The spatial pattern of trypanosomosis prevalence predicted with the aid of satellite imagery

G. HENDRICKX1*, A. NAPALA2, J. H. W. SLINGENBERGH3, R. DE DEKEN4, J. VERCRUYSSÈ3 and D. J. ROGERS6

1 FAO Trypanosomosis project GCP-RAF-347-BEL, BP 2034, Bobo Dioulasso, Burkina Faso
2 FAO Trypanosomosis project GCP-RAF-347-BEL, BP 114, Sokodé, Togo
3 FAO, AGAH, Via delle Terme di Caracalla, 00100 Rome, Italy
4 ITG/IMT, Nationale Straat 155, 2000, Antwerp, Belgium
5 Faculteit Diergeneeskunde, RUG, Salisburylaan, Merelbeke, Belgium
6 Department of Zoology, Oxford University, South Parks Road, Oxford OXI 3PS, England

(Received 22 March 1999; revised 19 August 1999; accepted 19 August 1999)

SUMMARY

Information on the spatial pattern of African animal trypanosomosis forms a prerequisite for rational disease management, but few data exist for any country in the continent. The present study describes a raster or grid-based Geographic Information System for Togo, a country representative of subhumid West Africa, with data layers on tsetse, trypanosomosis, animal production, agriculture and land use. The paper shows how trypanosomosis prevalence and packed cell volume (PCV) map displays may be predicted from correlations between representative field data and environmental and satellite data acquired from the National Oceanographic and Atmospheric Administration (NOAA) and Meteosat platforms. Discriminant analytical methods were used to assess the relationship between the amount of field data used and the accuracy of the predictions obtained. The accuracy of satellite derived predictions decreases from tsetse abundance to trypanosomosis prevalence to PCV value. The predictions improve when eco-climatic and epidemiological predictors are combined. In Togo, and probably elsewhere, the patterns of trypanosomosis prevalence and PCV are much influenced by animal husbandry and other anthropogenic factors. Additional predictor variables, incorporating these influences might therefore further improve the models.

Key words: trypanosomosis, multiple discriminant analysis, PCV, prediction, remote sensing, West Africa.

INTRODUCTION

Tsetse-transmitted trypanosomosis is a major constraint on increased livestock production in the subhumid eco-zone of sub-Saharan Africa. Moreover, tsetse infestations prevent the successful integration of crop and ruminant production (Winrock International, 1992). It is now generally accepted that an integrated approach to tsetse or disease suppression is a more realistic goal than tsetse or disease eradication (FAO, 1994), and there are many options and tools that may be used in such an approach. Central to many of these is a detailed knowledge of the spatial and temporal patterns of flies and disease, but such information is either not collected, or is outdated (Rogers & Randolph, 1988). In the last few years several attempts have been made to improve tsetse distribution and abundance maps using relationships between historical map data and recent satellite imagery (Rogers & Randolph, 1993; Rogers, Hay & Packer, 1996; Robinson, Rogers & Williams, 1997a, b). Only in the case of Togo have contemporary satellite and fly data been used (Hendrickx et al. 1993; Hendrickx, 1999; Rogers, Hendrickx & Slingenergh, 1994) as part of a wider study on tsetse, trypanosomosis, animal production, agriculture and land use in that country (Hendrickx & Napala, 1995). This study was also the first to investigate the effect of various degrees of subsampling the national database on the accuracy of the predictions of tsetse presence throughout the country (Hendrickx, 1999).

Fly numbers affect trypanosomosis risk in complex and non-linear ways and it may be preferable to use the methodology developed to predict fly numbers instead of predicting trypanosomosis prevalence directly or to use some other reliable measure of trypanosomosis impact such as the packed cell volume (PCV). This paper explores the utility of using eco-climatical satellite data to predict disease prevalence and PCV values, and investigates the extent to which satellite data may be used to minimize parasitological field surveys.

MATERIALS AND METHODS

The study area

Togo, between Ghana and Benin in West Africa, is 550 km long and up to 150 km wide with a total surface of 56,000 km². It can be divided into 4 areas with different seasonal characteristics (Hendrickx et
Fig. 1. The geography and climate of Togo. The small graphs show rainfall (histogram) and the greenness of vegetation (areas) depicted as satellite measured Normalized Difference Vegetation Indices (NDVI), for 3 sites from North to South: Dapaong, Sokodé and Tsevié, respectively referred to by a black square in Fig. 1a. Seasonal clusters were obtained after hierarchical clustering of a series of remotely sensed and ground measured eco-climatical variables (Hendrickx et al. 1999a). Major differences between northern clusters (A–B) and southern clusters (C–D) are a significantly higher ratio of humid over dry months in clusters C–D and a distinctly monomodal seasonal cycle in clusters A–B (i.e. a single long dry season).

a. General features.

b. Seasonal clusters.

