

FINAL TECHNICAL REPORT

Integrated satellite and field data analysis of tsetse and trypanosomiasis research and control programmes.

NRI Extra Mural Contract: ZC0012, R6626

Collaborating Institution: Department of Zoology, Oxford University, South Parks Rd., Oxford OX1 3PS

PI: Dr D.J. Rogers

Personnel: Dr Simon Hay, Ir Bernhard Bakker

Project Location: Department of Zoology, Oxford

NRI Liaison Officers: Professor M. Gill and Dr Wynn Richards

NRI Strategy Area: Resource Assessment and Farming Systems

Programme Area: Livestock Production

Problem Area: Livestock Protection

Date: September 1996 - September 1999

Total financial support £223,717

TABLE OF CONTENTS

Final Technical Report	1
Executive Summary	3
Background	5
Project Purpose	6
Research Activities	6
Data sets used in the study	6
Tsetse/trypanosomiasis data sets	7
Satellite data sets	7
SECTION 1	8
Deriving meteorological variables from passive sensor satellite data	8
Ground-based meteorological data	8
Satellite data - sources and initial processing	8
Satellite data - Temperature	9
Satellite data - Atmospheric moisture	9
Satellite data - Rainfall	10
Predictive accuracy of remote sensing	10
Spatial interpolation methods	10
Predictive accuracy of spatial interpolation	12
Comparisons of the accuracy of RS and spatial interpolation methods	12
SECTION 2	19
Deriving environmental information from active sensor satellite data	19
Instrument	20
Date	20
Bands	20
Polarisation	20
Resolution (m)	20
Swath width	20
Characteristics of radar signals	20
Band designation	20
Nominal frequency range	20
Wavelength (cm)	20
Characteristics of radar images	21
SAR imaging	23
Correlations between active and passive sensors for selected study sites	23
SECTION 3	25
Alternative time-series analytical methods	25
SECTION 4	32
Increasing the spatial resolution of AVHRR data using wavelet techniques	32
A brief guide to wavelet transforms	33
Correlating the imagery at the same spatial scales	34
Introducing spatial detail into the AVHRR image	34
SECTION 5	40
Analysis of tsetse population data from selected field sites	40
Predictive biological risk maps of tsetse	41
Analysis of trypanosome infection data from selected field sites	41
SECTION 6	47
Synthesis	47
Acknowledgements	50
Selected Project Outputs	51
APPENDIX	53
A simple model for trypanosome transmission.	53
BIBLIOGRAPHY	55

Executive Summary

The aims of the project were as follows:

- 1) to reach a biological understanding of trypanosomiasis transmission in a number of key study sites.*
- 2) to integrate high- and low-resolution satellite data for the study sites.*
- 3) to relate the biological models of Step 1 and the statistical models of Step 2.*
- 4) to model representative data sets using satellite imagery and to define the minimum field data necessary for acceptable accuracy.*
- 5) to verify the approach by using independent fly and disease data sets.*
- 6) to use the models to define optimal disease control strategies.*
- 7) to disseminate the results and run a Training Workshop for middle/advanced level personnel (workshop dependent upon additional funds).*

The first aim was achieved for The Gambia and Kenya study sites, within the limits of the data sets available (Section 5). Satellite derived Land Surface Temperature measures were the best predictors of monthly fly mortality rates in both sites, and the resulting models explained > 90% of the variance of fly numbers, as recorded by the traps. The transmission models developed within the project were able to explain > 50% of the variance in monthly infection prevalences in cattle at each site, a remarkably good figure given the simplifying assumptions that had to be made in the models. Extension of the tsetse models to larger areas seems feasible, and such models can form the basis for modelling intervention at a number of scales.

The second aim was achieved by showing how different sorts of data, with different spectral, spatial and temporal resolutions may be brought together to provide a unique, multi-variate picture of epidemiological landscapes (Sections 2 and 4), through the use of wavelet techniques, here applied for the first time to the combination of Landsat and AVHRR data.

The third aim was achieved by demonstrating how our statistical and biological approaches may be unified by a common set of multi-temporal satellite data that both describes statistically and helps to explain biologically the tsetse and disease data, in both space and time.

The fourth aim was achieved by applying generic models to tsetse and disease transmission, using standard sets of field data. There is no particular need for more types of fly data but the Report emphasises that it is important to ensure that the fly data apply to the cattle disease data. Usually these data sets are collected by different teams of workers, not always seeking a common epidemiological overview of the problem .

The fifth aim was difficult to achieve within the limitations of the data sets available to the Project. Ideally we need sets of data spanning a range of ecological conditions: the models will then allow us to fill in the gaps between representative sample sites.

The sixth aim may be realised within the context of particular project sites, and Section 6 gives a brief example of this.

The seventh aim of the project was realised during a brief work-shop held in Addis Abeba, Ethiopia, with the Regional Co-ordinators charged with gathering data on poverty nation-wide. Presentations were given high-lighting the potential importance of RS/GIS methods to poverty and health mapping.

During the project, collaborations with other organistaions and institutes were strengthened. TALA personnel are now actively involved with NASA's Interagency Partnership for Infectious Diseases (INTREPID) project and the pan-African MARA project mapping malaria risk in Africa. Collaboration with ERGO continues, and has resulted in a number of further FAO Reports and the electronic updating of the Ford & Katondo tsetse distribution maps, using TALA-processed satellite data.

Background

Work during the first Trypanosomiasis and Land-use in Africa (TALA) project ('TALA1', Reference No: X0239, *Identifying the constraints on livestock productivity and land-use in Africa imposed by trypanosomiasis*) concentrated on the *statistical* relationships between remotely sensed satellite variables and ground-based meteorological, ecological and epidemiological variables of the tsetse transmitted trypanosomiasis, with a view to extending these predictions in both space and time. The achievements of TALA1 included:

- Integration of ground-level and satellite-derived variables within a geographical framework.
- Development of methodologies for dimensional reduction of satellite data without loss of biological information. This is especially important for the application of high-temporal resolution data such as NOAA AVHRR imagery which, despite their relatively poor spatial resolution, contain useful information on habitat distribution and seasonality.
- Establishment of the statistical relationships between the criterion (dependent) variables (tsetse distribution and abundance and trypanosomiasis prevalence) and predictor (independent) variables (remotely sensed satellite data). As far as possible these statistical relationships, based on linear and non-linear discriminant analysis, were improved by the application of a simple understanding of the processes involved in generating the data, or through the use of new indices appearing regularly in the literature. These indices covered both the 'vegetation' channels (AVHRR Ch1 and Ch2; NDVI, GEMI and SAVI indices) and the thermal channels (Ch4 and Ch5; split window algorithms).
- Examination of areas where the best statistical predictions were apparently in error when compared with the original data. This usually led to a modification of the analytical methods and an improvement in the accuracy of the classifications.
- Application of the best statistical correlations to make predictions about the distributions of resources (*i.e.* vegetation type), vectors (tsetse) and disease (trypanosomiasis) in other places (usually adjacent countries or regions).

The results of TALA1 were most encouraging. We were able to predict the distribution of dominant vegetation types with accuracies up to 96 %, the distribution of tsetse with accuracies up to 95 % and the prevalence of the African animal trypanosomiasis with accuracies up to 74%.

The logical 'next step' was to integrate this statistical approach with a *biological* analysis of field data on tsetse/trypanosomiasis in Africa in order to establish the biological mechanisms underlying the statistical relationships. This integration should lead to process-based, mechanistic or biological models which are the only firm foundation on which sustainable, economic intervention strategies can be planned. This was the objective of the current project, entitled 'TALA2'.

Project Purpose

Purpose: To develop ecologically sound and sustainable tsetse and trypanosomiasis control programmes at local levels.

Objectively verifiable indicators: Local and regional maps identifying ecological areas within which alternative control strategies are optimal.

Means of verification: Uptake of results by NARs and NGOs. Changes in local approaches to tsetse control.

Assumptions: Tsetse/trypanosomiasis continues to be recognised as major constraint on livestock production in Africa. Tsetse control still supported.

Research Activities

TALA2 had a number of aims, as set out in the original project document. The way in which the project addressed these aims, and the results obtained, form the subject of this Report. How the results meet the project purpose is discussed at the end of this section of the Report.

The aims of the project were as follows:

- 1) to reach a biological understanding of trypanosomiasis transmission in a number of key study sites.*
- 2) to integrate high- and low-resolution satellite data for the study sites.*
- 3) to relate the biological models of Step 1 and the statistical models of Step 2.*
- 4) to model representative data sets using satellite imagery and to define the minimum field data necessary for acceptable accuracy.*
- 5) to verify the approach by using independent fly and disease data sets.*
- 6) to use the models to define optimal disease control strategies.*
- 7) to disseminate the results and run a Training Workshop for middle/advanced level personnel (workshop dependent upon additional funds).*

The original project purpose is encapsulated in the first aim (above) and the remaining project aims are concerned with the means by which this first aim can be reliably achieved, and the ends to which any resulting understanding might be put. The present report deals with each aim in turn, although it will be appreciated that a full understanding of the discussion of the first aim may depend upon the results of activities dealt with under later aims.

Data sets used in the study

This section describes the sources of the data used in this study, and the manipulation of these data sets for use by the project.

Tsetse/trypanosomiasis data sets

The first project activity involved gathering together data sets from past projects on tsetse and trypanosomiasis in a variety of habitat types in Africa, and establishing a common data format for data storage within the project. The project also gathered new sets of passive satellite data (both AVHRR and Landsat) and sets of active satellite (radar) data for many of the field sites from which the tsetse/trypanosomiasis data came.

We are most grateful to our collaborators for providing us with the tsetse/trypanosomiasis data used in the project. Without such careful data management we would not have been able to make any progress in TALA2. The data came to us in database or spreadsheet formats that were easily manipulated within MS Access and Excel for further analysis.

For a number of studies, information on the tsetse populations and infection rates and on local cattle populations and their infections were recorded on a monthly basis for periods of up to several years. The value of such data increases considerably with their duration. One or two years' of data are a minimum for any new area, to sample a variety of local seasons. Several more years are required to see how the interactions between flies and their hosts may be modulated by unusual weather patterns. Long period data sets involve another problem, however, of changing habitats or herd grazing patterns in relation to local fly challenge. In general, databases are rather poor at documenting such changes, and we have had to rely on conversations with the original data gatherers who alerted us to changing trap positions, fly suppression schemes, or novel herd management practices. When these occurred we have had to split longer period data sets before carrying out separate analyses.

Satellite data sets

NOAA AVHRR satellite data came from the public domain Pathfinder source (<http://edcwww.cr.usgs.gov/landdaac/>) and was routinely temporal Fourier processed to extract the 'seasonal fingerprints' in vegetation index and thermal channels (see the Final Technical Report of TALA1 for more details). METEOSAT Cold Cloud Duration (CCD) imagery was obtained from the ARTEMIS program at FAO, Rome. In addition Landsat MSS and TM data were obtained from the USGS (Souix Falls) via collaborative agreements with NASA Ames, and radar imagery was obtained from the European and Japanese radar satellite archives.

SECTION 1

Deriving meteorological variables from passive sensor satellite data

Investigations were undertaken in TALA1, correlating satellite and meteorological variables throughout Africa. One important activity of TALA2 was to investigate the reliability of using satellite data to estimate ground-level meteorological conditions quantitatively. This is especially important for the development of process-based biological models that depend upon laboratory or field-derived relationships between developmental and infection processes and air or water temperatures. For satellites to provide a useful proxy for these variables, their accuracy at estimating them remotely must first be assessed. Alternative approaches to estimating meteorological conditions on the ground often involve some sort of extrapolation/interpolation of ground level conditions at generally sparsely distributed meteorological station sites. Recent developments in spatial interpolation methods have significantly improved the accuracy of estimation of weather variables at sites far away from meteorological stations (Lennon and Turner, 1995) and such spatial interpolation may provide a better measure of local climates than direct measurement by satellites through the variable and occasionally unpredictable atmospheric column. The relative performance of remotely sensed (RS) and spatially interpolated (SI) estimates of local climate was investigated through collaboration between TALA2 (Dr. Simon Hay) and Dr. Jack Lennon using a variety of ground-based meteorological station records and data from the NOAA and METEOSAT satellites.

Ground-based meteorological data

Temperature, atmospheric moisture and rainfall surfaces were independently derived from spatial interpolation (SI) of measurements from the World Meteorological Organisation (WMO) member meteorological stations of Africa (from NOAA-NCDC), using the methods of (Lennon and Turner, 1995).

Satellite data - sources and initial processing

Daily NOAA-11 imagery for 1990 for Africa were obtained from the Pathfinder (PAL) site (James and Kalluri, 1994). The data include brightness temperatures (channels 4 and 5) and the Normalised Difference Vegetation Index (NDVI) (Myneni and Asrar, 1994). Data were extracted and scaled using information provided in (Agbu and James, 1994). Additional bands of ancillary information were used for quality control. Firstly, those pixels contaminated by clouds as determined by the "Clouds from AVHRR (CLAVR)" algorithm (Stowe *et al.*, 1991) were masked. Secondly, pixels viewed by the AVHRR sensor at an angle of greater than 42° were eliminated to reduce error introduced by observing the Earth at large view angles. Thirdly, those pixels recorded at solar zenith angles of greater than 80° were excluded because at these angles the twilight of dawn and dusk affects measurements. Land

surface temperature (LST) and VPD surrogate information were then derived from these corrected daily PAL data.

Satellite data - Temperature

As detailed in the Final Technical report of TALA1, AVHRR Channel 4 and 5 data may be used in Price's split-window approach (Price, 1984) to measure land surface temperature (LST), T_P (Kelvin) as follows:

$$T_P = Ch_4 + 3.33(Ch_5 - Ch_4)$$

where Ch_4 and Ch_5 refer to the brightness temperatures in the respective AVHRR channels. This equation has a stated LST accuracy of $\pm 2-3$ K.

The daily temperature data were then maximum value composited (MVC) by taking the maximum temperature for a given pixel over the monthly period. The rationale for this procedure is that clouds are generally colder than the land at tropical latitudes, so that the highest thermal value in the series will probably be the least cloud contaminated (Lambin and Ehrlich, 1995).

Satellite data - Atmospheric moisture

In order to estimate humidity, some estimate is required of the water content of the air at ground level. This is generally derived from the total precipitable water content U (kg/m^2) of the atmosphere (*i.e.* the water content of the entire atmospheric column), which may also be estimated from AVHRR data (Eck and Holben, 1994) where;

$$U = A + B(Ch_4 - Ch_5)$$

with A and B constants, generally given values of 1.337 and 0.837 respectively. U is first converted to its equivalent in cm, by dividing by 10 (since 1kg of water over 1m^2 will have a depth of 0.1 cm), before conversion to a near surface dew point temperature, T_d ($^{\circ}\text{F}$), using the following relationship (Smith, 1966);

$$T_d = \frac{\ln U - ((0.113 - \ln(\lambda + 1)))}{0.0393}$$

where λ is a variable that is a function of the latitude and the time of the year. In this analysis a mean value of $\lambda = 2.99$ was calculated from the annual mean λ presented by Smith (1966) for locations between 0 and ± 40 degrees of latitude.

The dew point temperature values were then converted into Kelvin and used with the Price (Price, 1984) estimate of land surface temperature, T_p (K), to calculate the vapour pressure deficit, VPD (KPa), using the equation provided in (Prince and Goward, 1995);

$$VPD = 0.611 \left[\exp\left(17.27 \times \frac{T_p - 273}{T_p - 36}\right) - \exp\left(17.27 \times \frac{T_d - 273}{T_d - 36}\right) \right].$$

These data were then subject to MVC by using the VPD from the date in the month when the NDVI was highest, because the NDVI value is generally reduced by cloud and other atmospheric contamination. The highest NDVI values recorded over any relatively short period therefore occur when cloud cover and atmospheric contamination are least and such values are taken to represent the least attenuated pixel value for the period ((Holben *et al.*, 1986; Stoms *et al.*, 1997)).

Satellite data - Rainfall

In the CCD imagery each pixel's value represents the number of hours for which the area it represents was covered by cold (*i.e.* rain-bearing) clouds during the compositing period. These values were therefore taken as a measure of rainfall, without further data manipulation.

Predictive accuracy of remote sensing

The latitude and longitude of each meteorological station was used to extract the corresponding RS pixel value from the satellite sensor data. Meteorological stations located within 20 km of the coast and large inland lakes and rivers were excluded from the RS analysis (leaving $n = 207$) because the corresponding pixel in the satellite image could have been contaminated by the signal originating from these water bodies.

The RS LST predictions were consistently higher than meteorological station screen temperatures. In order to quantify the accuracy with which RS predicts screen temperatures a linear regression of screen temperature against LST was applied. Similarly, for the purposes of accuracy assessment, rainfall was regressed against CCD and humidity against VPD. From these regressions, the root mean square error (RMSE) and the coefficient of determination (r^2) were used as goodness of fit statistics for each monthly comparison. The RMSE describes the accuracy with which climate at a new site may be predicted. The problem of missing meteorological station reports and the masking of cloud-affected data in the MVC imagery led to variation in the number of observations and the subset of sites investigated each month. For this reason the population adjusted coefficient of determination, the adjusted r^2 , are shown throughout (Sokal & Rohlf 1995).

Spatial interpolation methods

The interpolation method used has been described in (Lennon and Turner, 1995) and consists of a composite Digital Elevation Model (DEM) regression with variable selection and thin-plate spline algorithm. This is an elaboration of Hutchinson's implementation of thin-plate spline interpolation theory (Hutchinson, 1989). Interpolation is a two-stage procedure. First, the DEM and a system of spatial coordinates are used to account for spatial variation in the climatic factors associated

with simple linear spatial trends and topographic effects. Second, the thin-plate spline technique reduces the remaining trends in the residuals from the regression analysis.

Spatial co-ordinate data included in the multiple regression procedure comprised the latitude, longitude, and altitude co-ordinates, with the addition of two other co-ordinates: the cosine of latitude and the distance to the sea. It was considered necessary to include oceanic influences on climate, while the latitude transformation potentially represents the spatial rate of change of energy input more accurately than simple latitude alone. The three basic co-ordinates of latitude, longitude and altitude were forced to be present in each of the regression models.

Simple functions summarising the local topography were extracted from the original 1 x 1 km spatial resolution DEM. Around each site two quadrats were defined: a large quadrat (21 x 21 pixels) and, nested within this, a small quadrat (5 x 5 pixels). Within these quadrats four functions of altitude (the mean, maximum, minimum and standard deviation) of the regularly-spaced elevations were extracted. Additionally, the slopes to the south and to the west were calculated. This set of 10 DEM variables, together with the 5 spatial co-ordinate variables, were used as independent variables in a multiple regression analysis with the climatic factor as the dependent variable. Each of the three climate variables, for each of the 12 months, was individually regressed against the set of 15 covariates using a variable selection procedure (the stepwise algorithm, SAS 1990), in which the independent variables were entered and removed from the model at a basic significance level of $P < 0.05$. Since there were 12 independent variables (the three basic spatial co-ordinates were forced to be present), this translated to $P < 0.004$ for the entry and exit criteria after application of the Bonferroni correction (Sokal & Rohlf 1995).

The 12 months of the year were considered separately and the significant DEM variables changed between months because of variation in their individual contributions to the predictive power of the model. Those DEM variables that were significant in six or more months of the year were selected for subsequent analyses. In doing this a balance was struck between the use of a different set of topographic factors for each month and using all factors uncritically (Lennon and Turner, 1995). While a case may be made for using the former approach on the grounds that some topographic factors might only be important in particular months, there is a trade-off between including potentially non-significant factors (*i.e.* different topographic factors for each month) and using the same topographic factors each month: the latter may capture more completely the underlying causal trends than the former.

Application of the thin-plate spline routine to the residual variation left after the regression analyses automatically involves a generalised cross validation algorithm. This involves the attempted optimisation of the 'roughness' of the fitted surface, such that a good predictive fit is made. The reason for doing this is that it is easy to fit a surface that goes through all of the given data points exactly, but this is no guarantee of a good predictive fit. The thin-plate spline algorithm therefore smoothes the surface to obtain an optimal predictive fit. It achieves this by omitting each given data point in turn before fitting a provisional surface and adjusting the roughness parameter; the value of the roughness parameter that minimises the prediction error of the omitted data points is selected.

