

Can Remotely Sensed Meteorological Data Significantly Contribute to Reduce Costs of Tsetse Surveys?

Guy Hendrickx/⁺, Ayitou Napala*, David Rogers**, Patrick Bastiaensen*, Jan Slingenbergh***

FAO Trypanosomosis project, GCP-RAF-347-BEL, Burkina Faso *FAO Trypanosomosis project, GCP-RAF-347-BEL, Togo **Zoology Department, Oxford, United Kingdom ***Animal Health Division, FAO, Rome, Italy

A 0.125 degree raster or grid-based Geographic Information System with data on tsetse, trypanosomosis, animal production, agriculture and land use has recently been developed in Togo. This paper addresses the problem of generating tsetse distribution and abundance maps from remotely sensed data, using a restricted amount of field data. A discriminant analysis model is tested using contemporary tsetse data and remotely sensed, low resolution data acquired from the National Oceanographic and Atmospheric Administration and Meteosat platforms. A split sample technique is adopted where a randomly selected part of the field measured data (training set) serves to predict the other part (predicted set). The obtained results are then compared with field measured data per corresponding grid-square. Depending on the size of the training set the percentage of concurring predictions varies from 80 to 95 for distribution figures and from 63 to 74 for abundance. These results confirm the potential of satellite data application and multivariate analysis for the prediction, not only of the tsetse distribution, but more importantly of their abundance. This opens up new avenues because satellite predictions and field data may be combined to strengthen or substitute one another and thus reduce costs of field surveys.

Keywords : abundance - discriminant analysis - *Glossina tachinoides* - prediction - remote sensing - Togo

The negative impact of African animal trypanosomosis on African agriculture and economy does not need to be stressed here. A series of changing macro- and micro economical factors have induced, over the last decade, major changes in the development policy of donor and developing countries. Those include a tendency to favour mixed farming systems (i.e. integrating crops and livestock) and to transfer activities from the public to the private sector. Both existing constraints and changing development policies stress the need to identify priority areas where focused efforts towards integrated disease management will have a maximum impact on the development of those mixed farming systems.

To identify those priority areas a good knowledge of the spatial epidemiology, *sensu largo*, of dominant pathologies is essential. This knowledge has to be obtained through field studies, which are often expensive in terms of cash, personnel and time.

Such a detailed multivariate database has been established in Togo (Hendrickx & Napala 1998). A next step is to explore ways to reduce costs involved. One way might be to combine results of partial surveys with (cheaper) remotely sensed data using multivariate analysis techniques in order to produce reliable predictions of vector or disease occurrence.

To achieve this goal two major types of remotely sensed data are currently at our disposal:

A first type includes SPOT (Satellite pour l'observation de la terre) and Landsat imagery with a high spatial (10 to 30 m) but low temporal resolution (26 and 16 days). These data allow very detailed spatial studies such as described by de la Rocque (1997) during this conference or Kitron et al. (1996) in Kenya and are used, on a larger scale for a variety of purposes: thematic mapping, landuse maps, soil degradation studies etc. Unfortunately this type of imagery remains expensive.

A second type includes NOAA (National Oceanic and Atmospheric Administration) and Meteosat imagery with a lower spatial (1 to 7 km)

Work presented was conducted within the framework of the Belgian funded, FAO executed, project GCP-TOG-013-BEL. This publication also arises in part from funding provided by the Department for International Development (DfID), UK (Livestock Production Programme ZC0012, to DJR). However, the FAO nor DfID can accept responsibility for information provided or views expressed.

⁺Corresponding author: Fax: +226-973022. E-mail: trypfaso@fasonet.bf

Received 12 November 1998

Accepted 19 January 1999

but high temporal resolution (respectively 12 and 0.5 hr). This type of data is more suitable for our purpose. Applied to tsetse they have mainly been used to predict vector distribution (e.g. Rogers & Randolph 1993, Robinson et al. 1997) and to describe vector abundance (Hendrickx et al. 1993, Rogers et al. 1994, 1996).

