presence, using the same set of predictor variables. The resulting predictions of group membership are identical.)

In the present examples, the within-group co-variance matrix was calculated for a subsample of points (a 'training set') chosen randomly from the observed data set. Whilst a record of fly presence is indisputable evidence of the (at least temporary) suitability of that site for tsetse, a record of fly absence may arise because of the genuine unsuitability of that site, or because the site has not been surveyed. These unsurveyed sites, if they contain tsetse, will incorrectly contribute to the group of 'tsetse absence' and there will be a tendency to underestimate areas of fly presence. If possible, therefore, only sites known to have been surveyed should be included in the training set. In the absence of good 'tsetse absence' data we might resort to determining the tsetse's environmental envelope using packages such as BIOCLIM (Nix 1986).

Table 11.1 and Fig. 11.2 show the results of applying linear discriminant analysis to the distribution of the tsetse *Glossina morsitans* in Zimbabwe, Kenya and Tanzania. Table 11.1 includes the rank order of the variables as judged by the Mahalanobis distances (Marriott 1974), and determined in a step-wise fashion by first choosing the single variable that gave the greatest separation in multivariate space between sites of tsetse presence and absence (i.e. the highest squared Mahalanobis distance) then selecting a second variable, using the same criterion, to add to the first, and so on. The map of fly distribution in Kenya and Tanzania was compiled by Ford and Katondo (1977), but that for Zimbabwe is based on records prior to game elimination by European colonizers, the rinderpest panzootic at the end of the last century, and the activities of the Tsetse Control Division of the Department of Veterinary Services, Zimbabwe, in the present century, each of which has contributed to the eradication of flies from areas they previously inhabited (Ford 1971).

Several conclusions can be drawn from these analyses (Rogers & Randolph 1993). First the averages of the key predictor variables may differ by rather small amounts between areas of fly presence and absence.

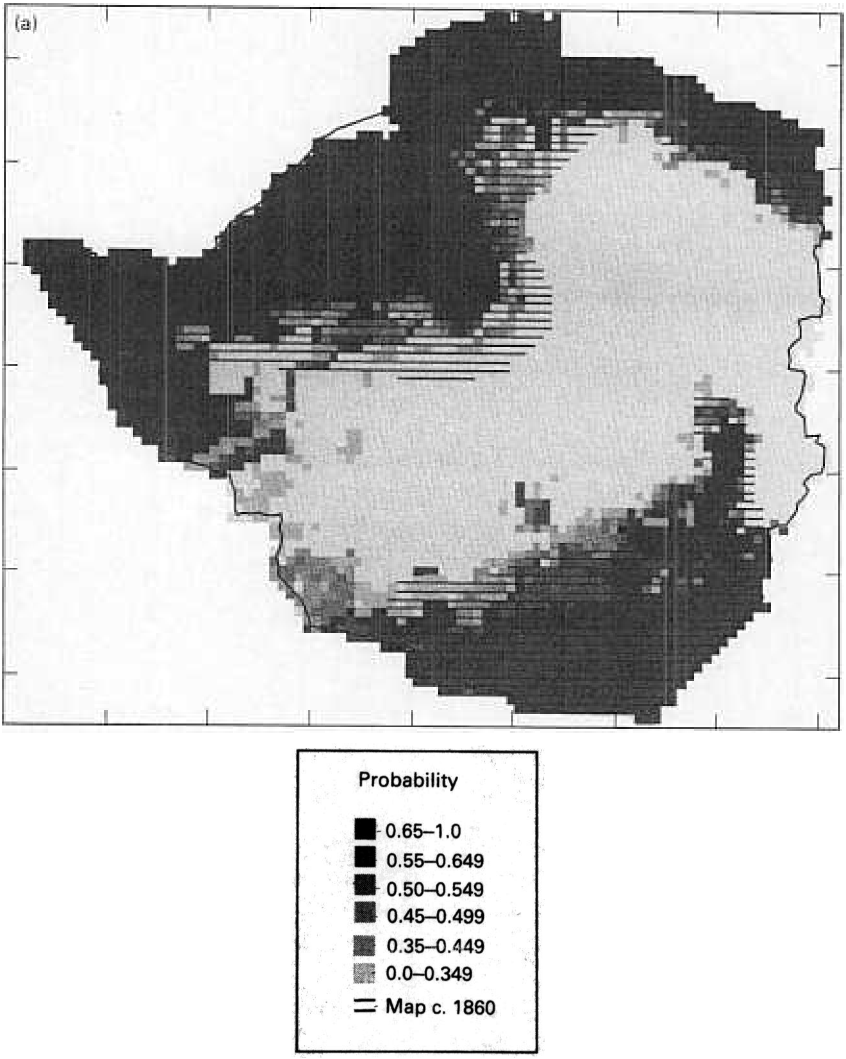

FIG. 11.2. Predicted distributions of *G. morsitans* in (a) Zimbabwe and (b) Kenya and Tanzania based on linear discriminant analysis using a subset of the observed distributions (also shown). The fly map for Zimbabwe (from Ford 1971) pre-dates a rinderpest panzootic that killed many vertebrate hosts of tsetse at the end of the last century, after which flies disappeared permanently from much of their previous range. The fly map for Kenya and Tanzania is the current distribution for this species, from Ford & Katondo (1977). Predicted distributions are on the probability scale shown in the figure. From Rogers & Williams (1993). (a) is drawn in the Plate Carrée projection and (b) in the Hammer–Aitoff projection (Snyder 1987).

Continued on p. 256

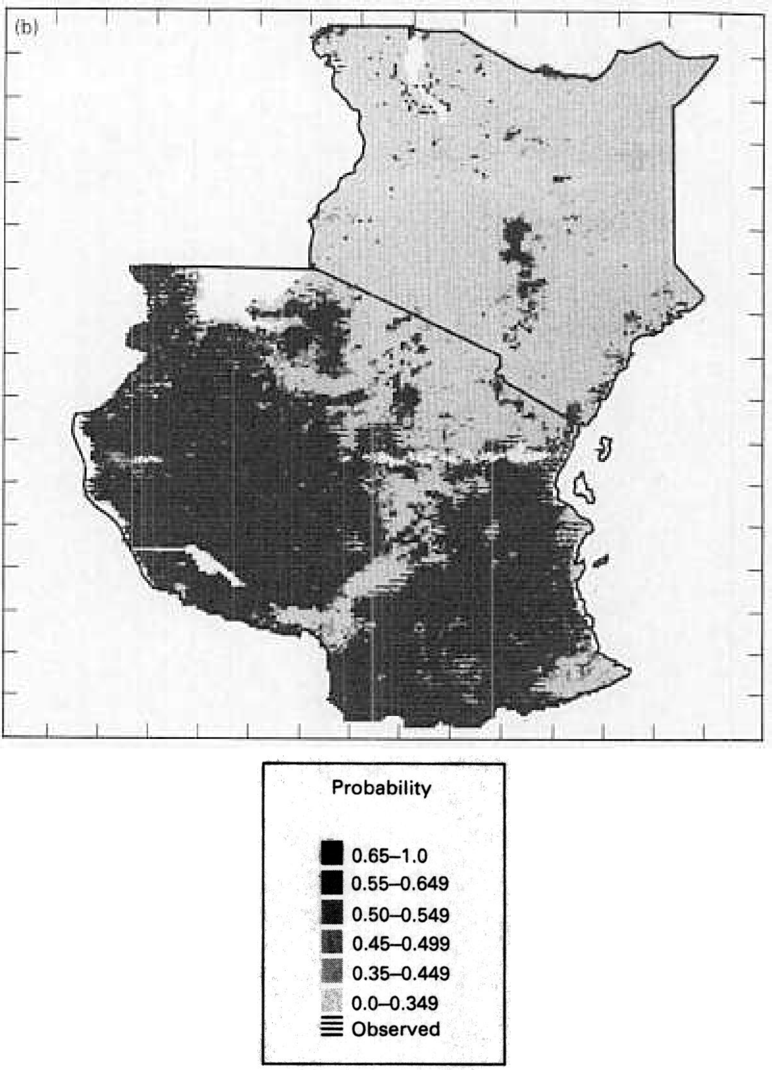

FIG. 11.2. *Continued.*

For example the average difference between *Txmm*, the maximum of the mean monthly temperature (the most important predictor), in areas of fly presence and absence in Zimbabwe is about 3°C whilst for the four key temperature variables for the same species in Kenya and Tanzania it is less than 0.4°C (details in Table 11.1). Making any predictions of changing

