The Macroepidemiology of Parasitic and Infectious Diseases: A Comparative Study Using Artificial Neuronal Nets and Logistic Regressions

J.F. Guégan · F. Thomas · T. de Meeüs · S. Lek · F. Renaud

14.1 Introduction

Of the about 270 species of helminths, protozoa and arthropods which may permanently or occasionally infect human populations, less than 45 species, or about 16 percent, are strictly dependent on humans for their survival (Ashford 1991; Petney and Andrews 1998). For the few West European countries providing reasonably reliable demographic data before the nineteenth century, epidemics, famines and wars are favoured as the three critical controlling mechanisms in human demographic crisis (Jones 1990). For instance, bubonic plague, one of the most dreadful epidemic killers, dominated the pattern of mortality variation from 1340 to its disappearance after 1670 in Europe, when smallpox epidemics may well have assumed a similar determining effect. Undoubtedly, mankind has experienced such disease effects along its evolution, leaving each time relatively resistant populations (Anderson and May 1991; Ewald 1994). Adopting a wider perspective, the "health" of man is determined essentially by his behaviour, his food and the nature of the world around him, and as such he is directly or indirectly influenced by different forms of parasitic and infectious diseases (Combes 1995). It may appear obvious that human conditions represent foci for a wide range of diseases (Anderson and May 1991). Unfortunately, the intimate interactions between different forms of diseases and human life-history traits have been virtually neglected (Immerman 1986). Life-history theory predicts that faced with virulent parasites, hosts should adjust their reproductive biology by increasing reproductive output and/or reducing age at maturity (Stearns 1992; Michalakis and Hochberg 1994; McNamara and Houston 1996; Sorci et al. 1996; Reeson et al. 1998; Kris and Lively 1998; Brooke et al. 1998). Intuitively, variation in parasite species composition across countries might be sensitive to human life-history traits, and vice versa as predicted by theory. In such a perspective, the determination of the exact relationships between abiotic and biotic characteristics and both presence/absence and spatial occupancy of diseases may appear crucial in that they might probably help to improve our understanding of the underlying processes that generate them. However, the nature of factors affecting the presence/absence of a given disease and its spatial distribution has been derived from a combination of expert opinion, limited data and the use of geographical and climate descriptors. This is largely due to a traditional individual-centred medical preoccupation in understanding these diseases (Jones 1990), and the disciplinary gap that exists between biomedical scientists, ecologists and evolutionary biologists (Petney and Andrews 1998). As recently pointed out by Craig et al. (1999), none has a clear and reproducible numerical definition of Malaria distribution in Africa, for instance; consequently, its comparative value is rather limited. Interestingly, large global data sets in-

cluding environmental and human population data are now available which make them suitable for comparative studies.

Given these above remarks, our primary focus in this contribution was to model, and then to predict the spatial representation of different human infectious and parasitic diseases, some of which with very deleterious effects on populations across countries. The traditional method for depicting distribution and temporal patterns of major diseases used by epidemiologists and geographers has been to map rates of change in spatial distribution, or to provide a series of static "snapshot" maps (Pedersen 1995). In practice, this facilitates a form of prediction in which we can calculate declines in site occupancy or colonizations of new sites by diseases. Even though these models are entirely relevant since they permit one to present the results to a wide audience (Martin 1996), they do not authorise crude predictions for risk assessment of disease re-emergence or colonization of new sites. Here, we conduct a comparative analysis on a global scale using two multivariate methods, i.e. logistic regressions and artificial neuronal networks, to precisely predict the spatial distribution of diseases. We then compare the performance of both logistic and artificial neuronal network models in predicting the actual spatial distribution of the infectious and parasitic diseases under study. Finally, we explore the utility of such predictive models in epidemiology, and notably how variations in time, say climate change, may affect the actual disease distribution.

14.2 Materials and Methods

14.2.1 Materials

We compiled data for a total set of 168 different countries located all over the world and for which all population, geographical and epidemiological information was available. Large global data sets are now available which make the modelling of disease spatial distribution entirely relevant. In doing so, one should make the maximum use of the available data, as accurate as possible, and be able to appreciate the potential inaccuracy in the results. Epidemiological data were obtained from two main sources, the World Health Organization (W.H.O.) and the Center for Disease Control and Prevention (C.D.C.), a quick and convenient method entirely reliable for such a purpose. Our models assume that all variables, e.g. presence of a given disease, have a homogeneous distribution across each source country. From the total data set of 168 countries, we considered a subset of 153 countries for a phylogeny-based comparison analysis (see below).

14.2.1.1 *Spatial Patterns*

Since geographical and ecological factors might strongly influence the variation of parasite species distribution across countries, we considered five ecogeographical variables for each country. These spatial descriptors are those which are probably the most usually invoked for explaining free-living species occupancy and distribution

on largest scale (Brown 1995; Rosenzweig 1995; Whittaker 1975). They are: *(1)* total surface area of a given country (in km^2), since larger land masses may harbour higher species diversity than smaller masses do, and thus the likelihood it incorporates a given parasite species is higher; *(2)* mean latitude (in degree and minutes which refers to the value taken at the geographical centre of each country), since higher species diversity is generally found under tropical areas, and many human infectious diseases are primarily concentrated in those regions, when compared to more septentrional provinces; *(3)* mean longitude (in degree and minutes, measured as previously) which takes into account the fact that parasite species might have dispersed along an east-west gradient from their centre of dispersion. These three environmental parameters were log-transformed in logistic regression models in order to minimize effects of nonnormality on statistics (Zar 1996). They have been kept unchanged in artificial neuronal net procedures. Furthermore, we considered whether or not a country was located *(4)* on the northern or southern hemisphere, since countries are more numerous in the northern part of the world which represents a statistical artefact, and *(5)* on a land mass or an island since island populations may present frequent fade-outs of infection (Rhodes and Anderson 1996) or a given disease may be extinguished there (Esch et al. 1990). These two variables were coded as categorical variables (0/1). Initially, the three continuous variables were incorporated into principal component analysis to reduce dimensionality and eliminate collinearity between these source variables (Sheldon and Meffe 1995; Oberdorff et al. 1998). Only one synthetic output principal component (PC*GEO*1, eigenvalue = 2 107.91) explained 99.58% of the total inertia, 89.98% of which depending on the latitudinal effect. Because of no effect of multicollinearity on final models, we decided to use raw variables. All this spatial information can be accepted as synthetic as possible reflecting other potential physical variables influencing parasite species distribution, e.g. temperature or ecosystem productivity as well. All this available data may contribute to the actual distribution of environmentally determined diseases. Spatial data were compiled from World Atlas v. 2.1.0 ©, on a Macintosh personal computer.