The aim of this paper is to illustrate our contribution to the development of such an operational approach, combining field surveys and NOAA and Meteosat data, to predict the distribution and abundance of *Glossina tachinoides* in Togo.

MATERIALS AND METHODS

The sampling area - Togo is a narrow (max. 135 km) West African country at the Golfe de Guinée between Ghana and Benin. It stretches from 6°10'N to 11°10'N and has a total surface of about 56 000 km². Its location makes it an ideal transect representative of a variety of eco-climatic conditions of coastal West Africa. To facilitate geo-referenced data gathering the country was divided in 311 sampling units (squares with 0.125-degree sides).

The dependant variable - Tsetse surveys, using standard bi-conical traps (Challier et al. 1977), covered in a first phase, 1990-92, the major part of the country. In a following phase, 1993-95, they were mainly concentrated on the known limits per fly species and on gaps left after the first survey. Survey sites were selected using 1:200,000 maps. On the field, mobile teams selected the most suitable tsetse habitats per survey site and sampled them with a minimum of 5 (average of 12) traps over a period of 24 or 48 hr. No odours were used. A total of 653 different survey sites were sampled in 305 of the 311 sample units. Data processing is explained and discussed in detail in Hendrickx and Napala (1998). After log₁₀ transformation, raw data were corrected for the seasonal variation occurring in the database. To address the problem of false negatives, data were pooled within defined fly-limits using a 3X3 rolling mean pattern; i.e. every sample unit is the average value of itself and its 8 surrounding squares. For the purpose of this study an abundance map of *G. tachinoides* was generated using four density classes: zero, low, medium and high. The limits between the three last classes were chosen to allow an approximately equal number of sample units per class.

The independent variables - Satellite data include three NOAA and one Meteosat derived variable. The NOAA data are: (i) the Normalised Vegetation Index (NDVI), a measure of the active vegetation biomass (Tucker et al. 1983); (ii) the Price thermal brightness index (Price 1984), a surrogate of the variation of ground temperature and (iii) the

AVHRR (Advanced Very High-Resolution Radiometer) Channel 3 value, recently shown to be one of the better predictor variables of landcover type in Nigeria (Rogers et al. 1997). In addition, the last variable used is the Cold Cloud Duration (CCD), from Meteosat. It is derived from the measured cloud-top temperatures. In West Africa cloud-top temperatures between -30 and -60°C have been related to ground measures of rainfall (Snijders 1991).

Data used are the monthly images from 1988 to 1990, except for CCD where data available covered 1989 to 1991. To reduce the size of the data base and extract seasonal information, variables were subjected to a Fourier time series analysis. The resulting phases and amplitudes of the annual, bi- and tri-annual cycles were included in the analysis, as were the calculated average values, minima, maxima and range. Some ground-measured data were added to the data bank: percentage agriculture, annual rainfall, number of dry months (rain < 30mm), maximum ground temperature and elevation.

The multivariate analysis approach - Multiple discriminant analysis is the most appropriate technique of multivariate analysis when the dependant variable is divided in more than two classes and the independent variables follow a metrical scale. First predictor variables are ranked per case to produce the best discriminating set of variables, then selected variables are used to compute discriminant functions and predictive maps of fly abundance are produced. To validate the obtained discriminant functions the total sample is divided in two sub samples. The first sub sample, called the analysis sample or trainingset, is used to rank the variables and compute the discriminant functions. The second sub sample, called the holdout sample or predicted set, is used to validate the computed functions, i.e. to produce "true predictions" based on the information contained in the analysis sample. This approach *de facto* offers the opportunity to simulate results obtained combining field surveys (analysis sample) and extrapolations (predictions made of the holdout sample), the purpose of this study. Sub samples were selected at random per identified seasonal cluster (Hendrickx & Napala 1998), assuming no prior knowledge of fly occurrence. The analysis was repeated with respectively 75%, 66%, 50% and 33% of the sample selected as analysis sample. Best results were obtained separately predicting fly presence-absence and fly abundance. Resulting maps are shown for best predictions.