Table 11.1. Average values of the predictor variables used in the discriminant analysis of the distribution of *Glossina morsitans* in Zimbabwe, Kenya and Tanzania

	Elevation*	NDmean	NDmin	NDmax	NDran	Txmx	Txmm	Tmmm	Tnmn	Tnmm
Zimbabwe										
Abs†	1097.1	0.29	0.19	0.38	0.19	30.51	23.38	19.98	6.0	14.66
Pres	790.0	0.29	0.16	0.40	0.24	33.09	26.32	22.23	8.36	16.94
Rank	5	10	3	4	8	9	1	6	7	2
% correct	84	85	83	85	86	85	82	86	86	84
Kenya and Tanzania										
Abs	953.7	0.24	0.02	0.48	0.46	30.78	24.63	22.72	15.49	20.95
Pres	963.4	0.44	0.1	0.7	0.6	30.47	24.56	22.58	14.59	20.53
Rank	4	7	10	1	8	5	6	2	3	9
% correct	78	83	84	69	84	81	82	75	77	84

* Variables (in order) are elevation (metres), Normalized Difference Vegetation Index (NDVI), monthly mean (NDmean), minimum (NDmin), maximum (NDmax) and range (NDran), maximum of the monthly maximum temperature °C (Txmx), maximum of the monthly mean temperature (Txmm), mean of the monthly mean temperature (Tmmm), minimum of the monthly minimum temperature (Tnmn) and minimum of the monthly mean temperature (Tnmm).

† Abbreviations: Abs, sites where the vectors are absent; Pres, sites where the vectors are present; Rank refers to the order of importance of each variable (as determined by the analysis) for the production of the predicted distribution maps; % correct, the percentage of sites where the analysis makes a correct prediction of presence or absence when all variables of this and all higher ranks are included.

vector distributions under different scenarios of global change will therefore require a degree of accuracy of Global Circulation Models (GCMs) that few presently possess. Second, the analysis suggests that the major limiting variable for a species may change from place to place (for tsetse it is temperature in Zimbabwe but Vegetation Index in Kenya and Tanzania). It may be important, therefore, to carry out separate analyses for different parts of a species' range to check for changes in the order of importance of the variables included in the analysis. Third, at the very edge of the range of tsetse, as in the case of Zimbabwe which is near the southern continental limits of tsetse in Africa, a single predictor variable may be sufficient to describe fly distribution, whilst throughout the continental range of flies several variables may play an important role. In Zimbabwe, adding all the other variables to *Txmm* improves the fit from 82% to only 85% correct. In Kenya and Tanzania adding temperature and elevation to *NDmax* (the maximum NDVI, the most important predictor for these two countries) improves the overall fit from 69% correct to 84% correct. It is likely that different variables are relatively more important in different areas of these two countries. Fourth, the rather small proportion of false negatives, an incorrect prediction of absence, to false positive results, an incorrect prediction of presence, suggests that whilst the analysis has correctly identified the major environmental constraints, the present tsetse distribution maps may underestimate the actual distribution of vectors. (However, the smaller proportion of false negatives to positives may arise from the fact that a smaller proportion of Kenya and Tanzania is inhabited by the flies rather than uninhabited by them.) False positive areas should be targeted by survey services since they may reveal the presence of vectors at low density. At the very least they represent 'ecological corridors' along which tsetse could move into new areas. Fifth, the statistical procedure of linear discriminant analysis and the potential biological significance of the results are transparent to the user in ways that 'black-box' techniques for analysing vector distributions are not. Some of these other techniques involve fitting a large number of arbitrary parameters. The final values of the parameters are chosen to give the best fit, but they may not, in the end, give a significantly better fit to vector distributions than does linear discriminant analysis (Rogers & Randolph 1993).

In this initial analysis those variables that are most useful in making statistical predictions about the distribution of tsetse also make biological sense. It has long been known that Zimbabwe represents the cold-temperature limits of tsetse, and fly distribution limits very often follow elevation contours. However, many of the temperature variables are highly correlated with each other, and the one that appears most important may be acting as a surrogate for another temperature variable. In Zimbabwe, for example, both *Txmm* and *Tnmm*, the maximum and minimum of the mean monthly temperature respectively, are higher in tsetse than in non-tsetse areas; *Tnmm* is less variable than *Txmm* and so statistically may be a slightly less accurate predictor.

Future developments of this technique will be used to compare linear discriminant analysis with non-linear approaches to predicting tsetse distributions, treading a delicate line between statistical accuracy and biological realism. What we wish to avoid is a statistically perfect prediction of a fly distribution map that we know is inaccurate. Applied carefully, these new techniques should continue to throw light on the environmental conditions on which tsetse depend.

SATELLITE SENSOR IMAGERY AND VEGETATION CLASSIFICATION

NDVIs provided several of the data layers used in the discriminant analysis of tsetse (maximum, minimum, average and range of NDVIs for each site), and are clearly of at least regional importance. In the past the information content of temporal sequences of NDVIs has been extracted using techniques such as principal component analysis (Tucker *et al.* 1985; Townshend & Justice 1986). As in other uses of this approach (e.g. Jeffers 1978) it is quite difficult to interpret biologically the principal axes obtained. The first principal axis contains more than 80% of the information in the 12 monthly NDVIs for an average year (D.J. Rogers, unpublished) and is strongly correlated with the annual average NDVI, whilst the second axis appears to be related to seasonal changes in the vegetation index (Tucker *et al.* 1985). In an attempt to extract biologically useful information from sequences of NDVIs, temporal Fourier analysis (Chatfield 1980) was performed on the 36 monthly images for the years 1987–89 (spatial Fourier analysis is often used in image processing packages to filter out periodic noise in the images arising from faults in the satellite sensors (Cracknell & Hayes 1991), but has previously been used, as here, for temporal analysis of NDVIs by Menenti *et al.* 1991). Fourier analysis describes the seasonal NDVI as the sum of sinusoidal components with frequencies of one to six cycles per year. The first term in the Fourier expansion (hereafter the 'first component') gives the best fit of a sinusoidal wave to the annual cycle of vegetation growth, the second term (the 'second component') gives the biannual cycle, and so on. The analysis gives both the phase and the amplitude of each term,