14.2.1.2
Economic, Social and Demographic Patterns

Because the human disease characteristics might differ so much across countries having more or fewer inhabitants, more or less urbanization, or more or fewer financial supports for health care campaigns, we also compiled demographic and economic data for the 153 countries. Data for population geography were essentially obtained from the 1992 world population data sheet (Jones 1990). Five demographic or economic parameters were retained for each country: *(1)* total population (in number of people per country), which represents the potential colonizing pool for any disease; *(2)* total population growth (per 1 000 people), which gives an estimate of the reproductive ability in growing populations; *(3)* population density (number of people per km^2), which permits one to separate countries on a continuum of populations with high aggregate behaviour (as for high urbanization areas) to lower level (in rural areas), which can strongly influence the likelihood of disease successful transmission; *(4)* death rate (per 1 000 people), which gives an estimate of differential mortality in the area, and which can be attributed in part to the deleterious effect some diseases may

actually have on human health; *(5)* per capita gross national product (GNP in US$ a year) to evaluate the resource, or income effect, on disease spatial distribution through financial supports granted by local politicians and governments in health care campaigns. The three variables, i.e. total population, population density and GNP, were log-transformed and the two parameters, i.e. total population growth and death rate, were arcsine-transformed to deal with nonlinearity before introduction into logistic regression models (Zar 1996). These variables were kept unchanged into artificial neuronal net methods. To avoid multicollinearity between all these variables, we proceeded as previously in using principal components analysis. Since final models did not change between using raw variables or principal components, we kept predictive variables unchanged. Conceptually, all these factors may reflect the probability of transmission occurring or not occurring from high (e.g. countries with high population density and low incomes) to low transmission intensity (e.g. countries with low population density and large incomes).

14.2.1.3 *Historical Patterns*

Generally, closely related taxa are more likely to exhibit similar traits than distant taxa since they have been subject to similar evolutionary constraints inherited from a common ancestor (Harvey and Pagel 1991). Intuitively, two groups of relative human ethnic groups might share similar traits. For instance, they might harbour the same infectious disease, or group of co-occurring diseases, which both represent a result of phylogenetic history. Immunodeficiency and reduced antibody levels to both related tribes may be invoked to explain this historical component, but such assumptions have seldom been elucidated completely. Although such aspects remain to be investigated, there is evidence indicating that two populations cannot be treated as statistically independent points (Martins 1996; Martins and Hansen 1997). To deal with the confounding effect of common history on parasitism, we used the human group phylogeny based on the findings of Cavalli-Sforza (1997). Unfortunately, it was impossible to directly use the entire phylogeny, mainly due to the difficulty of crossing this phylogenetic information based on the existence of more or less well-recognized ethnological groups with our data concerning political nations. From the entire database of 168 countries, we retained a subset of 153 countries for performing a phylogeny-based correction analysis. We decided to consider only the eight large divisions of ethnological groups as defined hereafter. Then, we used only countries for which at least 50% of inhabitants belong to one majority ethnological group. We omitted some countries, e.g. Brazil, *f*SU, South Africa, Chile, with a high human polymorphism. To control for the confundant effect exerted by common history on calculation, different comparative methods have been developed to take this nonindependence into account (see Martins 1996 for review). For the purpose of this work, we used a General Least-Square Model, which permits one to remove the variance due to common history using categorical codes (Grafen 1992). The choice of this phylogenetic method was based on its better robustness toward misspecification of our models (Martins and Hansen 1997). Phylogenetic coding variables, coded 0 or 1, were introduced into predictive models as dummy categorical variables. The eight main divisions of human groups considered in this work are:

I. Africans and Nilotics (except native people from the Maghreb);
II. Europeans (including people from the Middle-East);
III. Indians;
IV. Mongoloids, Japanese and Koreans;
V. Amerindians;
VI. New Guineans-Papous;
VII. Melanesians;
VIII. Mhongs, Khmers, Thais, Filipinos, Indonesians and related tribes.

The first coding variable separates the tribe division I from all other tribes, and it removes the differences between the means of these two hierarchical groups. Next, and nested within it, we separate the group formed by tribes II to V from the group including tribes VI to VIII, and so on down to the last bifurcating branch of the phylogeny.

14.2.1.4
Human Life-History Trait Patterns

Life-history theory assumes that reproduction is costly, and that strategic decisions should be selected by organisms over their lifetime in contrasting ecological and social environments (Alexander 1987; Shykoff et al. 1996; Lively 1987; Hochberg et al. 1992; Forbes 1993; Lafferty 1993). Parasitism may be such an underlying environmental condition, which could interfere with survival and reproduction capabilities in humans across distinct regions of the world. Since parasites use resources from their human hosts for their maintenance and own reproduction, costs of reproduction in humans can be predicted to increase in the presence of parasitism (see Møller 1997; Teriokhin 1998 for a theoretical viewpoint). In addition, differences in male and female life histories such as life span variability across sexes, i.e. constant shorter life span for man than for his congener world-wide (Teriokhin and Budilova 2000), may possibly interfere with varying levels of parasitism. In this study, three different human traits, susceptible to interfere with infection, were available for predicting disease spatial distribution: *(1)* fertility rate, which indicates the number of children that would be born to 1 000 women during their lifetime passing through the child-bearing ages; *(2)* female life expectancy at birth (in years); and *(3)* male life expectancy at birth (in years). Life expectancies at birth are substantially higher for females than males, and thus sex ratios are directly affected, and in turn both sexes may represent differential hosts for parasites. This refers to the three most current parameters used in evolutionary ecology to estimate the degree of organism fitness (see Teriokhin and Budilova 2000).