RESULTS

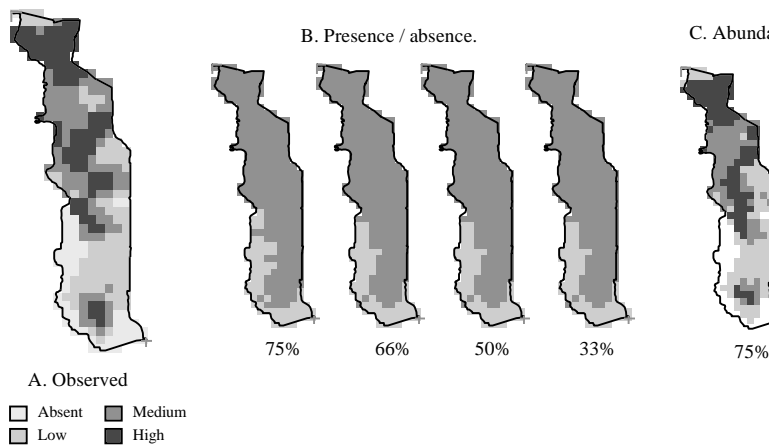
Maps given show a satisfactory overall fit, including the results obtained with a 33% analysis sample. It is remarkable to note that a single vari-

able, Price index amplitude 1 (ground temperature variation) is able to predict the distribution of *G. tachinoides* except for a rather uniform underestimation of 1 square, i.e. 13.6 km. Within fly limits temperature plays no role predicting abundance. The average values of the predictor variables per class (not given here) reflect a biological reality the fly being more abundant in a drier environment and in areas with lower agricultural pressure (Tables).

DISCUSSION

In this type of approach the acceptability of the obtained precision levels depends on what one is

prepared to pay to obtain the desired accuracy. From the point of view of selecting high, medium or low priority areas for control an overestimation of the tsetse problem might induce unnecessary costs and underestimating it might result in unidentified priority areas. In our case an average of 4% of this type of errors is considered acceptable. The relatively limited number of sample units for Togo (n = 311) did not allow us to test results below a 33% coverage of the country surveyed. It thus remains to be tested if a smaller proportion of a bigger area retains enough variability to allow satisfactory abundance predictions of the total area. Another aspect that will also have to be taken into consideration is



Best predictions of *Glossina tachinoides* in Togo using subsequently 75, 66, 50 and 33% of the total sample to predict the remaining part.

TABLE I

Best presence-absence predictions for the predicted set

Sample used as trainingset (%)	Variables yielding best prediction	Concordant absence predictions (%)	Concordant presence predictions (%)
75	Price amplitude 1 Channel 3 phase 1 Price amplitude 3 Price Range	92	91
66	Price amplitude 1	80	92
50	Price amplitude 1	87	94
33	Price amplitude 1	87	94

TABLE II

Best abundance prediction results for the predicted set

Trainingset (%)	Concordant predictions per class (%)				Major errors (%)	
	Low	Medium	High	Total	Overestimation	Underestimation
75	95	26	76	65	0	6
66	71	71	81	74	0	2
50	76	59	54	63	2	3
33	64	54	75	64	2	1

TABLE III
Best selected variables for abundance predictions

75%	66%	50%	33%
Channel 3 range	Channel 3 range	Channel 3 range	CCD phase 1
CCD amplitude 3	CCD amplitude 3	CCD amplitude 3	CCD amplitude 3
CCD amplitude 2	CCD amplitude 2	CCD phase 2	% Agriculture
Channel 3 minimum	Channel 3 minimum	Channel 3 amplitude 3	Channel 3 amplitude 3
% Agriculture	NDVI amplitude 1	CCD phase 1	Elevation
NDVI phase 1	% Agriculture	NDVI maximum	CCD amplitude 2
CCD phase 3	NDVI phase 2	CCD phase 3	CCD maximum
NDVI minimum	NDVI minimum	NDVI phase 1	NDVI phase 2
NDVI phase 2	NDVI phase 1	Price amplitude 1	
Channel 3 phase 1	CCD mean	NDVI minimum	
NDVI amplitude 1	Channel 3 phase 1	NDVI phase 1	
CCD phase 2	Channel 3 phase 2		
Price phase 2	CCD phase 1		

CCD: cold cloud duration; NDVI: normalized difference vegetation index.

that with a decreasing proportion of sample units surveyed it might be necessary to increase the sampling intensity within each sampled unit.