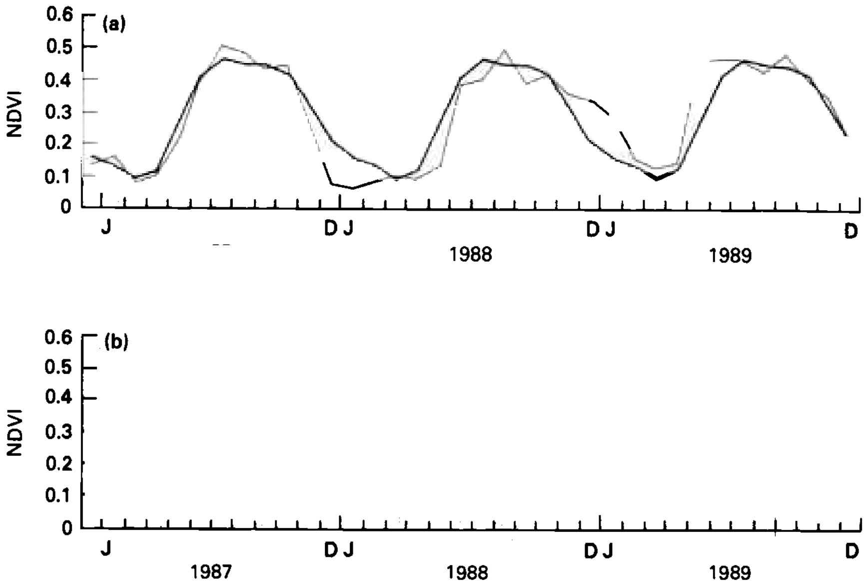

FIG. 11.3. An example of temporal Fourier analysis of the Normalized Difference Vegetation Index from a site in southern Sudan, east of Jonglei on the White Nile (31.6°E, 7°N), for the years 1987–89. The upper graph shows the observed NDVI (thin line) and the fitted Fourier curve, the sum of the first three terms ('components') of the Fourier expansion (thick line). Details of these components are shown in the lower graph. The first component (the solid line), with a frequency of one cycle per year, has a much larger amplitude than the second (dashed line) or third (dotted line), with frequencies of two and three cycles per year respectively, and so makes a major contribution to the overall fit (all three components are drawn around the mean NDVI). Thin vertical lines indicate the phase (= timing) of the first peak of each of the Fourier components.

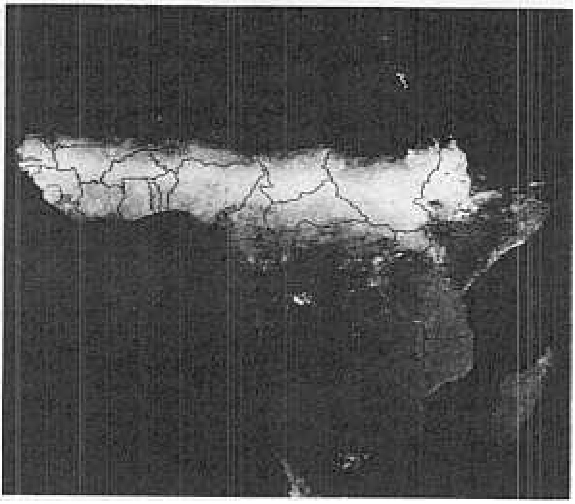

Plate 11.1. Fourier analysis of monthly NDVIs for the whole of Africa for the period 1987–89 inclusive. Analysis is based on the 36 monthly images, each the average of three, 10-day images produced by selecting for each pixel the maximum NDVI during the period (= maximum value compositing), to eliminate clouds. The average NDVI is put in the red gun of the computer screen, the phase of the first Fourier component in the green gun and the amplitude of this component in the blue gun. Values in each gun of the colour image were then stretched across the full range of intensities within the image processing system. International boundaries are also shown (Hammer–Aitoff projection).

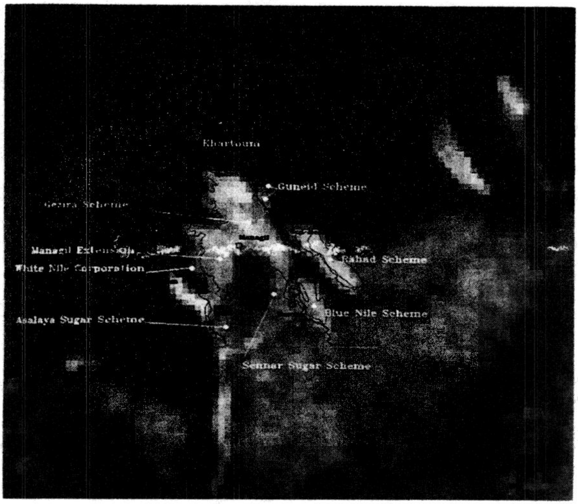

Plate 11.2. Close-up of part of Plate 11.1 showing the Gezira irrigation project area between the White and Blue Niles, just south of Khartoum. Different parts of the irrigation project are labelled (the green parallelogram marks the limits of the map. Scale: Managil to Kosti is *c.* 130 km) (from RIM 1987).

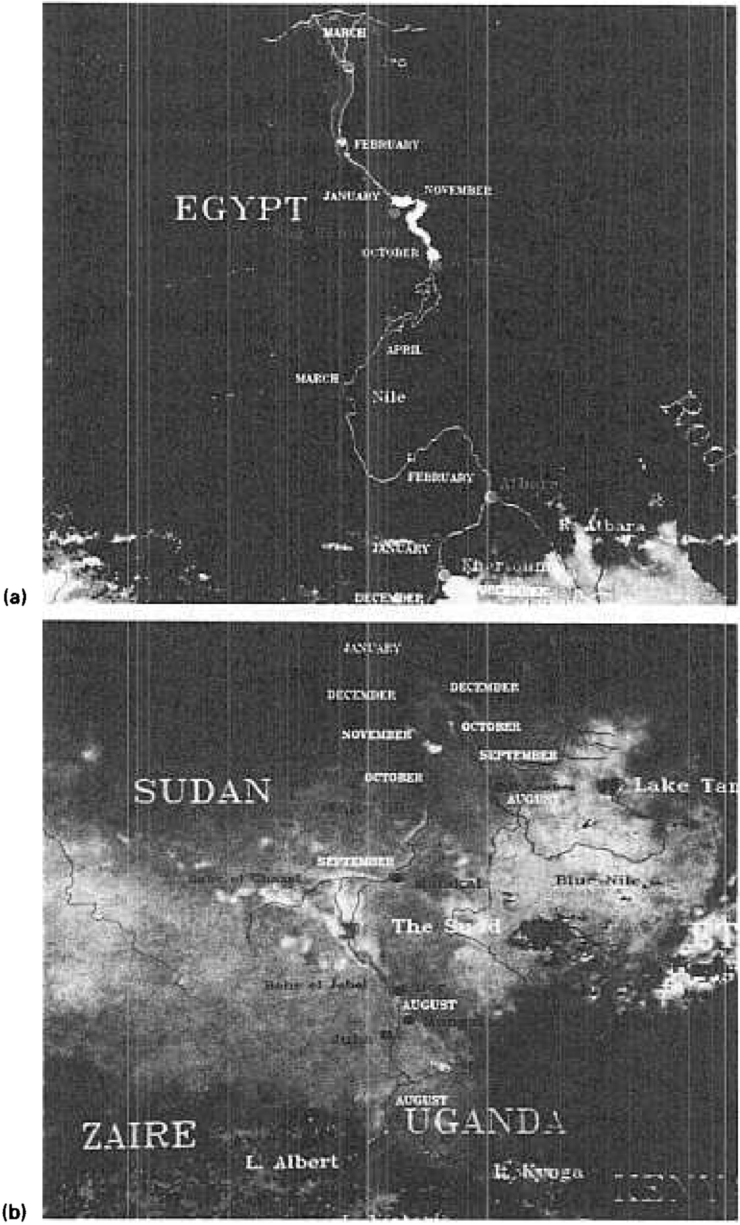

Plate 11.3. Enlargement of part of Plate 11.1 showing (a) the Upper and (b) the Lower Nile basin. Courses of the White and Blue Niles and the months of peak NDVI along the river, probably associated with the seasonal growth of vegetation associated with seasonal river flow, are shown. The long time interval between peak growth at the point of entry into Lake Nasser and peak growth just down river of the Lake is presumably determined by the irrigation schedule of the Aswan dam (scale: Cairo to Khartoum is *c.* 1600 km).

