Predicting the distribution of canine leishmaniasis in western Europe based on environmental variables

ANA O. FRANCO[1]†, CLIVE R. DAVIES[1]††, ADRIAN MYLNE[1], JEAN-PIERRE DEDET[2], MONTSERRAT GÁLLEGO[3], CRISTINA BALLART[3], MARINA GRAMICCIA[4], LUIGI GRADONI[4], RICARDO MOLINA[5], ROSA GÁLVEZ[5], FRANCISCO MORILLAS-MÁRQUEZ[6], SERGIO BARÓN-LÓPEZ[6], CARLOS ALVES PIRES[7], MARIA ODETE AFONSO[7], PAUL D. READY[8]* *and* JONATHAN COX[1]†

[1] *Faculty of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, London, UK*
[2] *Centre National de Référence des Leishmania, UMR MIVEGEC, Université Montpellier 1/Laboratoire de Parasitologie-Mycologie, CHU de Montpellier, Montpellier, France*
[3] *Laboratori de Parasitologia, Facultat de Farmàcia, Universitat de Barcelona, Barcelona, Spain*
[4] *Unit of Vector-borne Diseases and International Health, MIPI Department, Istituto Superiore di Sanità, Rome, Italy*
[5] *Laboratorio de Referenda de Leishmaniasis, Servicio de Parasitología, Centro Nacional de Microbiología, Instituto de Salud Carlos III, Madrid, Spain*
[6] *Departamento de Parasitología, Facultad de Farmacia, Universidad de Granada, Granada, Spain*
[7] *Unidade de Entomologia Médica/Unidade de Parasitologia e Microbiologia Médica, Instituto de Higiene e Medicina Tropical, Universidade Nova de Lisboa, Lisbon, Portugal*
[8] *Department of Entomology, Natural History Museum, London, UK*

(*Received 10 June 2011; accepted 1 August 2011; first published online 14 September 2011*)

SUMMARY

The domestic dog is the reservoir host of *Leishmania infantum*, the causative agent of zoonotic visceral leishmaniasis endemic in Mediterranean Europe. Targeted control requires predictive risk maps of canine leishmaniasis (CanL), which are now explored. We databased 2187 published and unpublished surveys of CanL in southern Europe. A total of 947 western surveys met inclusion criteria for analysis, including serological identification of infection (504, 369 dogs tested 1971–2006). Seroprevalence was 23·2% overall (median 10%). Logistic regression models within a GIS framework identified the main environmental predictors of CanL seroprevalence in Portugal, Spain, France and Italy, or in France alone. A 10-fold cross-validation approach determined model capacity to predict point-values of seroprevalence and the correct seroprevalence class (<5%, 5–20%, >20%). Both the four-country and France-only models performed reasonably well for predicting correctly the <5% and >20% seroprevalence classes (AUC >0·70). However, the France-only model performed much better for France than the four-country model. The four-country model adequately predicted regions of CanL emergence in northern Italy (<5% seroprevalence). Both models poorly predicted intermediate point seroprevalences (5–20%) within regional foci, because surveys were biased towards known rural foci and Mediterranean bioclimates. Our recommendations for standardizing surveys would permit higher-resolution risk mapping.

Key words: canine leishmaniasis, western Mediterranean, predicting distribution, serological records, seroprevalence.

INTRODUCTION

Zoonotic visceral leishmaniasis (ZVL) is endemic in Mediterranean Europe, where the causative agent *Leishmania infantum* is transmitted by phlebotomine sandflies of the subgenus *Phlebotomus* (*Larroussius*) (Ready, 2010), but its surveillance and control are neglected compared with research efforts (Dujardin *et al*. 2008). There are estimated to be only 700 new human cases of ZVL per year in southern Europe (Dujardin *et al*. 2008), and some people develop cutaneous lesions. However, the seroprevalence in the domestic dog, the only proven reservoir host (Quinnell and Courtenay, 2009), is often 20% (Dujardin *et al*. 2008), sufficiently high to pose a serious risk of re-emergence of human disease, especially in immuno-suppressed people (Ready, 2010). Human cases are not usually infectious to sandfly vectors, but the parasite has been exported from Europe in the canine reservoir hosts, historically to Latin America where it now causes much fatal infantile disease (Romero and Boelaert, 2010) and recently even to North America in fox hounds (Duprey *et al*. 2006).

The control of canine leishmaniasis (CanL) is considered to be the best way of reducing the

* Corresponding author: Department of Entomology, Natural History Museum, London SW7 5BD, UK. Tel: + 442079425622. Fax: + 442079425229. E-mail: P.Ready@nhm.ac.uk
† Except for these two, the authors are listed in alphabetical order of team leaders, each of whom is followed by any team member involved in databasing.
†† Clive Davies passed away in March 2009.

doi:10.1017/S003118201100148X

incidence of human disease in ZVL foci (Quinnell and Courtenay, 2009) and one way of controlling any spread northwards in Europe (Ready, 2010). Targeted control would be assisted by predictive risk maps of CanL, but unfortunately these do not exist, even in Europe where there have been numerous surveys and incidence has been mapped for positive point locations or administrative areas (Trotz-Williams and Trees, 2003). The current report investigates the challenges of producing a predictive risk map of CanL for western Europe, based on the authors' databasing of historical records of CanL during the EU EDEN project (Emerging Diseases in a changing European eNvironment; http://www.eden-fp6project.net). Records for Eastern Europe were included in our database, but they are not analysed in the current report because they are relatively few in number and the transmission cycles are biogeographically distinct (Ready, 2010). Only 2 sandflies, *Phlebotomus perniciosus* and *P. ariasi*, have been incriminated as vectors in Portugal, Spain and France, and the former is also the most widespread vector in Italy, where *P. perfiliewi* and *P. neglectus* pose regional threats. Further east, the vectors are *P. perfiliewi, P. neglectus* and *P. tobbi*.

Of all the CanL records we found in the literature and personal files, only those from random surveys of dog populations were targeted for our risk mapping, in order to lessen the chances of analysing populations of mixed geographical origins, including travel cases, or biasing prevalence estimates by missing asymptomatic seropositive dogs. Multivariate logistic regression models were built, using a GIS framework, to find the main predictors of CanL prevalence in western Europe or France alone based on a range of variables, including remotely-sensed climatological and vegetation indices.

MATERIALS AND METHODS

Developing a European database of CanL

The EDEN subproject on leishmaniasis developed a database of all surveys of CanL carried out in Europe and other Mediterranean countries from 1965. A standardized Microsoft® Office Excel worksheet included information related to: publication or dataset; survey type, date, location and environment; method of selecting dogs to be screened; method(s) of diagnosis of infection; dog travel history, breed, age and lifestyle; number of dogs tested and number positive.

For the present study, the data-gathering teams in Portugal (1), Spain (3), France (1) and Italy (1) performed a systematic search for published and unpublished reports in their regions, based on their expert knowledge of data sources. Data entries were standardized by the analytical team in the UK, based on clarifications provided by the data gatherers. The present analysis included only surveys that provided estimates of CanL prevalence based on serological diagnosis. Surveys were excluded from the analysis if 1 or more of the following criteria applied: CanL prevalence could not be estimated either directly or indirectly; the method of diagnosis was not serological (but was based solely on clinical signs, microscopy or molecular characterization); dogs were clearly not sampled 'randomly' from a settlement or administrative unit (excluded were positive cases reported by veterinary clinics or other cases possibly associated with passive detection and unknown combinations of localities); data were duplicated and reported more completely in another publication; location was missing or unable to be geo-positioned; and, environmental data were lacking.

Only a minority of reports included geographical coordinates. Most records mentioned the nearest settlement, which was geo-positioned using Google Earth (http://earth.google.com; accessed 2 February 2009). The survey date was potentially an explanatory variable, and so it was derived by (a) averaging the survey's start and end years, or (b) using one of those two dates if the other was missing, or (c) using the year of publication if survey dates were not reported.

Environmental data

Data for long-term climatic averages (based on the period 1961–1990) of monthly precipitation, relative humidity, mean temperature and mean diurnal temperature range were obtained at a spatial resolution of 10 arc-minutes from the Climatic Research Unit, University of East Anglia, UK (New *et al.* 2002). Monthly data were aggregated into averages for the warm months (May to September) and for the cold months (October to April). Remotely sensed (MODIS v4), Fourier-processed data (Rogers *et al.* 1996) for day-time (LSTD) and night-time (LSTN) land surface temperature, middle infrared reflectance (MIR) and Enhanced vegetation index (EVI) for the period 2001–2005 were produced by Professor David Rogers and his team (Spatial Ecology and Epidemiology Research Group, Department of Zoology, Oxford University) and obtained from the EDEN project data archive (http://ergodd.zoo.ox.ac.uk/EDEN/index.php?p=1; accessed 2 March 2009). These data have a spatial resolution of 1 km (250 m for EVI) and include separate products relating to the means, amplitudes and phases of annual, biannual and tri-annual cycles. Elevation data were derived from the GLOBE digital elevation model (spatial resolution 1 km) (US National Geophysical Data Center (NGDC), National Oceanic & Atmospheric Administration (NOAA) Satellite and Information Service. Available: http://www.ngdc.noaa.gov/; accessed 2 March 2009).