To quantify the accuracy of fit, in terms of the predictive power of the interpolation, this cross validation method was taken one stage further. Each station was omitted in turn before applying the thin-plate spline. This is quite distinct from the intrinsic cross-validation method within the thin-plate spline technique since it involved running the entire thin-plate spline procedure once for each available climate station. This procedure of dropping a station before fitting the spline surface ensures that the information is completely absent from the surface, and hence that a prediction for the omitted site based on this surface is completely independent of its known value. This is an advance over simply applying the spline surface over all stations at once and examining the residuals for accuracy, irrespective of the inherent cross validation calculations within the spline procedure.

Predictive accuracy of spatial interpolation

The predicted value for each site omitted from the thin-plate spline routine was compared in turn to the observed value. The RMSE and the population adjusted coefficient of determination of the predictions regressed against actual meteorological variables for the network of meteorological station sites were again the goodness of fit statistics.

Comparisons of the accuracy of RS and spatial interpolation methods

The accuracy of the alternative methods is summarised in Figs 1 - 3 and Table 1 (for more details see (Hay and Lennon, 1999)). Fig. 1 shows a comparison of the monthly RMSE of the satellite and SI estimates of temperature, rainfall and vapour pressure deficit, Fig. 2 shows continental maps of temperature predicted by the two methods for April, August and December 1990 and Fig. 3 shows continental maps of vapour pressure deficit and rainfall predictions for April 1990.

The predictive accuracy of SI temperature estimates (2.3 °C RMSE, fig. 1a) compares well with the accuracies reported by (Hulme *et al.*, 1996) on an averaged monthly data set derived from 30-year records (0.8 - 1.4 °C RMSE for maximum and minimum temperatures; they did not consider mean temperature). Shorter time series, as used here, are likely to show more relative stochastic variability.

The accuracies achieved using RS over continental Africa (RMSE 4 °C, fig. 1A) is approximately twice that reported for studies that have conducted daily comparisons of 1 x 1 km spatial resolution NOAA-AVHRR data, usually chosen for small areas on cloud free days (average RMSE ~2 °C; (Cooper and Asrar, 1989; Sugita and Brutsaert, 1993)). The sites used in the present study, however, were situated in habitats as diverse as deserts and rainforests and were at altitudes spanning a range of 2000m, so the greater variability of the RS measurements here is understandable.

Fig. 1.1. Comparison of RMSE of satellite (closed squares) and SI estimates (open squares) of A) temperature, B) rainfall and C) vapour pressure deficit.

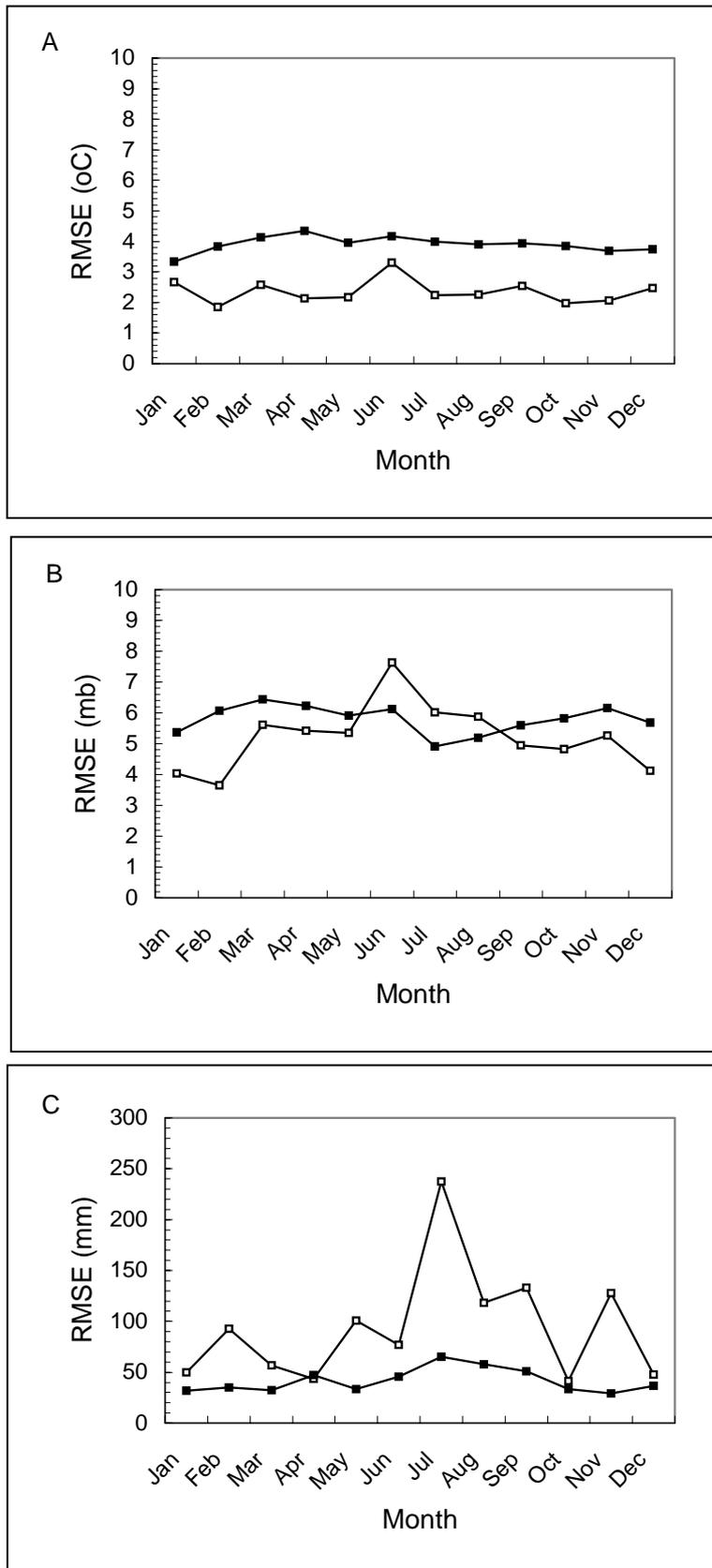


Table 1.1 The mean annual accuracy of spatial interpolation (SI) and remote sensing (RS) for predicting the land surface temperature, atmospheric moisture and rainfall across Africa in 1990.

	<i>RMSE</i>		<i>Coefficient of determination</i>	
	<i>RS</i>	<i>SI</i>	<i>RS</i>	<i>SI</i>
<i>Land surface temperature</i>	<i>4.0 °C</i>	<i>2.3 °C</i>	<i>0.52</i>	<i>0.83</i>
<i>Vapour pressure deficit</i>	<i>6.0 mbar</i>	<i>5.3 mbar</i>	<i>0.63</i>	<i>0.78</i>
<i>Rainfall</i>	<i>38 mm</i>	<i>93.7 mm</i>	<i>0.63</i>	<i>0.19</i>

The SI technique was therefore more accurate in its predictions of screen temperature, perhaps because of its use of additional variables such as altitude which can account for a considerable amount of residual variance (Hay *et al.*, 1996). Elevation has not been explicitly parameterised in any RS split-window equations but should be considered where application is required at the broad spatial scale, or where the range in local elevation is large. In the same way latitude, longitude and distance to the sea functions might be incorporated into the RS techniques to improve their performance relative to the SI techniques. A second difference between the two methods is that the satellite sensor records a surface temperature averaged over an 8 x 8 km area not at a single point of a Stevenson screen position. It remains to be established which measurement is more important to the population dynamics of disease vectors and how such measurements relate to the micro-climates actually experienced by the vectors.

There is little difference in the accuracy of RS and SI for predicting VPD (fig. 1b), although the satellite method has a less variable RMSE throughout the year. There appear to be no published attempts to validate VPD estimates from SI or RS at any spatial scale. Thirty year vapour pressure averages however, have been mapped for Africa south of the equator (Hulme *et al.*, 1996) with accuracies of 12% (N.B. RMSE is here expressed as a percentage of the mean). Expressed similarly the mean annual accuracy for SI of VPD in this study was 33%. This difference is probably again a consequence of the short time series of data, the larger geographical extent covered, the finer spatial resolution of the interpolations and the smaller number of stations used to interpolate the surfaces.

Finally, RS is substantially more accurate in predicting rainfall than is SI (fig. 1c). This is consistent with the large spatial variability of rainfall over short distances. For example, measurements over a season have been shown to vary by a factor of two

over distances of less than 10 km in West Africa (Beek *et al.*, 1992; Flitcroft *et al.*, 1989)).

Hulme *et al.* (1996) present accuracy figures for SI of rainfall of 17% for January and 106% for July (Hulme *et al.*, 1996). In similar percentage units, the annual accuracy for SI implemented in this study is 118%, with a large variation in predictive accuracy between months. Interestingly, the SI conducted here gives good results in January but also breaks down in July (fig. 1c), suggesting that some months are intrinsically more difficult to map accurately for precipitation than others, more or less regardless of the time series considered. Again the RS method shows less month-to-month variation in RMSE than does the SI method.

Pinping and Arkin have estimated monthly precipitation for the globe (1987-1995) at a $2.5 \times 2.5^\circ$ spatial resolution using outgoing longwave radiation estimates from the NOAA-AVHRR (Pinping and Arkin, 1998). They report a RMSE of 54% between the latitudes of 20° S and 20° N. (Herman *et al.*, 1997), have also recently established a technique that uses a combination of RS from Meteosat-HRR and ground data to obtain decadal (10 day rainfall estimates for Africa). Preliminary estimates of accuracy quoted for the Sahel region from June to September 1995 are 40 % of the measured precipitation value. Expressed in similar units the RMSE for precipitation estimates from Meteosat compares favourably at approximately 45%.

Fig. 1.2. A comparison of remotely sensed (above) and spatially interpolated (below) temperature estimates for Africa for April (left), August (middle) and December (right) for 1990. Background colour within Africa = no RS estimate for that period.

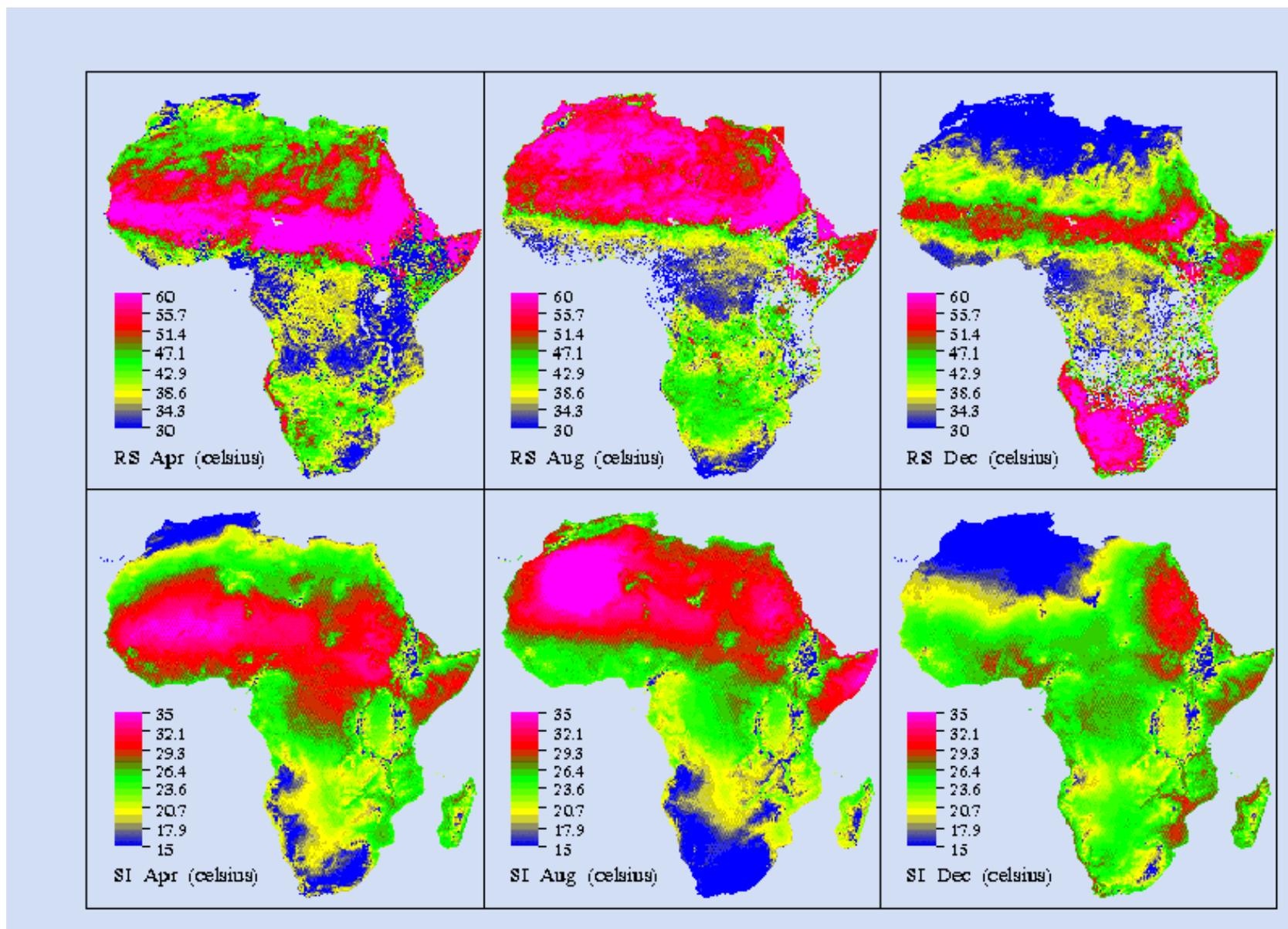
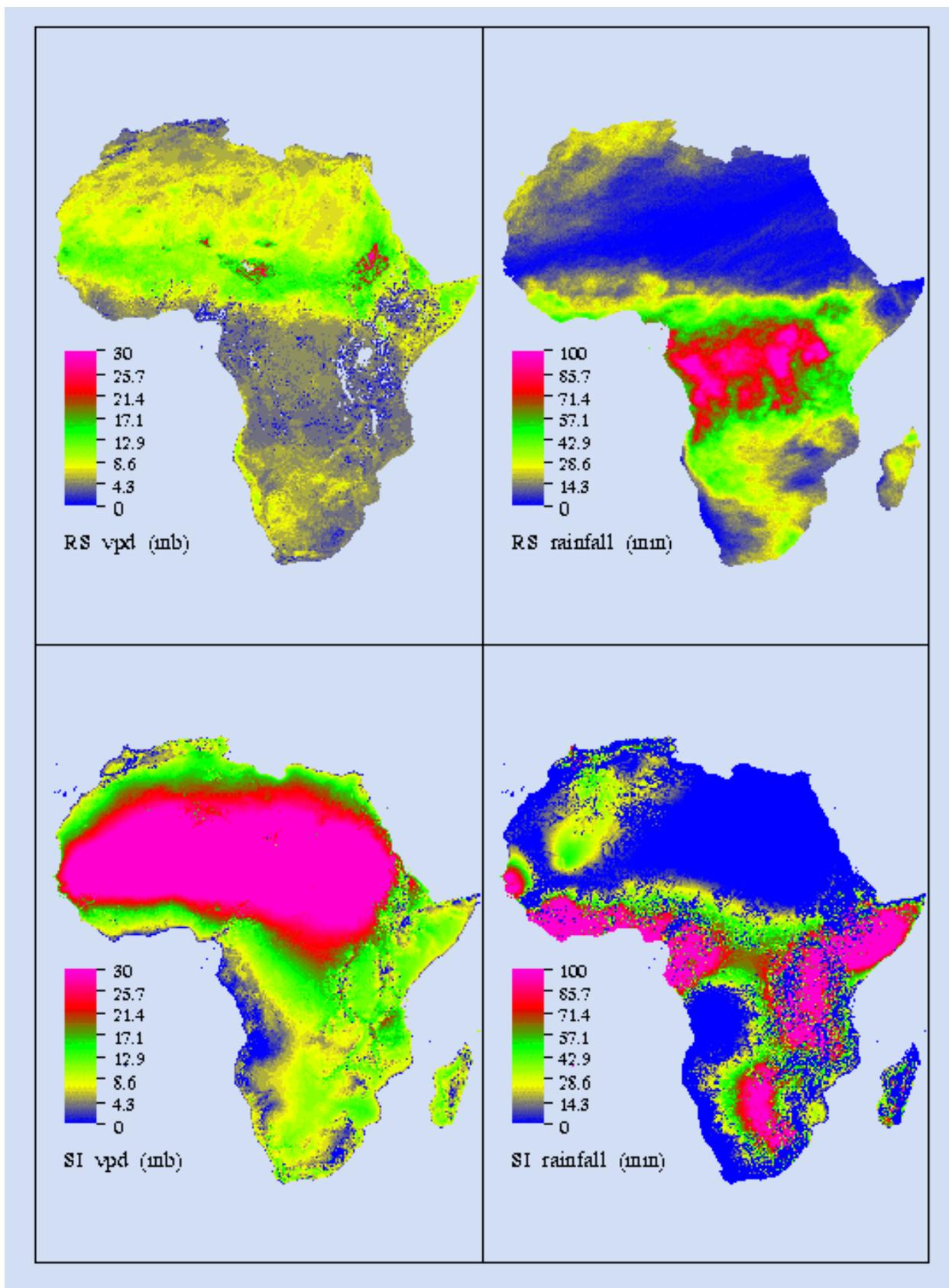


Fig. 1.3 A comparison of remotely sensed (above) and spatially interpolated (below) vapour pressure deficit (left) and rainfall (right) estimates for Africa for April 1990.



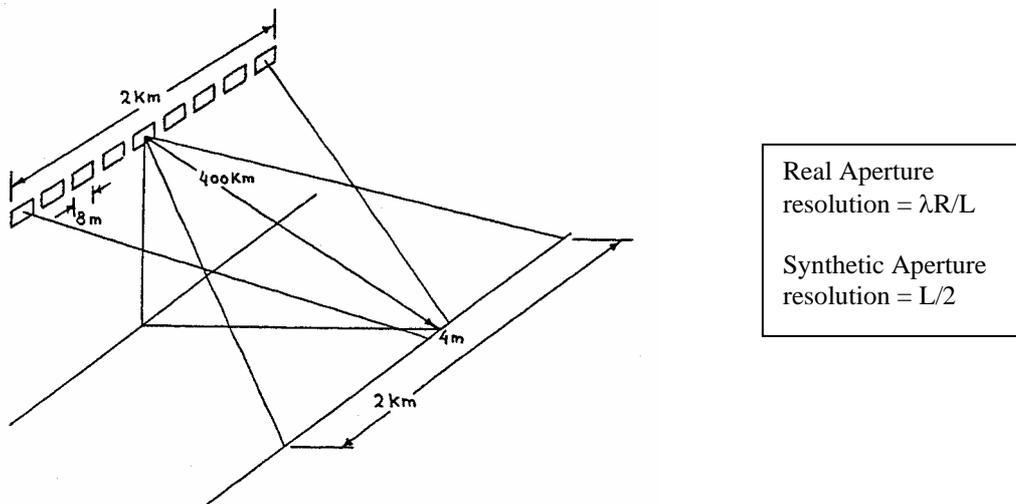
SECTION 2

Deriving environmental information from active sensor satellite data

One of the aims of TALA2 was to investigate the application of radar imagery to habitat and disease risk mapping. There are considerable potential advantages in using radar imagery, since it is unaffected by cloud contamination, although there are also numerous image processing difficulties involved in this application. Part of the project therefore investigated whether the benefits of using radar imagery exceed its costs in terms of labour invested in image processing.

Spaceborne (orbital) Synthetic Aperture Radar imagery (SAR), is a technology that allows prolonged observation of “reflectors” by synthesising a longer antenna and processing the signal as if it were observed from one point in both space and time. In other words, if over a period of time T , the returned signal is stored as amplitudes and phases, it is in principle possible to construct the signal which would have been obtained by an antenna of length vT , v being the speed of the platform. Since T can be made large, this “synthetic aperture” can also be made large.