ACKNOWLEDGEMENTS

To K Adomefa, D Batawui, B Dao, R De Deken, M Denke, N Doumandji and P Marchot for their permanent help and support.

REFERENCES

- Challier A, Eyraud M, Lafaye A Laveissière C 1977. Amélioration du rendement du piège biconique pour Glossines (Diptera, Glossinidae) par l'emploi d'un cône inférieur bleu. *Cah ORSTOM ser Entomol med Parasitol* 15: 283-286.
- de La Rocque S 1977. *Identification des Facteurs Discriminants Majeurs de la Présence des Glossines dans une Zone Agro Pastorale du Burkina Faso. Intérêt pour la Prévision du Risque Trypanosomien*, PhD thesis, Université de Montpellier II Sciences et Techniques du Languedoc, France, 212 pp.
- Hendrickx G, Napala, A 1998. Le contrôle de la Trypanosomose 'à la carte': une approche intégrée basée sur un système d'information géographique. *Acad R Belg Sci Outre Mer, Mem Cl Sci Nat Med* (in press).
- Hendrickx G, Rogers DJ, Napala A, Slingenbergh JHW 1993. Predicting the distribution of riverine tsetse and the prevalence of bovine trypanosomiasis in Togo using ground based and satellite data, p. 218-232. In International Scientific Council for Trypanosomiasis Research and Control (ISCTRC), Twenty Second Meeting, Kampala, Uganda, 1993, publ. 1995, Organization of African Unity, Scientific and Technical Research Commission (OAU-STRC), Nairobi, Kenya.
- Kitron U, Otieno LH, Hungerford LL, Odulaja A, Brigham WU, Okello OO, Joselyn M, Mohamed-Ahmed MM, Cook E 1996. Spatial Analysis of the Distribution of tsetse flies in the Lambwe Valley, Kenya, using landsat TM satellite imagery and GIS. *J Anim Ecol* 65: 371-380.
- Price JC 1984. Land surface temperature measurement for the split window channels of the NOAA 7 advanced very high resolution radiometer. *J Geophys Res* 89: 7231-7237.
- Robinson T, Rogers DJ, Williams B 1997. Mapping tsetse habitat suitability in the common fly belt of Southern Africa using multivariate analysis of climate and remotely sensed vegetation data. *Med Vet Entomol* 11: 235-245.
- Rogers DJ, Randolph SE 1993. Distribution of tsetse and ticks in Africa, past, present and future. *Parasitol Today* 9: 266-271.
- Rogers DJ, Hendrickx G, Slingenbergh JHW 1994. Tsetse flies and their control. *Rev Sci Tec Off Int Epizoot* 13: 1075-1124.
- Rogers DJ, Hay SI, Packer MJ 1996. Predicting the distribution of tsetse flies in West Africa using temporal, Fourier processed meteorological satellite data. *Ann Trop Med Parasitol* 90: 225-241.
- Rogers DJ, Hay SI, Packer MJ, Wint GRW 1997. Mapping land cover over large areas using multispectral data derived from NOAA-AVHRR: a case study of Nigeria. *Int J Remote Sensing* (accepted).
- Snijders FL 1991. Rainfall monitoring based on meteorological data - a comparison of techniques applied to the Western Sahel. *Int J Remote Sensing* 12: 1331-1347.
- Tucker CJ, Vanpraet C, Boerwinkel E, Gaston A 1983. Satellite remote sensing of total dry matter production in the Senegalese Sahel. *Remote Sensing Environ* 13: 461-474.