Statistical model building

CanL seroprevalence was modelled as a binomial variate: the number of dogs examined and dogs found exposed to *Leishmania* infection at each survey location. To overcome the dependency of observations at the same location, the association between risk of exposure to infection and the environmental covariates was determined using fixed-effects grouped logistic regression with robust standard errors (Rogers, 1993; Rabe-Hesketh and Everitt, 2004) in the statistical package Stata (version 10; StataCorp, College Station, Texas, USA). Binary logistic regression (*logit* command in Stata) is more frequently used for statistical modelling, but this would have required splitting CanL seroprevalence into 2 categories and, thereby, losing information (Rabe-Hesketh and Everitt, 2004). In addition, global spatial structures within the seroprevalence data and among model Pearson residuals from multivariate regression analysis were evaluated using semivariograms estimated using the R module GeoR (Ribeiro and Diggle, 2001). Semivariograms show the spatial dependence of the variable of interest as a scatter plot, and as such provide a means of assessing visually the presence of spatial autocorrelation (Pullan *et al*. 2008).

Prior to multivariate analysis, univariate binomial logistic regression analyses were carried out to test for associations between seroprevalence and each of the 46 environmental variables, as well as the potentially confounding variable of the survey country. Predictors that were significant at the 10% probability level were retained. The uncentred variance inflation factor (VIF) was used to screen for co-linearity between the retained environmental continuous variables (Rabe-Hesketh and Everitt, 2004). Within a group of themed environmental variables, those with VIF $\geqslant$10 were excluded one by one. The procedure was repeated to test for co-linearity between the variables of different groups. The remaining variables were included in a multivariate binomial logistic regression model, for which a backward stepwise selection procedure was performed, excluding the variables with the highest Wald test *P*-value until all the variables in the model had $P<0{\cdot}05$.

Some of the variables retained in the model were non-linear, and these were transformed if this improved the model fit. We found the best transformation for each predictor, starting with the one with the smallest *P*-value in the linear model, and then transforming the predictor with the next smallest *P*-value. The non-linearity of each variable (x) was explored by a quadratic polynomial transformation and by 7 fractional polynomial transformations of the first degree (FPs), namely $1/x^2$, $1/x$, $1/x^{0{\cdot}5}$, $\ln(x)$, $x^{0{\cdot}5}$, x^2, and x^3 (Royston *et al*. 1999; Hosmer and Lemshow, 2000). The best-fit transformation is the one that produces the largest reduction in the residual deviance. The quadratic polynomial transformation was selected if it decreased deviance by more than 5·99 compared with the best fit FP. Otherwise, the best fit FP was selected if it decreased deviance by more than 3·84 compared with the linear (i.e. untransformed) variable. If the decrease in deviance was less, then the variable remained linear. This process was repeated several times until no further FP transformations were indicated for the multivariate model without clustering, and then the same transformation was applied when fitting the final model with clustering (Hosmer and Lemshow, 2000).

The same method was used to build 2 additional models. For the four-country dataset, a separate model included diagnostic method and survey year, to assess whether they improved the accuracy of the model predictions. A separate model was also built for the data from France alone, in order to investigate better any effect of country on accuracy.

Model validation

The predictive performance of each model was assessed by using a 10-fold cross-validation approach (Hastie *et al*. 2001; Tibshirani *et al*. 2002). Briefly, the CanL seroprevalence dataset was split randomly into 10 approximately equal-sized parts. The final multivariate logistic model was fitted on 90% of the data points, and then used to predict the seroprevalence of the remaining 10% (the validation set). This procedure was performed 10 times, each time with 1 of the 10 dataset parts acting as the validation set. The seroprevalences predicted by all the validation sets were then compared with the observed values for the same locations. For each of the validation sets the predictive performance of the model was tested by determining the capacity of the model to predict either point-values of seroprevalence – using the correlation coefficient as a measure of linear association between the predicted and the observed values (Gething *et al*. 2008; Hay *et al*. 2009) – or the correct endemicity class. Three endemicity classes were considered: low (<5% seroprevalence), medium (5–20%), and high risk (>20%). The capacity of the model to predict the correct endemicity class was assessed using the area-under-curve (AUC) of a receiver-operating-characteristic (ROC) curve, which plots sensitivity versus one minus specificity for each endemicity class (Brooker *et al*. 2002; Hay *et al*. 2009). AUC values indicate the agreement between the observed and predicted endemicity class: poor, 0·51–0·70; reasonable, 0·71–0·90; and excellent, >0·9 (Brooker *et al*. 2002; Hay *et al*. 2009). The prediction accuracy was further assessed by a contingency table and by calculating the percentage of points classified in the correct endemicity class or

in a non-adjacent class (Hay *et al*. 2009). The validation statistics for all 10 dataset parts were averaged to compute the overall accuracy of the predictions.

Producing a risk map of CanL seroprevalence

Each statistical model was then used to estimate CanL seroprevalence for a country, based on the predictor values available from remotely sensed images for a 1 km-square grid. Finally, the predicted CanL risk was converted into 1 of 5 endemicity classes: low (<5% seroprevalence), medium-low (5–10%), medium-high (10·01–20%), high (20·01–30%) and very high risk (>30%), and the predicted seroprevalence was mapped using ArcGIS v. 9.2 (ESRI, Redlands, CA, USA). Seroprevalence was mapped only for the pixels where the environmental values were within the range of those in the survey locations.

RESULTS

Database of CanL in Europe

The historical CanL database included cross-sectional surveys, prospective surveys, laboratory records, cases reported at veterinary clinics and case reports. It contained 2187 surveys, including 33 from Croatia, 26 from Greece and 55 from Turkey. There was a total of 2073 reports from Portugal, Spain, France and Italy, of which 947 (45·7%) were included and 1126 were excluded from the current analysis based on the criteria reported above. The most frequent exclusion reasons were non-random sampling ($n=849$; 75·4%), missing data for the number of dogs tested or positive ($n=595$; 52·8%) and non-serological diagnosis ($n=258$; 22·9%). Of the latter, 117 were solely based on microscopy, 58 on culture, 47 on molecular characterization, and 36 on clinical signs. The datasets with the included and excluded CanL seroprevalence surveys can be obtained from the corresponding author.

The remaining 947 surveys were analysed, and the reports included 124 publications. The surveys were undertaken between 1971 and 2006 inclusive (75% after 1985, and 50% after 1992) and involved 504 369 dogs tested for exposure to *Leishmania* infections. The data analysed are summarized in Table 1 and mapped in Fig. 1. Of the analysed surveys, most were conducted in Italy (377; 40%), followed by Spain (213; 22%), Portugal (188; 20%) and France (169; 18%). The median number of dogs tested per survey was 67, being highest in Italy (168) and lowest in Spain (39). The frequency distribution of the number of dogs per survey was highly right skewed, with the maxima being 30 001 for Italy, 7067 for France, 1803 for Spain and 1024 for Portugal. Surveys were conducted between latitudes 35·5° and 47·5° North and longitudes 9·3° West and 17·6° East. The analysed records covered altitudes of 1–1838 m above sea level (median 204 m a.s.l.). Most surveys (99%) were conducted below 1000 m a.s.l. (3 at 1000–1500 m a.s.l. in France; and 6 at 1000–1500 m a.s.l. and 1 at 1838 m a.s.l. in Spain). The diagnostic method most commonly used was the indirect fluorescent antibody test (IFAT), both overall (77%) and in each country, and the most frequent IFAT cut-off was equal to or above the current minimum standard of 1:80. The proportion of surveys using an IFAT cut-off <1:80 was 35·4% in Italy, 19·2% in Portugal, 2·4% in Spain and 0% in France. The other serological diagnostic methods most frequently used were ELISA, DOT-ELISA and the direct antibody test (DAT).

The overall CanL seroprevalence was 23·2% (116,968/504,369), and 5% of the surveys recorded seroprevalences >40%. Point seroprevalences >80% were recorded in Italy, Portugal and Spain, but the maximum in France was only 43%. The median seroprevalence was 10%, being highest in Italy (17·7%), followed by France (8%), Portugal (7·3%) and Spain (5·9%). Zero seroprevalences were recorded in 14·7% of the surveys, being more frequent in Spain (25%), followed by Portugal (20%), France (14%), and Italy (6%).