Fig. 2.1: Advantage of a SAR over SLAR system in a space application. The length (L) of the Real Aperture is 8 m. The length of the Synthetic Aperture is 2 km. In this example the wavelength (λ) is 4 cm and R is 400 km. The resolution of the Real Aperture in the along-track direction is 2 km, The resolution of the Synthetic Aperture in along track is 4m.



Thus the longer the aperture of the antenna the finer the detail it can resolve, and therefore the higher the spatial resolution of the final image. The net effect is a system able to achieve high resolution independent of its altitude [Curlander, 1991 #510]. A summary of the most commonly used orbital SAR systems is given in Table 2.1.

Table 2.1: Specification summary of commonly used SAR satellites.

<i>Instrument</i>	<i>Date</i>	<i>Bands</i>	<i>Polarisation</i>	<i>Resolution (m)</i>	<i>Swath width</i>
ERS-1/2	1991/1995	C	VV	30	100 km
JERS-1	1992	L	HH	18	75 km
Radarsat	1994	C	HH	10 – 28	50-500 km
SIR-C/X-SAR	1996	L, C, X	Quad	10 – 60	15-90 km
EOS-2	2000?	L, C, X	Quad	10 – 120	15-700 km
ENVISAT	1999?	C	VV, HH	30 – 100	100-500 km

Characteristics of radar signals

Radar signals are characterised by wavelength (λ), polarisation, amplitude (A) and phase, all determined at the design stage of the sensor. The plane in which the signal “vibrates” is called its polarisation which may occur in the horizontal (H) or vertical (V) directions (achieved by filtering the outgoing and incoming waves). Interactions with the objects of observation may change the polarisation, amplitude and/or phase of the returned signal. These changes may be object specific and are therefore an important source of information for the discrimination of objects in radar images. In theory, radars can operate in 4 modes: H transmit and H receive (HH), V transmit and V receive (VV) (both called like-polarisations), H transmit and V receive (HV) and V transmit and H receive (VH) (cross-polarisations). In practice, current orbital radar systems are limited to the H or V like-polarised waves. More advanced polarimetric radar systems are required to record all of the information in both polarisations and phase; these are currently available only on airborne, not satellite platforms.

As with the passive satellite sensors, wavelength determines the extent to which objects in the environment may be discerned and also affects signal propagation through the atmosphere. The extent to which the signal is returned depends on the size, shape, orientation and dielectric constant (ϵ) of the object in relation to the wavelength used. Longer wavelengths are less susceptible to atmospheric attenuation and/or dispersal than shorter wavelengths (e.g. wavelengths < 3 cm). The specific wavelengths used in radar platforms are often referred to by a letter code (which stems from its military history). A number of frequencies are specifically reserved for Earth observation, thus reducing interference with other applications such as telecommunications and air traffic control [Hoekman, 1990 #13]. Some of the most commonly used radar bands are listed in Table 2.2.

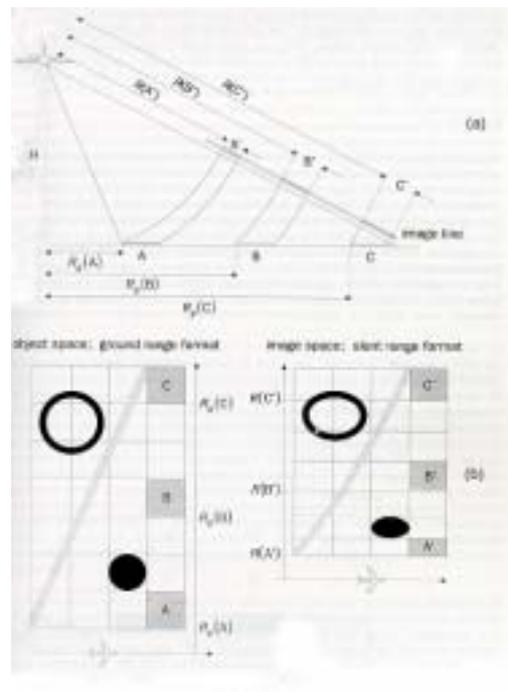
Table 2.2: Standard radar frequency letter-band nomenclature. The frequency range shown are based on the 1979 International Telecommunications Union (ITU) assignment.

<i>Band designation</i> <i>n</i>	<i>Nominal frequency range</i>	<i>Wavelength (cm)</i>
X	8-12GHz	3.47 - 3.51
C	4-8 GHz	5.61 - 5.71
L	1-2 GHz	23.08 - 24.69

Characteristics of radar images

A radar image is a grey scale tonal representation of the strength (intensity) of the backscattered (returned) signals from a single radar pulse, with the grey scale becoming lighter as intensity increases. The most striking feature that makes radar imagery different from any other form of remotely sensed imagery is its geometry. Due to the side-looking configuration, objects are measured in slant range distance from the sensor (in contrast to ground range) and do not therefore have a constant scale.

Fig. 2.2. Image geometry: slant range versus ground range. (a) SLR- systems measure the slant range R rather than ground distance R_g . (b) such configuration leads ultimately to images with distortions in range direction images. (adapted from [van der Sanden, 1997 #5])



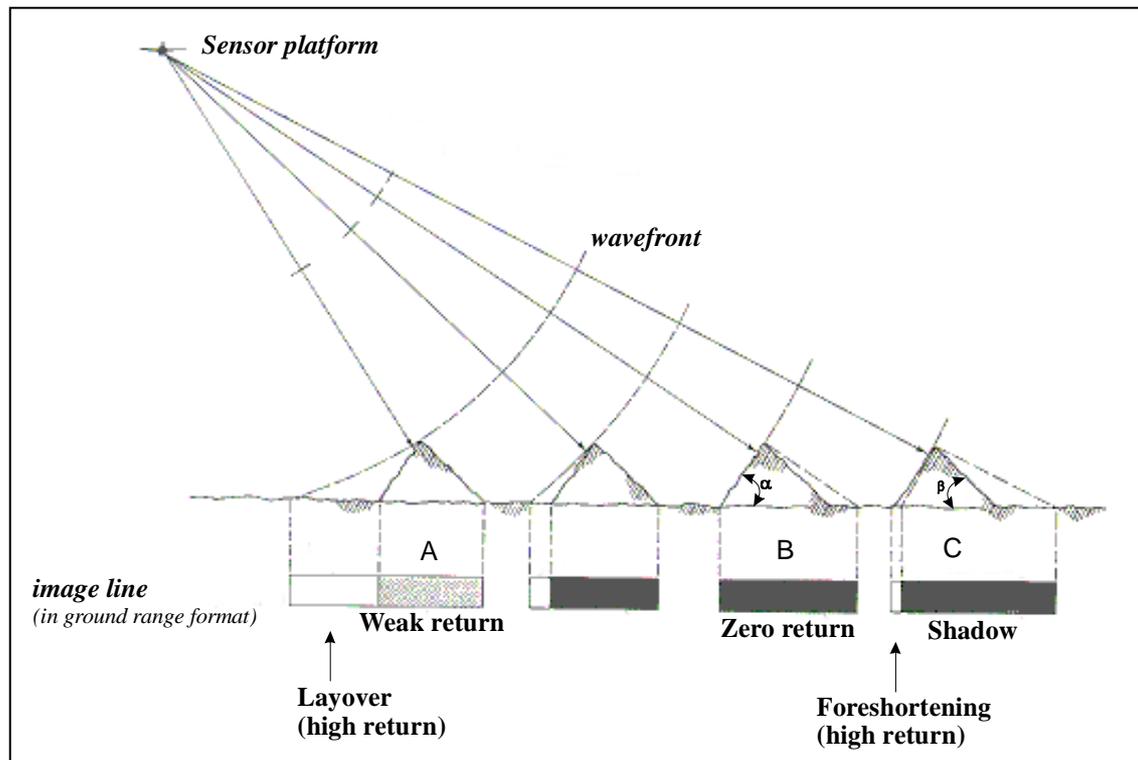
As is illustrated in fig. 2.2 the slant range scale increases with the slant range distance. As a result the image exhibits distortions in range direction. These distortions pose problems for images that require accurate geometry.

In addition, radar images of terrain with varying relief may show different forms of shadowing and layovers (fig. 2.3). Shadowing is an effect of reduced radar return, when slopes face away from the sensor (back slopes). Enhanced returns are received when slopes face the sensor (front slopes) and are referred to as layover areas. The return signals from each slope are therefore affected by the slope's aspect and by its displacement towards the flight direction. For front slopes the aspect causes enhanced reflection and leads to the reflected power being confined to small image regions. Back slopes show weak reflections towards the sensor and result in image spread over a wider area. As a result, areas of foreshortening appear bigger (viewed longer) and areas of layover appear to be smaller than they are in an orthogonal map projection.

The time delay of the returned radar pulse contains the information from which the across-track (range direction) coordinates of scatterers are deduced. This means

that the across-track coordinates of scatterers is represented by its distance (slant range) from the radar's path, and this distance is determined not only by the 'horizontal' coordinates (e.g. latitude and longitude) of the scatterers but also by their altitudes. In addition, distortion in the along-track (azimuth) direction is caused by variations across the swath width of the relative velocity of the Earth's surface and the radar, due to the Earth's rotation.

Fig. 2.3. Geometrical effects in SAR imagery (Adapted from ESA, 1998)



Correction for these effects, generally termed geocoding, involves differentially moving the pixels in a SAR image so that they correspond to a given map projection. The simplest type of geocoding corrects the data for the effects of the Earth's shape assuming that this can be described by an ellipsoid. These corrections are generally made available by the space agency operating the SAR. While such images are suitable for the study of areas having little or no relief, they are inadequate where there is significant topographic variation. A scatterer located at a height h above the ellipsoid will appear displaced by a distance of approximately $h \cot \theta$ towards the sub-satellite track, where θ is the local incidence angle (the angle between the incident radiation and the local vertical). As an example, for the ERS-1 SAR satellite (minimum incidence angle of 19°) terrain exhibiting a range of heights of 1000m may show differential displacements of almost 300 m. Correction for this effect requires a Digital Elevation Model (DEM) of the imaged area, with a horizontal resolution comparable to the spatial resolution of the SAR.

One common feature of SAR imagery is the occurrence of a grainy appearance even of homogeneous areas. This is called 'speckle' and is caused by interference of backscattered signals from individual scattering elements present within one resolution cell. Each of these individual scatterers will produce a certain phase and amplitude of the returned signal and these may interact constructively or destructively

giving a stronger or weaker backscatter, respectively. Radar backscatter can often be improved by "speckle reduction", which is currently achieved by averaging the signals from several different 'looks' at the same area. Since speckle is effectively stochastically random, this averaging process results in images with a less grainy appearance. Unfortunately averaging also results in a reduction in spatial resolution, and so new techniques to overcome this negative effect are being developed.

SAR imaging

As geometrical and electrical characteristics of object vary with time, the challenges of extracting useful information from SAR data reside in how to deconvolve these effects on radar backscatter [Dobson, 1995 #301]. Questions such as "What are the main scatterers (and their characteristics)?" and "What was the dominant mechanism that led to the formation of the measured backscatter signal?" can only be answered through theoretical modelling involving both the geometrical and electrical properties of the backscatterers. Geometrical properties include the size, shape and orientation of objects. Large objects, relative to the wavelength, lead to more direct backscattering patterns than small objects, which exhibit rather broad backscattering patterns. Smooth surfaces cause signals to scatter in the specular (random) direction. The smoother the surface, the more energy is scattered away from the sensor. Thus, structural properties exert considerable influence on the net backscatter [Dobson, 1995 #301]. Electrical properties of backscatterers are determined by their dielectric constant ϵ (the relative electric conductivity of a medium) which is strongly dependent on moisture conditions. As a consequence the solar cycle and atmospheric conditions have pronounced effects on radar backscattering and attenuation. Values of ϵ can range from 3 - 8 for dry surfaces to 80 for water [Dobson, 1995 #301].

As with the visible/ir satellites, radar satellites encounter limitations for land-cover classification due to signal saturation at high levels of biomass and ambiguities between various land-cover types. These difficulties may be overcome by combining information from the different satellite series or by using, in addition, information from polarimetric airborne SARs.

For present purposes we were concerned first to establish the correlations between radar imagery and the more usual LANDSAT imagery of the same sites and then to see if the radar imagery is able to add to the information about tsetse habitats that we can gain from the more widely used (passive) remotely sensed data.

Correlations between active and passive sensors for selected study sites

Radar imagery for The Gambia was obtained from the European and Japanese radar satellites and geo-registered with a Landsat image for the same area around Bansang. An example of the radar imagery is shown in Fig. 2.4. Correlations were calculated between the optical and radar channels, with the result shown in Table 2.3. There are often strong correlations between the channels from any particular satellite type: those between satellites are weaker, the best being the ERS-1 and Landsat TM correlations.

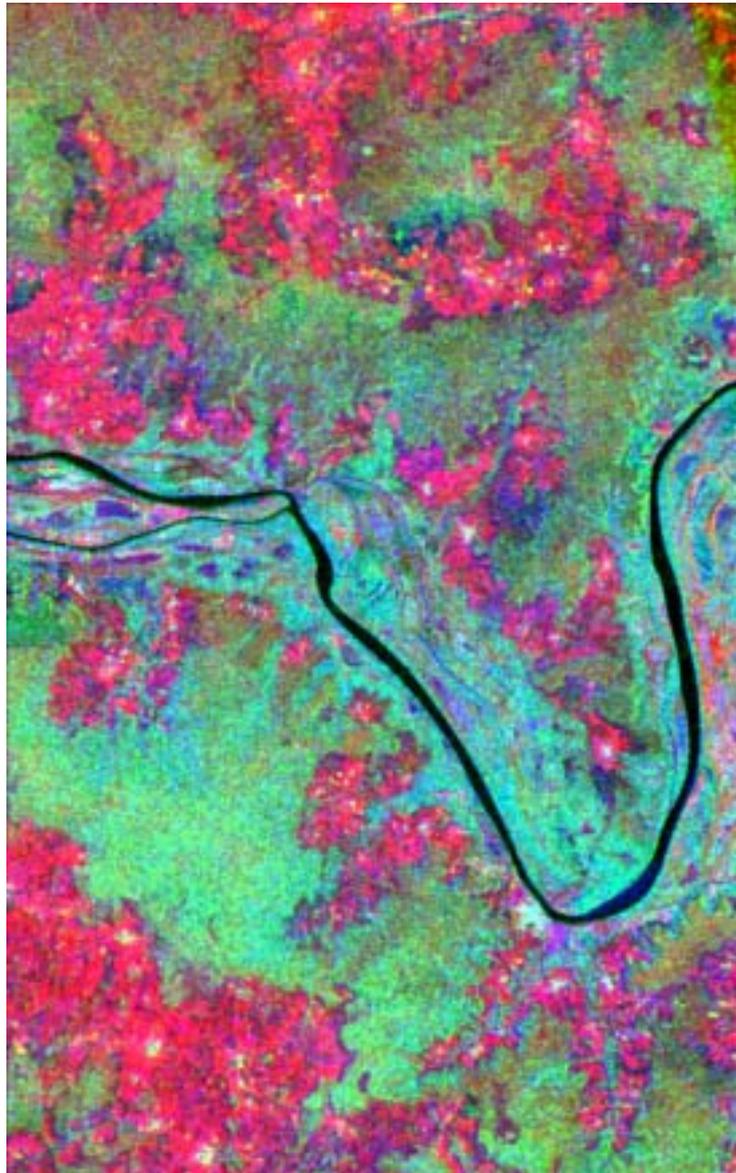
Table 2.3. Correlations between JERS, ERS and Landsat TM bands for the Bansang region of The Gambia.

	JERS-1	ERS-1	TM-1	TM-2	TM-3	TM-4	TM-5	TM-6	TM-7	VNIR-1	VNIR-2	VNIR-3
JERS-1	1	0.5*	-0.2	-0.3*	-0.3	0.4*	-0.0	-0.2	-0.1	-0.1	-0.1	0.5
ERS-1	0.6*	1	-0.4*	-0.5*	-0.5*	0.2	-0.3*	-0.5*	-0.5*	-0.4*	-0.3*	0.2
TM-1	-0.2	-0.4*	1	0.9*	0.8*	0.1	0.5*	0.5*	0.6*	0.5*	0.5*	0.1
TM-2	-0.3*	-0.5*	0.9*	1	0.9*	0.2	0.2	0.4	0.6*	0.5*	0.5*	0.1
TM-3	-0.3	-0.5*	0.8*	0.9*	1	0.2	0.7*	0.6*	0.8*	0.6*	0.6*	0.1
TM-4	0.4*	0.2	0.1	0.2	0.2	1	0.4*	-0.1	0.2	0.3*	0.3*	0.5*
TM-5	-0.0	-0.3*	0.5*	0.5*	0.7*	0.4*	1	0.5*	0.9*	0.5*	0.6*	0.3*
TM-6	-0.2	-0.5*	0.5*	0.4*	0.6*	-0.1	0.5*	1	0.7*	0.3*	0.5*	0.2
TM-7	-0.1	-0.5*	0.6*	0.6*	0.8*	0.2*	0.9*	0.7*	1	0.5*	0.6*	0.2*
VNIR-1	-0.1	-0.4*	0.5*	0.5*	0.6*	0.3*	0.5*	0.3*	0.5*	1	0.9*	0.3*
VNIR-2	-0.0	-0.3*	0.2	0.5*	0.6*	0.3*	0.6*	0.5*	0.6*	0.9*	1	0.4*
VNIR-3	0.5*	0.2	0.1	0.1	0.1	0.5*	0.3*	0.2	0.2	0.3*	0.4*	1

* Correlation are significant at the .05 level base (sample size = 90 pixels).

TM-n: Landsat-TM band number. VNIR-n: JERS-1 VNIR band number

Fig.2.4 Example of radar imagery for the Bansang region of The Gambia. Bansang is the pale region on the lower bend of the River.



SECTION 3

Alternative time-series analytical methods

In TALA1 and much of TALA2 we have routinely used temporal Fourier processing to extract characteristic habitat fingerprints from multi-temporal satellite data. The temporal Fourier series represents one of a number of alternative methods of orthogonal signal decomposition into a series of waves, the sum of which is the signal itself. Since temporal Fourier analysis produces information on the power contained within any particular resolved frequency (the Fourier 'harmonics'), the decomposition can be examined to discover what appear to be the main driving forces within each environment's seasonal cycle. Within Africa, the obvious driving forces are annual and bi-annual cycles of the local weather systems. In addition to these, however, there are other longer period cycles, the impact of which is important to discover, since what appears to be a trend on one time scale may become a cycle on a longer time scale. Currently we are concerned that human-induced trends are causing a one-way movement of habitat type and quality to some less desirable situation (e.g. of greater desertification or deforestation). Clearly much stronger corrective action is required if a change is indeed a trend rather than part of a cycle.

Extracting information about long-period cycles from relatively short term data involves two major problems. The first is that statistically it is difficult to detect a cycle in data sets of length less than six times that of the cycle period. The second is that the cycle may not be as regular as the Fourier series assumes. The first problem is recognised by all time series analysts: the second is of particular concern to geographers and meteorologists.

Today a variety of alternatives to the regular harmonic decomposition of the Fourier series is available. The alternatives retain the important property of orthogonality, but abandon the assumption of smooth, sine or cosine harmonics. As part of TALA2 we investigated one recently developed addition to this family of alternatives, the Hilbert spectral decomposition method (Huang *et al.*, 1999; Huang *et al.*, 1998).

Hilbert spectral analysis assumes that there may be variation in signal strength, frequency and phase over time and estimates each from the data set. Analysis generally requires some pre-treatment of the data set that both removes trends and splits the data into a series of orthogonal components, called Intrinsic Mode Functions (IMFs), which are separately analysed for their amplitude and frequency characteristics. Hilbert spectral methods are long-established, but their combination with IMFs appears to be relatively recent (Huang *et al.*, 1998). An IMF is a component of the original signal that has the same number of maxima or minima as there are crossing points (*i.e.* where $y = 0$). Empirical Mode Decomposition (EMD) extracts IMFs from the signal by a process of sequential averaging and differencing. Spline curves are first fitted to the local maxima and minima and then averaged: the signal is then differenced from this average, and the averaging and differencing process is continued until the resulting curve has the required characteristic of no more maxima or minima than crossing points. This resulting curve is the first Intrinsic Mode Function. This is then subtracted from the original data series and the

result is subjected to further averaging and differencing to extract the second IMF; and so on. Eventually the original data series is broken up into a number of IMFs and a residual trend line which generally crosses $y = 0$ once or twice only. Successive IMFs have progressively lower fundamental frequencies and the sum of all the IMFs and the residual trend line is the original data series, *i.e.* EMD achieves orthogonal decomposition of the data series. Much information about underlying processes may be gained by examining the IMFs without any further analysis. Obviously IMFs with larger amplitudes are making more of a contribution to total signal variation than are IMFs with smaller amplitudes: what is the average period of these important IMFs and does it correspond to any known environmental periodicity? In regions of anthropogenic impacts are some of the IMFs characteristically different from those where there is no impact?