(Hendrickx et al. 1999b). Briefly, a total of 26,481 blood samples were collected during a cross-sectional study in 200 out of the 217 grid-cells where sedentary cattle herds were found present. On average 4-7 different herds were sampled per grid. After microcentrifugation and measurement of the PCV, a wet smear of the buffy coat was examined for trypanosomes (Murray, Murray & McIntyre, 1977).

Since surveys were carried out during different times of the year in different places, seasonal adjustments were calculated on the basis of monthly indices obtained by pooling the monthly prevalence data both for the northern (seasonal clusters A and B) and southern (seasonal clusters C and D) part of the country. Monthly indices were computed sep-
Fig. 2. The influence of eco-climatical variables on the epidemiology of trypanosomosis. Anthropogenic variables are depicted in boxes with a grey background.

Spatial smoothing was carried out through averaging each grid value with that of its 8, or fewer (in the case of border squares), neighbouring grid-squares. Missing grid-values were also calculated in this way. The resulting seasonally adjusted, smoothed data were re-classed into 3 categories for later analysis, of low, medium or high prevalence or herd PCV. Class limits were set to obtain an approximately equal number of grid squares per class.

**Predictor variables**

The epidemiology of African animal trypanosomosis is influenced by a variety of external factors (Fig. 2), many of which are directly or indirectly influenced by temperature, rainfall, vegetation or altitude. These variables for Togo were obtained either from digital elevation surfaces, meteorological records, published maps, or from remotely sensed data. Those used in the present study are listed in Table 1.

**Eco-climatical data.** Rainfall figures were derived from 30 year data held by the meteorological unit of the Ministère de l’Agriculture de l’Elevage et des Pêches; Lomé. The average annual rainfall (mm) and the number of months with rainfall less than 30 mm (dry months) were used in the present study.

Temperatures (°C) were interpolated from point records (HMSO, 1983) using a contouring programme in a commercial software package (SURFER, Golden Software Inc.); the annual temperature range used here is the difference between the mean monthly maximum and minimum temperatures.

An elevation figure was calculated for each grid cell using the Digital Elevation Model (DEM) for Africa produced by the Global Land Information System of the United States Geological Survey, Earth Resources Observation Systems. The original 0.083° resolution data were resampled and averaged to fit the 0.125° resolution data base.

**Epidemiological data.** The agricultural intensity (expressed as the percentage of land brought into the cultivation cycle, including fallow) was computed per grid square from published maps derived originally from aerial photographs (PNUD, 1984).

Human population data were extracted from the UNEP population density map (UNEP, 1991). Densities are given in eight classes: 0–1, 1–5, 5–25, 25–50, 50–100, 100–250, 250–500, > 500 people/km².

The production of comprehensive, 0.125 degree resolution, tsetse distribution and abundance maps covering the whole of Togo is detailed elsewhere (Hendrickx et al. 1999a). Tsetse were sampled in 305 of the 311 grid-squares using an average of 12
Table 1. Predictor variables

(Rain, rainfall; DryM, months with less than 30 mm rain; Trange, range between the annual average monthly minimum and maximum temperature; DEM, digital elevation model. %Agri, percentage agriculture; Tachi, pooled adjusted for seasonality abundance of G. tachinoides; HpopD, human population density; CattleD, cattle density; %Taur, percentage of taurine cattle; %TaurF, percentage of taurine fertile females; %TaurM, 3 × 3 pooled percentage of taurine fertile males. Own, number of owners per herd; %ORur, percentage of rural owners; %ORAgr, percentage of urban owners being agriculturists; %ORHer, percentage of rural owners being herdsmen; %OAgri, percentage of owners being agriculturists; %OURb, percentage of urban owners; %OUCiv, percentage of urban owners being civil servants; %OUTrad, percentage of urban owners being tradesmen. NDVI, normalized difference vegetation index; Ch3, Channel 3 value; Price, Price thermal brightness index; CCD, cold cloud duration. P1–2–3 & a1–2–3, phase and corresponding amplitude of the annual; bi-annual or tri-annual Fourier cycle; x, maximum; m, minimum; r, range = maximum − minimum.)