14.2.1.5
Parasite Patterns

Disease occurrences in the 153 different countries were compiled from information available on mainly two different web sites, the World Health Organization (Geneva, Switzerland at http://www.WHO.int/) and the Center for Disease Control and Prevention (Atlanta, USA at http://www.CDC.gov/) sites. Many of the investigations of human parasites have been based on the microscopic examination of the patients' stool

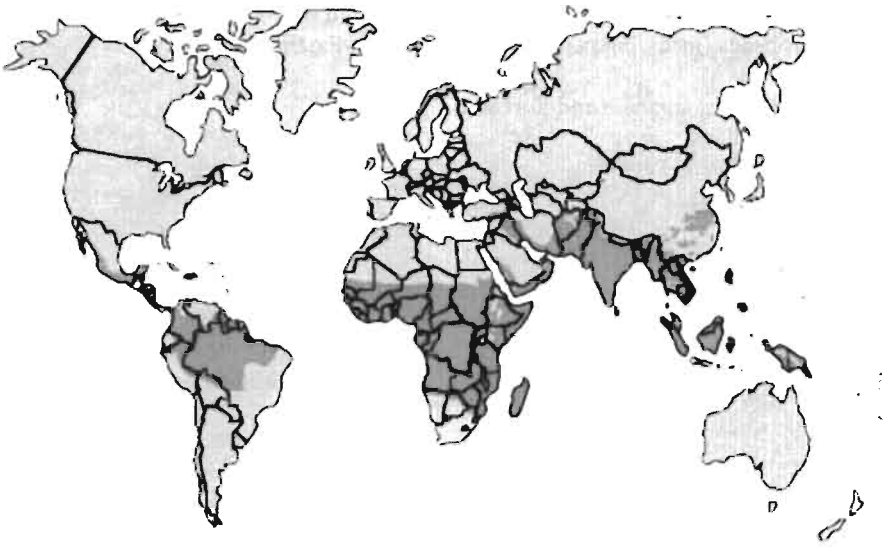

Fig. 14.1. Actual distribution of malaria in the world. A total of 111 different countries (66.1%) on 168 present in our total database are affected by paludism (after W.H.O./C.T.D. 1997)

for helminth eggs. Thus, the estimates of the overall human parasitofauna are certainly underevaluated. Therefore, we were able to collect data on presence/absence for a set of 15 categories of human diseases known to have a more or less large impact on human health. When information at the species level was not available, we decided to pool these data by category of diseases. Disease categories are as follows: Hepatitis A, Hepatitis B, Malaria (see Fig. 14.1 for illustration), Schistosomiasis, Filariosis, Meningococcosis, Yellow fever, Dengue fever, Cholera, Trypanosomiasis, Dracunculosis, Chagas, Lyme, cutaneous Leishmaniosis and visceral Leishmaniosis. Other diseases were available, e.g. Typhoid fever, but they were widespread species with a range size of 153 countries. We are absolutely conscious that these values are subject to some sources of error, e.g. some parasite species may have been recently introduced into countries and thus they do not yet appear into the available check-lists, or they have been annihilated for a couple of decades, but their potential effect on human populations is still effective, or the presence/absence of diseases have not been declared to the two international organizations, or this refers to only a sub-sample of what really exists, but these data are what is really available to us today! In many ways, many of these infectious and parasitic diseases represent the actual most dreadful killers occurring on earth.

14.2.2 Methods

From the variety of multivariate statistics that can be used to predict a presence/absence event, we opted for two distinct techniques for estimating the probability that an event of presence (or absence) of a parasitic or infectious disease occurs across

countries. There are logistic regressions (Jongman et al. 1995; Norusis 1997) and artificial neuronal networks (Rumelhart et al. 1986; Edwards and Morse 1995). A third potential method, i.e. multiple discriminant analysis, was not relevant here since the data incorporate categorical variables as independent parameters (Jongman et al. 1995).

14.2.2.1 *Logistic Regression Models*

Presence, or absence, of a given infectious disease across the 153 countries was fitted to the 20 independent variables listed above (see Section 14.2.1) using logistic regression procedure. The general linear model can be written as follows:

$$prob(1) = 1 / (1 + e^{-z})$$

$$prob(0) = 1 - prob(1)$$

with e is the base of the natural logarithms, $prob(1)$ the associated probability that a given disease occurs in a country, $prob(0)$ the associated probability that a given disease does not occur, and z the linear combination of the independent variables of the form

$$z = b_0 + b_1(\text{surface area}) + b_2(\text{mean latitude}) + \ldots + b_{20}(\text{ethnyVIII})$$

The logistic model in terms of the log of odds, or log*it*, can be written as follows:

$$\text{log}it\,[prob(1) / prob(0)] = z \text{ or } prob(1) / prob(0) = e^z$$

The parameters of the logistic regression model were estimated using the maximum likelihood method. The nine categorical variables, i.e. hemisphere, landmass/island, and the seven ethnic groups, were entered into regressions as indicator-variable coding. The other variables were considered as continuous variables. The Wald's statistic and its associative significance level were used to detect significant independent variables within the logistic model, using a significance level of 0.10. R statistics were used for determination of partial correlation between the disease occurrence dependent variable and each of the independent parameters. Accuracy of fit of the logistic models was tested using −2 times of the likelihood (−2LL) with a model perfectly fitting data having a score of 0. The proportion of total explained variation in logistic regression was given by the Nagelkerke R-square statistics.

We then compare predictions obtained from logistic models to the observed outcomes using contingency tables with cut values of 0.50. In the case of a cut value of 0.50, this indicates whether the estimated probability is greater or less than one-half.

All the independent variables were first entered into logistic regression models which permits one to control for the effect of other independent variables on a given descriptor variable (general model). Second, we proceeded to backward elimination procedure in order to identify minimal models with a subset of independent variables as good predictors as the total set of independent variables entered into general models.