The set of IMFs are then subjected to Hilbert spectral analysis that essentially extracts the instantaneous amplitude and frequency of each IMF by looking at the rate of change of the signal at each instant of time (Huang et al., 1999; Huang et al., 1998). The results from all IMFs can be plotted on the same graph of amplitude at each frequency for each instant of time (in general frequency is plotted on the y-axis, time on the x-axis and amplitude is colour coded within the x-y plane). Examination of such graphs often shows that certain signal components remain constant over the entire duration of the observation, whilst others vary considerably. Such variation is common in climatic/oceanographic applications and renders invalid the more traditional Fourier approach (which will 'force' all variation into one of the precisely regular harmonics).

Fig. 3.1 shows an example of EMD analysis of temperature and dengue haemorrhagic fever (DHF) incidence (per 100,000 population) in Bangkok for the period January 1966 to December 1998 (there are unfortunately no long-term trypanosomiasis records for analysis). In the case of the temperature records IMF2 captures the annual cycle of variability, whilst IMF3 captures a semi-regular cycle with a period of about 2.5 years (note the changing scales of the IMFs). IMF1 captures within-year variation (*i.e.* variation around the normal seasonal cycle).

In the case of the disease data, IMF2 again appears to capture the average annual variation, whilst IMF3 captures important variation with a period of between 2.75 and 3.3 years (*i.e.* 12 or 10 cycles in the 33-year period of the data). This variation was particularly pronounced around 1986/87 when there was a major increase in disease incidence (all IMFs except the first have high values at around this time: many also have high values near the end of the data series, coinciding with another major increase in DHF).

The Hilbert spectra of these two sets of IMF are shown in Fig. 3.2, together with the spectra for rainfall in Bangkok and for a multi-variate index of the Southern Ocean oscillation, the MEI (the ENSO has recently been associated with outbreaks of other vector-borne diseases, notably malaria). The temporal variations in frequency and power that Hilbert spectral analysis allows are colour coded on these images, with orange as the background (= zero power), and increasing power on a rainbow colour scale from blue (low) to red (high). The temperature and rainfall images show most power concentrated in the region corresponding to a 12-month period - the annual cycle. Although there is some variation on longer time scales (especially for rainfall)

this is random and not particularly pronounced. The dengue image also shows a band of high power periodicities at 1 year period and, in addition, another band at 2 - 3 year periodicity, and a more diffuse, lower power band at a 10 - 15 year periodicity. The MEI image does not show the annual band, but has a range of important periodicities between 2 and 5 years.

Fig. 3.2. reveals the variability of the periodicities inherent within each signal. Whilst the temperature and rainfall images always show pronounced power at a period of 12 months, corresponding to the annual cycle, the MEI image has no consistently pronounced periodicity, reflecting the intrinsic variability of the ENSO signal.

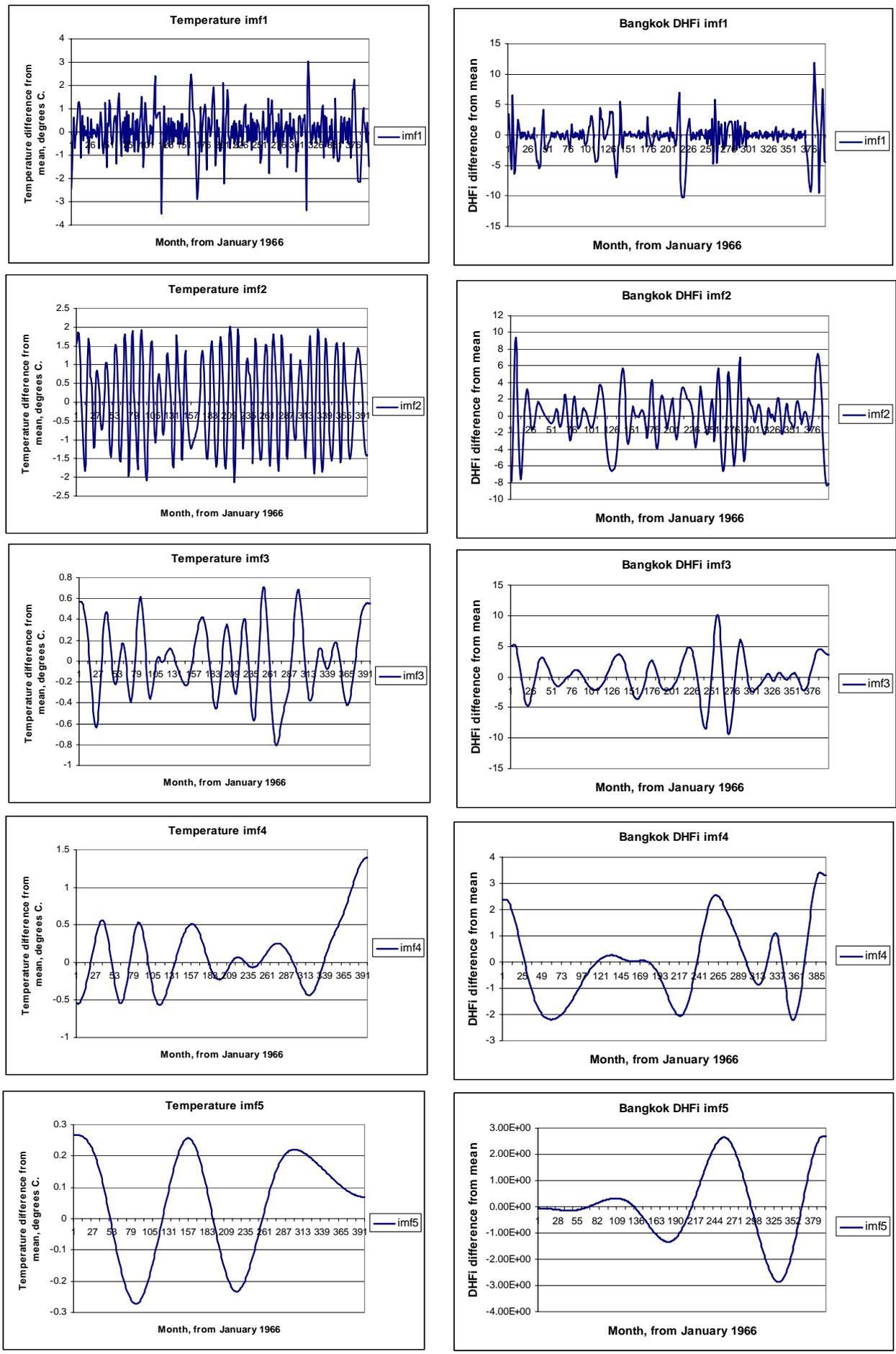
The mean power at each of the periodicities within images such as Fig. 3.2 can be estimated by averaging across each image row. This essentially piles all the information within the image onto the y-axis. The result is shown in Fig. 3.3. This figure also includes a measure of the variability of the signal around the mean power at each period; this is given as twice the standard error of the signal, assuming no serial correlation between successive data points. Data are likely to be serially correlated for two reasons. First although Hilbert analysis makes no assumptions about the constancy of cycle periodicity, it is unlikely that this will vary entirely at random from one observation period to the next. Secondly the images in Fig. 3.2 were obtained by 7x7 Gaussian filtering of the original Hilbert spectral images, which makes visual interpretation easier; this filter essentially smooths point observations in both the x- and y-directions (*i.e.* across observation periods and across periodicities respectively) and the former will increase the intercorrelation between adjacent data points. Thus the error limits in Fig. 3.3 must be interpreted cautiously. (The unfiltered images were also treated in the same way to see the effect of filtering on the following conclusions: the shapes of the spectra were very similar but, as expected, showed more variation from one periodicity to the next. Hence Gaussian filtering of the images in Fig. 3.2 has an effect analogous to the windowing that must be carried out on the power spectral density graphs of the traditional Fourier approach (Chatfield, 1980; Diggle, 1990)).

Fig. 3.3. confirms the impression from Fig. 3.2 that most variation in the temperature and rainfall signals occurs with a periodicity of 12 months. There is very little sign of variation on a multi-annual time scale (except, perhaps, for a 24-month cycle in the rainfall data). In contrast the spectra for both the DHF and MEI data show a number of pronounced peaks. Fig. 3.3. shows quite clearly that there is a shared peak at c. 30 months periodicity in both data sets. There are also matched peaks at a period of about 4 years, but this is much stronger in the MEI data series than in the DHF series. There is quite a pronounced mis-match at periods of about one year. The DHF data has its major peak at 13 months (only 5% higher than the value at 12 months), but the less pronounced peak in the MEI data occurs at a periodicity of 17 months; the value here is 37% greater than that at 12 months. Even allowing for serial correlations, the standard error bars suggest that many of the differences in Fig. 3.3 are significant.

It is clear from this analysis that the Hilbert spectral approach overcomes many of the objections to the application of Fourier analysis to quasi-periodic time series data. Two obvious problems remain. The first concerns the spline fitting involved in

EMD of the original signal: some assumption must be made about the slope of the fitted spline curve at the start and end of the data series (*i.e.* at the very first and last points). This analysis assumed a default value of zero, but Fig. 3.2. suggests that this is not always appropriate. The second problem, mentioned above, is the estimation of the confidence intervals of spectral plots derived from the Hilbert spectral images, such as those shown in Fig. 3.3.

Fig. 3.1. Empirical Mode Decomposition of temperature (left) and dengue incidence data (right) for Bangkok, from 1966 to 1998 inclusive. Each panel shows one Intrinsic Mode Function (IMF). Notice how each IMF captures a different range of frequencies.



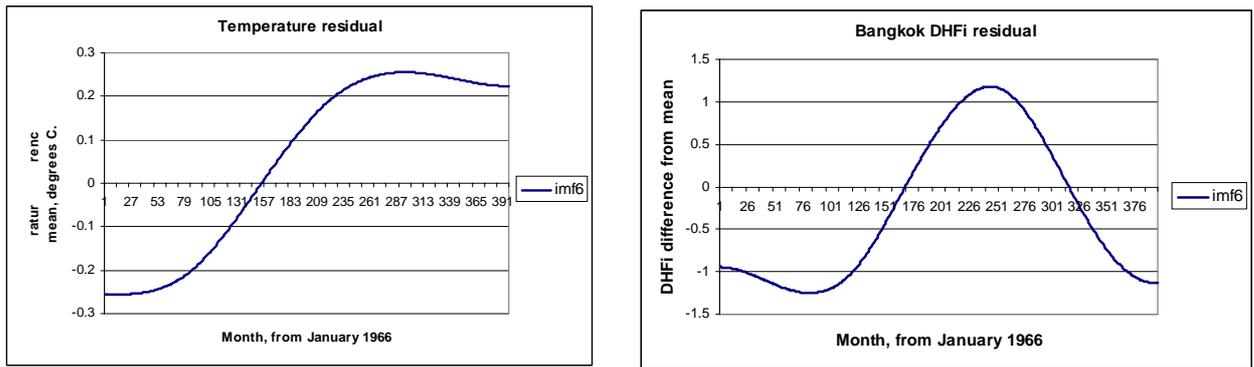


Fig. 3.2 (below). Hilbert spectra of the IMFs shown in Fig. 3.1. for temperature (top left), rainfall (top right), dengue incidence data (lower left) in Bangkok and for the MEI, a multi-variate index of the Southern Ocean oscillation (an ENSO measure) (lower right), from 1966 to 1998 inclusive. The images show how the signal power is concentrated at particular frequencies (expressed here as periods, in years on the y-axes); signal power (within each image) is rainbow scale colour coded (blue = low, red = high power, with zero power in orange). Original images were filtered with a 7x7 Gaussian filter. Notice the strong and constant annual signal in the temperature data, a slightly less constant one in the rainfall data and a more variable one in the DHF data. In the latter there is also a variable 2-3-year signal and a more diffuse 9-15 year signal. The MEI data shows an irregular band at 1.5 - 6 year periodicity (with one peak coinciding with the 2-3 year DHF peak) and signs of a lower frequency cycle of c. 14-17 years.

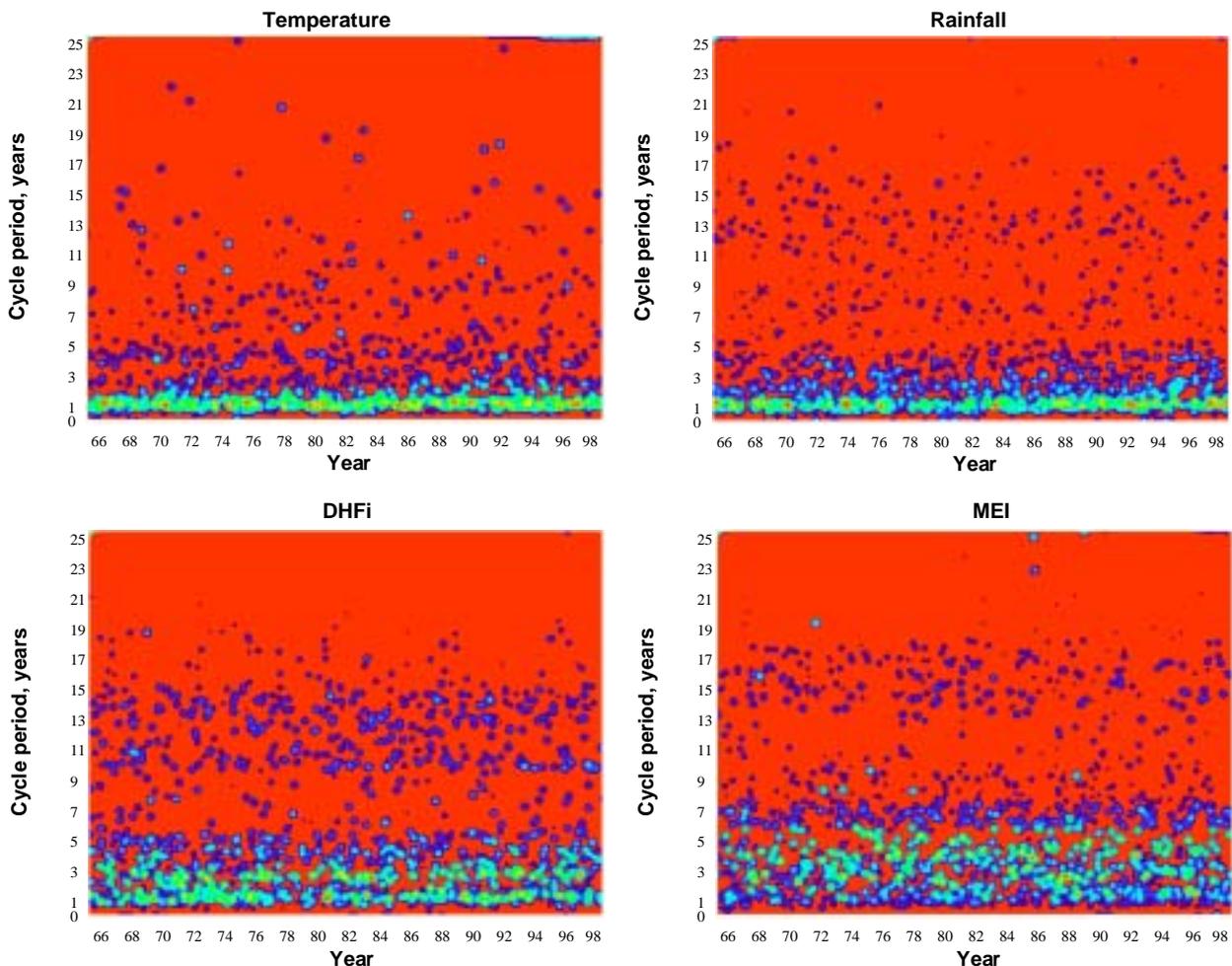
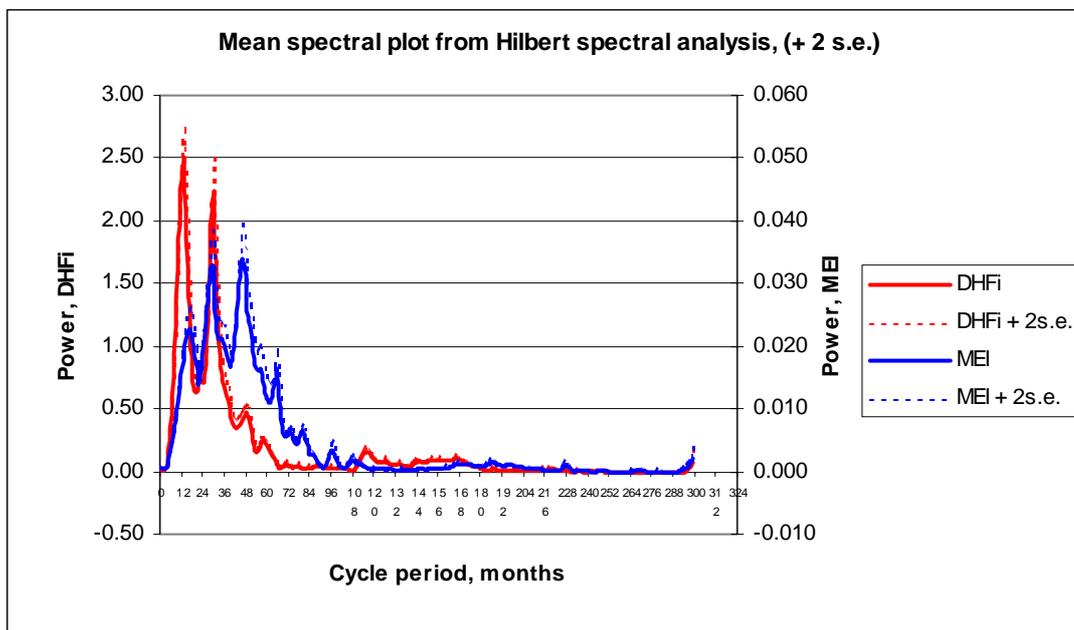
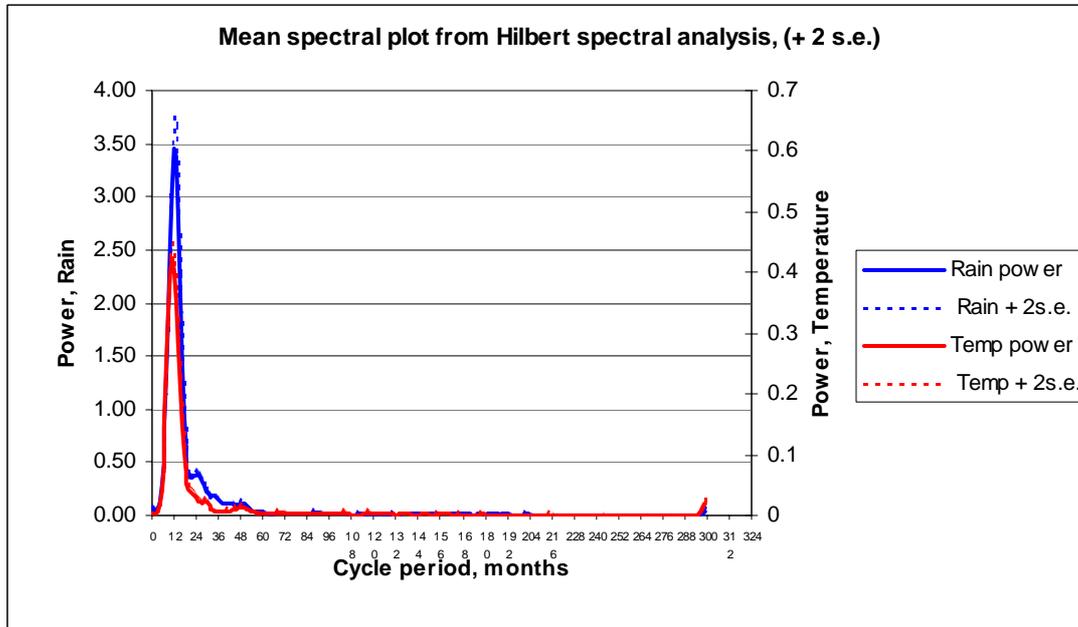


Fig. 3.3. Mean spectral density plots from the Hilbert spectra of Fig. 3.2. Upper, rainfall and temperature; lower, DHF and MEI. Dashed lines show the +2 standard error limits of each spectrum assuming no serial correlation between successive data points (see text for further details).