<table>
<thead>
<tr>
<th>Ground measured variables</th>
<th>Epidemiological variables</th>
<th>Remotely sensed variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eco-climatical</td>
<td>Epidemiological</td>
<td>Remotely sensed variables</td>
</tr>
<tr>
<td>Rain</td>
<td>%Agri</td>
<td>Own</td>
</tr>
<tr>
<td>DryM</td>
<td>HpopD</td>
<td>%ORur</td>
</tr>
<tr>
<td>Trange</td>
<td>Tachi</td>
<td>%ORAgr</td>
</tr>
<tr>
<td>DEM</td>
<td>Cattle D</td>
<td>%ORHer</td>
</tr>
<tr>
<td></td>
<td>%Taur</td>
<td>%OAgri</td>
</tr>
<tr>
<td></td>
<td>%TaurF</td>
<td>%OURb</td>
</tr>
<tr>
<td></td>
<td>%TaurM</td>
<td>%OUCiv</td>
</tr>
<tr>
<td></td>
<td></td>
<td>%OUTrad</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NDVI m</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NDVI r</td>
</tr>
</tbody>
</table>

standard biconical traps (Challier et al. 1977) per trapping site. On average, 2–1 sites were sampled per grid during 24 or 48 h. A study of the relationships between tsetse distribution, tsetse abundance and disease prevalence revealed that, out of the 5 different tsetse species present, Glossina tachinoides was the main vector of the disease in cattle (Hendrickx et al. 1999b). An abundance map of this essentially riverine species was generated for use in the present study. Logarithmically (log₁₀) transformed fly density data were first adjusted for seasonality and smoothed spatially, following the same procedure as for disease prevalence and PCV.

Data on cattle density, breeds and ownership were gathered during a country-wide cattle census (Hendrickx et al. 1999b). The survey was guided by the results of a previous census (PROPAT, 1990) and the local knowledge of veterinary field staff. All identified sedentary herds were visited and, for each one, a questionnaire was completed with inputs from the herdsman. The data layers resulting from this survey are given in Table 1. Because this survey was both exhaustive and unlikely to show marked seasonal changes (at least in herd numbers), no seasonal adjustment or smoothing was performed on the data, except for the fertile taurine male layer where a 3 × 3 spatial smoothing was adopted to reflect the mobility of genitors (Hendrickx et al. 1999b).

Remotely sensed data. Remotely sensed data originated from 2 sources, the Advanced Very High Resolution Radiometer (AVHRR) on board the polar orbiting National Oceanic and Atmospheric Administration (NOAA) satellites and the Visible and Infrared Spin Scan Radiometer (VISSR) on board the geo-stationary METEOSAT satellite. Data from these platforms were processed either at source or specifically for this study to produce a variety of products as follows. (i) The Normalized Difference Vegetation Index (NDVI) is an index of active photosynthetic vegetation (Tucker et al. 1983) and is obtained from measurements made in the visible red and near infrared wavebands of the AVHRR instrument (Channels 1 and 2 respectively) using the equation:

\[
NDVI = \frac{(Ch2 - Ch1)}{(Ch2 + Ch1)}.
\]

The index is based on the principal that active plant tissue (mainly chlorophyll pigments) absorbs light in the visible red waveband (Ch1) and mesophyll tissue reflects light in the near-infrared waveband (Ch2) (Sellers, 1985). Healthy plants will thus look darker in the visible and lighter in the near-infrared bands than unhealthy or dried out plants, yielding a higher NDVI.

(ii) The Land Surface Temperature, AVHRR Channel 3, measuring the middle infrared radiation, both reflected and emitted, has been related to vegetation canopies, especially in forested areas, and was recently recognized to be one of the better predictor variables of landcover type in Nigeria (Rogers et al. 1997).

(iii) The Price Thermal Brightness Index, Ts (Price, 1984), simplified by Hay et al. (1996), is
phenomena, have been given by Hay and their application to measuring land-surface
(temporally). The applicable threshold (which varies spatially and
each month that cloud-top temperatures were below
1991). CCD are expressed as the number of hours in
related to ground measures of rainfall (Snijders,
cloud-top temperatures. In West Africa such tem-
from METEOSAT, is derived from measured
variables (i.e. the value least affected by cloud or other
study used the monthly maximum value of each
variable (i.e. the value least affected by cloud or other
atmospheric contamination), in imagery with an
original resolution of 8 km provided by the NOAA
PATHFINDER program (James & Kalluri, 1994)
and for the period 1988–1990. The METEOSAT
data, with the same temporal and spatial resolution,
were available only from September 1988, so the 3-
year period 1989–1991 was used in this study. All
8 km imagery was subjected to temporal Fourier pro-
cessing which extracts from the time-series a satellite
‘finger-print’ of vegetation phenology or meteorolo-
gical data seasonality (Rogers & Williams, 1994;
Rogers et al. 1996). This information was stored as
the mean value of each variable as well as the
amplitudes (i.e. peaks) and phases (i.e. seasonal
timing) of the annual, bi-annual and tri-annual
cycles that contribute to each ‘fingerprint’.
Finally, all Fourier-processed 8 km imagery was
extracted as a series of 2 × 2 pixel arrays coincident
with the Togo 0°125 grid square centres. These pixel
arrays were averaged and assigned to their geo-
 graphical location in the database.