Statistical models

Four-country model. Among the 49 variables screened by univariate analysis (Table 2), 32 were significantly associated with CanL seroprevalence at the 10% probability level: diagnostic method, survey year, country, altitude, latitude, 21 Fourier-transformed remotely sensed (RS) variables from 4 environmental groups, and 6 variables from 4 climatic groups. Of the latter 27 variables, 19 were excluded due to co-linearity, leaving 11 variables to be tested for inclusion in the multivariate model. These consisted of country, altitude, latitude and 8 Fourier-transformed RS variables relating to night-time land surface temperature (minimum (LSTNmn), amplitude of the tri-annual cycle (LSTNa3)), day-time land surface temperature (amplitude of the bi-annual cycle (LSTDa2), phase of the tri-annual cycle (LSTDp3)), and enhanced vegetation index (minimum (EVImn), amplitude of the annual cycle (EVIa1), and phases of the annual and tri-annual cycles (EVIp1, EVIp3)). Of these, 7 variables remained in the final multivariate model (Table 3). The association with seroprevalence was inversely U-shaped for some predictors (i.e. the highest seroprevalence was observed at intermediate values), namely altitude, LSTNmn, LSTNa3 and to a much lesser degree EVIp1, and so all were used with a quadratic transformation. When testing for the confounding effect of diagnostic method and survey year, only the latter dropped from the multivariate

Table 1. Seroprevalences of canine leishmaniasis and number of dogs tested for all surveys analysed

	All countries (n = 947)			France (n = 169)			Italy (n = 377)			Portugal (n = 188)			Spain (n = 213)		
No. surveys	Prevalence		%	Prevalence		%	Prevalence		%	Prevalence		%	Prevalence		%
Survey category	Median	(Range)		Median	(Range)		Median	(Range)		Median	(Range)		Median	(Range)	
Total	10·0	(0–100)	100[a]	8·0	(0–43·3)	17·8[a]	17·7	(0–100)	39·8[a]	7·3	(0–81·1)	19·9[a]	5·9	(0–100)	22·5[a]
Survey year															
1971–1980	6·3	(0–100)	13·8	4·4	(0–40·0)	42·6	11·1	(0–100)	12·5	0·0	(0–3·6)	6·4	—	(—)	—
1981–1990	10·2	(0–66·7)	28·9	13·2	(0·7–43·3)	14·8	19·1	(2·5–66·7)	9·3	10·4	(0–39·8)	57·4	6·9	(0–45·5)	49·8
1991–2000	11·7	(0–100)	45·0	9·1	(0–27·6)	34·9	21·1	(0–100)	53·6	6·3	(0–81·1)	33·0	5·2	(0–100)	48·4
2001–2006	11·1	(0–100)	12·2	11·1	(4·1–17·7)	7·7	10·7	(0–100)	24·7	17·6	(0–21·4)	3·2	16·9	(3·8–66·1)	1·9
Altitude															
1–<100	12·5	(0–100)	32·4	7·3	(0–43·3)	28·4	16·8	(0–100)	44·8	7·1	(0–34·1)	35·6	5·9	(0–34·6)	10·8
100–<500	11·5	(0–100)	45·1	9·4	(0–40·0)	45·6	20·0	(0–100)	46·4	10·1	(0–80·0)	43·6	9·1	(0–66·1)	43·7
500–<1000	5·6	(0–100)	20·3	7·3	(0–23·7)	21·3	20·0	(0·4–80·0)	7·7	4·5	(0–81·1)	19·7	4·3	(0–100)	42·3
1000–<1500	5·0	(0–11·1)	1·0	0·0	(0–5·0)	1·8	—	(—)	—	—	(—)	—	6·4	(0–11·1)	1·0
1500–1838	12·3	(12·3–12·3)	0·1	—	(—)	—	—	(—)	—	—	(—)	—	12·3	(12·3–12·3)	0·1
Diagnostic															
IFAT	11·3	(0–100)	76·6	4·8	(0–43·3)	64·5	18·4	(0–100)	96·6	8·3	(0–39·8)	66·5	4·4	(0–100)	59·6
ELISA	11·1	(0–100)	8·4	11·1	(0–27·6)	34·9	24·4	(7·1–100)	1·1	6·5	(6·5–6·5)	0·5	13·0	(0–34·1)	7·5
Dot ELISA	7·7	(0–36·4)	7·3	—	(—)	—	—	(—)	—	—	(—)	—	7·7	(0–36·4)	32·4
DAT	6·6	(0–81·1)	5·8	25·7	(25·7–25·7)	0·6	—	(—)	—	6·3	(0–81·1)	28·7	—	(—)	0·0
Other[b]	4·0	(0–66·1)	1·9	—	(—)	—	1·6	(0–37·1)	2·4	6·1	(0–15·8)	4·3	66·1	(66·1–66·1)	0·5
No. dogs tested	All countries (n = 504 369)			France (n = 39 259)			Italy (n = 423 831)			Portugal (n = 15 896)			Spain (n = 25 383)		
	Median	(Interquartile Range)	%[c]	Median	(Interquartile Range)	%[c]	Median	(Interquartile Range)	%[c]	Median	(Interquartile Range)	%[c]	Median	(Interquartile Range)	%[c]
No. dogs tested /survey	67	(31–209)	100	60	(32–112)	7·8	168	(43–612)	84·0	52	(34–104)	3·2	39	(20–93)	5·0

[a] Percentage of surveys out of the total of surveys in all countries. Percentages for categories total 100%, except for altitude because this was not found for 11 surveys (5 in France, 4 in Italy and 2 in Portugal).
[b] Including CIE, IFAT and CIE, LST. [IFAT: Indirect Fluorescence Antibody Test. ELISA: Enzyme-Linked Immunosorbent Assay. DAT: Direct Agglutination Test. CIE: Counter-immunoelectrophoresis. LST: Leishmanin Skin Test].
[c] Percentage of dogs tested calculated out of the total of dogs tested in all countries.

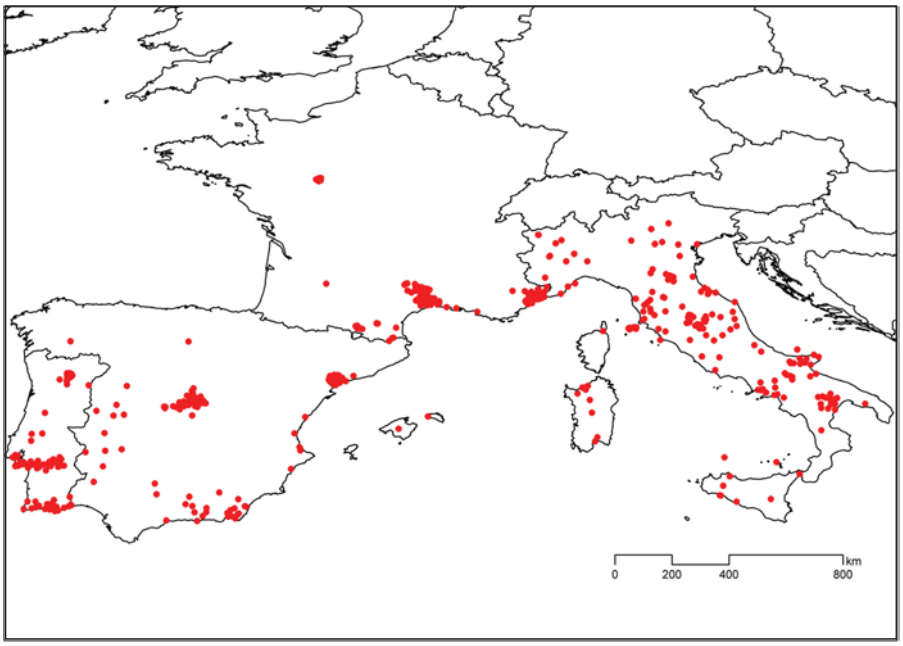

Fig. 1. Distribution of all 947 serological surveys of canine leishmaniasis included in the analyses and modelling of its seroprevalence.

model. However, diagnostic method was not retained, because it did not substantially improve the accuracy of the model predictions. Semivariograms were estimated on the basis of both observed prevalence and Pearson residuals from multivariate models, to investigate the removal of spatial trends potentially related to environmental variables. Both analyses produced unbounded semivariograms (not shown) in which semivariance increased steadily with increasing distance (spatial lags) between observations. In the absence of any clear spatial structure in the prevalence data, modelling was restricted to a non-spatial regression analysis. The final model equation was: CanL seroprevalence = (exp(p)/(1 + exp(p)) × 100; where p = − 2·67413 − 0·5203425 × country_2 − 1·486901 × country_3 − 1·627514 × country_4 + 0·0015933 × altitude − 8·76 × 10^{-7} × $(\text{altitude})^2$ + 0·1632941 × LSTNmn − 0·0167179 × $(\text{LSTNmn})^2$ + 4·346963 × LSTNa3 − 2·562916 × $(\text{LSTNa3})^2$ + 0·0874507 × $1/\sqrt{\text{EVIa1}}$ + 0·136483 × EVIp1 − 0·0361457 × $(\text{EVIp1})^2$ − 0·1942139 × (1/EVIp3); and where country_2, country_3 and country_4 were dummy variables for Italy, Portugal and Spain, respectively.

France-only model. Following the same method, the only variables remaining in the final multivariate model were altitude, LSTNmn, amplitude of the bi-annual cycle of the night-time land surface temperature (LSTNa2), LSTNa3, amplitude of the annual cycle of middle infrared reflectance (MIRa1) and amplitude of the tri-annual cycle of middle infrared reflectance (MIRa3) (Table 4). The final model equation was: CanL seroprevalence = (exp(p)/(1 + exp(p)) × 100; where p = − 5·849738 + 0·0526301 × $\sqrt{\text{altitude}}$ + 32·16849 × MIRa1 − 129·4284 × MIRa3 + 1·429675 × $1/\sqrt{\text{LSTNa2}}$ + 1·130128 × LSTNa3 + 0·2185278 × LSTNmn.

Validation of four-country model

This was performed with 931 surveys, because at least 1 of the environmental variables in the model was missing for 16 records (5 in France, 8 in Italy and 3 in Portugal).

Predicting point-values of CanL seroprevalence: The global correlation coefficient (r) between the predicted and observed seroprevalences was 0·341, and the weakness of this linear agreement was clear in a scatter plot. Predicted seroprevalences were concentrated at ⩽40%, while observed seroprevalences ranged from 0 to 100%. The overall fit did not improve significantly (r = 0·352) if only the observed seroprevalences ⩽40% (95% of records) were considered. The country level correlation was considerably stronger for Italy (r = 0·432), which had about twice as many surveys as the other 3 countries (r = 0·233–0·262).