SECTION 4

Increasing the spatial resolution of AVHRR data using wavelet techniques

Currently only the AVHRR instrument on board the NOAA series of satellites can give us the multi-temporal imagery of terrestrial surfaces from which the characteristic seasonality profiles of different vegetation types may be extracted by Fourier processing. The maximum resolution of this instrument is 1km for beneath-track recording stations: imagery is often degraded to 4km or 8km spatial resolution by the time it reaches the end-users.

The next generation of multi-temporal instruments will have a spatial resolution of 0.25 km (MODIS/TERRA) which will provide imagery of about the right spatial scale for many vector-borne disease studies, since 0.25km is about the size of the smallest habitat unit within which many vector population can survive, and also approximates the daily flight range of many vectors.

Until these higher resolution images become available, however, it is possible artificially to increase the spatial resolution of the AVHRR imagery by introducing spatial details from higher resolution imagery such as that of Landsat or SPOT. The principle here is that many spatial features of habitats are seasonally invariant. For example, a piece of riverine forest neither expands nor contracts during the year, although its spectral characteristics will change. If we can fuse the spatial characteristics as detected by Landsat with the seasonal spectral characteristics as detected by AVHRR, then we should have a higher spatial resolution, seasonal image which would be our 'best guess' of the distribution of different vegetation types in the habitat being studied.

A variety of wavelet techniques ((Starck *et al.*, 1998; Stollnitz *et al.*, 1996)) allows us to do this with satellite imagery. As the name implies, wavelets are filters of fixed length (which might be in units either of time or space) that are passed over (*i.e.* convolved with) imagery to extract temporal or spatial features at the temporal/spatial scale of the wavelet. By changing the size of the wavelet relative to the image, different scale features are extracted in different 'passes' of the filter. In general fine features are extracted first with wavelets of small physical size, followed by coarser scale features as the wavelet size increases. It follows that as long as two images of the same physical ground area, with equivalent spectral characteristics but of different maximum spatial resolution, are correctly geo-registered the details of one may be applied to the other artificially to increase its spatial resolution.

There are three problems in this exercise. The first is to choose the types of imagery that will be fused in this way; the second is to achieve accurate geo-registration of the two images; and the third is to choose the sort of wavelet transform to use.

To investigate wavelet fusion techniques we selected Landsat and 1km AVHRR NDVI imagery for one of our field sites, at Nguruman, SW Kenya. The Landsat equivalent of the NDVI is derived from Landsat channels 3 and 4 which are spectrally

closest to the AVHRR channels 1 and 2. For present purposes fine spatial resolution imagery was required which could be aggregated up block-wise to achieve a spatial resolution equivalent to that of the coarser imagery: to achieve this the Landsat imagery was first re-sampled to a pixel size of 32.5 metres. The Landsat and AVHRR images used in the following analysis are shown in Fig. 4.1.

There is a voluminous literature on co-registration of imagery, emphasising either co-registration of data values (when the images are of the same type, as here) or else co-registration of the Fourier phase information (when images are of different types). In theory it is possible to use Fourier techniques to co-register images of different spatial resolution, as long as one spatial resolution is a 2-tuple (*i.e.* a two times multiple) of the other (the Fourier coefficients are cross-correlated, missing out those from the finer-scale spatial resolution imagery that are not present in the coarser resolution imagery). In this study it was decided first to aggregate the Landsat pixels to a 1km spatial scale equal to that of the AVHRR image, before subsequent cross-correlation. The latter was achieved using a Fast Fourier Transform (FFT) technique (Press *et al.*, 1995) that output the cross correlation at all possible values of both positive and negative lags. The AVHRR image was shifted relative to the Landsat image until the cross correlation was highest at lag zero in both the x- and y- image directions. Finally the shifted image was clipped to the same size as the target image. At this point the best registration between the two original images had been achieved. The original, full resolution Landsat image was then used for the next, wavelet stage.

There are many wavelet transforms available for use in image analysis. The present study was essentially a feasibility exercise so the Haar wavelet transform, one of the simplest available, was chosen for its tractability. The principle of its operation is explained in the following section.

A brief guide to wavelet transforms

Consider a 1-dimensional image of four pixels

$$9 \quad 7 \quad 3 \quad 5$$

Averaging pairwise, this can be reduced to a two-pixel (lower spatial resolution) image

$$8 \quad 4$$

but some information has been lost. This 'lost' information can be stored in 'detail coefficients', shown as the last two figures in the following:

$$8 \quad 4 \quad 1 \quad -1$$

Thus the first pair of numbers in the original set (9 and 7) is obtained by addition and subtraction of the first average (8) with the first detail coefficient (1), $8+1 = 9$, $8 - 1 = 7$. And so on.

The averages can be processed further to give lower spatial resolution means, with further sets of detail coefficients. Thus in the next round of wavelet filtering the

image mean values of 8 and 4 become a new mean (6) with a new detail coefficient (2).

$$6 \quad 2$$

This process continues recursively until the 'wavelet decomposition image' consists of a single average value (the overall image average) and series of detail coefficients applied at different spatial scales.

$$6 \quad 2 \quad 1 \quad -1$$

The following comments apply:

- The wavelet image is exactly the same size as the original image, although most pixels are storing detail coefficients and only a single one is storing information about the overall image mean.
- The transform is easily extended to 2-dimensional images: alternate (block-wise) row and column transforms are the basis for the 'non-standard' image decomposition method used in the present analysis.
- Addition and subtraction to generate two pixels from a single average and detail coefficients is the basis of Haar wavelet decomposition. More complex wavelets average over several detail coefficients using modified B-spline wavelets.
- Image re-construction is an exact reversal of the image decomposition process, and can be halted at any level of spatial resolution.

An example of a partially wavelet decomposed image is shown in Fig. 4.2. In general it is quite difficult to display such imagery since most pixels have very low values whilst the single average pixel has a very high value. Automatic image contrast stretching shows an image which is apparently uniformly grey. Only by stretching across the values of the detail coefficients can the details be revealed.

Correlating the imagery at the same spatial scales

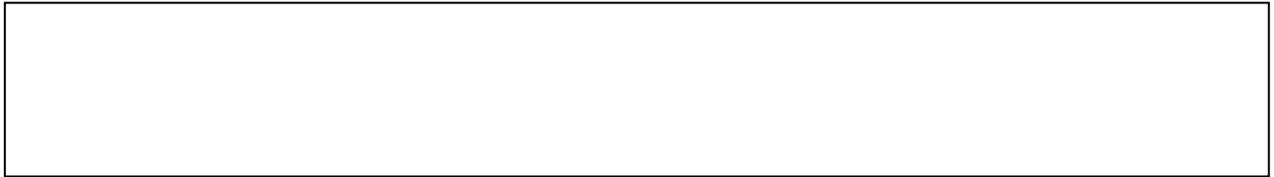
To investigate whether it is appropriate to use the detail coefficients from one image in the other it is first necessary to demonstrate that there are significant correlations between the detail coefficients of the two images at the same spatial scales. This was done by extracting from each fully decomposed image the detail coefficients at the 2, 4, 8 etc km stage of aggregation. The results are shown in Fig. 4.3. The correlations between the AVHRR and Landsat images are significant at all spatial scales, and increase as the mean pixel size increases.

Introducing spatial detail into the AVHRR image

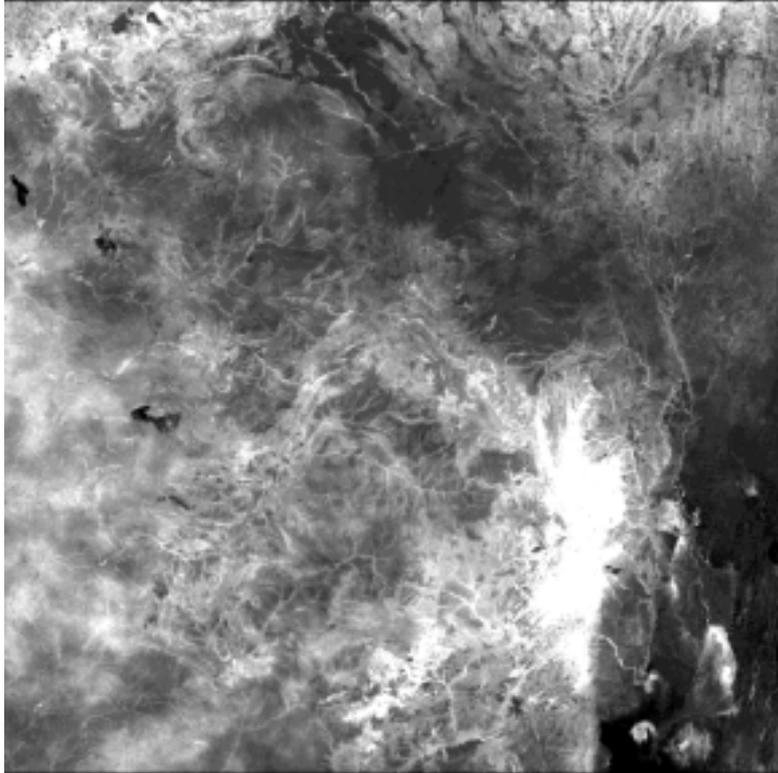
Having established the above correlations, the patterns of change in the values of the (major-axis) regression coefficients and intercepts of the relationships shown in Fig. 4.3 were calculated, in order to predict the expected relationships at finer spatial scales than are present in the AVHRR imagery (*i.e.* of 0.5 km and 0.25km). From these relationships could be predicted the expected detail coefficients of the AVHRR

imagery at these finer spatial scales. Finally these detail coefficients were applied to the AVHRR imagery to produce 0.5 km and 0.25km spatial resolution AVHRR images. An example is shown in Fig. 4.4 (upper).

Fig. 4.4, as expected, shows considerably more spatial detail than the original 1km AVHRR imagery: it includes features such as water courses that are otherwise only visible on the Landsat image. By producing a series of such images for each month of the year, we hope to be able to produce seasonal Fourier images at the same spatial scales (e.g. Fig. 4.4, lower), and these images will form the basis of higher resolution, more accurate predictions of tsetse habitats.



Landsat TM NDVI, 4096 by 4096 pixels, 32.5x32.5 m.



AVHRR NDVI, 128 by 128 pixels, 1km by 1km.



Fig. 4.2. Upper. Non-standard Haar wavelet decomposition of the Landsat TM image shown in Fig. 4.1. This is an intermediate image, where the original 32.5m. pixels have been aggregated to 1km pixels into the top left-hand corner of the image (the white square, which is therefore at the same spatial scale as the AVHRR image in Fig. 4.1). The rest of the image shows detail coefficients produced during this aggregation. The large red squares show the detail coefficients at the first level of aggregation, the yellow squares show those at the next level of aggregation, and so on. In each of the three quadrants of detail coefficients at any particular level of aggregation, the top-right quadrant captures vertical features of the image, the bottom-left quadrant captures horizontal features and the bottom-right quadrant captures diagonal features. This image has been contrast-stretched to reveal the detail coefficients rather than the mean values of the 1km pixels (which are very much higher). Lower. Full wavelet decomposition of the 1km AVHRR (left) and 1km TM (right) images (the TM image is that within the white square of the upper image). Notice the similarity of the detail coefficients in the two images.

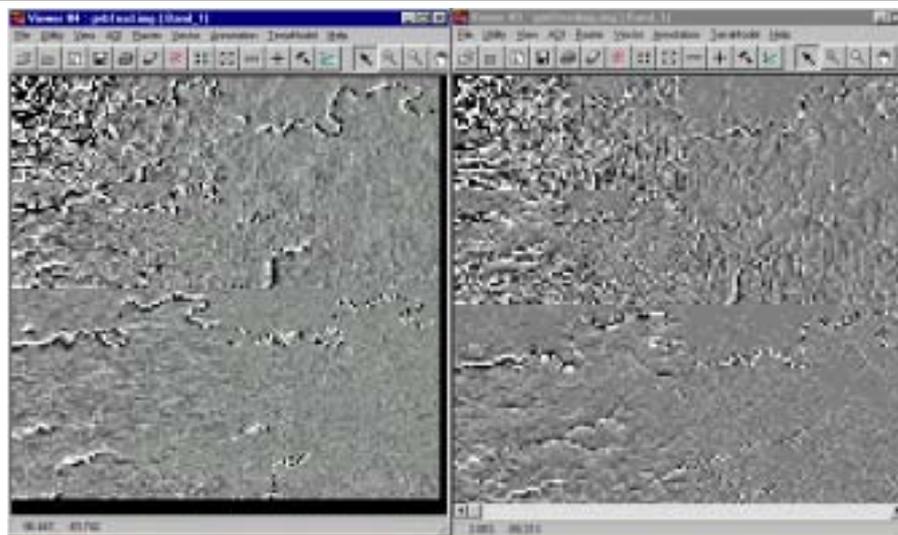
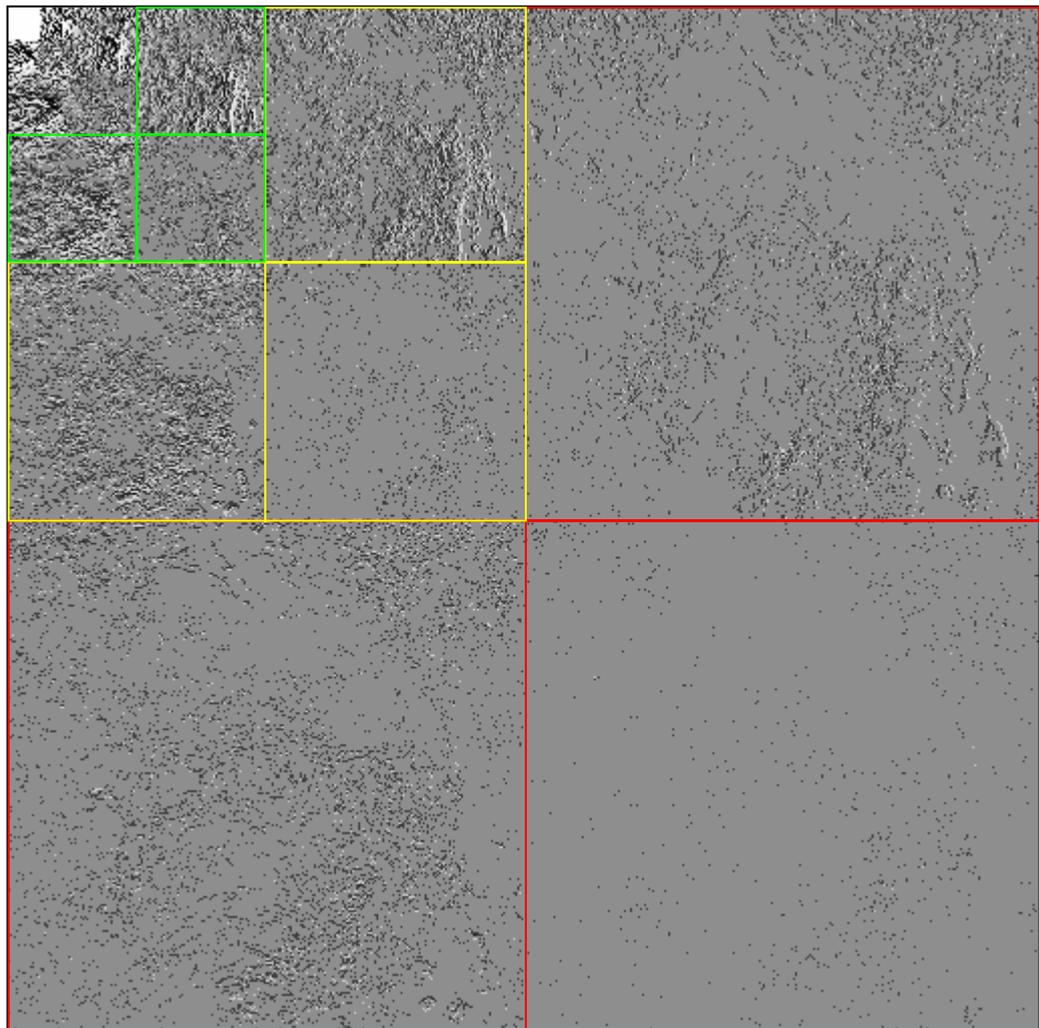


Fig. 4.3. Relationship between the TM and AVHRR detail coefficients of wavelet images at 5 different levels of aggregation of both, starting at the stage of aggregation from 1km to 2km pixels ('level 6'). Each graph shows the standard least squares regression line, but the analysis eventually used the geometric mean of the least squares regressions of y on x and x on y (since there are errors in each variable). x- and y-values are in DN units of the original images.