Discriminant analysis

Multiple discriminant analysis is traditionally the
most appropriate technique of multivariate analysis
in situations where the dependent variable has more
than 2 categories and the predictor variables are a
series of multi-variate normally distributed metrical
variables (the technique can also extend to dummy-
coded categorical variables). An earlier version of the
model adopted in this paper has been described in
detail (Rogers et al. 1996). Predictor variables were
ranked individually in a step-wise fashion, and the
best variables selected to compute discriminant
functions and produce predictive maps of fly abun-
dance.

Before analysis outlying data points, defined as
observations or values that differed from their class
means by more than 6 standard deviations, were
filtered from the data sets. Such outliers obviously
adversely affect the performance of any technique
that assumes multi-variate normality, and were
generally attributable to known errors (e.g. the
inclusion of satellite data from sea pixels in the
coastal zones). The selection of discriminant variables
was based on the calculation of Mahalanobis dis-
tances (D²) between the groups that were to be
distinguished by the analysis. The variable selection
rule was designed to maximize D² between group
centroids, first computing the best single-variable
model and subsequently adding more variables.

Analysis of satellite-derived predictions of tsetse
abundance (Hendrickx, 1999) has shown that the
accuracy of the prediction was optimal when the
variable selection rule described above concentrated
only upon fly abundance within the fly distribution
limits and ignored grid squares from which no flies
were recorded. A similar approach was adopted here
for the disease data.

Several analyses were carried out with different
sets of predictor variables, in order to examine their
relative contributions to describing the field data: (i)
epidemiological variables only, (ii) epidemiological
variables plus remotely sensed vegetation-related
data (NDVI, Ch3), (iii) all predictor variables and
(iv) ground-measured plus remotely sensed eco-
climatical data.

To test the accuracy of the approach adopted here
the field data were divided into 2 subsamples, one
(the training set) to determine the important pre-
dictor variables (i.e. to compute discriminant func-
tions and also to define prior probabilities of class
membership) and the other independently to assess
the accuracy of predictions made using these
variables with a different data set (the ‘hold out’ or
‘predicted’ set). Training sets should be of a
minimum sample size depending upon the number
and distributions of the variables used, and the
number of classes involved. In this study we kept a
minimum of 100 samples in the training set (Hair
et al. 1995). The main aim of the present study was to
explore ways to economize field surveys and yet
produce accurate or even improved spatial dis-
tribution layers.

The training sets were randomly selected for each
seasonal cluster (Fig. 1), or satellite derived ecozone,
and comprised 100, 66 or 33 % of the total data set.
Specific discriminant functions were computed each
time for the separate trypanosome species, for both
species combined and for the PCV.

Map displays of predictions were compared with
the originally observed prevalences and PCV. For
each predicted map the percentage of predicted grid-
square values in agreement with the observed values
were calculated per class (%)correct, or ‘hit-ratio’).
The percentage correct figures given here for the
results obtained with a 66 or 33 % trainingset are
Fig. 3. Best predictions obtained for the total prevalence. Results were obtained using different sets of predictor variables (see text and Table 2 for more details).
Satellite prediction of trypanosomosis prevalence

Fig. 4. Best predictions obtained for the prevalence of *Trypanosoma congolense*. Results were obtained using different sets of predictor variables (see text and Table 3 for more details).

Fig. 5. Best predictions obtained for the prevalence of *Trypanosoma vivax*. Results were obtained using different sets of predictor variables (see text and Table 3 for more details).

Based only upon the grid squares not included in the training set whilst those given for the 100% training set reflect the ability of the model to describe the whole data set ‘known to the analysis’.

The computed classification matrices, in which the distribution of observed data appears in rows, and predicted data in columns, allow a study of the error distribution. Given the aim of the study, emphasis was placed on the occurrence of major errors, i.e. of more than one class, since these could generate major decision-taking errors during application of the methodology for surveying. Errors were considered to be over- or under-estimations of prevalence or PCV values. The predictive accuracy was further recorded as a single value, Kappa Coefficient, or Index of Agreement (Cohen, 1960), widely used for this purpose (Congalton, 1991):

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

where $p_o$ = the sum of the proportions of predictions in concordance with observed values $p_e$ = the sum of proportions of concordant predictions expected by...
Fig. 6. Best predictions obtained for the mean herd PCV. Results were obtained using different sets of predictor variables (see text and Table 4 for more details).
Table 2. Best predictions of total prevalence (predicted set)