Predicting endemicity classes of CanL seroprevalence: When comparing the model predictions with the observed seroprevalences in the 931 surveys, 41% of the records were correctly classified to 1 of the 3 endemicity classes (Italy: 48%; Spain: 40%; Portugal: 37%; France: 32%). Only 5·5% were misclassified to a non-adjacent class (Italy: 3·8%; Spain: 4·2%; Portugal: 2·2%; France: 14·6%). Overall AUC statistics for the two extreme seroprevalence classes (<5%, >20%) exceeded the 0·7 threshold for fair to

Table 2. Univariate binomial logistic regression analyses to test associations between seroprevalence of canine leishmaniasis in the four countries and each potential explanatory variable

Variable		No./947	Range	Odds Ratio	(95% confidence intervals)	*P*-value
Country	France	169		1·00		<0·00001*
	Italy	377		1·02	(0·689–1·518)	
	Portugal	188		0·38	(0·247–0·571)	
	Spain	213		0·34	(0·221–0·519)	
Diagnostic	Dot ELISA	69		1·00		<0·00001*
	IFAT	725		3·91	(2·331–6·569)	
	ELISA	80		1·56	(0·910–2·666)	
	DAT	55		1·12	(0·620–2·014)	
	Other (Table 1)	18		0·21	(0·118–0·392)	
Surv. date		947	(1971–2006)	1·02	(1·001–1·047)	0·044*
Latitude		947	(35·5011–47·5267)	0·90	(0·840–0·966)	0·004*
Altitude		936	(1–1838)	1·00	(0·999–1·000)	0·074*
MIR[a]	MIRa0	947	(0–0·243)	2·34	(0·105–52·276)	0·591
(Middle	MIRa1	947	(0–0·104)	20·76	(0·006–73975·1)	0·467
Infrared	MIRa2	947	(0–0·034)	$8{\cdot}36\times10^{-11}$	($3{\cdot}12\times10^{-26}$–$2{\cdot}24\times10^{5}$)	0·200
Reflectance)	MIRa3	945	(0–0·017)	$8{\cdot}87\times10^{8}$	($1{\cdot}11\times10^{-8}$–$7{\cdot}07\times10^{25}$)	0·299
	MIRmn	947	(0–0·187)	0·97	(0·011–86·242)	0·990
	MIRmx	947	(0–0·317)	2·84	(0·277–29·118)	0·380
	MIRp1	947	(0–7·620)	1·32	(1·097–1·580)	0·003*
	MIRp2	947	(0–5·710)	1·01	(0·959–1·069)	0·658
	MIRp3	947	(0–3·960)	1·02	(0·868–1·201)	0·805
LSTD[a]	LSTDa0	938	(11·96–29·58)	1·04	(0·985–1·102)	0·153
(Day-time	LSTDa1	941	(5·48–17·36)	0·88	(0·822–0·935)	0·0001*
land	LSTDa2	941	(0·20–3·08)	0·73	(0·565–0·950)	0·019*
surface	LSTDa3	941	(0·06–1·78)	1·09	(0·652–1·820)	0·744
temperature)	LSTDmn	940	(−15·22–17·18)	1·15	(1·096–1·199)	<0·00001*
	LSTDmx	938	(22·56–46·08)	0·98	(0·946–1·011)	0·187
	LSTDp1	940	(6·07–7·00)	8·15	(4·070–16·321)	<0·00001*
	LSTDp2	941	(0·61–3·10)	0·62	(0·442–0·856)	0·004*
	LSTDp3	941	(0·03–3·98)	1·20	(1·017–1·428)	0·0313*
LSTN[a]	LSTNa0	938	(3·98–15·56)	1·10	(1·045–1·169)	0·0005*
(Night-time	LSTNa1	941	(5·10–11·80)	0·83	(0·742–0·919)	0·0004*
land	LSTNa2	941	(0·10–2·24)	1·07	(0·830–1·381)	0·599
surface	LSTNa3	941	(0·04–1·64)	1·86	(1·166–2·967)	0·009*
temperature)	LSTNmn	938	(−7·40–9·90)	1·11	(1·071–1·156)	<0·00001*
	LSTNmx	938	(13·76–26·78)	1·07	(1·010–1·129)	0·021*
	LSTNp1	940	(6·41–7·26)	28·73	(11·335–72·810)	<0·00001*
	LSTNp2	941	(0·03–5·86)	0·51	(0·333–0·790)	0·003*
	LSTNp3	941	(0·14–2·91)	0·98	(0·745–1·292)	0·891
EVI[a]	EVIa0	945	(−0·035–0·430)	0·18	(0·056–0·610)	0·006*
(Enhanced	EVIa1	945	(0·002–0·247)	0·01	($6{\cdot}86\times10^{-5}$–0·792)	0·040*
vegetation	EVIa2	922	(0·002–0·085)	0·01	($2{\cdot}32\times10^{-7}$–$6{\cdot}93\times10^{-2}$)	0·433
Index)	EVIa3	944	(0·001–0·054)	$5{\cdot}27\times10^{-6}$	($5{\cdot}00\times10^{-16}$–$5{\cdot}55\times10^{4}$)	0·302
	EVImn	945	(−0·093–0·371)	0·24	(0·044–1·286)	0·095*
	EVImx	945	(0·005–0·653)	0·34	(0·129–0·911)	0·032*
	EVIp1	945	(0·530–10·530)	0·90	(0·827–0·974)	0·010*
	EVIp2	945	(0·040–5·920)	0·83	(0·747–0·913)	0·0002*
	EVIp3	945	(0·040–3·990)	1·21	(1·078–1·349)	0·001*
Climate[b]	DTR summer	925	(5·98–16·20)	0·85	(0·817–0·893)	<0·00001*
	DTR winter	925	(4·79–11·50)	0·84	(0·789–0·897)	<0·00001*
	TMP summer	925	(11·36–23·50)	1·18	(1·070–1·295)	0·001*
	TMP winter	925	(0·63–14·77)	1·17	(1·108–1·244)	<0·00001*
	PRE summer	925	(9·12–107·98)	0·98	(0·971–0·988)	<0·00001*
	PRE winter	925	(28·30–146·40)	1·00	(0·989–1·001)	0·126
	REH summer	925	(38·30–62·42)	1·05	(1·031–1·078)	<0·00001*
	REH winter	925	(61·90–86·20)	1·01	(0·951–1·070)	0·767

[a] The last 2 characters denote the output from Fourier processing:– a0: mean, mn: minimum, mx: maximum, a1: amplitude of annual cycle, a2: amplitude of bi-annual cycle, a3: amplitude of tri-annual cycle, p1: phase of annual cycle, p2: phase of bi-annual cycle, p3: phase of tri-annual cycle; [b] DTR: diurnal temperature range, TMP: mean temperature, PRE: rainfall, REH: relative humidity, summer: May-September, winter: October-April; * P <0·1 in the univariate analysis.

Table 3. Multivariate binomial logistic regression analysis of the associations between seroprevalence of canine leishmaniasis and each explanatory variable (or risk factor) for the four-country model

Explanatory variable	Category	Ref. point	No. dogs tested	No. surveys	Mean prevalence	(Standard deviation)	OR[b]	(95% Cl)[c]
Country	France		39 259	169	10·3	(9·9)	1·00	
	Italy		423 831	377	20·6	(17·8)	0·59	(0·40–0·87)
	Portugal		15 896	188	12·0	(13·6)	0·23	(0·14–0·37)
	Spain		25 383	213	9·5	(12·0)	0·20	(0·13–0·31)
Altitude	1–99	50	390 292	307	15·7	(15·3)	1·00	
	100–499	300	74 578	427	15·8	(15·5)	1·38	(1·18–1·61)
	500–999	750	32 297	192	10·6	(15·1)	1·87	(1·27–2·74)
	1000–1838	1400	1906	10	5·3	(4·9)	1·55	(0·74–3·25)
LSTNmn[a]	−7·40–0·34	−4·00	36 642	191	11·7	(16·0)	1·00	
(Night-time	−0·33–1·10	0·40	148 263	183	12·0	(14·6)	2·67	(1·89–3·78)
land surface	1·11–2·30	1·70	78 823	190	18·3	(17·2)	3·16	(2·05–4·88)
temperature)	2·31–4·40	3·50	142 838	188	14·5	(13·8)	3·62	(2·14–6·15)
	4·41–9·90	7·00	92 977	186	15·4	(13·7)	3·47	(1·89–6·38)
LSTNa3[a]	0·04–0·30	0·15	16 231	211	10·1	(10·0)	1·00	
(Night-time	0·31–0·37	0·35	20 283	156	11·3	(13·5)	1·85	(1·27–2·68)
land surface	0·38–0·50	0·45	43 745	206	12·0	(14·0)	2·32	(1·40–3·87)
temperature)	0·51–1·00	0·75	354 105	338	20·1	(17·9)	3·40	(1·78–6·49)
	1·01–1·64	1·15	65 376	30	16·1	(15·5)	2·76	(2·52–3·02)
EVIa1[a]	0·002–0·025	0·015	330 482	234	15·8	(15·0)	1·00	
(Enhanced	0·026–0·050	0·030	104 380	322	14·9	(15·7)	0·81	(0·71–0·98)
vegetation	0·051–0·075	0·045	22 709	179	15·2	(16·8)	0·74	(0·61–0·95)
Index)	0·076–0·100	0·065	21 528	98	14·0	(14·8)	0·69	(0·54–0·93)
	0·101–0·247	0·165	25 097	112	10·0	(12·5)	0·61	(0·44–0·89)
EVIp1[a]	0·53–3·00	1·5	18 641	180	14·4	(17·3)	1·00	
(Enhanced	3·01–4·50	3·5	28 261	91	15·4	(16·0)	0·92	(0·61–1·38)
vegetation	4·51–6·00	5·5	186 709	280	15·6	(14·7)	0·63	(0·36–1·10)
Index)	6·01–7·00	6·5	268 213	362	13·9	(15·0)	0·47	(0·28–0·79)
	7·01–10·53	8·5	2372	32	10·5	(10·8)	0·21	(0·16–0·27)
EVIp3[a]	0·40–0·70	0·50	89 441	191	14·2	(16·5)	1·00	
(Enhanced	0·71–1·00	0·80	83 086	192	15·9	(14·1)	1·16	(1·11–1·21)
vegetation	1·01–1·40	1·20	123 799	173	12·4	(13·9)	1·25	(1·18–1·34)
Index)	1·41–3·00	2·20	159 632	205	15·9	(15·4)	1·35	(1·24–1·47)
	3·01–3·99	3·50	48 238	184	13·9	(16·5)	1·40	(1·27–1·53)