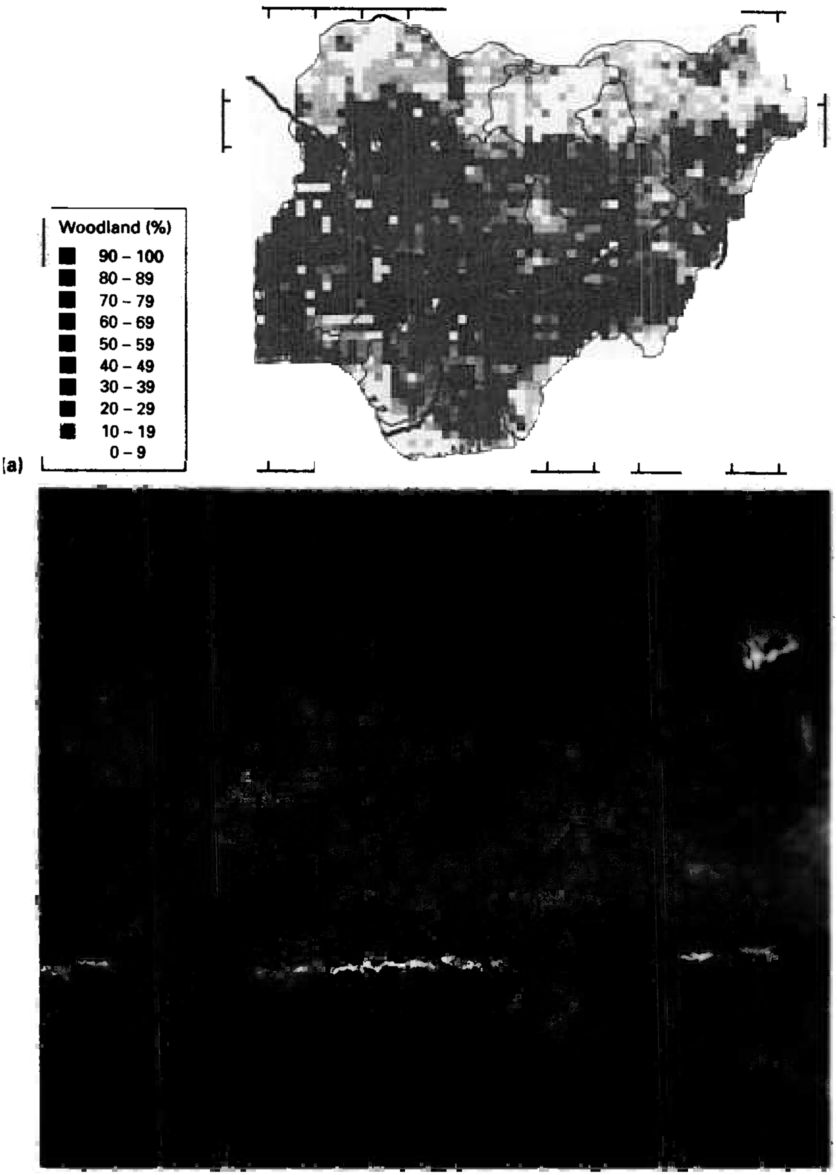

Plate 11.4. (a) Results of an aerial survey of woodland in Nigeria carried out by the Environmental Resource Group Oxford in 1991 (RIM 1992). Each grid square is *c*. 20 km on a side. (b) enlargement of part of Plate 11.1 for Nigeria, with the same state boundaries as in Plate 11.4a. Notice the correspondence between the heavily wooded areas of Plate 11.4a and the blue areas of Plate 11.4b, i.e. areas where the amplitude of the first Fourier component is particularly pronounced (Plate Carrée projection). (The different colour shading of Plates 11.2, 11.3 and 11.4 in comparison with Plate 11.1 is due to colour stretching of only parts of the image, to reveal the features of interest.)

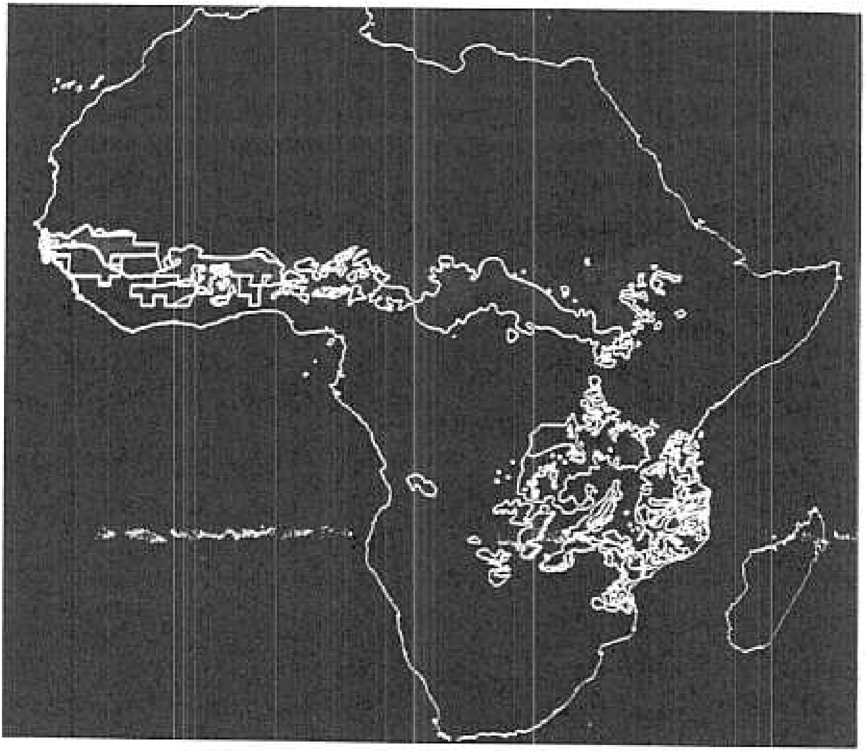

PLATE 11.5. The amplitude of the first Fourier component is shown on a map of Africa (produced by turning off the red and green guns of the computer screen that gave Plate 11.1) and the pan-African distribution of the tsetse *G. morsitans* is shown (within the white continental boundary lines). Plate 11.4 suggests that the brighter blue areas are associated with seasonal woodlands, the habitat of the savannah species of tsetse. This conclusion is here supported on a pan-African scale (Hammer–Aitoff projection).