14.2.2.2
Artificial Neuronal Networks

Artificial neuronal nets are known for their capacity to process nonlinear relationships (Rumelhart et al. 1986; Freeman and Skapura 1992). In the present work, we used one of the general principles of artificial nets, i.e. the back propagation algorithm (Gallant 1993) for training the database with a typical three-layer feed-forward 20-3-2 network (Fig. 14.2), that is, 20 input neurons corresponding to the 20 independent parameters introduced into the model, 3 hidden neurons determined as the optimal configuration to obtain a best compromise between bias and variance and 2 output neurons for disease occurrence, i.e. one for presence of the disease and one for its absence. The performance of the artificial neuronal nets was analysed using two different techniques of partitioning the total data set into a first subset for training the neuronal model and a second set for testing its real predictive power. Since there is

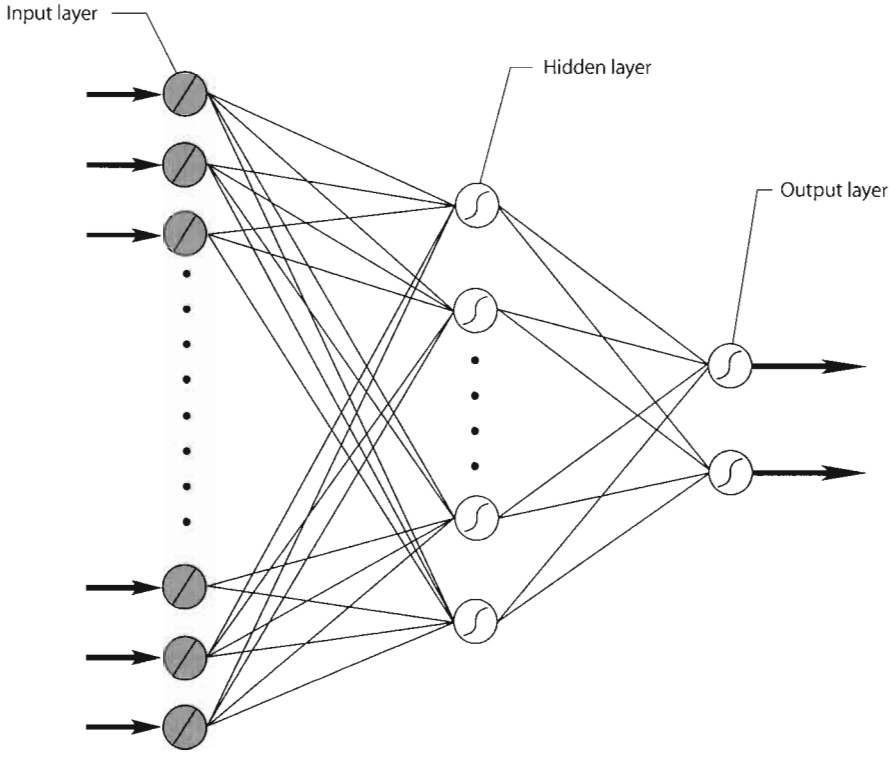

Fig. 14.2. Schematic representation of a three-layered feed-forward neuronal network, with one input layer, one hidden layer and one output layer as used in the present work. The left side shows the input parameters used in back propagation network models, and the right side illustrates the output parameters, i.e. presence or absence of a given disease

actually a large debate on the adequacy of partitioning methods and its effect on model error rates (Fielding and Bell 1997) without any special consensus, we opted to develop both "leave-one-out" (Efron 1983; Kohavi 1995) and "hold-out" (Efron 1983; Kohavi 1995; Friedman 1997) cross-validations for our data-set.

The "leave-one-out", or jack-knife procedure, leaves out a test set (one country × 20 inputs) from the training set (152 countries × 20 inputs), and this is repeated for each country. Then, the model run with the training set may be used to predict the presence/absence of a given disease in the test set. This was repeated with a maximum of 1 000 iterations for each country.

In the "hold-out" procedure also called *k*-fold cross-validation, two random sets are extracted from the total data set: a trained set (¾, i.e. 115 countries) and a test set (¼, i.e. 38 countries). Similar to the jack-knife procedure, the model is first adjusted with the training set, and then it is used for prediction of presence/absence in the remaining test set (Kohavi 1995; Friedman 1997). This procedure was repeated 10 times to provide a better compromise prediction on the random test sets.

As for logistic regressions, we compared predictions obtained from neuronal net models to the observed outcomes using classification tables with cut values of 0.50.

Readers will be able to find further details on the different neuronal network procedures in the first chapter of this book.

All statistical analyses were performed with SPSS 7.5 and MatLab 5.0 for a personal computer

14.2.2.3 *Comparative Analysis*

We evaluated the efficacy of both multivariate models for classifying the presence, or absence, of the 15 infectious and parasitic diseases by plotting the ratios of sensibility values (i.e. the true positive fraction) versus 1-specificity values (i.e. false positive fraction) against the different levels of occurrence frequency, i.e. prevalence observed across the total set of diseases. In fact, as pointed out by Fielding and Bell (1997) and more recently by Manel et al. (1999), a decreasing frequency of occurrence may be responsible for an exaggerated inflation of positive prediction errors. It is well accepted that logistic regressions are sensitive to such biases (Norusis 1997), but unfortunately we do not have any specific ideas how neuronal networks may be affected by the frequency of events occurring.

Additionally, the use of an arbitrary threshold probability, or cut-off value, which discriminates between predicted probabilities, of saying 0.50 and greater to be classified as having positive nodes, and of 0.50 and smaller as having negative ones, may be strongly influenced by prevalence values, i.e. the number of positive occurrences in the total data-set. To deal with the effect of selecting a specific cut-off value on prediction error, we compared the predictive performance of both logistic regression and artificial neuronal net models using ROC (Received Operating Characteristics) curves across different levels of threshold probabilities as recognized by Zweig and Campbell (1993) and Manel et al. (1999).

14.3 Results

14.3.1 Logistic Regressions and Artificial Neuronal Networks Face to Face

14.3.1.1 *Logistic Regressions*

Table 14.1 illustrates the classification results obtained for the 15 infectious diseases using logistic regression methods. The percentages of good classification for the 15 diseases were strongly high, varying from 90.2 to 100% (mean 97.3%, $SD = \pm 3.5$) of countries well classified when all independent variables were kept in regressions. True absence scores, which determine the negative predictive power of data, varied from 90.6 to 100% (mean 91.9%, $SD = \pm 2.9$), and true presence scores, which represent the positive predictive power, ranged from 85.2 to 100% (mean 95.8%, $SD = \pm 5.2$). Interestingly, these results show that there is no substantial difference between classification performance of positive and negative cases.