(Pred. var, used predictor variables (see also Table 1): Epi, epidemiological variables; Veg, remotely sensed vegetation linked data (NDVI and Channel 3); All, all predictor variables; Eco.Clim, ground measured and remotely sensed eco-climatical variables; TS, percentage of grids used as trainingset (100%, n = 204–266%, n = 135 and 33%, n = 70); #var, number of predictor variables giving the best prediction; L/M/H, low, medium and high category; Tot, hit ratio off total sample; Major errors: overestimation (+) and underestimation of observed values with more than one class; K, Kappa index of agreement, \( z = z \)-statistic (K is significant with 95% probability when \( z > 1.96 \) and with a 99% probability when \( z > 2.58 \).)

<table>
<thead>
<tr>
<th>TS</th>
<th>Pred. var.</th>
<th>#var.</th>
<th>L</th>
<th>M</th>
<th>H</th>
<th>Tot.</th>
<th>+</th>
<th>–</th>
<th>K</th>
<th>z</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>Epi</td>
<td>12</td>
<td>74</td>
<td>57</td>
<td>70</td>
<td>67</td>
<td>3</td>
<td>5</td>
<td>0.505</td>
<td>7.16</td>
</tr>
<tr>
<td></td>
<td>Epi.Veg.</td>
<td>15</td>
<td>85</td>
<td>60</td>
<td>85</td>
<td>76</td>
<td>1</td>
<td>1</td>
<td>0.641</td>
<td>7.78</td>
</tr>
<tr>
<td></td>
<td>All</td>
<td>11</td>
<td>82</td>
<td>64</td>
<td>84</td>
<td>76</td>
<td>4</td>
<td>1</td>
<td>0.640</td>
<td>7.78</td>
</tr>
<tr>
<td></td>
<td>Eco.Clim.</td>
<td>19</td>
<td>92</td>
<td>64</td>
<td>87</td>
<td>80</td>
<td>0</td>
<td>1</td>
<td>0.719</td>
<td>7.69</td>
</tr>
<tr>
<td>66</td>
<td>Epi</td>
<td>8</td>
<td>85</td>
<td>14</td>
<td>67</td>
<td>50</td>
<td>0</td>
<td>4</td>
<td>0.274</td>
<td>2.91</td>
</tr>
<tr>
<td></td>
<td>Epi.Veg.</td>
<td>9</td>
<td>85</td>
<td>34</td>
<td>62</td>
<td>57</td>
<td>1</td>
<td>4</td>
<td>0.368</td>
<td>3.56</td>
</tr>
<tr>
<td></td>
<td>All</td>
<td>13</td>
<td>85</td>
<td>45</td>
<td>62</td>
<td>61</td>
<td>1</td>
<td>4</td>
<td>0.426</td>
<td>3.87</td>
</tr>
<tr>
<td></td>
<td>Eco.Clim.</td>
<td>16</td>
<td>90</td>
<td>45</td>
<td>52</td>
<td>60</td>
<td>1</td>
<td>6</td>
<td>0.406</td>
<td>3.77</td>
</tr>
<tr>
<td>33</td>
<td>Epi</td>
<td>4</td>
<td>74</td>
<td>35</td>
<td>45</td>
<td>50</td>
<td>1</td>
<td>6</td>
<td>0.249</td>
<td>3.58</td>
</tr>
<tr>
<td></td>
<td>Epi.Veg.</td>
<td>3</td>
<td>72</td>
<td>30</td>
<td>43</td>
<td>46</td>
<td>3</td>
<td>4</td>
<td>0.195</td>
<td>2.92</td>
</tr>
<tr>
<td></td>
<td>All</td>
<td>3</td>
<td>72</td>
<td>30</td>
<td>43</td>
<td>46</td>
<td>3</td>
<td>4</td>
<td>0.195</td>
<td>2.92</td>
</tr>
<tr>
<td></td>
<td>Eco.Clim.</td>
<td>5</td>
<td>74</td>
<td>31</td>
<td>52</td>
<td>50</td>
<td>2</td>
<td>8</td>
<td>0.266</td>
<td>3.83</td>
</tr>
</tbody>
</table>

Table 3. Best predictions of the prevalence of *Trypanosoma congolense (T.c.)* and *T. vivax (T.v.)* (predicted set)

(Pred. var, used predictor variables (see also Table 1): Epi, epidemiological variables; All, all predictor variables; TS, percentage of grids used as trainingset (100%, n = 204–266%, n = 135 and 33%, n = 70); #var, number of predictor variables giving the best prediction; L/M/H, low, medium and high category; Tot, hit ratio off total sample; Major errors: overestimation (+) and underestimation of observed values with more than one class; K, Kappa index of agreement, \( z = z \)-statistic (K is significant with 95% probability when \( z > 1.96 \) and with a 99% probability when \( z > 2.58 \).)