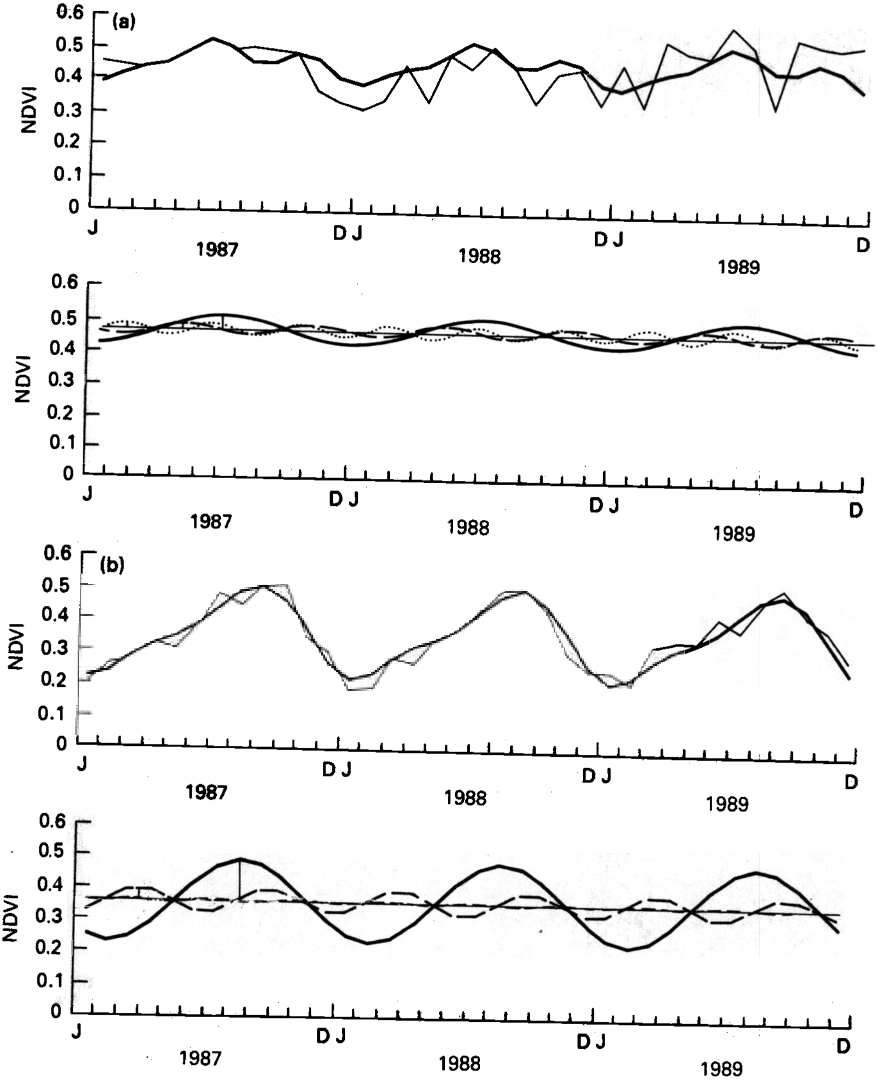

FIG. 11.4. Examples of Fourier analyses of a series of NDVIs for sites from the coast of West Africa, inland to Mali for 1987–89. (a) 9°W, 6°N, in the Liberian rainforest, (b) 6°W, 10°N, near Korhogo, Côte d'Ivoire, (c) 1°W, 13°N, near Ougadougou, Burkina Faso and (d) 0°E, 16°N, near Gao, Mali. Notice changes in the mean NDVI and the changing relative importance of the first and second Fourier components (details as in Fig. 11.3).

Continued on p. 262

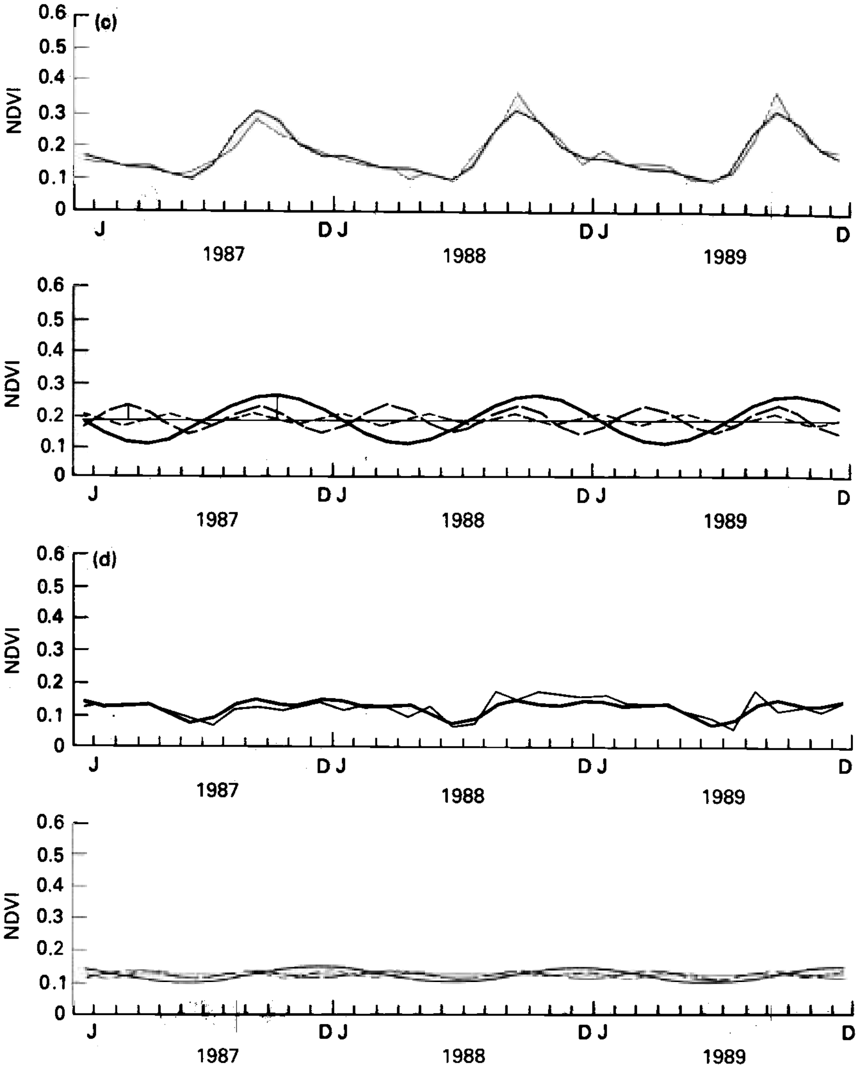

FIG. 11.4. *Continued.*

amplitude fluctuations in the savannah zone of Africa (Fig. 11.4b and c). Thus this analysis confirms, but presents in a different way, the conclusions drawn from principal components analysis of similar images (Tucker *et al.* 1985). The annual cycle of vegetation growth appears to be well captured by Fourier analysis, as can be seen from the examples in Fig. 11.4.

The Fourier analysis of the NDVIs for the whole of Africa for 1987–89 is shown in Plate 11.1 (facing p. 260), where the average NDVI (one of the outputs of the analysis) has been put in the red gun of the screen, the phase of the first component in the green gun and its amplitude in the blue gun. Areas of the image in Plate 11.1 which are predominantly red indicate sites with a large, but unvarying NDVI (= 'forests'); areas which have a blue tinge have seasonally highly variable NDVIs, whilst those which are bright green have a late peak of maximum vegetation growth. The fact that they are green means that neither the average NDVI nor the amplitude of the annual cycle of vegetation growth is very pronounced in such areas. Areas with a high average NDVI plus a late peak of vegetation growth appear either yellow in Plate 11.1 (since red plus green = yellow in colour monitors), or white if the seasonality is also very pronounced (red + green + blue = white).