[a] The last 2 characters denote the output from Fourier processing:— mn: minimum, a1: amplitude of annual cycle, a3: amplitude of tri-annual cycle, p1: phase of annual cycle, p3: phase of tri-annual cycle.

[b] Odds Ratio (OR) was calculated for the reference point of the categories of each variable X_i from the final model: Log Odds $= -2{\cdot}67413 + \beta_i X_i$. Where $\beta_i X_i$ is: for country, $-0{\cdot}5203425 \times$ country_2 $- 1{\cdot}486901 \times$ country_3 $- 1{\cdot}627514 \times$ country_4 (where the baseline country is France, and country_2, country_3 and country_4 are dummy variables for country = Italy, Portugal or Spain, respectively); for altitude: $0{\cdot}0015933 \times \text{altitude} - 8{\cdot}76 \times 10^{-7} \times (\text{altitude})^2$; for LSTNmn: $0{\cdot}1632941 \times \text{LSTNmn} - 0{\cdot}0167179 \times (\text{LSTNmn})^2$; for LSTNa3: $4{\cdot}346963 \times \text{LSTNa3} - 2{\cdot}562916 \times (\text{LSTNa3})^2$; for EVIa1: $0{\cdot}0874507 \times 1/\sqrt{\text{EVIa1}}$; for EVIp1: $0{\cdot}136483 \times \text{EVIp1} - 0{\cdot}0361457 \times (\text{EVIp1})^2$; and for EVIp3: $-0{\cdot}1942139 \times (1/\text{EVIp3})$.

[c] The 95% confidence interval (CI) was calculated according to Pullan *et al*. (2008).

good discrimination (AUC = 0·71 and 0·73, respectively), but the AUC for the intermediate class (5–20%) was below the threshold (AUC = 0·55). When looking at the results by country, similar results applied to France and Italy, where the model performed best, while in Portugal and Spain the AUC for all CanL seroprevalence classes was below 0·6.

Validation of France-only model

This was performed with 164 of the 169 surveys carried out from 1971 to 2006, because at least 1 of the environmental variables in the model was missing for 5 records.

Predicting point-values of CanL seroprevalence: The global correlation coefficient (r) between the predicted and observed seroprevalences was 0·648. The range of predicted seroprevalences was 2–38%, which is comparable to the observed range of 0–43%.

Predicting endemicity classes of CanL seroprevalence: When comparing the model predictions with the observed seroprevalences in the 164 surveys, 54% of the records were correctly classified to 1 of the 3 endemicity classes, substantially more than those correctly classified for France (32%) by the four-country model. Only 3% were misclassified to a non-adjacent class, substantially fewer than those misclassified for France (14·6%) by the four-country

Table 4. Multivariate binomial logistic regression analysis of the associations between seroprevalence of canine leishmaniasis and each explanatory variable (or risk factor) for the France-only model

Explanatory variable	Category	Ref. point	No. dogs tested	No. surveys	Mean prevalence	(Standard deviation)	OR[b]	(95% CI) [c]
Altitude	1–99	50	30444	48	13·2	(14·0)	1·00	
	100–249	175	3247	41	9·8	(7·8)	1·38	(1·15–1·67)
	250–499	375	2751	36	11·1	(7·3)	1·91	(1·31–2·78)
	500–1838	750	2121	39	6·8	(6·5)	2·91	(1·57–5·42)
LSTNmn[a]	−5·16–0·34	−3·00	3089	48	5·4	(4·9)	1·00	
(Night-time	−0·33–1·10	0·40	2660	42	7·4	(7·9)	2·10	(1·34–3·31)
land surface	1·11–2·30	1·70	3545	38	11·1	(7·2)	2·79	(1·49–5·22)
temperature)	2·31–4·66	3·50	29965	41	18·4	(12·7)	4·14	(1·74–9·83)
LSTNa2[a]	0·180–0·50	0·35	27794	32	17·4	(14·4)	1·00	
(Night-time	0·510–0·66	0·60	2455	28	10·2	(7·4)	0·57	(0·42–0·76)
land surface	0·661–0·80	0·75	3195	40	9·7	(8·8)	0·46	(0·31–0·69)
temperature)	0·801–1·00	0·90	2566	38	6·2	(7·1)	0·40	(0·25–0·65)
	1·010–2·24	1·50	3249	31	9·2	(6·5)	0·29	(0·15–0·55)
LSTNa3[a]	0·040–0·30	0·15	3343	38	8·8	(8·9)	1·00	
(Night-time	0·301–0·37	0·35	3342	39	10·4	(7·3)	1·25	(1·02–1·55)
land surface	0·371–0·50	0·45	3221	39	8·5	(6·3)	1·40	(1·02–1·92)
temperature)	0·501–0·65	0·55	28276	30	18·3	(15·2)	1·57	(1·03–2·39)
	0·651–1·54	1·15	1077	23	5·6	(5·3)	3·10	(1·08–8·87)
MIRa1[a]	0·0060–0·016	0·010	1980	35	7·7	(6·8)	1·00	
(Middle	0·0161–0·019	0·017	2932	31	9·3	(6·9)	1·25	(1·08–1·46)
Infrared	0·0191–0·025	0·023	2807	35	10·0	(6·7)	1·52	(1·15–2·01)
reflectance)	0·0251–0·032	0·030	3049	34	9·8	(9·0)	1·90	(1·23–2·93)
	0·0321–0·040	0·036	28491	34	14·9	(15·7)	2·31	(1·31–4·05)
MIRa3[a]	0·0005–0·0014	0·001	28139	29	20·3	(13·7)	1·00	
(Middle	0·0015–0·0026	0·002	3461	32	12·6	(8·0)	0·87	(0·78–1·00)
Infrared	0·0027–0·0042	0·003	2800	37	8·0	(5·7)	0·77	(0·60–0·99)
reflectance)	0·0043–0·0056	0·005	2914	35	6·8	(7·9)	0·59	(0·36–0·98)
	0·0057–0·0106	0·008	1945	36	6·1	(6·5)	0·40	(0·17–0·97)

[a] The last 2 characters denote the output from Fourier processing: – mn: minimum, a1: amplitude of annual cycle, a2: amplitude of bi-annual cycle, a3: amplitude of tri-annual cycle.

[b] Odds Ratio (OR) was calculated for the reference point of the categories of each variable Xi from the final model: Log Odds = −5·849738 + βiXi. Where βiX is, for altitude: 0·0526301 × $\sqrt{}$altitude; for LSTNmn: 0·2185278 × LSTNmn; for LSTNa3: 1·130128 × LSTNa3; for LSTNa2: 1·429675 × 1/$\sqrt{}$LSTNa2; for MIRa1: 32·16849 × MIRa1; and for MIRa3: −129·4284 × MIRa3.

[c] The 95% confidence interval (CI) was calculated according to Pullan *et al*. (2008).

model. Overall AUC statistics for the two extreme seroprevalence classes (< 5%, >20%) exceeded the 0·7 threshold for fair to good discrimination (AUC = 0·77 and 0·81, respectively), but the AUC for the intermediate class (5–20%) was below the threshold (AUC = 0·64).

Risk maps

The predicted CanL seroprevalence was mapped for the four-country model (Fig. 2) and for the France-only model (Fig. 3) using 5 endemicity classes. Pixels were unclassified if their environmental values were outside the range of those in the survey locations. When estimating seroprevalence for the risk maps, a minimum increment was added to each value of EVIp3, EVIa1 and LSTNa2, in order to eliminate any zero values. In the four-country model, EVIp3 and EVIa1 were replaced by EVIp3 + 0·01 and EVIa1 + 0·001, respectively; and, in the France-only model, LSTNa2 was replaced by LSTNa2 + 0·02.