<table>
<thead>
<tr>
<th>TS</th>
<th>Pred. var.</th>
<th>#var.</th>
<th>L</th>
<th>M</th>
<th>H</th>
<th>Tot.</th>
<th>+</th>
<th>–</th>
<th>K</th>
<th>z</th>
</tr>
</thead>
<tbody>
<tr>
<td>T.c.</td>
<td>100 Epi</td>
<td>14</td>
<td>74</td>
<td>70</td>
<td>67</td>
<td>71</td>
<td>3</td>
<td>2</td>
<td>0.563</td>
<td>7.46</td>
</tr>
<tr>
<td></td>
<td>All</td>
<td>16</td>
<td>91</td>
<td>81</td>
<td>78</td>
<td>83</td>
<td>0</td>
<td>1</td>
<td>0.749</td>
<td>7.54</td>
</tr>
<tr>
<td></td>
<td>66 All</td>
<td>10</td>
<td>79</td>
<td>84</td>
<td>35</td>
<td>68</td>
<td>3</td>
<td>1</td>
<td>0.512</td>
<td>4.07</td>
</tr>
<tr>
<td></td>
<td>33 All</td>
<td>6</td>
<td>63</td>
<td>43</td>
<td>54</td>
<td>53</td>
<td>8</td>
<td>4</td>
<td>0.292</td>
<td>4.03</td>
</tr>
<tr>
<td>T.v.</td>
<td>100 Epi</td>
<td>8</td>
<td>77</td>
<td>60</td>
<td>53</td>
<td>64</td>
<td>1</td>
<td>6</td>
<td>0.462</td>
<td>6.79</td>
</tr>
<tr>
<td></td>
<td>All</td>
<td>20</td>
<td>87</td>
<td>84</td>
<td>88</td>
<td>89</td>
<td>0</td>
<td>1</td>
<td>0.813</td>
<td>7.04</td>
</tr>
<tr>
<td></td>
<td>66 All</td>
<td>9</td>
<td>58</td>
<td>76</td>
<td>28</td>
<td>53</td>
<td>3</td>
<td>10</td>
<td>0.303</td>
<td>3.06</td>
</tr>
<tr>
<td></td>
<td>33 All</td>
<td>4</td>
<td>66</td>
<td>45</td>
<td>38</td>
<td>50</td>
<td>3</td>
<td>11</td>
<td>0.242</td>
<td>3.40</td>
</tr>
</tbody>
</table>

Results

Map displays of observed and predicted distributions are given in Figs 3–5 for disease prevalence, and in Fig. 6 for PCV. The associated accuracy or ‘hit ratio’ resulting from respective discriminant analyses is given in Table 2 for total prevalence, in Table 3 for *T. congolense* and *T. vivax* separately and in Table 4 for PCV. Tables 5 and 6 summarize the contribution of the different predictor variables applied to the prevalence and PCV predictions respectively.

The description of the total prevalence (TS 100%, Table 2) improved markedly when vegetation and eco-climatical data were added to the epidemiological predictor data. Best results (highest hit ratio and lowest percentage of major errors) were obtained using ground-based and remotely sensed eco-climatic predictor variables alone. With a 66% training-set eco-climatical data alone did not improve predictions obtained with all predictor variables. With a 33% training-set the best results were obtained using

chance alone. Values are given with 95% confidence limits and statistical significance was tested by evaluating the normal curve deviate \( Z \), where \( K \) is considered significantly larger than zero at a 95% probability level when \( Z > 1.96 \) and at a 99% probability level when \( Z > 2.58 \).
only epidemiological variables and the addition of further variables did not translate into better prediction. The hit-ratio remained low. As a rule, major errors, which rarely exceed 10% of the predicted prevalence, tend to underestimate disease prevalence, i.e. high prevalence grids are predicted as low prevalence.

The prevalence of *T. vivax* was better described than that of *T. congolense* (Table 3, 100% TS). In both cases results are better than for the total prevalence. Also, with a reduced trainingset (66 and 33%) the overall hit ratio was higher, especially for *T. congolense*, but the percentage of major errors was more pronounced for the individual trypanosome species than for the total prevalence.