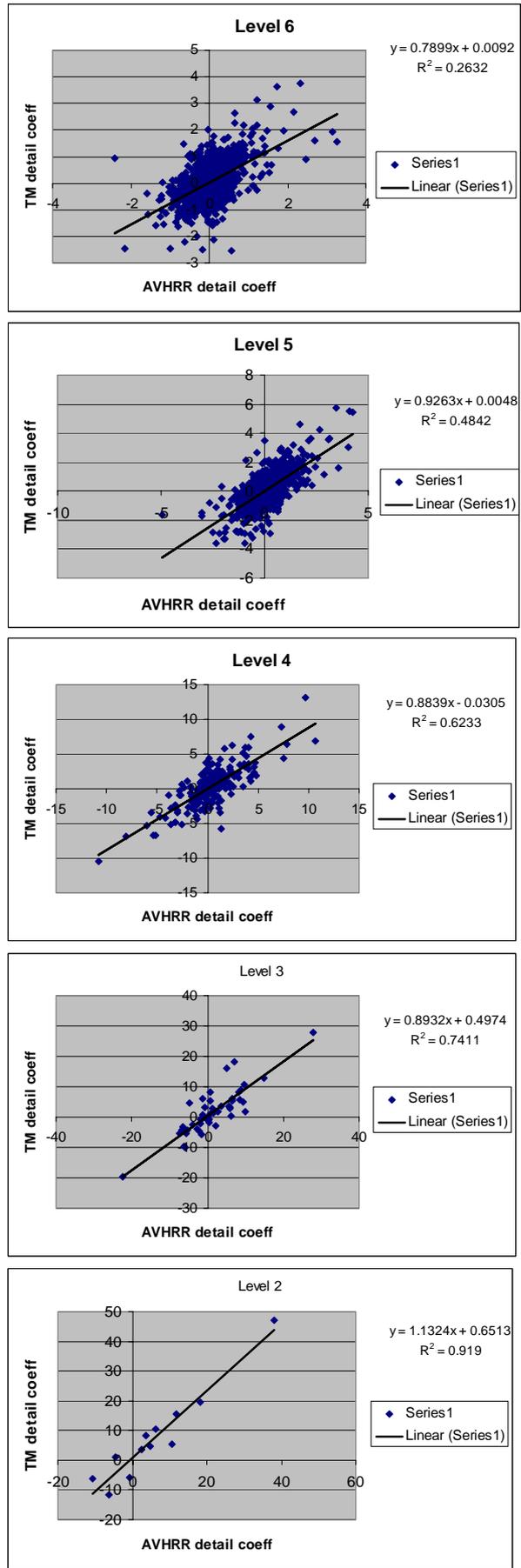
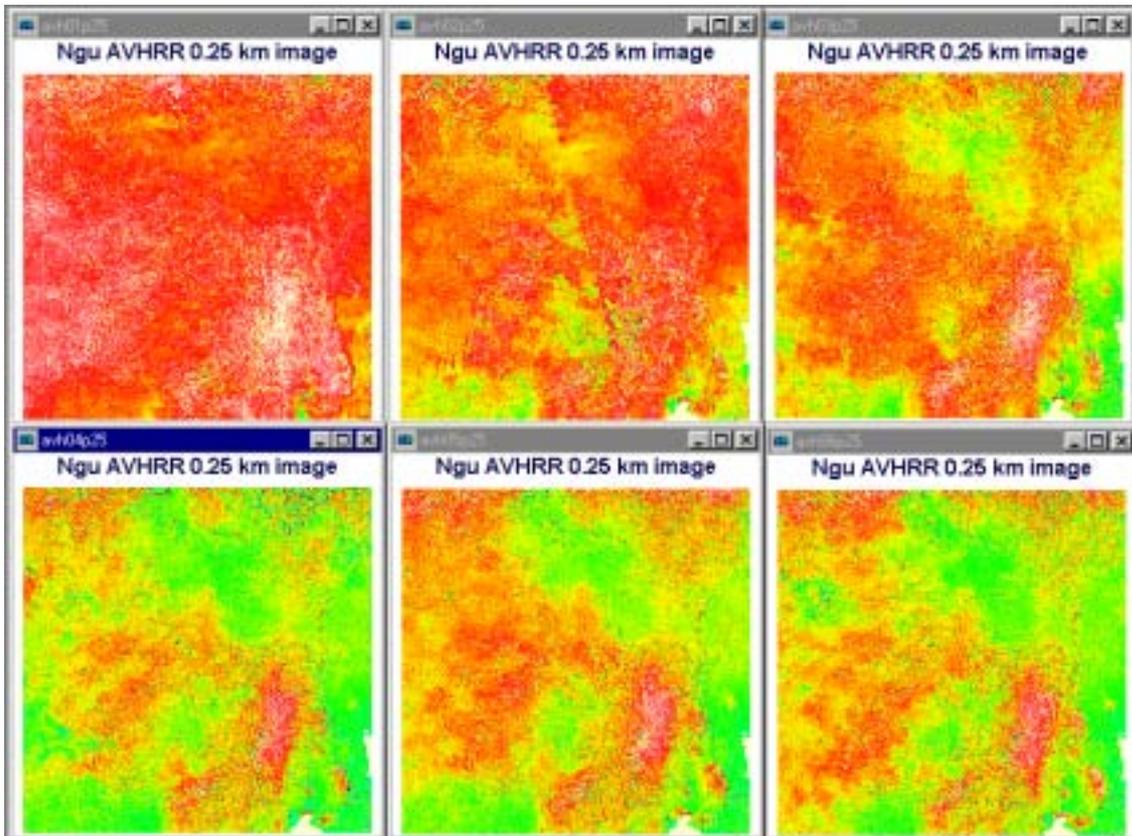
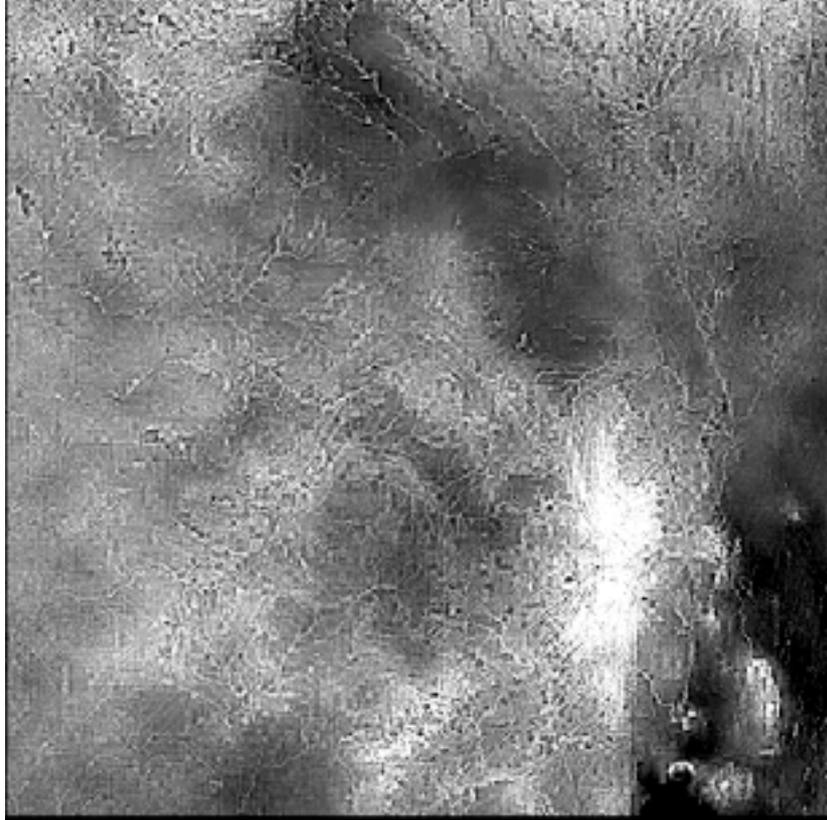


Fig. 4.4. Upper. Wavelet-enhanced AVHRR NDVI image for part of SW Kenya, each pixel representing 0.25 km on the ground. Wavelets have introduced into the original 1km imagery (Fig. 4.1, lower) the spatial detail from the Landsat image (Fig. 4.1, upper) as captured by the coefficients displayed as an image in Fig. 4.2. Wavelet analysis retains the overall mean value of each 1km pixel, but distributes this mean between the sub-pixels according to the values of the detail coefficients. Lower. False colour-coded, wavelet enhanced images of the same area, for January to June, showing details of seasonal changes.



SECTION 5

Analysis of tsetse population data from selected field sites

Generic models for vectors involve components of birth and death, at least one of which must be density dependent. The temperature and humidity dependence of demographic rates may be investigated in the laboratory (e.g. (Buxton and Lewis, 1934)), but these studies provide only an approximate guide to such rates in field conditions. Tsetse apparently carefully select micro-habitats that are considerably moister (Bursell, 1959) and, at least in the hottest conditions, cooler than ambient (Hargrove and Packer, 1993). The remarkably slow rate of offspring production by tsetse have been described by simple equations involving only air temperature (Glasgow, 1963; Hargrove, 1994). These equations may be used in predictive models using either meteorological or satellite data. One problem of the latter is that satellite sensors record directly only the thermal radiance (reflected or emitted) of the soil and vegetation cover, which is often much higher than air temperature (Hay et al., 1996). Various manipulations of the satellite data can give air temperature estimates with accuracies of a few degrees Celsius (Prihodko and Goward, 1997; Prince *et al.*, 1998), sufficiently good for initial models.

In sharp contrast to birth rates, tsetse death rates appear to depend on both temperature and atmospheric moisture (Rogers and Randolph, 1986), and there is also strong evidence for density dependence at both the puparial and adult stages (Rogers, 1974; Rogers and Randolph, 1984; Rogers *et al.*, 1984).

A combination of birth and death rates, each described by the locally appropriate meteorological variables, and with a variable amount of density dependence, successfully described tsetse population changes in both West and East Africa (Rogers, 1990; Rogers *et al.*, 1994). It is now possible to use satellite data as a surrogate for the standard meteorological data. For example, in the Yankari Game Reserve in Nigeria the correlation between the bi-monthly mortality rate of *Glossina morsitans submorsitans* and Land Surface Temperature (LST) estimates derived from satellites orbiting more than 800 km above the earth's surface is stronger than the correlation between mortality and saturation deficit (the best ground-based correlate) calculated from meteorological data collected about 50km from the field site (Fig. 5.1).

A satisfactory description of seasonal changes in tsetse populations is achieved by fine tuning several critical parameter values in the biological models. This fitting process can be automated by steepest descent search methods (Hargrove and Williams, 1998) although a careful check must be kept on parameter values since there appear to be many locally stable equilibria when models are fitted to population data. One example, using the satellite data to predict the monthly mortality rate of *G. m. submorsitans* in Nigeria, is shown in Fig. 5.2. The model also requires some estimate of air temperature for predicting both inter-larval and puparial developmental periods. Rather than predicting temperature from published formulae relating land surface radiance to air temperature, the model in Fig. 5.2 included an additional fitted parameter to relate satellite (LST) and air temperature directly. This parameter was

varied along with all the others to achieve a least-squares fit of the model to the field data.

Figs. 5.3 shows the same generic model applied to data for *G. morsitans* from the Gambia and for *G. pallidipes* from the Nguruman area of Kenya. The least squares method manages to explain >90% of the variance of the mean monthly trap catches at each site.

These and other fitted tsetse population models suggest that substantial density dependence operates on tsetse in the field (Rogers and Randolph, 1984), although the agents of these mortalities have never been sufficiently investigated. Fly abundance is a product of both the density independent, abiotic mortalities (which may be predicted from satellite data) and density dependent biotic ones, and hence models developed for one area may not be easily extended to others unless the density dependent components are in some way described by satellite data.

Predictive biological risk maps of tsetse

The tsetse model produced from the ITC trap catch data for Bansang was then used with the mean LST satellite data layer to produce a predicted density surface for tsetse for the Bansang region and beyond. This is shown in Fig. 5.4. This appears to be the first satellite derived risk map for any species. Once tested for accuracy, such maps can be the starting point for modelling intervention strategies on an area-wide basis, using any means of fly suppression, as long as the impact of the chosen method on fly fertility or mortality rates is known.

Initial inspection of Fig. 5.4 suggests that the tsetse model has not yet captured the constraining influence of the higher temperatures North of The Gambia: the model appears to over-estimate fly population density in such areas, probably because the higher temperatures cause a higher tsetse reproductive rate in the model without any counter-balancing extreme mortality in these dry areas. Clearly data for this species from other sites are required, to investigate the hot temperature limits tsetse population growth. Nevertheless, this first attempt shows that it is possible to produce spatially detailed risk maps using satellite imagery.

Analysis of trypanosome infection data from selected field sites

Once the tsetse model has been fitted to field data, the disease transmission component may be added to it. A simple transmission model for the African trypanosomiasis, based on the standard susceptible, infected, recovered-immune model (Anderson and May, 1991) is described in the Appendix. This model contains equations describing changes in the proportions of vectors and hosts that are currently incubating infections, and of hosts that have recovered and are immune to re-infection for a period of time (these proportions are usually set to their equilibrium values in models that predict only equilibrium disease prevalences (Rogers, 1988)).

Fig. 5.5 shows the results first of fitting the tsetse model and then of adding the trypanosomiasis component to field observations from The Gambia (data from (Rawlings *et al.*, 1991a)). and Kenya (data from collaboration with KETRI). The Gambia model applies to locations where the major vector species is *G. morsitans*

submorsitans and the hosts are the local trypanotolerant N'Dama cattle: the Kenya model assumes that *G. pallidipes* is the only vector of trypanosomiasis to the local zebu cattle (*G. longipennis* is also present in the area, but at very much lower apparent densities). The various parameters of the transmission equations were varied to provide a good fit to the field data both in terms of the average level of infections in both the vectors and hosts and in the seasonal changes in infection rates. Fly infection rates are highest when large proportions of the vectors are old; this does not occur at peak tsetse population levels, which include many young flies. In The Gambia, fly infection rates show two peaks in the model, in February and August, whilst field infection rates were variable and showed no significant month-to-month differences (at around the same mean level). Model host infection rates are highest in April, compared with May in the field, and show similar seasonal changes (Fig. 5.5). In the Nguruman model, predicted host infection rates lag behind observed rates by one month, and there is only a single peak in fly infection rates two months after the peak in fly numbers and cattle infection rates. The generic trypanosomiasis model explains 50 - 60% of the variance of field infection rates, *i.e.* lower than that of the tsetse population models.

The right-hand panels in Fig. 5.5. show some further outputs of the model, in terms of the age-prevalence curves of fly and host infections, and of host immunity. The Gambia model used a long duration of immunity in individual N'Dama hosts (800 days) and a relatively slow build up to high levels of population (*i.e.* 'herd') immunity (relatively low effective challenge): this explains why the level of host immunity is still rising in the oldest age class considered by the model (1000 days, Fig. 5.5): the variation in the host immunity curve around this overall trend occurs because the oldest animals have lived through 3 years' of seasonal fly challenge. The host age prevalence curve for The Gambia shows an unpronounced peak at an intermediate age, characteristic of systems with temporary rather than life-long immunity: the peak is not pronounced since herd immunity has not reached its equilibrium value.

The situation is quite different for Nguruman (lower panels in Fig. 5.5). Infection rates are higher overall, immune duration is shorter (98 days) and population immunity reaches a plateau at about 20% in cattle of about 1 year of age or older. In this situation host infection prevalences peak within the first year of life, as many susceptible animals become infected together, before they enter the immune category from which they are eventually re-cycled into the susceptible category. Beyond one year of age herd immunity ensures that there is no recurrence of the high levels of infection seen in younger animals.

Fitting both the tsetse and transmission models to field data revealed the following:

- tsetse population models tend to be more accurate if driven by relationships between bi-monthly, rather than monthly, tsetse mortality rates and satellite data. Calculation of bi-monthly rates smooths the mortality rate data and results in estimates that are more appropriate (as are the satellite data) for the mid-points of the respective months. Monthly mortality rate data apply to the ends of the months concerned, since they are calculated as the difference between the logarithms of the population sizes in consecutive months.

- least squares fitting of the tsetse population model revealed a number of local minima, each characterised by a different set of parameters values. Thus, for the moment, we cannot be sure we have the best fits to the field data. During fitting, attention was paid to the changing values of all the parameters that were allowed to vary in the model. When particular parameters did not change during an entire round of least squares fitting, they were adjusted by large amounts before the next run of the model. This sometimes resulted in finding a better overall least squares fit.
- obviously some assumptions had to be made in fitting the trypanosomiasis model to effectively incomplete field data. The proportion of tsetse blood meals taken from the domestic animals was fixed at reasonable levels (0.3 in the Gambia, 0.2 in Nguruman); other hosts of tsetse were assumed not to be involved in disease transmission, *i.e.* all model fly infections were derived only from the domestic animal hosts; finally, a model-variable parameter related trap catches of flies - a relative measure of fly abundance - to the total population of flies that bit hosts within the model;
- the model seemed rather insensitive to large variations in certain parameters, e.g. the duration of immunity. This may have been because average herd immunity was relatively low in both models, or because the least squares fit was applied to infection rate data, not to data on host immunity (which were not available). Obviously any field information about the age-specific model outputs shown in the right hand panels of Fig. 5.5 would be extremely useful in parameterising the model more accurately in the future.
- it was not always clear that the tsetse data related to the cattle infection data. Although flies and cattle infections came from the same general areas in both The Gambia and Nguruman, the Gambian fly and cattle infection data published in (Rawlings *et al.*, 1991b) were the means from several sites, whilst the traps in Nguruman nearest to the place where the cattle were normally kept showed an annual cycle of variation rather different from that modelled, which was the average from traps spread over a wider area.

In view of the various assumptions in the trypanosomiasis model, and the reservations about the applicability of the modelled fly data to the transmission model, the overall results of this biological modelling are very encouraging. We conclude that remotely sensed satellite data, selected on the basis of current understanding of tsetse dynamics in the field, may be used to drive a fully integrated disease transmission model. The challenge for the future is to see if such models give realistic predictions when extended to other areas, and when subjected to variations mimicking those of natural (climate) and anthropogenic (intervention) changes.

Fig 5.1. Relationships between the mean bi-monthly density independent mortality of *Glossina morsitans submorsitans* in the Yankari Game Reserve, Nigeria and monthly saturation deficit calculated from meteorological records from Bauchi, c. 50km away from the Reserve (left) or Land Surface Temperature derived from AVHRR receivers on the NOAA series of satellites >800 km above the Reserve. The bi-monthly mortality rate (the average of the previous and present month's mortality rate) is the mean estimated mortality at the mid-point of the month.

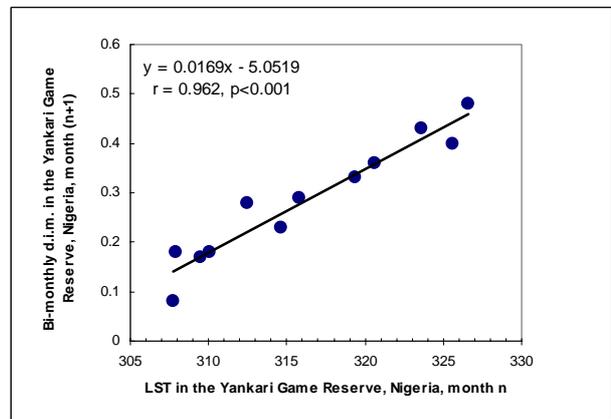
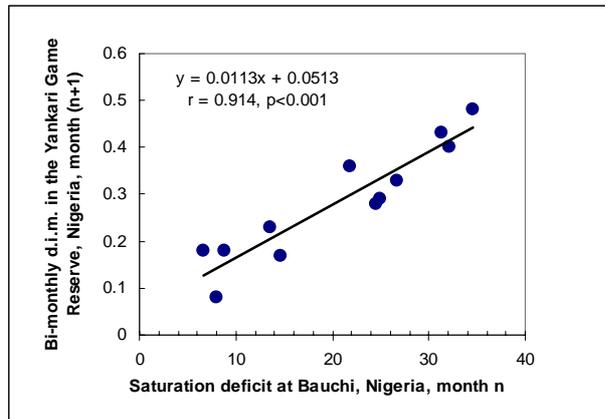
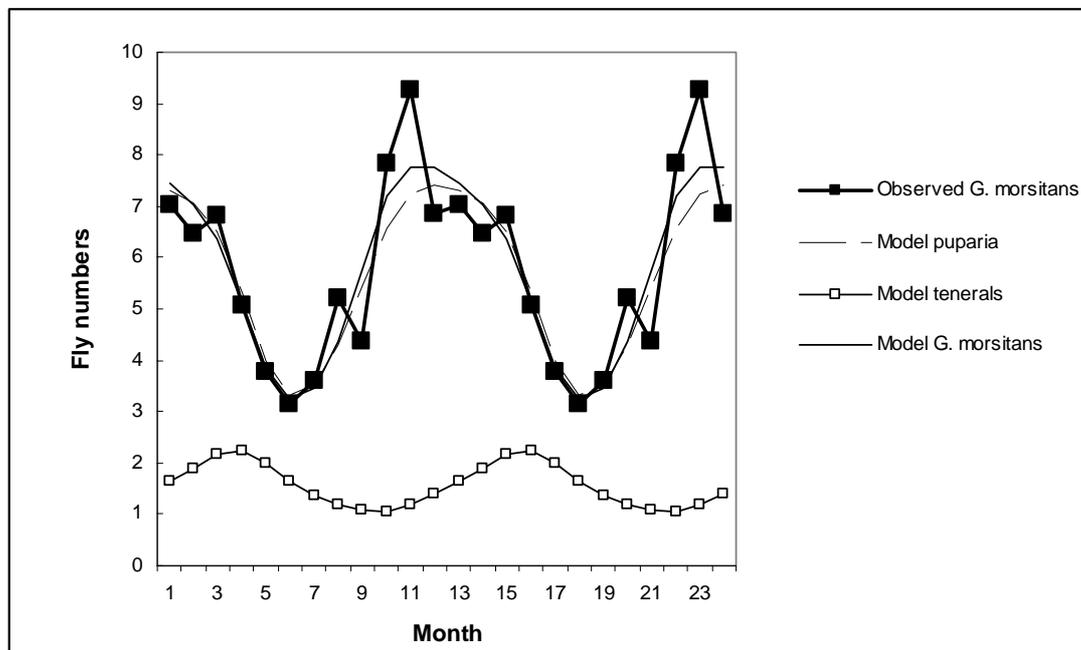


Fig 5.2. Least squares fitted tsetse population model for *Glossina morsitans submorsitans* in the Yankari Game Reserve in Nigeria. This model was driven by satellite data, using the relationship shown in Fig. 5.1 (right), and all other relevant parameters were fitted by iteratively varying them one at a time, to arrive at the least squares solution.