Unclassified pixels were relatively few in number for the four-country model, but considerably higher for the France-only model. For the four-country map, a total of 4034 (6·31%) pixels were unclassified (Italy: 1969 (14·88%); Spain: 661 (3·12%); Portugal: 55 (1·47%)), with 1711 of these pixels being outside the altitude range of the survey locations in the four countries (254 <1 m a.s.l.; 1457 >1838 m a.s.l.), and 50% of the unclassified pixels were located <1492 m a.s.l. For France mapped using the four-country model, a total of 1349 (5·24%) pixels were unclassified, with 575 of these pixels being outside the altitude range of the survey locations in the four countries (81 < 1 m a.s.l.; 494 >1838 m a.s.l.), and 50% of the unclassified pixels were located <1457 m a.s.l. In contrast, for France investigated by the France-only model, a total of 7998 (31·05%) pixels were unclassified, with 1333 of these pixels being outside the altitude range of the survey locations in France (81 < 1 m a.s.l.; 1252 > 1193 m a.s.l.), and 50% of the unclassified pixels were located <257 m a.s.l.

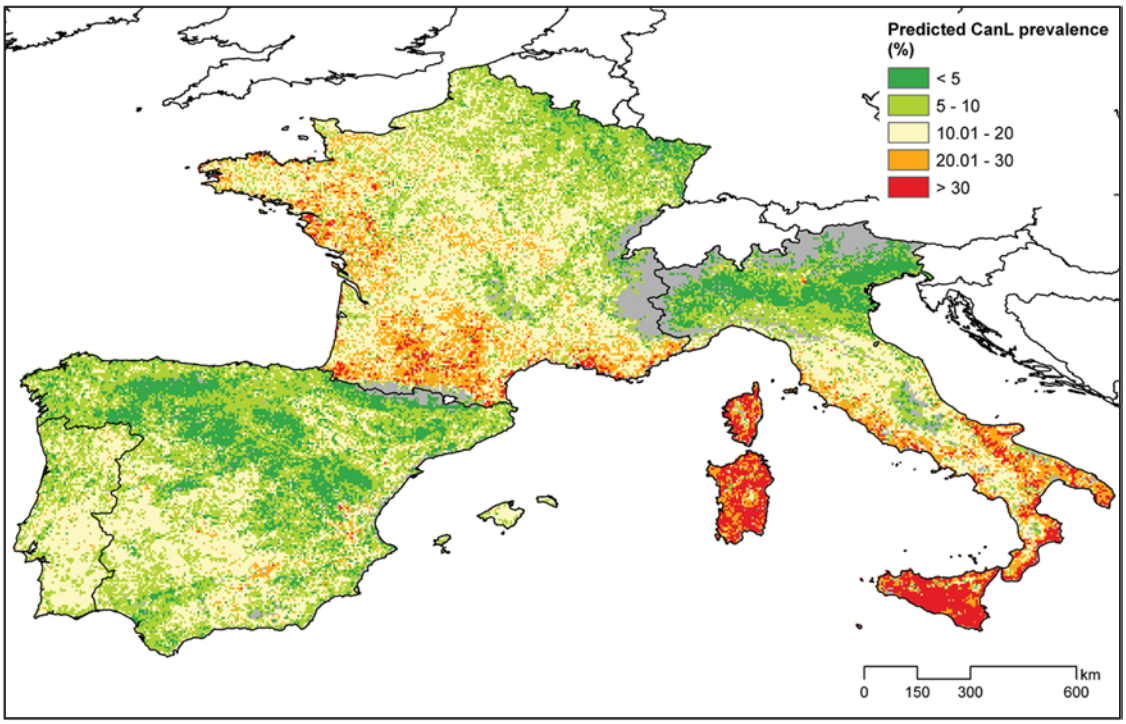

Fig. 2. Risk map for canine leishmaniasis (CanL) in Portugal, Spain, France and Italy based on the four-country model. Predicted CanL seroprevalence was mapped only for the pixels where the predictive environmental values were within the range of those in the survey locations. Other pixels, those outside this mask, are shaded in grey.

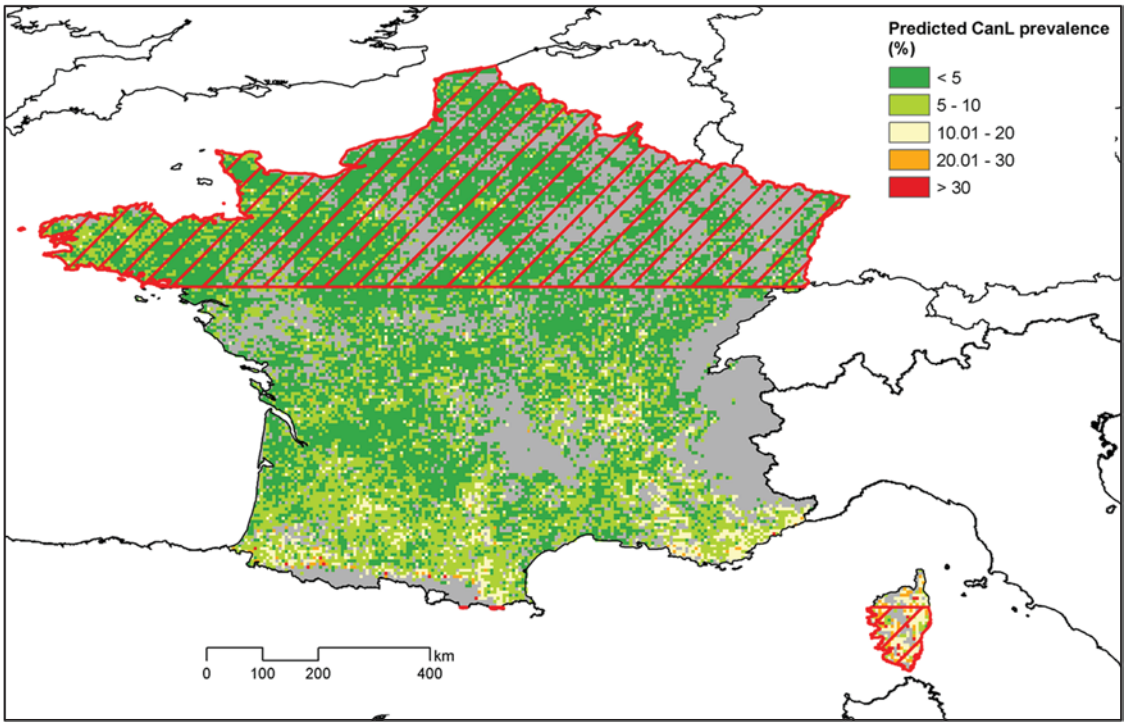

Fig. 3. Risk map for canine leishmaniasis (CanL) in France based on the France-only model. Predicted CanL seroprevalence was mapped only for the pixels where the predictive environmental values were within the range of those in the survey locations. Other pixels, those outside this mask, are shaded in grey. The areas marked by red diagonals show France outside the latitudinal range of all the surveys analysed. Latitude was not a predictive variable.

For France, overall, there were almost 13 times more pixels not classified at altitudes ⩽1193 m a.s.l. in the France-only risk map compared with the four-country one.

DISCUSSION

Our database is unique. It is the first time that detailed historical records of CanL in endemic European countries have been collated in this standardized way for such a long period (1965–2006), and the database is an important resource for future eco-epidemiological analyses at country and regional levels. However, like all such databases, its entries mirror the different histories of European teams working with changing diagnostic techniques and public health priorities (Dujardin *et al*. 2008), including the period (1990–1998) when epidemiological investigations of ZVL focused on co-infections with HIV. Partly for these reasons, many records were not suitable for our specific analyses, and we hope this finding will prompt better standardization of prospective surveys when cost effective.

Our use of this database in the current report has produced 2 sets of novel findings. One set arises from the construction of the first explicit risk map for CanL in western Mediterranean Europe and consideration of the lessons to be learned from its apparent deficiencies. The second set of findings relates to the identification of shortcomings in the scope and standardization of CanL surveys, and how these can inform guidelines for future collection of prospective data.

Our best model for western Europe, the four-country model, predicted the seroprevalence of CanL based on 7 variables, namely country, altitude, minimum night-time land surface temperature (LSTNmn), amplitude of the tri-annual cycle of the night-time land surface temperature (LSTNa3), amplitude of the annual cycle of the Enhanced vegetation index (EVIa1), phase of the annual cycle of the Enhanced vegetation index (EVIp1) and phase of the tri-annual cycle of the Enhanced vegetation index (EVIp3). The association with CanL seroprevalence was inversely U-shaped for 4 of the predictors (altitude, LSTNmn, LSTNa3, and to a lesser degree EVIp1), and this can be explained by the ecological requirements of the nocturnal sandfly vectors being met in rural settings on the lower and middle slopes of hills and mountains (often 100–800 m a.s.l.) during the dry Mediterranean summer (Ready, 2008; Gálvez *et al*. 2010*a*; Mahamdallie *et al*. 2011). The more linear associations with EVIa1 and EVIp3 were negative and positive, respectively, and these findings should be explored by sandfly ecologists. There is no published risk map for CanL in the western Mediterranean to compare with our own. However, our risk map does appear to be inconsistent with the incidence map of Trotz-Williams and Trees (2003), principally by not predicting more high risk areas near the coastal plains of Spain and Portugal, and also by predicting too many high risk areas in southwest, northwest and central France. Chamaillé *et al*. (2010) used the machine-learning Maxent method to produce an ecological niche model and a risk map for CanL in southern France, and this was based on presence data from our database, including many records from veterinary clinics. Our risk maps differ from theirs by identifying more low-high risk areas in central France and the inland parts of southwest France, as well as by identifying fewer medium-high risk areas in the more coastal parts of southwest France, where the Atlantic influences the climate.