The Fourier image in Plate 11.1 reveals information about both patterns and processes connected with vegetation types in Africa. One example of a pattern is shown in Plate 11.2 (facing p. 260), a close-up of the Gezira irrigation scheme in Sudan, between the White and Blue Niles just before they join at Khartoum. This scheme is one of the oldest large-scale irrigation projects in the continent and is supplied mostly by the pure waters of the Blue Nile which, through a series of irrigation canals, move down a gradient towards the White Nile. (The history of the Gezira project is described briefly in RIM 1987.) The original Gezira scheme was initially expanded into the Managil Extension, and other schemes were later added around these. Much of the cultivated area is presently under a 4-year crop rotation (of cotton, wheat, legumes such as ground nuts, and fallow), but some of the schemes are more permanently cropped with sugar cane. Some areas, such as the White Nile Corporation Schemes, have much lower yields than others, and this can be partly attributed to poorer irrigation. The non-cropped areas outside the irrigation scheme provide grazing for the considerable numbers of livestock owned by the workers within the scheme, and the whole is a rather intricate, interdependent patchwork (RIM 1987). A map of the Gezira scheme is superimposed on the image in Plate 11.2, and a strong correspondence can be seen between the map and the underlying image. The brightest areas in Plate 11.2 are those that are best irrigated, while the sugar schemes have a speckled appearance.

An example of an ecological process revealed by Fourier analysis is shown in Plate 11.3a and b (facing p. 260) for the upper and lower Nile basin respectively. The White Nile and the main Nile can be detected in these images (the map overlay shows the course of the river) through a combination of differences between the riverine vegetation and the surrounding vegetation. These differences are of several types. In southern Sudan the Sudd can be picked out because it has a high and constant average NDVI throughout the year whereas the surrounding vegetation has a lower average and shows strong seasonality. Further north it is the riverine vegetation which shows strong seasonal variation whilst the surrounding semi-desert or desert shows little seasonal change. The seasonal variation in the NDVI along the course of the river is presumably related to the seasonal flow of water along the river (Shahin 1985; Howell *et al.* 1988), determined by regional rainfall patterns. Of particular interest is the phase (= month) of peak vegetation growth of the annual cycle of vegetation change, which is indicated along the course of the river in Plate 11.3. The outflow of water from Lake Victoria shows little seasonality and there appears to be little seasonality in the NDVI along the Victoria Nile. Beyond Lake Albert, however, the river develops a seasonal flow because of water received from various rain-fed tributaries. Peak vegetation growth occurs during July/August. Continuing along the Nile, the peak of vegetation growth occurs later in the year the further along the river it is measured, so that at Khartoum it occurs around December of the same year. Here the highly variable influx of the Blue Nile contributes significantly to the total volume of water in the main Nile (Shahin 1985). Peak seasonality nevertheless appears to follow a reasonable timing up to Lake Nasser where all signs of a seasonality different from that of the surrounding vegetation disappear. However, beyond the Aswan dam there is a remarkably strong seasonal signal of vegetation growth with a peak in September/October, approximately half a year later than the peak at the entry point into Lake Nasser. This is associated with the timing of the release of waters by the Aswan dam authorities (subject to international agreements with other countries along the Nile). Another dam at Nag Hammadi appears to delay the seasonal peak still further, after which the peak vegetation growth along the remainder of the Nile follows a sensible sequence, ending in the Nile delta in March, approximately 20 months after it was first detected near Lake Albert, almost 3000 miles away but only 750 m higher. It appears therefore that a ('statistical') drop of water from Lake Victoria takes almost 2 years to reach the Mediterranean!

It is possible to look at the Fourier analysis of the NDVI for points along the course of the Nile, and some examples are shown in Fig. 11.5. The slight difference in seasonality of the first Fourier component detected

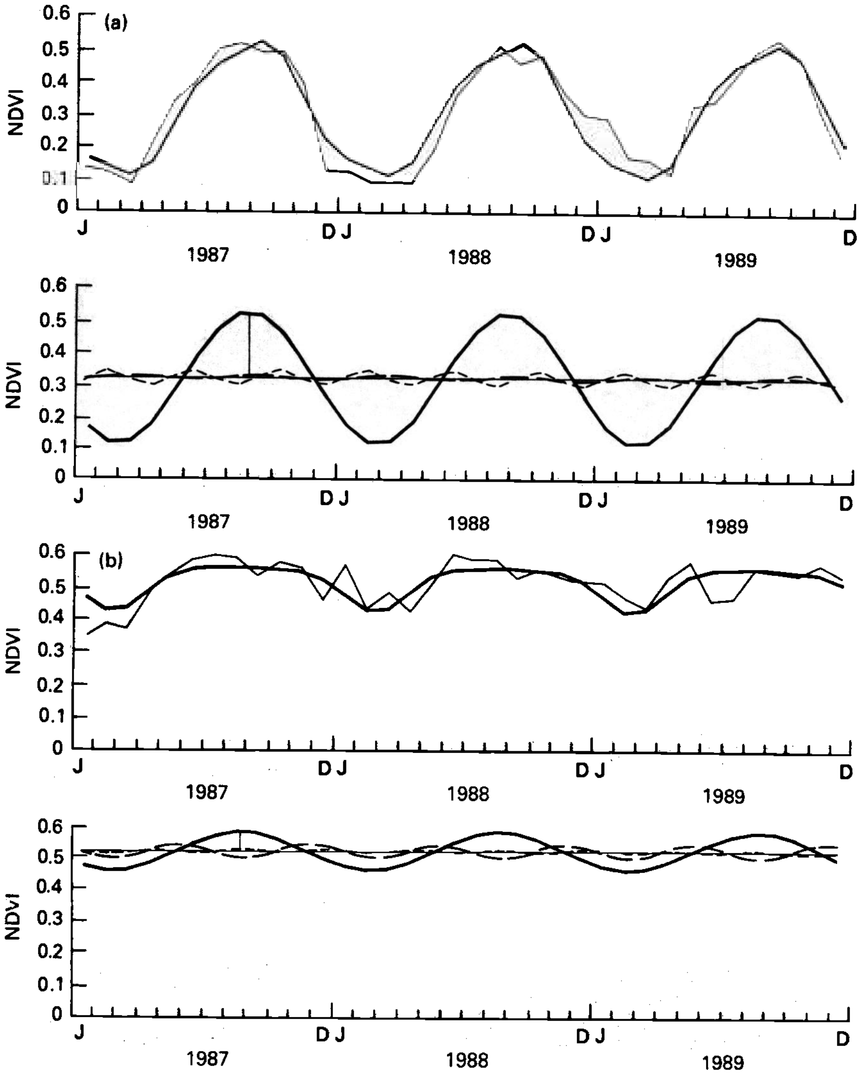

FIG. 11.5. Fourier analysis from selected points near to, or on, the Nile River. (a) West of Jonglei (i.e. just downriver from Bor) (30.6°E, 7°N), (b) on the Nile near Jonglei (30.8°E, 7°N), i.e. in the Sudd, (c) in the desert east of Qena (32.95°E, 25.85°N), on the sharp bend in the Nile between the Aswan Dam and Nag Hammadi, (d) at Qena (32.75°E, 25.85°N) on the Nile and (e) at Tahta (31.45°E, 26.75°N) just downriver from Nag Hammadi (for most place names see Plate 11.3).