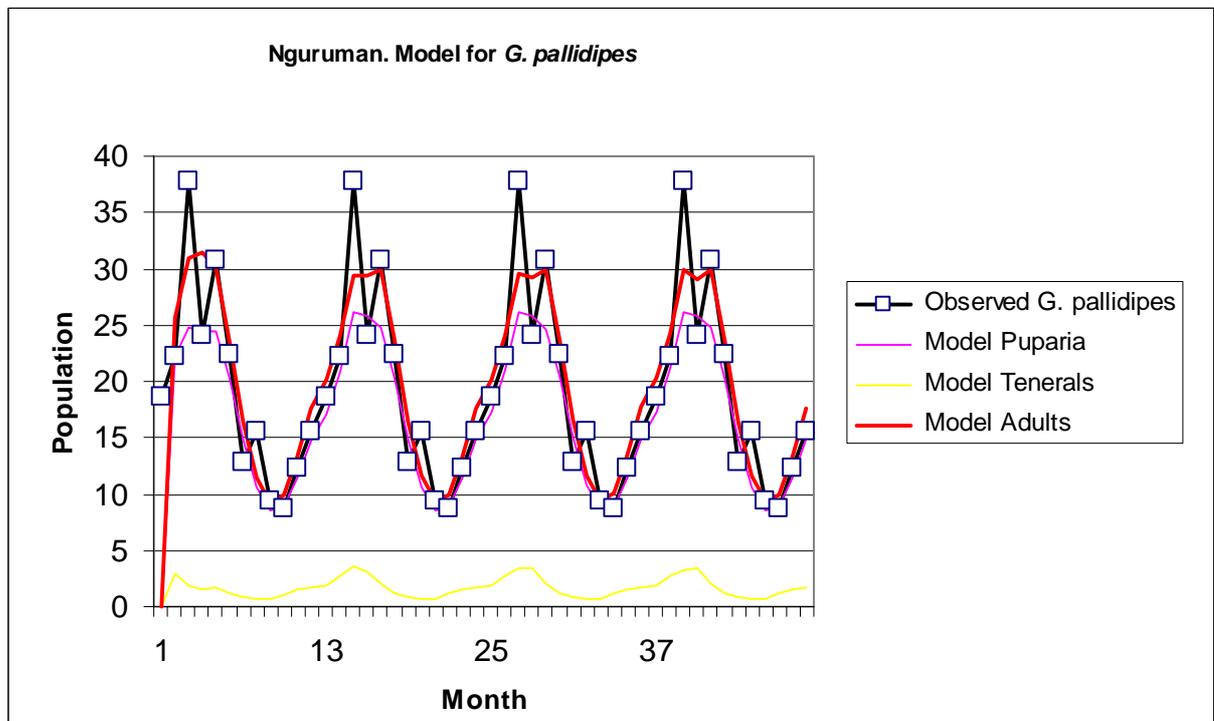
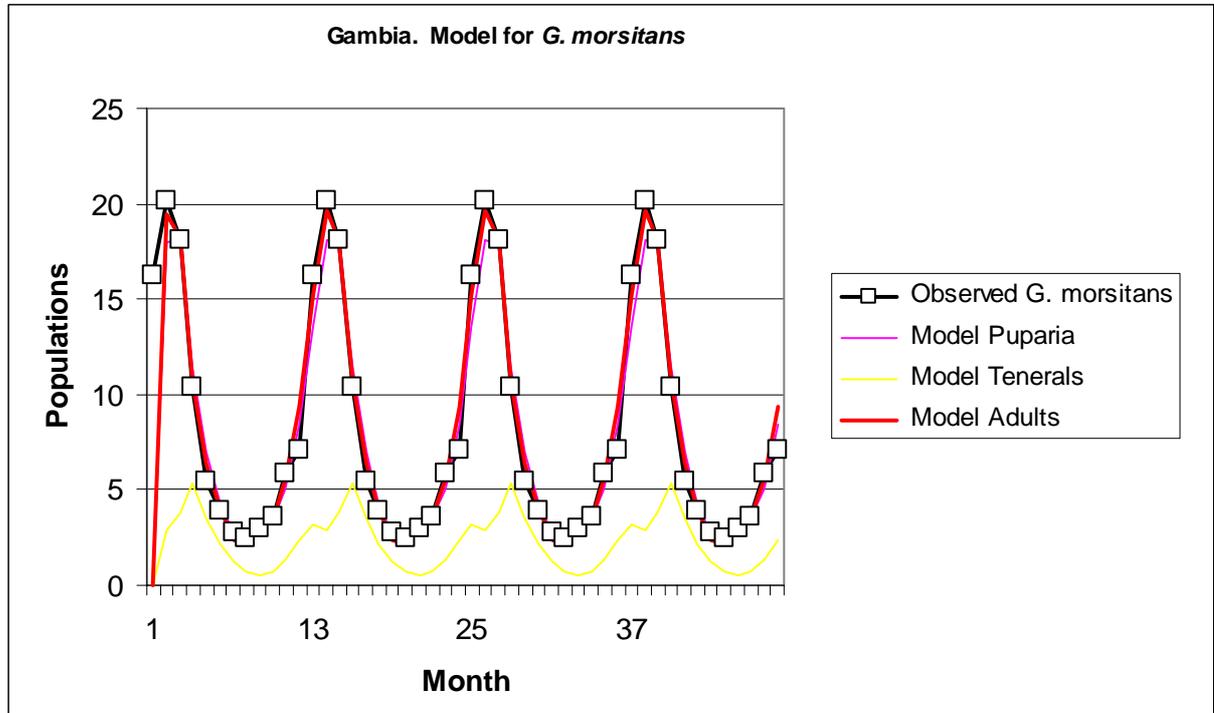


Fig. 5.4. Predicted density surface for *G. m. submorsitans* in The Gambia. The model for this species, developed for the Bansang site (the trap system is visible near the centre of the image) was applied to the region around Bansang and beyond, using the mean LST surface from the AVHRR sensor. Mean predicted tsetse densities are colour coded from black (no flies) through blues, reds and yellows to green (dark green is the highest density). This is the first satellite-derived biological risk map for any species.

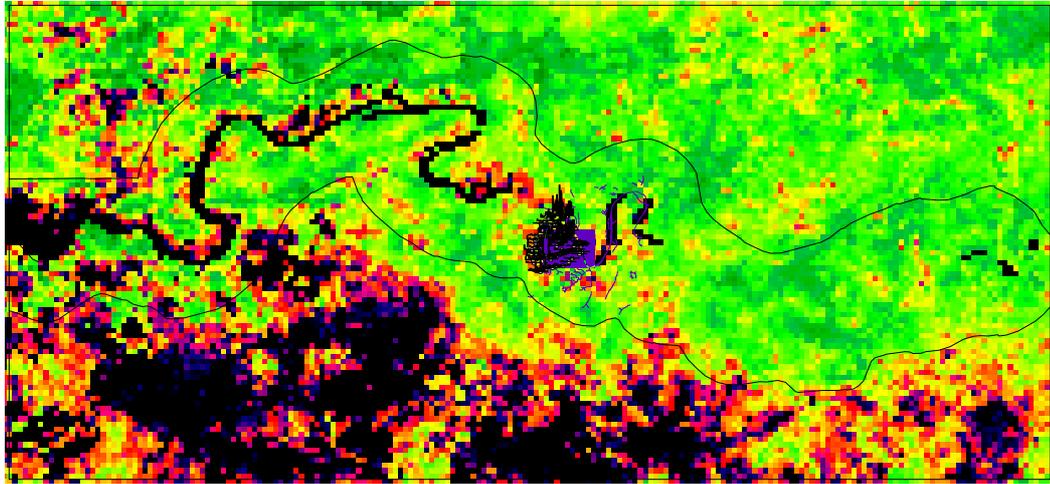
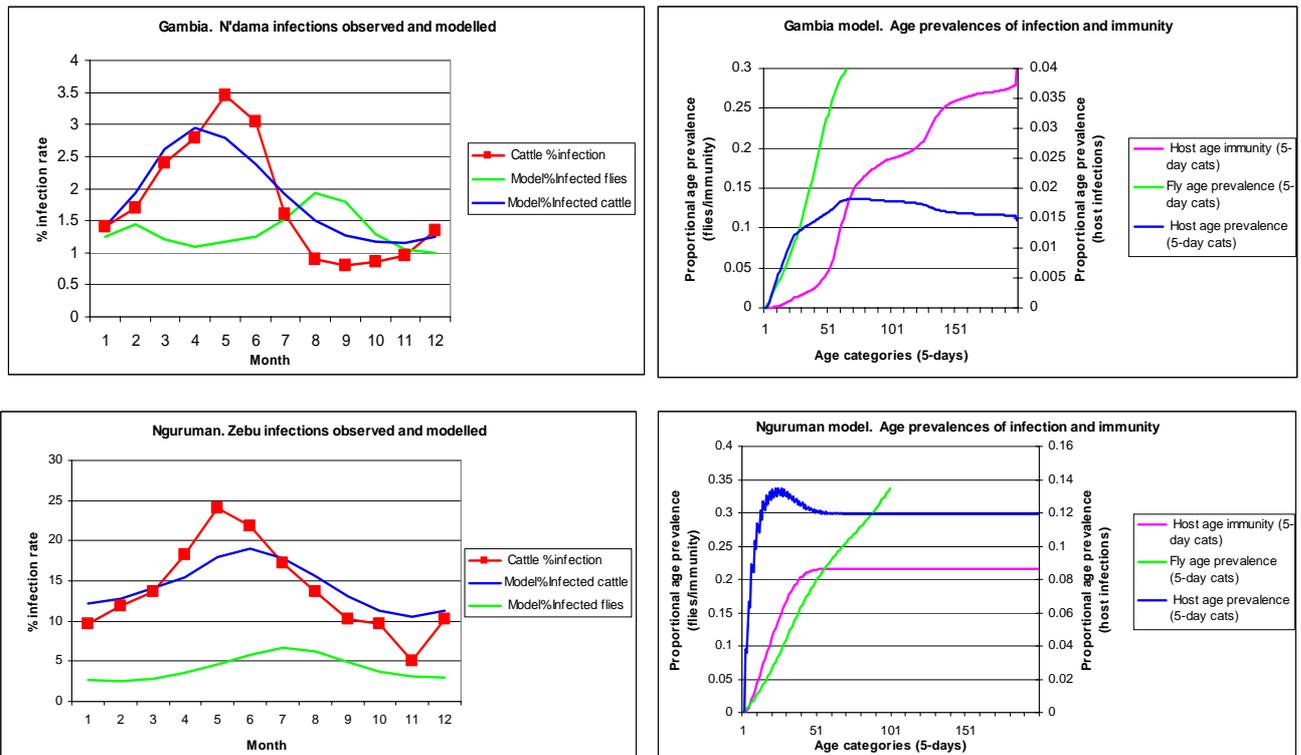


Fig 5.5. Least squares fitted trypanosomiasis transmission model involving *Glossina morsitans submorsitans* in The Gambia (above) and *G. pallidipes* in Kenya (below). In the left-hand panels observed cattle infection rates are shown in red, with filled boxes: modelled infection rates are shown in blue (for cattle) and green (for flies). In each case the model fly population was as shown in Fig. 5.3, *i.e.* driven by satellite data. The right-hand panels show the respective age-specific disease prevalences (blue for hosts, green for tsetse) and age-specific host-immunity profiles (in purple). For these panels, the x-axis shows successive, 5-day age categories (*i.e.* when $x = 100$, age = 500 days). Although tsetse prevalences reach very high values in old flies within the model, very few flies reach such an age, and the average population infection rate is much lower, as the left-hand panels show.



SECTION 6

Sythesis

In this final section we summarise the results of this project in the light of the original objectives.

1) to reach a biological understanding of trypanosomiasis transmission in a number of key study sites.

This aim was achieved for The Gambia and Kenya study sites, within the limits of the data sets available (Section 5). Satellite derived Land Surface Temperature measures were the best predictors of monthly fly mortality rates in both sites, and the resulting models explained > 90% of the variance of fly numbers, as recorded by the traps. The transmission models developed here were able to explain > 50% of the variance in monthly infection prevalences in cattle at each site, a remarkably good figure given the simplifying assumptions that had to be made in the models. Extension of the tsetse models to larger areas seems feasible, and such models can form the basis for modelling intervention at a number of scales. Obviously the tsetse population model can be interrogated for the relative importance locally of density dependent and density independent mortalities, and these are a guide to the ease or difficulty of tsetse eradication from each site.

Future work needs to concentrate more on the links between the fly population changes and the changing prevalence of disease in domestic animal hosts. This would be helped if we could be certain that the tsetse populations as sampled by the trapping systems are responsible for all the disease transmission that is occurring locally. Averaging tsetse catches across several trap systems reduces the noise in the trap data but may give an inaccurate picture of the fly challenge to local cattle.

2) to integrate high- and low-resolution satellite data for the study sites.

This study has shown how different sorts of data, with different spectral, spatial and temporal resolutions may be brought together to provide a unique, multi-variate picture of epidemiological landscapes (Sections 2 and 4). The AVHRR sensor continues to provide the best source of information about seasonality over very large areas, and the processed images may be clustered to identify characteristic ecozones, related to vegetation type or cover. High spatial resolution data from Landsat or SPOT may be incorporated via wavelet methods, to refine the spatial details of the seasonal maps (Section 4)

Future work could profitably anticipate the arrival of MODIS/TERRA data that will have an ideal spatial resolution for areal studies and sufficient spectral and temporal resolution to continue to provide the vital information on habitat seasonality that is the most useful feature of the current AVHRR data.

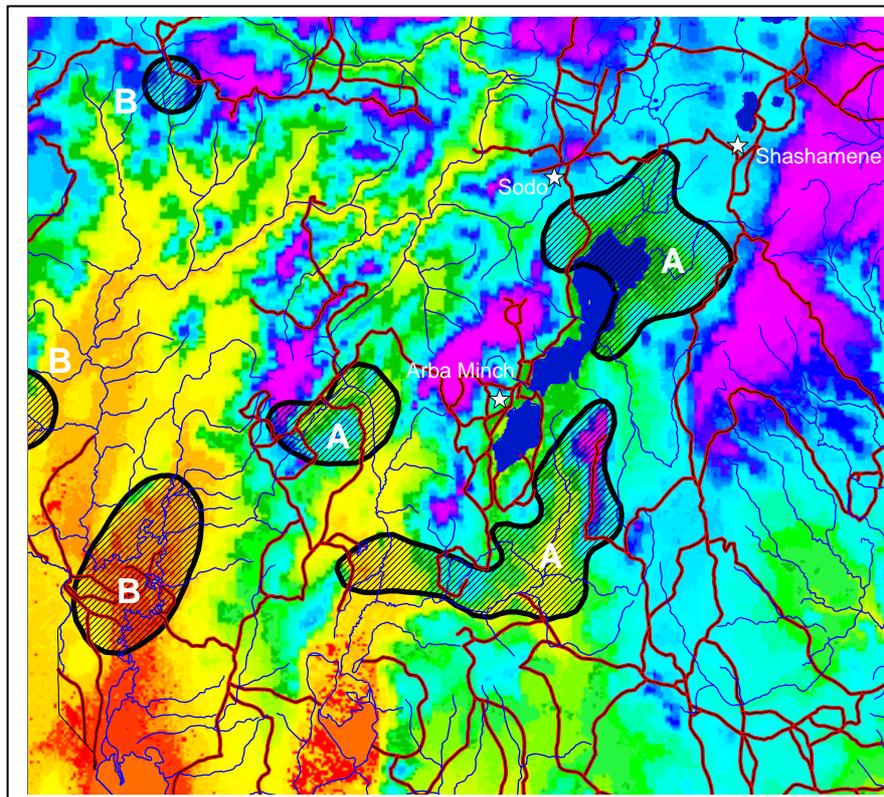
3) to relate the biological models of Step 1 and the statistical models of Step 2.

Our statistical and biological approaches are unified by a common set of satellite data that both describes and helps to explain the tsetse and disease data. An example showing the potential of linking the statistical and biological approaches is shown in Fig. 6.1, which is a processed satellite image of part of southern Ethiopia that is being considered for a SIT tsetse project by the International Atomic Energy Agency.

The underlying image in Fig. 6.1 is a false-colour coded result of unsupervised classification of Fourier processed 1km AVHRR data. The different colour swaths represent different vegetation types, and it is easy to see from the river overlays how these are probably determined by local topography and water availability. Thus this purely statistical exercise begins to make sense biologically. The satellite image may be interrogated to determine the differences between the mean conditions within each vegetation type.

The cross-hatched black overlays in Fig. 6.1 show the local distribution of tsetse (*G. pallidipes*). These overlays occur on several different vegetation types in the classified image. The vegetation patches are much more widespread than are the tsetse patches: this suggests that the tsetse maps might be inaccurate in this area (as has long been suspected). The conclusion follows that the existing fly maps should not be used as the sole guide to choosing areas for tsetse fly suppression. More importantly the maps also suggest the routes along which flies might move, to

Fig. 6.1. Unsupervised classification of Fourier processed 1km AVHRR data for Southern Ethiopia (Lake Turkana at bottom left) showing how the distribution of major vegetation types is apparently determined by topology and water availability. The black overlay shows the local distribution of *G. pallidipes* as recorded by Ford & Katondo. This map suggests the local distribution of this species is poorly understood, and that none of the fly blocks (except possibly B) is isolated from any of the others - a result with important implications for fly suppression or local eradication schemes.



threaten a suppression area or re-invade an eradication area (*i.e.* areas coloured the same as those where the flies are known to occur).

With minimal but regular trap catch data from several sites within this region it would be possible to produce biological models, to generate a density surface similar to that of The Gambia (Fig. 5.4). Such models could indicate the local demographic vigour of the fly population, and therefore the ease with which it could be suppressed by any appropriate control measure. Accurate fly models could also generate a challenge map for the region, and this could be an input into an economic assessment of the value of the suppression scheme.

During the course of TALA 2 our existing collaboration with Guy Hendrickx in Togo continued. The Togo project now provides us with both fly and disease data to test this approach in West Africa, and this is the subject of on-going research.

4) to model representative data sets using satellite imagery and to define the minimum field data necessary for acceptable accuracy.

The generic models of Section 5 have shown how this can be done. There is no particular need for more types of fly data but, as emphasised above, it is important to ensure that the fly data apply to the cattle disease data that are often collected by different teams of people working at different times and often in different places from the tsetse personnel.

Models could be much improved by gathering more information from the infected animals, including age-prevalence data of the sort modelled in Fig. 5.5 (*i.e.* age prevalence of both infection and immunity): even ball-park figures would be better than nothing at all.

5) to verify the approach by using independent fly and disease data sets.

The current modelling has shown that extrapolation from sites of intensive studies may create problems when predictions are made for areas that are climatically very different. It is here that the wide-area approach of the statistical models could contribute to refining the biological models.

Data from both The Gambian and Togo series of sites that sampled a wide range of the environmental conditions within each country are the basis for development of this approach. For various reasons we were not able to obtain Gambian data from more than two of the study sites (Bansang and Missira) until very near the end of this project.

6) to use the models to define optimal disease control strategies.

As the original project document explained, the word 'optimal' has ecological, biological, sociological and economic dimensions, and they are rarely coincident. What is optimal biologically may be (and usually is) sub-optimal economically. It is therefore important to establish along how many dimensions, and using what criteria, the optimal strategies need to be defined.

It is now possible to define the local environmental constraints of flies and disease both statistically and biologically, and to investigate the efficacy of intervention methods with specified levels of disease transmission suppression (e.g. via fly control, or prophylactic/curative drugs).

It is now also possible to investigate in a rather general way, the economic benefits of disease suppression through use of the FAO PAATIS software developed by ERGO and based on a number of ERGO/TALA consultancy Reports to FAO (Rogers and Wint, 1996; Wint *et al.*, 1997). It is hoped that future modelling developments will be able to refine this approach.

It seems to the writer that the integration of satellite data, models and mapping and prediction capabilities is ideal to address the current poverty-focus of DFID. Poverty is an *n*-dimensional constraint on development, many of the dimensions depending on environmental factors that may be investigated using the techniques developed within TALA 2.

7) to disseminate the results and run a Training Workshop for middle/advanced level personnel (workshop dependent upon additional funds).

A short work-shop was held by the PI in Addis Abeba in spring 1998, and involved all the regional planning officers of the new Government of Ethiopia. The focus of the work-shop was to explain how satellite data and GIS could contribute to an ambitious Ethiopian scheme to map poverty country-wide.