In a backward stepwise selection procedure generating a minimal logistic model for each disease, we obtained comparable scores of classification for countries (data not illustrated for clarity of the manuscript): overall good classification between 49.4 and 100% (mean 90.8%, $SD = \pm 13.8$); true absence scores between 46.9 and 100% (mean 86.7%, $SD = \pm 27.3$); true presence scores between 33.3 and 99% (mean 71.6%, $SD = \pm 33.7$). Two diseases, Chagas and Hepatitis A, strongly affected overall performance through lower scores obtained for these diseases, i.e. 49.4 and 65.5% respectively, which contributed to the relatively high standard deviations observed across all different minimal models. For illustration, Table 14.2 shows the results of both general and minimal logistic models for Schistosomiasis with the contribution of the different significant factors for explaining the presence or absence of this disease across countries.

14.3.1.2 *Artificial Neuronal Nets*

Table 14.1 shows results obtained from artificial neuronal networks with the total set of 20 inputs for the 15 diseases using jack-knife procedure. As previously observed for logistic regression (see above), the percentages of good classification scores were very high, ranging from 88.9 to 100% (mean 96.3%, $SD = \pm 3.6$) when all input variables were kept in models. True absence and true presence scores varied from 0 to 100% (mean 90.3%, $SD = \pm 25.3$) and from 0 to 100% (mean 80.6%, $SD = \pm 29.7$), respectively. Three diseases (Cutaneous and Visceral Leishmaniosis and Hepatitis A) strongly affected classification scores in that true positive performances for the two Leishmaniosis, i.e. 20% and 0% respectively, and true negative performances for Hepatitis A, i.e. 0%, were extremely low. Moreover, predictions obtained with minimal neuronal networks formed with only 3 input parameters, i.e. the 3 human life-history traits, performed nearly as well as global nets to model the disease occupancy per country: 79.2 to 98.2% (mean 92.7%, $SD = \pm 5.5$) of total good prediction scores, 0 to 100% (mean 82.8%,

Table 14.1. Results obtained from both logistic regressions and neuronal nets for predicting the presence/absence of 15 infectious and parasitic diseases across a set of 153 countries when all independent parameters are kept into models. True positive and true negative estimates with their respective percentage values and overall good classification percentage values were derived from leave-one-out procedure. Data show that both methods converge in efficiently predicting the occurrence of the different diseases across countries, with a slight higher performance of good achievement of prediction for logistic regression procedure (but see Tables 14.2 and 14.3). See text for further explanation on statistics. C. Leishmaniosis refers to Cutaneous Leishmaniosis and V. Leishmaniosis to Visceral Leishmaniosis.

Model	Logistic								Neuronal Net				
Disease	True Negatives	%	True Positives	%	Overall Performance	$-2LL$	R	p	True Negatives	%	True Positives	%	Overall Performance
Chagas	136	100	17	100	100	0.0	1.00	<0.0001	136	100	16	94.1	99.3
Cholera	87	90.6	51	89.5	90.2	77.6	0.762	<0.0001	87	90.6	49	85.9	88.9
C. Leishmaniosis	148	100	5	100	100	>100	1.00	<0.0001	148	100	1	20.0	97.4
Dengue fever	87	94.6	52	85.2	90.8	57.1	0.840	<0.0001	86	93.5	51	83.6	89.5
Dracunculosis	133	99.2	18	94.7	98.7	13.5	0.917	<0.0001	134	100	17	80.5	98.7
Filariosis	90	94.7	53	91.4	93.5	58.6	0.832	<0.0001	90	94.7	53	91.4	93.5
Hepatitis A	1	100	152	100	100	0.0	1.00	ns	0	0.0	152	100	99.3
Hepatitis B	3	100	150	100	100	0.0	1.00	ns	3	100	150	100	100
Lyme	142	100	11	100	100	0.0	1.00	<0.0001	142	100	9	81.8	98.7
Malaria	49	96.1	101	99.0	98.0	26.4	0.927	<0.0001	49	96.1	99	97.1	96.7
Meningococcosis	107	98.2	39	88.6	95.4	31.3	0.902	<0.0001	106	97.2	37	84.1	93.5
Schistosomiasis	91	95.8	53	91.4	94.1	43.2	0.882	<0.0001	92	96.8	54	93.0	94.1
Trypanosomiasis	81	100	72	100	100	0.0	1.00	<0.0001	78	96.3	72	100	98.0
V. Leishmaniosis	150	100	3	100	100	0.0	1.00	ns	150	100	0	0.0	98.1
Yellow fever	114	99.1	37	97.4	98.7	20.5	0.931	<0.0001	114	99.1	37	97.4	98.7
Mean		91.9		95.8	97.3					90.3		80.6	96.3
SD		2.9		5.2	3.5					25.3		29.7	3.6

Table 14.2. Results of logistic regression procedure for modelling the effect of continuous and discrete independent variables on one discrete dependent variable such as the Schistosomiasis occurrence (presence or absence) across the 153 countries. Table illustrates the results obtained in a global linear model (**a**) and after a backward stepwise elimination procedure (**b**). Both models converge in that the same set of explanatory variables is retained. Estimation parameters for both models are highly significant (Model a: $-2LL = 43.169$, Goodness of fit = 49.917, Nagelkerke $R^2 = 0.882$, Chi-square model = 159.897, $d.f. = 20$, $p = 0.0000$, Overall classification = 94.12%, true negative score = 95.79%, true positive score = 91.38% with a cut-off value of 0.50; Model b: $-2LL = 52.007$, Goodness of fit = 70.815, Nagelkerke $R^2 = 0.854$, Chi-square model = 151.059, $d.f. = 7$, $p = 0.0000$, Overall classification = 93.46%, true negative score = 96.84%, true positive score = 87.83% with a cut-off value of 0.50). Details on statistics are given in the text

a **Independent factors**	*B*	**Wald's**	*d.f.*	*p*	*R*
Spatial characteristics					
Surface area	<0.0001	1.3049	1	ns	0.0000
Mean latitude	−0.1447	3.8996	1	0.0483	−0.0967
Mean longitude	−0.0590	2.8664	1	0.0904	−0.0653
Land mass vs. island	−0.2281	0.0162	1	ns	0.0000
North/South	1.2638	0.6156	1	ns	0.0000
Human traits					
Fertility rate	0.8045	3.0667	1	0.0799	0.0725
Female life span	0.6544	2.8330	1	0.0923	0.0640
Male life span	−0.8534	3.2046	1	0.0734	−0.0770
Demographic characteristics					
Population growth	0.3337	0.9190	1	ns	0.0000
Population density	−0.0005	0.0725	1	ns	0.0000
Death rate	−2.1056	0.7520	1	ns	0.0000
GNP	−0.0003	1.5130	1	ns	0.0000
Total population	<0.0001	0.0234	1	ns	0.0000
Historical patterns					
Dummy Group I	−0.8507	0.2568	1	ns	0.0000
Dummy Group II	0.0389	0.0002	1	ns	0.0000
Dummy Group III	8.0851	0.0009	1	ns	0.0000
Dummy Group IV	5.8435	0.0044	1	ns	0.0000
Dummy Group V	7.1893	0.0071	1	ns	0.0000
Dummy Group VI	13.8095	0.0141	1	ns	0.0000
Dummy Group VII	12.2174	0.0308	1	ns	0.0000
Constant	−34.3221	0.0110	1	ns	