Continued on p. 266

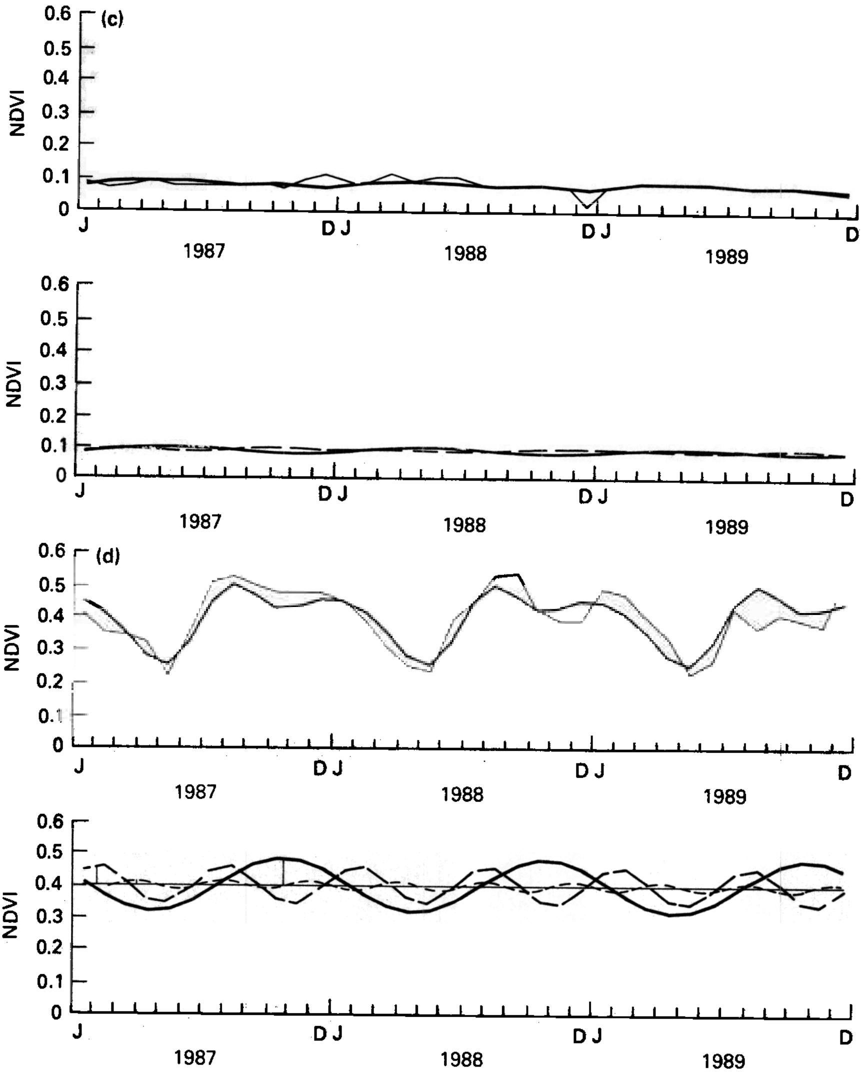

FIG. 11.5. *Continued.*

between the Aswan dam and Nag Hammadi and beyond Nag Hammadi itself can be seen in Fig. 11.5d and e to hide a much more dramatic difference between these two sites revealed by looking at both the first and second Fourier components. Whereas the annual signal dominates at the first site, the biannual signal is stronger than the annual signal at the

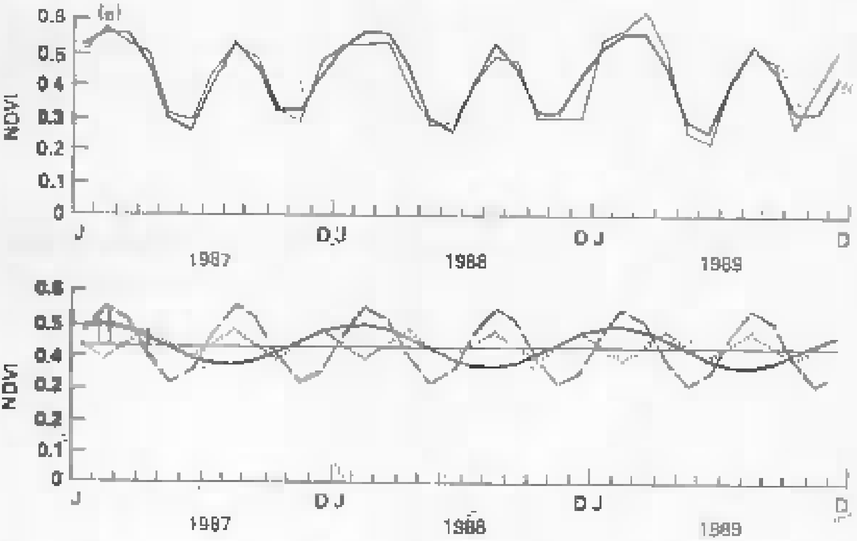

FIG. 11.5. *Continued*

second site. This suggests either a quite different irrigation schedule at the two dams, or a quite different cropping pattern, or some combination of the two.

Finally we return to the problem of predicting tsetse distributions. The seasonal variation of vegetation growth picked out by the amplitude of the first Fourier component appears to be strongly correlated with savannah woodland areas of Africa. Plate 11.4 (facing p. 260) shows the comparison for Nigeria where the Environmental Resource Group Oxford (ERGO) has recently carried out an aerial survey of livestock and land-use categories for the Federal Government of Nigeria (RIM 1992). Plate 11.4a shows a map of the percentage woodland seen in the 20-km grid squares of the country-wide survey whilst Plate 11.4b shows an enlarged part of Plate 11.1, with the Nigerian state boundaries from Plate 11.4a superimposed. The correspondence between the heavily wooded areas of Plate 11.4a and the blue areas of Plate 11.4b is remarkable. *G. morsitans* is a species of tsetse associated with savannah woodland and Plate 11.5 (facing p. 261) shows the agreement between its pan-African distribution and regions of the continent showing strong annual seasonality of vegetation growth. Not only are the northern limits of this species in West Africa well described by the Fourier analysis, but so too are the southern limits in this region. It has long been acknowledged that the northern limits are strongly correlated with rainfall (or vegetation determined by rainfall), but the southern limits have been much more difficult to predict (Nash 1948).

CONCLUSIONS

It is relatively easy to extract biological conclusions from the two methods of analysis presented here although, as emphasized in Chapter 7, it must be remembered that correlations do not necessarily imply causation. When correlations are consistent, however, and when they are supported by intensive biological studies, we may be more certain of their biological significance.

The statistical analyses presented here should help to refine research questions to be addressed by future field work, the need for which becomes more urgent as indications of long-term climatic change are more widely demonstrated and accepted. The stimulus to discover the determinants of animal (and plant) distributions is ironically greatest at the moment when long-term distributions are on the point of disruption by human activity, in much the same way that diseases stimulate the study of the healthy human body, or mechanical breakdowns show the need to understand how machines work properly. Statistical analysis works best in a constant world and, in a changing environment, the same methods of analysis may mislead more than they direct. Nevertheless the option to investigate the biological determinants of distribution and abundance is no longer open, because of a lack of time to carry out such studies. Answers are required quickly and they will need to be based on statistical analysis.

This study has shown how satellite sensor imagery can have an important role in detecting patterns and processes in biological systems. Temporal Fourier analysis appears to be a novel and exciting way to extract hitherto unused aspects of NDVI information. The output of Fourier analysis is more easily interpreted in terms of the seasonality and amplitude of vegetation growth and is therefore more understandable biologically than the output from other methods of analysing NDVIs. The images used in this study had a spatial resolution of *c.* 8 km, much coarser than the maximum of 1.1 km of which the NOAA series of satellites is capable. Ground-based receivers near to, or beneath the track of, the orbiting satellites can receive the full resolution 1.1 km data in all five wave bands of these satellites. This would allow the processing of the NDVI Channels 1 and 2 in different ways to give a variety of vegetation indices, some of which are probably more useful for heavily vegetated areas than is the NDVI, which was originally formulated to detect changes in the sparsely vegetated regions of the Sahelian zone of Africa (Jackson & Huete 1991). Other potentially useful information is available in the non-NDVI infra-red bands that have been used in the past to study wetlands (Van de Griend *et al.* 1985; Xue & Cracknell 1992). Novel combinations of the spectral signals and the much finer spatial resolution of the 1.1 km data at the high temporal frequency of the NOAA satellites must be explored if full advantage is to be taken of these satellite sensor data for biology and conservation. Problems of sensor calibration, atmospheric aerosol effects and view angle must also all be overcome (Goward & Huemmrich 1992) before such imagery can usefully complement the higher spatial resolution LANDSAT and SPOT imagery.