The work-shop identified four main categories contributing to poverty (political, educational, agricultural and structural) and suggested about 30 indicators that could be recorded during the data collection phase. The Government of Ethiopia has committed more funds to this project than to any other equivalent Ministry, an indication of the importance it attaches to identifying the many different local reasons for poverty.

The regional representatives at the meeting were very enthusiastic about the potential of both RS and GIS techniques in helping to make sense of the poverty and health data that are to be collected nation-wide. The present PI subsequently submitted proposals to both DFID and the Burroughs Wellcome infectious diseases initiative to use satellite data to investigate the disease and poverty burdens in Africa, concentrating initially on Ethiopia. Unfortunately these were not supported and so the current TALA work in this area comes to an end.

Acknowledgements

The following generously provided data and invaluable advice from past and present tsetse and trypanosomiasis projects:

Prof. Leo Dempfle and Prof. Dieter Mehlitz, Directors General & Dr. Bill Snow, Senior Entomologist, International Trypanotolerance Centre, PMB 14, Banjul, The Gambia.

- Dr. Guy Hendrickx, Project Manager and Trypanosomiasis Specialist, & Dr. A. Napala Directeur National, Projet GCP-TOG-013-BEL, BP 114 Sokodé, Togo.
- Dr. J. Omuse, and Dr. J. Ndung'u, Directors, & Dr. Peter Stevenson, ODA Project Manager, Kenya Trypanosomiasis Research Institute, PO Box 362 Kikuyu, Kenya. Tel (+254)
- Dr. Guy D'Iteren, Director, and Dr. Stephen Leak, Senior Entomologist, International Livestock Research Institute, P.O. Box 30709, Nairobi, Kenya.
- Dr. Jan Slingenbergh, Senior Animal Health Officer (Tsetse) & Brian Hursey Senior Animal Health Officer (Tryps), FAO, Via delle Terme di Caracalla, Roma 00100, Italy.
- Dr. G.R.W. Wint, Managing Director, ERGO, Department of Zoology, South Parks Rd., Oxford, U.K.

We also gratefully acknowledge the assistance of many other people during our data gathering field trips abroad, particularly the ITC field staff in Gambia, Dr. Woudyalew Mulatu (Ethiopia) and George Oloo (Kenya)

Selected Project Outputs

The following are outputs of project staff in peer-reviewed journals

- Bourn, D., Gibson, C., Augeri, D., Wilson, C.J., Church, J. & Hay, S.I. (1999) The rise and fall of the Aldabran giant tortoise population. *Proceedings of the Royal Society London B* **266**, 1091-1100.
- Green, R.M. & Hay, S.I. (2000) Mapping of climate variables using meteorological satellite data. *Remote Sensing of Environment* (in press).
- Hay, S.I. & Lennon, J.J. (1999) Deriving meteorological variables across Africa for the study and control of vector-borne disease: a comparison of remote sensing and spatial interpolation of climate. *Tropical Medicine and International Health* **4**, 58-71.
- Hay, S.I. (1997) Remote sensing and disease control: past, present and future. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **91**, 105-106.
- Hay, S.I. (2000) An overview of remote sensing and geodesy for epidemiology and public health application. *Advances in Parasitology* (in press)
- Hay, S.I., Omumbo, J., Craig, M. & Snow, R.W. (2000) Earth Observation, Geographic Information Systems and Plasmodium falciparum Malaria in sub-Saharan Africa. *Advances in Parasitology* (in press)
- Hay, S.I., Packer, M.J. & Rogers, D.J. (1997) The impact of remote sensing on the study and control of invertebrate intermediate hosts and vectors for disease. *International Journal of Remote Sensing* **18**, 2899-2930.
- Hay, S.I., Randolph, S.E. & Rogers, D.J. (eds.) (2000) *Remote Sensing and GIS in Epidemiology*. Academic Press, London. (in press).
- Hay, S.I., Rogers, D.J., Toomer, J.F. & Snow, R.W. (2000). Annual Plasmodium falciparum entomological inoculation rates (EIR) across Africa. I. Literature survey, internet access and review. *Transactions of the Royal Society of Tropical Medicine and Hygiene* (in press)
- Hay, S.I., Snow, R.W. & Rogers, D.J. (1998) From predicting mosquito habitat to malaria seasons using remotely sensed data: Practice, problems and perspectives. *Parasitology Today* **14**, 306-313.
- Hay, S.I., Snow, R.W. & Rogers, D.J. (1998) Predicting malaria seasons in Kenya using multitemporal meteorological satellite sensor data. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **92**, 12-20.
- 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., Napala, A., Slingenbergh, J.H.W., De Deken, R., Vercruyssen, J. & Rogers, D.J. (2000). The spatial patterns of trypanosomiasis predicted with the aid of satellite imagery. *Parasitology* (in press).

- Hendrickx, G., Napala, A., Slingenbergh, J.H.W., De Deken, R. and Rogers, D.J. (2000) A contribution towards reducing the extend of area-wide tsetse surveys using medium resolution meteorological satellite data. *Bulletin of Entomological Research* (Accepted).
- Johnson, D.D.P., Hay, S.I. & Rogers, D.J. (1998) Contemporary environmental correlates of endemic bird areas derived from meteorological satellite sensors. *Proceedings of the Royal Society of London B* **265**, 951-959.
- Myers, M.F., Rogers, D.J., Cox, J., Flauhalt, A. & Hay, S. I. Forecasting disease risk for increased epidemic preparedness in public health *Advances in Parasitology* (in press)
- Patz, J.A., Strzepek, K., Lele, S., Hedden, M., Greene, S., Noden, B., Hay, S.I. , Kalkstein, L. & Beier, J.C. (1998) Predicting key malaria transmission factors, biting and entomological inoculation rates, using modelled soil moisture in Kenya. *Tropical Medicine & International Health*, **3**, 818-827.
- Randolph, S.E. & Rogers, D.J. (1997) A generic population model for the African tick *Rhipicephalus appendiculatus*. *Parasitology* **115**, 265-279.
- Randolph, S.E. & Rogers, D.J. (2000) Fragile transmission cycles of tick-borne encephalitis virus may be disrupted by predicted climate change. *Proceeding of the Royal Society of London B* (submitted).
- Randolph, S.E., Green, R.M., Labuda, M. & Rogers, D.J. (2000) Pan-European focal distribution of tick-borne encephalitis predicted and explained by satellite imagery. *Nature* (under review).
- Randolph, S.E., Green, R.M., Peacey, M.F. & Rogers, D.J. (2000) Seasonal synchrony: the key to tick-borne encephalitis foci identified by satellite data. *Parasitology* (in press).
- Randolph, S.E., Miklisová, D., Lysy, J., Rogers, D.J. & Labuda, M. (1999) Incidence from coincidence: patterns of tick infestations on rodents facilitate transmission of tick-borne encephalitis virus. *Parasitology* **118**, 177-186.
- Robinson, T.P., Rogers, D.J. & Williams, B.G. (1997) Mapping tsetse habitat suitability in the common fly belt of Southern Africa using multivariate analysis of climate and remotely sensed vegetation data. *Medical and Veterinary Entomology* **11**, 235-245.
- Robinson, T.P., Rogers, D.J. & Williams, B.G. (1997) Univariate analysis of tsetse habitat in the common fly belt of Southern Africa using climate and remotely sensed vegetation data. *Medical and Veterinary Entomology* **11**, 223-234.
- Rogers, D.J. (2000) Satellites, space, time and the African trypanosomiasis *Advances in Parasitology* (in press).
- Rogers, D.J. & Randolph, S.E. (2000) The global spread of malaria in a future, warmer world: under-quantified and over-stated? *Science* (under review).
- Rogers, D.J., Hay, S.I. & Packer, M.J. (1996) Predicting the distribution of tsetse flies in West Africa using temporal Fourier processed meteorological satellite data. *Annals of Tropical Medicine and Parasitology* **90**, 225-241.
- Rogers, D.J., Hay, S.I., Packer, M.J. & Wint, G.R.W. (1997) Mapping land-cover over large-areas using multispectral data derived from the NOAA-AVHRR: a case study of Nigeria. *International Journal of Remote Sensing* **18**, 3297-3303.

APPENDIX

A simple model for trypanosome transmission.

Changes in the proportions of the vertebrate hosts that are susceptible (*s*) infected but not yet infectious (*f*), infectious (*x*) and recovered, immune (*i*) are described by the following set of equations,

$$\frac{ds}{dt} = - abmx's + tN + wi - \mu s$$

$$\frac{df}{dt} = abmx's - vf - \mu f$$

..7

$$\frac{dx}{dt} = vf - rx - \mu x$$

and

$$\frac{di}{dt} = rx - wi - \mu i$$

where *a* is the biting rate of vectors on hosts,
b is the transmission coefficient from vector to vertebrate,
m is the ratio of vectors to hosts ($= M/N$),
x' is the proportion of infected vectors,
t is the birth rate of the hosts,
N is host population size (the present equations describe proportions, so that $N=s + f + x + i = 1$),
w is the rate of loss of immunity, returning immune animals to the susceptible category, *s*,
μ is the host's natural death rate,
v is the incubation rate of the disease in the vertebrate hosts,
and *r* is the rate of recovery of the vertebrates from infection.

The equivalent equations for the vectors, continuing to use prime to indicate vector parameters and variables analogous to those of the hosts, are as follows:

$$\frac{ds'}{dt} = - acxs' + t'M - \mu' s'$$

$$\frac{df'}{dt} = acxs' - v'f' - \mu' f' \quad \text{..8}$$

$$\frac{dx'}{dt} = v'f' - \mu' x'$$

where, in addition,
c is the transmission coefficient from vertebrate to vector,
t' is the birth rate of the tsetse population,
M is the tsetse population size (as in the case of the hosts, $M=s' + f' + x' = 1$),
and *μ'* is the tsetse mortality rate ($= t'$ at equilibrium).

Vector mortality rate appears explicitly in these equations since it is assumed that fly infections can only be lost when infected flies die. These losses are balanced by births, which introduce new, susceptible flies into s' . In the case of the vertebrate hosts, animals which lose their immunity re-cycle into the susceptible category.

The sum of each set of equations, for both hosts and vectors, is zero, indicating that although each subpopulation (susceptible, infected etc.) may change over time there is no net change in the summed proportions, which must always be equal to 1.0.

The above equations, which apply to the simple situation of a single-vector, single-host disease may be applied to the African trypanosomiasis assuming all other hosts fed upon by flies are negligible sources of infection compared to the modelled hosts. The equations are therefore simplified versions of those that have been written for the African trypanosomiasis (Rogers, 1988) and may be taken to apply to the situation of trypanosomiasis in domestic animals in areas with few alternative hosts; this simplification allows us to model disease transmission quickly, and to estimate the need for more complex transmission models in field situations. Human sleeping sickness involving domestic and wild reservoir hosts will not be satisfactorily modelled in this simple way.

In making the output of the tsetse population model one of the inputs into the disease transmission model additional scaling parameters are required to relate the tsetse numbers to the vertebrate host numbers. Since the vector-host ratio (m) always and only appears in the above equations with the transmission coefficient b , estimates of these two quantities will vary together so that their product remains the same.

BIBLIOGRAPHY

- Agbu, P.A. and James, M.E. (1994). *The NOAA/NASA Pathfinder AVHRR land data set user's manual.*, Goddard Distributed Active Archive Center Publication (GCDC), Greenbelt, MD.
- Anderson, R.M. and May, R.M. (1991). *Infectious Diseases of Humans: dynamics and control.* Oxford, Oxford University Press.
- Beek, E.G., Stein, A. and Janssen, L.L.F. (1992). Spatial variability and interpolation of daily precipitation amount. *Stochastic Hydrology and Hydraulics* **6**, 209-221.
- Bursell, E. (1959). The water balance of tsetse flies. *Transactions of the Royal Entomological Society of London* **3**, 205-235.
- Buxton, P.A. and Lewis, D.J. (1934). Climate and tsetse flies: laboratory studies upon *Glossina submorsitans* and *tachinoides*. *Philosophical Transactions of the Royal Society of London (B)* **224**, 175-240.
- Chatfield, C. (1980). *The Analysis of Time Series: An Introduction.* London, Chapman and Hall.
- Cooper, D.I. and Asrar, G. (1989). Evaluating atmospheric correction models for retrieving surface temperatures from the AVHRR over a Tallgrass prairie. *Remote Sensing of Environment* **27**, 93-102.
- Diggle, P.J. (1990). *Time Series: A Biostatistical Introduction.* Oxford, Clarendon Press.
- Eck, T.F. and Holben, B.N. (1994). AVHRR split window temperature differences and total precipitable water over land surfaces. *International Journal of Remote Sensing* **15**, 567-582.
- Flitcroft, I.D., Milford, J.R. and Dugdale, G. (1989). Relating point to area average rainfall in semiarid West Africa and the implications for rainfall estimates derived from satellite data. *Journal of Applied Meteorology* **28**, 252-266.
- Glasgow, J.P. (1963). *The Distribution and Abundance of Tsetse.* Oxford, Pergamon Press.
- Hargrove, J.W. (1994). Reproductive rates of tsetse flies in the field in Zimbabwe. *Physiological Entomology* **19**, 307-318.
- Hargrove, J.W. and Packer, M.J. (1993). Nutritional states of male tsetse flies (*Glossina* spp.) (Diptera: Glossinidae) caught in odour-baited traps and artificial refuges: models for feeding and digestion. *Bulletin of Entomological Research* **83**, 29-46.
- Hargrove, J.W. and Williams, B.G. (1998). Optimized simulation as an aid to modelling, with an application to the study of a population of tsetse flies, *Glossina morsitans morsitans* (Diptera: Glossinidae). *Bulletin of Entomological Research* **88**, 425 - 435.
- Hay, S.I. and Lennon, J.J. (1999). Deriving meteorological variables across Africa for the study and control of vector-borne disease: a comparison of remote sensing and spatial interpolation of climate. *Tropical Medicine and International Health* **4**, 58-71.
- Hay, S.I., Tucker, C.J., Rogers, D.J. and 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.

- Herman, A., Kumar, V.B., Arkin, P.A. and Kousky, J.V. (1997). Objectively determined 10-day African rainfall estimates created for famine early warning systems. *International Journal of Remote Sensing* **18**, 2147-2159.
- Holben, B., Kimes, D. and Fraser, R.S. (1986). Directional reflectance response in AVHRR red and near-IR bands for three cover types and varying atmospheric conditions. *Remote Sensing of Environment* **19**, 213-236.
- Huang, N.E., Shen, Z. and Long, S.R. (1999). A new view of nonlinear water waves: the Hilbert spectrum. *Annual Review of Fluid Mechanics* **31**, 417-457.
- Huang, N.E., Shen, Z., Long, S.R., Wu, M.L. and Shih, H.H. (1998). The Empirical Mode Decomposition and Hilbert spectrum for nonlinear and nonstationary time series analysis. *Proceedings of the Royal Society of London, Series A.* **454**, 903-995.
- Hulme, M., Conway, D., Joyce, A. and Mulenga, H. (1996). A 1961-90 climatology for Africa south of the equator and a comparison of potential evapotranspiration estimates. *South African Journal of Science* **92**, 334-343.
- Hutchinson, M.F. (1989). A new procedure for gridding elevation and stream line data with automatic removal of spurious pits. *Journal of Hydrology* **106**, 211-232.
- James, M.E. and Kalluri, S.N.V. (1994). The pathfinder AVHRR land data set - an improved coarse resolution data set for terrestrial monitoring. *International Journal of Remote Sensing* **15**, 3347-3363.
- Lambin, E.F. and Ehrlich, D. (1995). Combining vegetation indexes and surface temperature for land-cover mapping at broad spatial scales. *International Journal of Remote Sensing* **16**, 573-579.
- Lennon, J.J. and Turner, J.R.G. (1995). Predicting the spatial distribution of climate: temperature in Great Britain. *Animal Ecology* **64**, 370-392.
- Myneni, R.B. and Asrar, G. (1994). Atmospheric effects and spectral vegetation indices. *Remote Sensing of Environment* **47**, 390-402.
- Pinping, X. and Arkin, P.A. (1998). Global monthly precipitation estimates from satellite-observed outgoing longwave radiation. *Journal of Climate* **11**, 137-164.
- Press, W.H., Teukolsky, S.A., Vetterling, W.T. and Flannery, B.P. (1995). *Numerical Recipes in Fortran 77: the art of scientific computing*. Cambridge, Cambridge University Press.
- Price, J.C. (1984). Land surface temperature measurements from the split window channels of the NOAA 7 advanced very high resolution radiometer. *Journal of Geophysical Research* **89**, 7231-7237.
- Prihodko, L. and Goward, S.N. (1997). Estimation of air temperature from remotely sensed surface observations. *Remote Sensing of the Environment* **60**, 335-346.
- Prince, S.D., Goetz, S.J., Dubayah, R.O., Czajkowski, K.P. and Thawley, M. (1998). Inference of surface and air temperature, atmospheric precipitable water and vapour pressure deficit using Advanced Very High-Resolution Radiometer satellite observations: comparison with field observations. *Journal of Hydrology* **212-213**, 230-249.
- Prince, S.D. and Goward, S.N. (1995). Global primary production: a remote sensing approach. *Journal of Biogeography* **22**, 815-835.
- Rawlings, P., Dwyer, R.H. and Snow, W.F. (1991a). An analysis of survey measurements of tsetse challenge to trypanotolerant cattle in relation to aspects of analytical models of transmission. *Parasitology* **102**, 371-377.
- Rawlings, P., Dwyer, R.H. and Snow, W.F. (1991b). An analysis of survey measurements of tsetse challenge to trypanotolerant cattle in relation to aspects of analytical models of trypanosomiasis. *Parasitology* **102**, 371-377.

- Rogers, D. (1974). *Ecology of Glossina: natural regulation and movement of tsetse fly populations*. Paris, I.E.M.V.T.
- Rogers, D.J. (1988). A general model for the African trypanosomiasis. *Parasitology* **97**, 193-212.
- Rogers, D.J. (1990). A general model for tsetse populations. *Insect Science and its Application* **11**, 331-346.
- Rogers, D.J., Hendrickx, G. and Slingenbergh, J. (1994). Tsetse flies and their control. *Revue Scientifique et Technique de l'Office International des Epizooties* **13**, 1075-1124.
- Rogers, D.J. and Randolph, S.E. (1984). A review of the density-dependent processes in tsetse populations. *Insect Science and its Application* **5**, 397-402.
- Rogers, D.J. and 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. and Kuzoe, F.A.S. (1984). Local variation in the population dynamics of *Glossina palpalis palpalis* (Robineau-Desvoidy) (Diptera: Glossinidae). I. Natural population regulation. *Bulletin of Entomological Research* **74**, 403-423.
- Rogers, D.J. and Wint, G.R.W. (1996). *Towards Identifying Priority Areas for Tsetse Control in East Africa*. Rome, FAO.
- Smith, W.L. (1966). Note on the relationship between total precipitable water and surface dewpoint. *Journal of Applied Meteorology* **5**, 726-727.
- Starck, J.-L., Murtagh, F. and Bijaoui, A. (1998). *Image Processing and Data Analysis: the multiscale approach*. Cambridge, Cambridge University Press.
- Stollnitz, E.J., Derosé, T.D. and Salesin, D.H. (1996). *Wavelets for Computer Graphics*. San Francisco, Morgan Kaufmann.
- Stoms, D.M., Bueno, M.J. and Davis, F.W. (1997). Viewing geometry of AVHRR image composites derived using multiple criteria. *Photogrammetric Engineering and Remote Sensing* **63**, 681-689.
- Stowe, L.L., McClain, E.P., Carey, R., Pellegrino, P., Gutman, G.G., Davia, P., Long, C. and Hart, S. (1991). Global distribution of cloud cover derived from NOAA/AVHRR operational satellite data. *Advances in Space Research* **11**, 51-54.
- Sugita, M. and Brutsaert, W. (1993). Comparison of land surface temperatures derived from satellite observations with ground truth during FIFE. *International Journal of Remote Sensing* **14**, 1659-1676.
- Wint, W., Robinson, T.P. and Rogers, D.J. (1997). *Ecozones, Farming Systems and Priority Areas for Tsetse Control in East, West and Southern Africa*. Rome, FAO.