$SD = \pm 30.5$) of true negative scores and 16.7 to 100% (mean 73.1%, $SD = \pm 28.8$) of true positive scores (Table 14.3). Except for the lowest negative scores obtained for the two Hepatitis A and B, and the lowest positive scores obtained for the Chagas disease and

Table 14.2. *Continued*

b Variables in the equation	*B*	Wald's	*d.f.*	*p*	*R*
Spatial characteristics					
Surface area	<0.0001	4.5876	1	0.0322	0.1129
Mean latitude	−0.1917	14.1435	1	0.0002	−0.2445
Mean longitude	−0.0547	11.4365	1	0.0007	−0.2156
Human traits					
Fertility rate	0.5692	5.2436	1	0.0220	0.1264
Demographic characteristics					
Population growth	0.6483	7.3891	1	0.0066	0.1629
Historical patterns					
Dummy Group VI	11.9461	0.0257	1	ns	0.0000
Dummy Group VII	11.3130	0.0756	1	ns	0.0000
Constant	−26.5672	0.0972	1	ns	

Model if constant removed	**Log likelihood**	**2 log *LR***	***d.f.***	***p***
Spatial characteristics				
Surface area	28.316	4.625	1	0.0315
Mean latitude	41.055	30.103	1	0.0000
Mean longitude	38.099	24.191	1	0.0000
Human traits				
Fertility rate	−29.077	6.148	1	0.0132
Demographic characteristics				
Population growth	−31.112	10.216	1	0.0014
Historical patterns				
Dummy Group VI	30.977	9.947	1	0.0016
Dummy Group VII	34.949	17.890	1	0.0000

the two Cutaneous and Visceral Leishmaniosis, we obtained similar prediction scores compared to neuronal net models with the 20 input parameters (see Table 14.1). These findings particularly demonstrate that minimal neuronal nets perform as well as general models to predict and model the distribution of those infectious diseases. In addition, all these results corroborate those obtained using logistic regressions. Again, artificial neuronal nets, as well as logistic regressions, were not influenced, or slightly perturbed, by a presence/absence effect for prediction of true negative plots versus true positive plots.

Table 14.3. Results obtained with artificial neuronal network modelling for predicting the presence/absence of the 15 infectious and parasitic diseases across the total set of 168 countries with only the three human-life history traits, i.e. fertility and life spans for both sexes. Since historical information was excluded from this analysis, it was possible to use the total number of countries available in our database. True positive and negative scores with their respective percentage values and overall good classification percentage values were derived from leave-one-out procedure (see text). Data show that modelling efficiently predicts the occurrence of most of diseases solely using human characteristics as input parameters

Disease	True negatives	(%)	True positives	(%)	Prediction performance
Chagas	147	100	5	23.8	94.4
Cholera	97	91.5	48	77.4	86.3
Cut. Leishmaniosis	162	100	1	16.7	97.0
Dengue fever	90	90.9	60	87.0	89.3
Dracunculosis	149	100	16	84.2	98.2
Filariosis	90	91.5	46	74.2	79.2
Hepatitis A	0	0.0	165	100	98.2
Hepatitis B	1	20.0	163	100	97.6
Lyme	153	98.1	7	58.3	95.2
Malaria	43	75.4	107	96.4	89.3
Meningococcosis	120	97.6	35	77.8	92.3
Schistosomiasis	102	95.3	51	83.6	91.1
Trypanosomiasis	128	96.2	33	94.3	95.8
Visc. Leishmaniosis	164	100	1	25.0	98.2
Yellow fever	107	84.9	41	97.6	88.1
Mean		82.8		73.1	92.7
$\pm SD$		30.5		28.8	5.5

Hold-out modelling gave similar results of good recognition patterns. Table 14.4 shows results of the ten tests for two diseases, Schistosomiasis and Yellow fever. Scores of overall correctly classified countries were between 81.6 and 94.7% (mean 87.6%, $SD = \pm 3.73$) for Yellow fever disease, and 78.9 to 92.1% (mean 85.3%, $SD = \pm 4.86$) for Schistosomiasis.

14.3.2 Occurrence and Threshold Effects

14.3.2.1 *Occurrence Effects*

As illustrated in Fig. 14.3, we did not observe any occurrence effect (range values from 25 to 93% for the 15 parasite species across the 153 countries) on sensitivity/1–specificity ratios for both neuronal network models (Fig. 14.3a) and logistic regressions (Fig. 14.3b). This means that, at least in this study, infectious and parasitic diseases

Table 14.4. Results of artificial neuronal network modelling after the partitioning of the data set into a training set (¾ of countries) and a test set (¼ of countries), or hold-out procedure. Table shows results of the ten tests for two diseases, i.e. Schistosomiasis (**a**) and Yellow fever (**b**). Also given are mean values of the ten trials and their corresponding standard deviations

a	**True negatives**	**(%)**	**True positives**	**(%)**	**Prediction performance**
Trial 1	17	80.9	15	88.2	84.2
Trial 2	21	91.3	10	66.7	81.6
Trial 3	20	87.0	10	66.7	78.9
Trial 4	22	91.7	13	92.8	92.1
Trial 5	21	87.5	12	85.7	86.8
Trial 6	19	86.3	14	87.5	86.8
Trial 7	22	95.6	13	86.7	92.1
Trial 8	18	78.3	12	80.0	78.9
Trial 9	19	86.4	15	93.7	89.5
Trial 10	21	87.5	11	78.6	84.2
Mean	20	87.3	12	82.7	85.3
SD	1.70	5.03	1.84	9.65	4.86