The results for PCV (Table 4) followed a similar pattern as for total prevalence. With a reduced trainingset, neither epidemiological variables alone nor eco-climatical variables alone gave satisfactory results. With a 66% trainingset best results were obtained with the complete predictor set and the accuracy level was similar to that for prevalence. With a 33% trainingset best results were obtained combining epidemiological with vegetation predictor.
Table 6. Best predictors selected from ground based and remotely sensed eco-climatical variables

<table>
<thead>
<tr>
<th>No.</th>
<th>Variable</th>
<th>L</th>
<th>M</th>
<th>H</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Trange</td>
<td>4.1</td>
<td>4.6</td>
<td>4.9</td>
</tr>
<tr>
<td>2</td>
<td>CCD p2</td>
<td>3.1</td>
<td>3</td>
<td>3.1</td>
</tr>
<tr>
<td>3</td>
<td>Ch3 a1</td>
<td>6.8</td>
<td>7.6</td>
<td>7.8</td>
</tr>
<tr>
<td>4</td>
<td>CCD a3</td>
<td>63</td>
<td>65.3</td>
<td>69.8</td>
</tr>
<tr>
<td>5</td>
<td>CCD av</td>
<td>80.3</td>
<td>74.6</td>
<td>70.4</td>
</tr>
<tr>
<td>6</td>
<td>CCD p3</td>
<td>2.2</td>
<td>1.8</td>
<td>1.5</td>
</tr>
<tr>
<td>7</td>
<td>CCD a1</td>
<td>204</td>
<td>214</td>
<td>216</td>
</tr>
<tr>
<td>8</td>
<td>Price a2</td>
<td>1.6</td>
<td>1.7</td>
<td>2.4</td>
</tr>
<tr>
<td>9</td>
<td>Ch3 a3</td>
<td>0.67</td>
<td>0.7</td>
<td>0.69</td>
</tr>
<tr>
<td>10</td>
<td>Ch3 a1</td>
<td>1.8</td>
<td>1.8</td>
<td>1.7</td>
</tr>
<tr>
<td>11</td>
<td>Price n</td>
<td>306</td>
<td>307</td>
<td>308</td>
</tr>
<tr>
<td>12</td>
<td>NDVI a2</td>
<td>0.06</td>
<td>0.04</td>
<td>0.04</td>
</tr>
<tr>
<td>13</td>
<td>Price p2</td>
<td>2.7</td>
<td>3</td>
<td>3.2</td>
</tr>
<tr>
<td>14</td>
<td>Rain</td>
<td>1221</td>
<td>1237</td>
<td>1189</td>
</tr>
<tr>
<td>15</td>
<td>Ch3 a2</td>
<td>1.4</td>
<td>1.4</td>
<td>1.7</td>
</tr>
<tr>
<td>16</td>
<td>NDVI av</td>
<td>0.31</td>
<td>0.29</td>
<td>0.26</td>
</tr>
<tr>
<td>17</td>
<td>NDVI x</td>
<td>0.44</td>
<td>0.42</td>
<td>0.4</td>
</tr>
<tr>
<td>18</td>
<td>NDVI r</td>
<td>0.32</td>
<td>0.32</td>
<td>0.32</td>
</tr>
<tr>
<td>19</td>
<td>Ch3 av</td>
<td>310</td>
<td>312</td>
<td>313</td>
</tr>
<tr>
<td>20</td>
<td>Price r</td>
<td>13.5</td>
<td>16.2</td>
<td>18.1</td>
</tr>
</tbody>
</table>

Grey shade; variables selected in both cases. See Table 1 for the definition of the different variables. L, low; M, medium; H, high.

data. Despite the fact that the mapped outputs broadly reflect the observed trends (Fig. 6) results might be too poor to become meaningful in practical terms.

Table 5 shows that while the trends in the values of the predictor variables for prevalence and PCV were broadly identical, there were some differences, particularly for agriculture and cattle densities, and to a lesser extent for fly densities. Prevalence was highest and PCV accordingly lowest in areas where fly densities were high, taurine cattle predominant, herds belonged to several owners and owners originated from rural areas. When the best predictors were selected from the complete set, 4 out of 11 concerned epidemiological variables, in the case of prevalence, and 2 out of 20 in the case of PCV. In both cases agriculture and fly density appear in the list of selected variables.

Table 6 shows that more than half of the selected eco-climatical variables (12 out of 20) described both prevalence and PCV. Ten out of those 12, changed in parallel – high prevalences and low PCV became more pronounced in the drier parts of the country. Out of the remaining variables 4 parallel changes in prevalence and 6 parallel changes in PCV were observed. No particular type of predictor data contributes disproportionately except perhaps the ground-measured variables for PCV. The Price thermal brightness index scored less than expected (1/20) for PCV.