One of the predictors of CanL in our four-country model was country. It is often the case that models differ between nearby countries, and this is true for anthroponotic visceral leishmanisis caused by *Leishmania donovani* in East Africa (Ready, 2008). Model variation between European countries might be explained by geographical differences in climate and seasonality associated with the bioclimatic and ecological requirements of the sandfly species serving as regional vectors. These considerations prompted us to construct a separate model and risk map for France, where both vectors (*P. perniciosus* and *P. ariasi*) have overlapping altitudinal ranges throughout much of the Mediterranean region (Ready, 2008; Chamaillé *et al*. 2010; Hartemink *et al*. 2011), unlike in Italy where 1 of the 2 main vectors (*P. perfiliewi*, not *P. perniciosus*) occurs only in the centre-south of the country (Maroli *et al*. 2008). Our best model for France alone, the France-only model, predicted the seroprevalence of CanL based on 6 variables, namely altitude, LSTNmn, amplitude of the bi-annual cycle of the night-time land surface temperature (LSTNa2), LSTNa3, amplitude of the annual cycle of middle infrared reflectance (MIRa1) and amplitude of the tri-annual cycle of middle infrared reflectance (MIRa3). LSTNa2 and MIRa3 were negatively associated with CanL seroprevalence, while the other 4 variables were positively associated with it. The France-only model differs from the four-country model mainly in the replacement of EVI, an index that estimates plant structural variation based on the reflectance of red and near-infrared wavelengths, with MIR, an index that distinguishes better between active vegetation on the one hand and senescent vegetation, soils, rocks or anthropogenic surfaces on the other (Scharlemann *et al*. 2008).

The France-only model was found to be better than the four-country model for predicting the seroprevalence of CanL in France, probably because the latter was based mostly on surveys and dogs from Italy (40% and 84%, respectively) rather than France (18% and 7·8%, respectively). The four-country model was a poor global predictor of point values of

CanL seroprevalence (correlation coefficient, r = 0·341), although the correlation was stronger for survey-rich Italy (r = 0·432) than for the other 3 countries (r = 0·233–0·262). In comparison, the France-only model gave a much better correlation between the predicted and observed CanL seroprevalences (r = 0·648). Based on AUC statistics, both models did provide fair-to-good discrimination (AUC >0·7) for the 2 extreme classes of CanL endemicity (<5%, >20% seroprevalence) but not for the intermediate class (5–20%). Many observed seroprevalences fell within this intermediate class (360 out of 947 overall, and 84 out of 169 in France), which is also true for much of the Mediterranean region (Dujardin *et al*. 2008), and so it is important to consider whether this lack of resolution results from the modelling approach or inadequate sampling.

The statistical methods commonly used to predict the occurrence or distribution of parasitic diseases in relation to environmental variables include discriminant (Rogers and Randolph, 2006) and Bayesian methods (Clements *et al*. 2006) as well as logistic regression, both binary (Brooker *et al*. 2001) and binomial (Rabe-Hesketh and Everitt, 2004). Binary logistic regression was used successfully for the spatial modelling of anthroponotic visceral leishmaniasis in East Africa (Thomson *et al*. 1999). Non-linear discriminant analysis (NLDA) might provide better predictions than logistic regression (Rogers and Randolph, 2006), and it would be interesting to apply it to our datasets. However, we do not believe that NLDA would much improve the resolution of our predictions for those locations where CanL seroprevalence is in the intermediate range of 5–20%. This conclusion stems from the limitations of our climate and RS datasets as well as of the sampling procedures used to obtain our survey data. In common with most spatial models, ours relied on using widely available sets of climate and RS data, which unfortunately cover different time periods and have varying spatial resolutions. None of the resolutions fully captured the variation in the microclimates inhabited by sandflies, which would hinder the inclusion of sandfly density as an explanatory variable in any integrated model (Hartemink *et al*. 2011). We could have modelled CanL seroprevalence over shorter time periods using matching climate surfaces from the EU ENSEMBLES project (http://www.ensembles-eu.org/), in order to improve the risk maps and investigate any effects of climate change. However, this requires the availability of CanL seroprevalence datasets with fewer limitations than those currently available, as now explained.

Our models' poor predictions of CanL seroprevalence in the intermediate range of 5–20% relate to shortcomings in the scope and standardization of CanL surveys. Firstly, there has been a bias towards known rural and peri-urban foci in Mediterranean bioclimate zones, and consequently zero seroprevalence was recorded in only 14·7% of the surveys (14–25% in France, Portugal and Spain, but just 6% in Italy). We recommend that future surveys do not neglect environments at more extreme altitudes (<100 m, >1000 m a.s.l.) and northerly latitudes. Secondly, the most used serological test (IFAT) has not been standardized, e.g. the proportion of surveys using a cut-off of <1/80 was 35·4% in Italy and 19·2% in Portugal, but negligible in Spain (2·4%) and France (0%). The use of low IFAT cut-offs in many surveys in Italy could possibly have produced for this country the highest median seroprevalence (17·7%; 5·9–8% elsewhere) and the lowest frequency of zero seroprevalence (6%; 14–25% elsewhere). We strongly recommend that all published surveys include the proportion of dogs seropositive for each of the serial dilutions of the sera, not just for those making the cut-off, so that records from different periods can be better compared. Improvements in fluorescence microscopy have led to changes in the consensus cut-off, first 1/40 in the 1970s, then 1/80 in the 1990s, and sometimes 1/160 recently. Thirdly, serology only detects a small fraction of resistant dogs and the susceptibles progressing to clinical disease after long incubation (Maia and Campino, 2008). Therefore, ideally, serology should be complemented by molecular screening in horizontal surveys (Lachaud *et al*. 2002; Quinnell and Courtenay, 2009). Lastly, the resolution of the sampling is inadequate, because a point location was very often an assembly point in a large village, to which dogs were brought for screening from a radius of 10 km or more (Rioux and Golvan, 1969; Maroli *et al*. 2008; Martín-Sánchez *et al*. 2009; Gálvez *et al*. 2010*b*). To improve the sampling resolution, we recommend including in the model 'dog factors', such as the age and lifestyle of individual dogs recorded as living most of the time in specific geo-referenced habitations. Concerning lifestyle, guard dogs and hunting dogs often have a higher prevalence of CanL than pet dogs that sleep indoors (Lanotte *et al*. 1978; Martín-Sánchez *et al*. 2009; Gálvez *et al*. 2010*b*). A potentially important explanatory variable missing from our modelling is dog density. Estimates have been used for ecological niche modelling (Chamaillé *et al*. 2010) and R_0 modelling (Hartemink *et al*. 2011) of CanL or its vectors in southern France, but these estimates of dog density were based on human population densities that were low and varied little per km^2 in the rural areas characteristic of CanL.

As a result of these shortcomings, our risk maps can not be expected to have much predictive power within many CanL foci. It might not always be cost effective to produce high resolution risk maps within CanL foci, for which it will be necessary to follow all our recommendations, but the consequences of not doing so ought to be assessed when agreeing guidelines for collecting prospective data. In

contrast, it should be less challenging to improve spatial models so that different geographical regions can be compared. Then, our relatively straightforward method of producing risk maps would be more useful for predicting any northward emergence of CanL and helping to plan barrier methods of control, based currently on topical insecticides (deltamethrin-impregnated dog collars and pour-ons) not dog culling (Quinnell and Courtenay, 2009) and hopefully on vaccines (Ready, 2010). Already, we have demonstrated that country-level risk maps can distinguish well between areas with high (>20%) and low (<5%) CanL seroprevalences (only 3% misclassified using the France-only model) and that seroprevalences are invariably low in areas of CanL emergence. For example, seroprevalence is usually <5% throughout northern Italy above latitude $45{\cdot}5^0$ N, where CanL has emerged within the last 20 years (Maroli *et al*. 2008). The improvement of spatial models for predicting CanL emergence in Europe depends on carrying out new standardized serological and molecular surveys in central and northern Europe as well as at extreme altitudes (<100 m, >1000 m a.s.l.) in Mediterranean locations with comparable land covers. Without such sampling, predictions will not be possible for many regions of interest, such as much of France above $47{\cdot}5^0$ N, which is the approximate northern limit of our records and of the risk map produced by ecological niche modelling (Chamaillé *et al*. 2010). The complexities of predicting the effects of climate change (Kovats *et al*. 2001) lessen the likelihood of risk maps being used to predict the spread of CanL in response to global warming.

ACKNOWLEDGEMENTS

This work is dedicated to Professor Clive R. Davies, who initiated the research but passed away in March 2009. We thank the other members of the EDEN-Leishmaniasis subproject who helped to construct the database, namely Maria Antoniou and the team at the University of Heraklion (Greece), Robert Farkas and his team in the Faculty of Veterinary Science of Budapest (Hungary), and Yusuf Ozbel and his team in the University of Ege (Turkey). We are grateful for advice on statistical analyses from Francesco Checchi and Simon Brooker of the London School of Hygiene and Tropical Medicine, London, as well as Archie Clements and Ricardo Magalhães of the School of Population Health, University of Queensland, Australia.

FINANCIAL SUPPORT

This work was funded by a grant (GOCE-2003-010284) of the European Union awarded to the EDEN (Emerging Diseases in a changing European eNvironment) project. The publication is catalogued by the EDEN Steering Committee as EDEN0262 (www.edenfp6project.net). Its contents are the sole responsibility of the authors and do not necessarily reflect the views of the European Commission. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. The authors have declared that no competing interests exist.