b	**True negatives**	**(%)**	**True positives**	**(%)**	**Prediction performance**
Trial 1	21	87.5	11	78.6	84.2
Trial 2	19	79.2	12	85.7	81.6
Trial 3	21	87.5	12	85.7	86.8
Trial 4	23	92.0	11	84.6	89.5
Trial 5	20	83.3	13	92.9	86.8
Trial 6	22	91.7	12	85.7	89.5
Trial 7	23	95.8	13	92.9	94.7
Trial 8	19	82.6	13	86.7	84.2
Trial 9	20	90.9	14	87.5	89.5
Trial 10	22	88.0	12	92.3	89.5
Mean	21	87.9	12	87.3	87.6
SD	1.49	5.03	0.95	4.46	3.73

having a very low spatial occupancy on earth, i.e. endemic species with a very limited area (e.g. Chagas disease restricted to some countries of Southern America), may be as best predicted as widespread diseases since the occurrence effect plays a poor or a very slight role when considering all the predictive variables (inputs) into modelling. Nevertheless, minimal models built with the 3 human life-history traits only were more sensitive to variation in prevalence values (see Table 14.3). Considering minimal models, both logistic regressions and neuronal nets were affected by those variations in

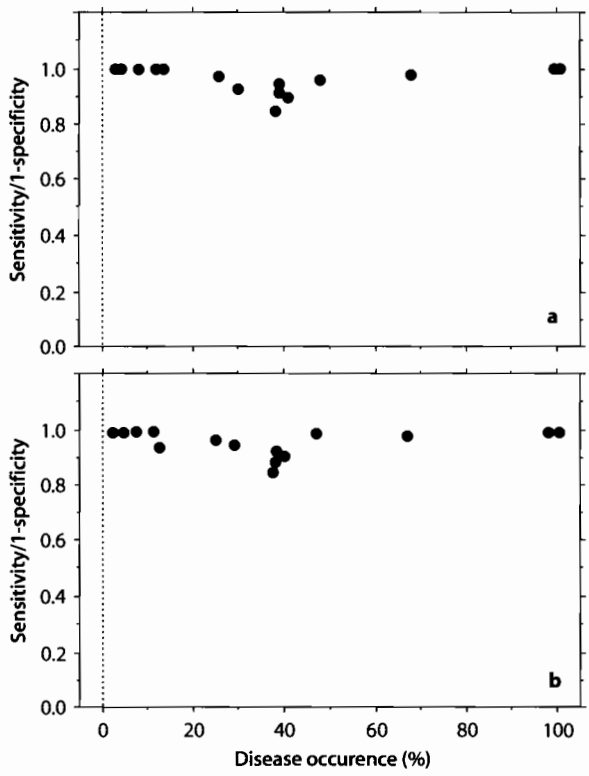

Fig. 14.3. Sensitivity/1-specificity ratios against occurrence values for the 15 infectious and parasitic diseases: **a** scatter diagram obtained using artificial neuronal network procedure; **b** scatter diagram obtained after logistic regression method. Both statistical techniques converge in that sensitivity/1-specificity profiles are not or very slightly influenced by occurrence effect (see text for further details)

prevalence values, and we found both low prevalence and high prevalence effects on classification scores. This can be observed in Table 14.3 for the 5 diseases, i.e. the Hepatitis A and B (high prevalence effect observed on true negative scores), the Cutaneous and Visceral Leishmaniosis and the Lyme disease (low prevalence effect on true positive scores). However, the lowest sensitivity/1–specificity ratios were observed for the average values of prevalence, i.e. diseases having a prevalence value of around 35–40%, i.e. the Cholera and Dengue fevers and the two helminthic Schistosomiasis and Filariosis diseases. We did not find any special explanations to these findings. These results conflict in part with previous studies showing low prevalence values of an event which might be responsible for better prediction of true absences than true presences (see Manel et al. 1999).

14.3.2.2 *Threshold Effects*

Figure 14.4 plots the relationship between sensitivity values (i.e. the true positive fraction) against 1-specificity values (i.e. the false positive fraction) across different thresholds of cut-off-values obtained by jack-knife procedures for both classifying methods. The procedure was repeated for each of the 15 infectious diseases. These results show that a sufficiently high rate of true positives and a low rate of false positives may

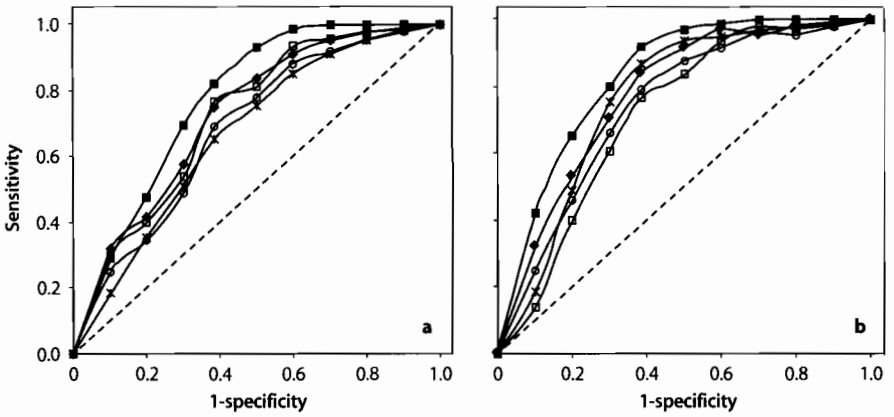

Fig. 14.4. Received operating characteristic curves; **a** artificial neuronal networks; **b** logistic regressions illustrating the relationship between sensitivity values (i.e. the true positive fraction) against 1-specificity values (i.e. the false positive fraction) across different thresholds of cutoff-values obtained by jack-knife procedures. These results show that a sufficiently high rate of true positives and a low rate of false positives may be achieved at a threshold probability of 0.5 for both logistic regression and neuronal net methods. Curve behaviours illustrate the general tendency observed across the 15 different diseases. The *dotted line* corresponds to x equals y

be achieved by both neuronal net (Fig. 14.4a) and logistic regression (Fig. 14.4b) methods. However, sensitivity estimates increase sharply to larger values for logistic regression than for neuronal nets, which means that the former method more correctly classifies new cases irrespective of the threshold value of probability to accept presence when compared to the latter. Artificial neuronal network procedure was more sensitive in that it classified false presence with a high probability of event. Our present findings correspond to those obtained by Manel et al. (1999) in that we observed a general but slight tendency for better performance in logistic regressions for classifying countries across distinct threshold probabilities than neuronal nets did.