New space platforms are currently being planned which increase both the spatial and spectral resolution of remotely sensed information. The wealth of data that these promise to provide could overwhelm biologists unless a start is made soon on developing techniques of dimension reduction that provide useful, large-area information for each of the many problems of research and conservation that will need to be addressed in the very near future. Literally and metaphorically we would like to see both the wood *and* the trees.

ACKNOWLEDGEMENTS

We thank Dr William Wint of the Environmental Resource Group Oxford (ERGO) for help and advice with the interpretation of the Fourier images, for the loan of slides used during the presentation of the talk at the British Ecological Society Symposium, for reading and commenting on the manuscript and for permission to use information in two reports from Resource Inventory and Management Ltd (RIM 1987 and 1992), for the map of Plate 11.2 and for Plate 11.4a.

The climatic data for Zimbabwe were kindly provided by Trevor Booth, CSIRO, Australia, and for East Africa by Dr Tim Robinson. Satellite sensor data were provided by the FAO Remote Sensing Centre, Rome, courtesy of Jelle Hielkema and by USAID/FEWS NASA GSFC, courtesy of Barry Henricksen.

Financial support was provided by the UNDP/World Bank/WHO Special Programme for Research and Training in Tropical Diseases, project 860010, and all image processing was carried out using equipment provided by the NERC, project GR3/7524 (to DJR).

REFERENCES

Bartholomé, E. (1988). Radiometric measurement and crop yield forecasting: some observations over millet and sorghum experimental plots in Mali. *International Journal of Remote Sensing*, **9**, 1539–1552.

Buxton, P.A. (1955). *The Natural History of Tsetse Flies.* Memoir of the London School of Hygiene and Tropical Medicine No. 10. H.K. Lewis, London.

Buxton, P.A. & Lewis, D.J. (1934). Climate and tsetse flies: laboratory studies upon *Glossina submorsitans* and *G. tachinoides*. *Philosophical Transactions of the Royal Society, London, Series B*, **224**, 175–240.

Chatfield, C. (1980). *The Analysis of Time Series: an Introduction.* Chapman & Hall, London.

Cracknell, A.P. & Hayes, L.W.B. (1991). *Introduction to Remote Sensing.* Taylor & Francis, London.

Ford, J. (1971). *The Role of the Trypanosomiases in African Ecology: a Study of the Tsetse Fly Problem.* Clarendon Press, Oxford.

Ford, J. & Katondo, K.M. (1977). *The Distribution of Tsetse Flies* (Glossina) *in Africa 1973.* Organization of African Unity – Scientific and Technical Research Commission. Cook, Hammond & Kell, London.

Goward, S.N. & Huemmrich, K.F. (1992). Vegetation canopy PAR absorptance and the normalised difference vegetation index: an assessment using the SAIL model. *Remote Sensing of Environment*, **39**, 119–140.

Green, P.E. (1978). *Analyzing Multivariate Data.* The Dryden Press, Hinsdale, Illinois.

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

Hoare, C. (1972). *The Trypanosomes of Mammals.* Blackwell Scientific Publications, Oxford.

Howell, P., Lock, M. & Cobb, S. (Eds) (1988). *The Jonglei Canal: Impact and Opportunity.* Cambridge University Press, Cambridge.

Huete, A.R. & Jackson, R.D. (1987). Suitability of spectral indices for evaluating vegetation characteristics on arid rangelands. *Remote Sensing of Environment*, **23**, 213–232.

Hugh-Jones, M. (1989). Applications of remote sensing to the identification of the habitats of parasites and disease vectors. *Parasitology Today*, **5**, 244–251.

Huh, O.K. (1991). Limitations and capabilities of the NOAA satellite advanced very high resolution radiometer (AVHRR) for remote sensing of the Earth's surface. *Preventive Veterinary Medicine*, **11**, 167–183.

Jackson, R.D. & Huete, A.R. (1991). Interpreting vegetation indices. *Preventive Veterinary Medicine*, **11**, 185–200.

Jeffers J.N.R. (1978). *An Introduction to Systems Analysis: with Ecological Applications.* Edward Arnold, London.

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

Menenti, M., Azzali, S., Verhoef, W. & van Swol, R. (1991). Mapping agro-ecological zones and time lag in vegetation growth by means of Fourier analysis of time series of NDVI images. Report No. 32, SC-DLO. Winand Staring Centre, Wageningen, The Netherlands.

Molyneux, D.H. & Ashford, R.W. (1983). *The Biology of* Trypanosoma *and* Leishmania, *Parasites of Man and Domestic Animals.* Taylor & Francis, London.

Nash, T.A.M. (1948). *Tsetse Flies in British West Africa.* HMSO, London.

Nix, H. (1986). A biogeographic analysis of Australian elapid snakes. In: *Atlas of Elapid Snakes of Australia.* (Ed. by R. Longmore), pp. 4–15. Australia Flora and Fauna Series No. 7. Australian Government Publishing Service, Canberra, Australia.

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.* Technical Centre for Agricultural and Rural Cooperation, Commission of the European Communities.

RIM (1987). *Gezira Livestock Integration Study, Final Report Vol. 4. Livestock, Human and Environmetal Resources.* Devco, Dublin.

RIM (1992). *Nigerian Livestock Resources.* Resource Inventory and Management, St Helier, Jersey.

Rogers, D.J. (1979). Tsetse population dynamics and distribution: a new analytical approach. *Journal of Animal Ecology*, **48**, 825–849.

Rogers, D.J. & Randolph, S.E. (1985). Population ecology of tsetse. *Annual Review of Entomology*, **30**, 197–216.

Rogers, D.J. & Randolph, S.E. (1986). The 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**, S77–S92.

Shahin, M. (1985). *Hydrology of the Nile Basin.* Elsevier, Amsterdam.

Snyder, J.P. (1987). *Map Projections – a Working Manual.* US Geological Survey Professional Paper 1395. US Government Printing Office, Washington.

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

Ter Braak, C.J.F. & Prentice, I.C. (1988). A theory of gradient analysis. *Advances in Ecological Research*, **18**, 271–317.

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.

Tucker, C.J. & Sellers, P.J. (1986). Satellite remote sensing of primary production. *International Journal of Remote Sensing*, **7**, 1395–1416.

Tucker, C.J., Townshend, J.R.G. & Goff, T.E. (1985). African land-cover classification using satellite data. *Science*, **227**, 369–375.

Van de Griend, A.A., Camillo, P.J. & Gurney, R.J. (1985). Discrimination of soil physical parameters, thermal inertia and soil moisture from diurnal surface temperature fluctuations. *Water Resources Research*, **21**, 997–1009.

Xue, Y. & Cracknell, A.P. (1992). Thermal inertia mapping: from research to operation. *Proceedings of the 18th Annual Conference of the Remote Sensing Society* (Ed. by A.P. Cracknell & R.A. Vaughan), pp. 471–480. Remote Sensing Society, Nottingham.