REFERENCES

Brooker, S., Hay, S. I., Issae, W., Hall, A., Kihamia, C. M., Lwambo, N. J., Wint, W., Rogers, D. J. and Bundy, D. A. (2001). Predicting the distribution of urinary schistosomiasis in Tanzania using satellite sensor data. *Tropical Medicine & International Health* **6**, 998–1007.

Brooker, S., Hay, S. I. and Bundy, D. A. (2002). Tools from ecology: useful for evaluating infection risk models? *Trends in Parasitology* **18**, 70–74.

Chamaillé, L., Tran, A., Meunier, A., Bourdoiseau, G., Ready, P. and Dedet, J.-P. (2010). Environmental risk mapping of canine leishmaniasis in France. *Parasites & Vectors* **3**, 31.

Clements, A. C., Moyeed, R. and Brooker, S. (2006). Bayesian geostatistical prediction of the intensity of infection with *Schistosoma mansoni* in East Africa. *Parasitology* **133**, 711–719.

Dujardin, J. C., Campino, L., Cañavate, C., Dedet, J. P., Gradoni, L., Soteriadou, K., Mazeris, A., Ozbel, Y. and Boelaert, M. (2008). Spread of vector-borne diseases and neglect of Leishmaniasis, Europe. *Emerging Infectious Diseases* **14**, 1013–1018.

Duprey, Z. H., Steurer, F. J., Rooney, J. A., Kirchhoff, L. V., Jackson, J. E., Rowton, E. D. and Schantz, P. M. (2006). Canine visceral leishmaniasis, United States and Canada, 2000–2003. *Emerging Infectious Diseases* **12**, 440–446.

Gálvez, R., Descalzo, M. A., Miró, G., Jiménez, M. I., Martín, O., Dos Santos-Brandao, F., Guerrero, I., Cubero, E. and Molina, R. (2010*a*). Seasonal trends and spatial relations between environmental/meteorological factors and leishmaniosis sand fly vector abundances in Central Spain. *Acta Tropica* **115**, 95–102.

Gálvez, R., Miró, G., Descalzo, M. A., Nieto, J., Dado, D., Martín, O., Cubero, E. and Molina, R. (2010*b*). Emerging trends in the seroprevalence of canine leishmaniasis in the Madrid region (central Spain). *Veterinary Parasitology* **169**, 327–334.

Gething, P. W., Noor, A. M., Gikandi, P. W., Hay, S. I., Nixon, M. S., Snow, R. W. and Atkinson, P. M. (2008). Developing geostatistical space-time models to predict outpatient treatment burdens from incomplete national data. *Geographical Analysis* **40**, 167–188.

Hartemink, N., Vanwambeke, S. O., Heesterbeek, H., Rogers, D., Morley, D., Pesson, B., Davies, C., Mahamdallie, S. and Ready, P. (2011). Integrated mapping of establishment risk for emerging vector-borne infections: a case study of canine leishmaniasis in southwest France. *PLoS ONE* **6**, e20817.

Hastie, T., Tibshirani, R. and Friedman, J. (2001). Model assessment and selection. In *The Elements of Statistical Learning; Data Mining, Inference and Prediction*, pp. 214–217. Springer, New York, USA.

Hay, S. I., Guerra, C. A., Gething, P. W., Patil, A. P., Tatem, A. J., Noor, A. M., Kabaria, C. W., Manh, B. H., Elvazar, I. R., Brooker, S., Smith, D. L., Moyeed, R. A. and Snow, R. W. (2009). A world malaria map: *Plasmodium falciparum* endemicity in 2007. *PLoS Med* **6**, e1000048.

Hosmer, D. W. and Lemshow, S. (2000). *Applied Logistic Regression*. John Wiley & Sons, New York, USA.

Kovats, R. S., Campbell-Lendrum, D. H., McMichael, A. J., Woodward, A. and Cox, J. S. (2001). Early effects of climate change: do they include changes in vector-borne disease? *Philosophical Transactions of the Royal Society of London, B* **356**, 1057–1068.

Lachaud, L., Chabbert, E., Dubessay, P., Dereure, J., Lamothe, J., Dedet, J.-P. and Bastien, P. (2002). Value of two PCR methods for the diagnosis of canine visceral leishmaniasis and the detection of asymptomatic carriers. *Parasitology* **125**, 197–207.

Lanotte, G., Rioux, J.-A., Croset, H. and Vollhardt, Y. (1978). Ecology of leishmaniasis in southern France. *9. Sampling methods in the study and analysis of canine enzootic leishmaniasis*. Annales de Parasitologie Humaine et Comparée **53**, 33–45.

Mahamdallie, S. S., Pesson, B. and Ready, P. D. (2011). Multiple genetic divergences and population expansions of a Mediterranean sandfly, *Phlebotomus ariasi*, in Europe during the Pleistocene glacial cycles. *Heredity* **106**, 714–726.

Maia, C. and Campino, L. (2008). Methods for diagnosis of canine leishmaniasis and immune response to infection. *Veterinary Parasitology* **158**, 274–287.

Maroli, M., Rossi, L., Baldelli, R., Capelli, G., Ferroglio, E., Genchi, C., Gramiccia, M., Mortarino, M., Pietrobelli, M. and Gradoni, L. (2008). The northward spread of leishmaniasis in Italy: evidence from retrospective and ongoing studies on the canine reservoir and phlebotomine vectors. *Tropical Medicine & International Health* **13**, 256–264.

Martín-Sánchez, J., Morales-Yuste, M., Acedo-Sánchez, C., Barón, S., Diaz, V. and Morillas-Marquez, F. (2009). Canine leishmaniasis in southeastern Spain. *Emerging Infectious Diseases* **15**, 795–798.

New, M., Lister, D., Hulme, M. and Makin, I. (2002). A high-resolution data set of surface climate over global land areas. *Climate Research* **21**, 1–25.

Pullan, R. L., Bethony, J. M., Geiger, S. M., Cundill, B., Correa-Oliveira, R., Quinnell, R. J. and Brooker, S. (2008). Human helminth co-infection: analysis of spatial patterns and risk factors in a Brazilian community. *PLoS Neglected Tropical Diseases* **2**, e352.

Quinnell, R. J. and Courtenay, O. (2009). Transmission, reservoir hosts and control of zoonotic visceral leishmaniasis. *Parasitology* **136**, 1915–1934.

Rabe-Hesketh, S. and Everitt, B. (2004). *A Handbook of Statistical Analyses Using Stata, 3rd Edn*. Chapman & Hall/CRC, Boca Raton, FL, USA.

Ready, P. D. (2008). Leishmaniasis emergence and climate change. In *Climate Change: Impact on the Epidemiology and Control of Animal Diseases* (ed. De la Roque, S.) *Revue Scientifique et Technique, Office International des Épizooties/ Scientific and Technical Review, World Organization for Animal Health* **27**, 399–412.

Ready, P. D. (2010). Leishmaniasis emergence in Europe. *Eurosurveillance* **15**, e19505.

Ribeiro, P. J. Jr, and Diggle, P. J. (2001). geoR: a package for geostatistical analysis. *R-NEWS* **1**, 15–18.

Rioux, J. A. and Golvan, Y. J. (1969). *Épideémiologie des leishmanioses dans le sud de la France*. Institut National de la Santé et de la Recherche Médicale, Paris.

Rogers, W. H. (1993). Regression standard errors in clustered samples. *Stata Technical Bulletin* **13**, 19–23. (Available: http://www.stata.com/products/stb/journals/stb13.pdf. Accessed 11 October 2009)

Rogers, D. J., Hay, S. I., and 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. and Randolph, S. E. (2006). Climate change and vector-borne diseases. In *Global Mapping of Infectious Diseases: Methods, Examples and Emerging Applications* (ed. Hay, S. I., Graham, A. J. and Rogers, D.J.), pp. 345–381. Academic Press, London, UK.

Romero, G. A. and Boelaert, M. (2010). Control of visceral leishmaniasis in Latin America – a systematic review. *PLoS Neglected Tropical Diseases* **4**, e584.

Royston, P., Ambler, G. and Sauerbrei, W. (1999). The use of fractional polynomials to model continuous risk variables in epidemiology. *International Journal of Epidemiology* **28**, 964–974.

Scharlemann, J. P., Benz, D., Hay, S. I., Purse, B. V., Tatem, A. J., Wint, G. R. and Rogers, D. J. (2008). Global data for ecology and epidemiology: a novel algorithm for temporal Fourier processing MODIS data. *PLoS One* **3**, e1408.

Thomson, M. C., Elnaiem, D. A., Ashford, R. W. and Connor, S. J. (1999). Towards a kala azar risk map for Sudan: mapping the potential distribution of *Phlebotomus orientalis* using digital data of environmental variables. *Tropical Medicine & International Health* **4**, 105–113.

Tibshirani, R., Hastie, T., Narasimhan, B. and Chu, G. (2002). Diagnosis of multiple cancer types by shrunken centroids of gene expression. *Proceedings of the National Academy of Science,s USA* **99**, 6567–6572.

Trotz-Williams, L. A. and Trees, A. J. (2003). Systematic review of the distribution of the major vector-borne parasitic infections in dogs and cats in Europe. *Veterinary Record* **152**, 97–105.