The results presented in Fig. 7 show that the optimal number of predictor variables increased with the size of the trainingset (r = 0.765, n = 12, P < 0.005), both in the case of prevalence and PCV (see also Tables 2–4). Predictions, obtained with eco-climatic predictors alone, for (i) the abundance of G. tachinoides (Hendrickx, 1999); (ii) the total trypanosomosis prevalence and (iii) the associated PCV level are summarized in Table 7.

**Discussion**

As may be expected from the flow chart given in Fig. 2, the accuracy of satellite-derived predictions decreases from vector abundance to trypanosomosis prevalence to PCV level. Whilst eco-climatic variables directly relate to vector habitat, the (satellite observed) environmental link is less obvious
and direct for trypanosomosis prevalence or PCV. Both prevalence and PCV are also influenced by other variables, mainly anthropogenic in nature; such as herd and pasture management (Rawlings, Wacher & Snow, 1994), breed choice (Murray et al. 1981) and direct vector or parasite control (i.e. veterinary management). PCV values are furthermore influenced by concurrent parasite infections (Kaufmann et al. 1992), nutritional factors (Agyemang et al. 1990) and the physiological status of the host (Agyemang et al. 1992).

Whilst these factors diminish the contribution of satellite data to the accuracy of predicting prevalence and PCV, this does not imply that the relationships between tsetse abundance, trypanosomosis prevalence and PCV values are weak. On the contrary, it has been shown with the Togo data set (Hendrickx et al. 1999b) that there are strong positive associations between fly challenge, trypanosomosis prevalence, PCV change and breed of cattle. Furthermore, the average herd PCV appears to be a measure of the indirect impact of tsetse and trypanosomosis on agriculture: cattle densities decrease with low PCV values as do the levels of integration cattle/crop production.

This suggests that further avenues may be explored to improve further the predictions of prevalence and PCV using satellite data, perhaps by first using the same data to demarcate different mixed crop-cattle-farming systems, as a step to quantifying the trypanosomosis constraint (FAO, 1997).

Another aspect, the vector–host contact rate, is not included in this type of area-wide, cross-sectional studies, yet is known to influence levels of disease in both humans and cattle (e.g. Laveissiere, Harvouet & Couret, 1986; Rawlings et al. 1994; Wacher et al. 1994). Vector–host contact encompasses several factors acting on the transmission rate; the parasite reservoir in the hosts, fly infection rates, fly preferences in feeding, etc. While these variables are often included in longitudinal studies at a micro-scale, it is not feasible to include these in area-wide studies of the present type. Whilst the map outputs clearly show general trends they often fail to highlight more subtle differences at a local level. This suggests that (i) additional variables might be needed and/or that (ii) one might need to combine the approach described here with studies at a local level depending on the level of accuracy needed. Such a complementary approach is currently being developed in Burkina Faso (de La Rocque, 1997) and aims at identifying discriminant factors of tsetse presence and trypanosomosis risk at a village level using SPOT (Satellite pour l’Observation de la Terre) imagery. In Togo, studies are presently undertaken to investigate how socio-economic variables locally influence disease management (Basteiens, personal communication).

The finding that the optimal number of predictor variables varies with the size (or degree of completeness) of the training set confirms an earlier established pattern for G. tachinoides and G. palpalis abundance predictions, also plotted in the same graph of Fig. 7 (grey dots). In theory the size of the training set required to define a multi-variate distribution depends upon the number of variables involved. If the absolute size of the training set available is limited, as it is here by the fixed total number of grid squares available, then the optimum number of variables is likely to change with the degree of subsampling involved. It is therefore difficult with this particular data set to draw firm conclusions about the minimum percentage coverage of a training set that will result in predictions with a definable accuracy: in a physically larger country, a smaller percentage of the total area may give results as accurate as those obtained for Togo only with a larger percentage of the total available grid squares.

In conclusion, the results obtained in the present study show the potential of remotely sensed data to assist not only in mapping vectors but also the disease they transmit. Whilst this paves the way for rapid disease mapping the approach could still benefit from further refinement concentrating upon the remotely sensed predictors (if any) of important anthropogenic influences that affect the expression of the important determinants of these and other tropical diseases.

All field activities were conducted within the framework of the Belgian-funded FAO project, GCP-TOG-013-BEL, of the Ministère Togolais de l’Agriculture, de l’Elevage et des Pêches. The authors are particularly grateful to K. Adoméa, B. Bauer, M. Denke, N. Dowmandji and J. Mortelmans for their help and support at different stages of this work. This publication arises in part from funding provided by the Department for International Development (DFID), UK (Livestock Production Programme ZC0012, to D J R). However, the DFID can accept no responsibility for information provided or views expressed.

REFERENCES


ROGERS, D. J., HAY, S. I., PACKER, M. J. & WINT, G. R. W.