14.4 Discussion

There have been very scarce comparison studies of the ability of logistic regressions and neuronal networks to discriminate between the presence and the absence of an event (de Garine-Wichatitsky et al. 1999). We provide in this work an attempt to compare the ability of these two methods to make accurate predictions of the spatial occupancy and occurrence for 15 human diseases, some of which can be considered as endemic on a continent, e.g. Chagas disease, and some others have a widespread distribution world-wide, e.g. Hepatitis A and B. Logistic regressions have been widely used and proposed for several applications relevant to ecological and evolutionary investigations (Schoener and Adler 1991; Trexler and Travis 1993; Veltman et al. 1996; Gaston 1998; Sorci et al. 1998). Several texts provide informative discussions on logistic regressions (McCullagh and Nelder 1989; Norusis 1997). Unfortunately, very little attention has been paid to comparing the efficiency of these two methods, and we wish to aid

in this present investigation the development of a comparative analysis of statistical techniques when an event is expected to occur or not.

First, defining the precise spatial distribution of parasitic and infectious diseases is of prime interest, but it remains difficult. On smallest spatial scales ecological variability and temporal changes make disease distributions not easily definable in space (Sutherst 1998). Many infectious diseases have their distribution in time, which waxes or wanes with the natural periodicity of events. This is particularly true for Malaria distribution in Sub-Saharan Africa, for instance (Craig et al. 1999).

We have demonstrated in this work that simple models may be used to predict the actual distribution of different diseases on a global scale. The data sets and the two methodological approaches we used are entirely relevant at this largest scale, but they do not consider small-scale anomalies that evidently affect distribution, such as local extinction of vectors, arid zones, deforestation, etc. As preconised by Craig et al. (1999), the modelling of disease distribution may be viewed as a four-tier approach: *(1)* the first level, at a large scale, defines the broad distribution of diseases (this work); *(2)* the second level, at a southcontinental scale, takes into account differences between ecological zones; *(3)* the third level, at a regional or national scale, defines the transmission intensity within a given zone of transmission ecology, such as perennial, seasonal or bi-seasonal transmission; *(4)* the fourth level, at a local scale of say several km-squares, which allows more detailed precisions. On largest scales, prediction of a disease in terms of presence according to very global environmental and biological descriptors is highly appreciable. Interestingly, such models are entirely relevant at smallest scales, since the inclusion of other smaller-scale data sets should permit more detailed predictions.

Second, logistic regression procedure tends to be the method of choice for classifying an event of presence or absence, simply because it is used widely and is generally understood. It is worth noting here that logistic regressions gave similar results to artificial neuronal nets. In addition, logistic regressions were less time-consuming than neuronal nets. On the practical side, logistic regressions generally provide an information about the importance of predicting factors on the likelihood of an event. Astonishingly, we did not find any major effect of occurrence, i.e. the proportion of absence scores on presence scores, on classification using general models. Contrastingly, minimal models for both methods were more sensitive to the occurrence effect. One possible explanation is that in both statistical techniques, data are assumed to implicitly contain the information necessary to establish the relation, an assumption more probable when performing a general model than in a minimal model. The very high scores of classification across the 15 infectious and parasitic diseases on a global scale we obtained with general models probably reveal the fact that there is good reason for using these factors to indicate the probable presence of a given disease, at least at this scale of investigation. In addition, the examples we show confirm that human life-history traits allow very good predictive scores of presence of a disease. This may be particularly true when faced with virulent parasites, hosts should adjust their reproductive biology by increasing reproductive output and/or reducing age at maturity as suggested by theory. Thus, these findings would tend to show that some diseases, e.g. schistosomiasis, responsible for high incidences of morbidity and mortality in human populations, might be associated with their host life history characters, and vice versa. Although it is well known that some diseases cause severe impacts

on human populations, they do not necessary imply a subtle response in the adjustment of human life-history traits. Prediction does not demonstrate causation between the presence of a given disease and shifts in life-history parameters. According to Møller (1998), parasites should impose stronger selection pressures on their hosts in the tropics compared to nontropical climatic zones, and thus parasite constraints on man should be stronger around the Equator line. Three alternative hypotheses could explain our findings: *(1)* both parasite species occurrence and human characteristics may be related to a third surrogate variable, or group of surrogate variables, not encountered into our models, which would act simultaneously on both these parameters, *(2)* parasite impacts should select for optimal human responses, or *(3)* the presence of one disease, or one group of co-occurring diseases, might be determined by humankind. As of present, we cannot evaluate the relative importance of these three mechanisms, but just apprehend them by modelling. Nevertheless, we demonstrate that human-parasite biological systems are probably characterized by nonlinear spatially extended relationships, which might reflect the connectedness that would exist in real communities between man and his infectious and parasitic diseases.

14.5 Conclusion

It is remarkable to see how such simple predictive models approximate the actual distribution of the different infectious and parasitic diseases across the world so well. Adopting a wider perspective, i.e. macroepidemiology, we show here that this modelling can be easily repeated and manipulated in combination with other new available data sets and combining data from multiple sources. Collations of data for infectious and parasitic diseases on the field are numerous, but the modelling of diseases is still in its infancy. Probably, more global data sets including various abiotic and biotic parameters are needed in conjunction with geographical information systems. Thus, we view the modelling of infectious and parasitic diseases as a promising avenue of research.

Our results suggest the existence of nonlinear laws for the spatial distribution of some diseases and human life-history traits. Many factors determine why an area in a particular region harbours a disease (or not). Classical theories of biogeographers explain the spatial distribution of diseases by cultural, geographical and socio-economical factors, with southtropical areas harbouring the bulk of parasitic and infectious diseases on earth (Anderson and May 1991). Because a wide array of factors may contribute to the actual variation in space of human diseases, it is actually difficult to disentangle the respective effect of the different variables involved. However, our findings tend to show the existence of a correlation between diseases and human characteristics, with fertility being the most important factor.

In free-living organisms, there is evidence of rare-common species differences (see Kunin and Gaston 1993; Gaston et al. 1998; Blackburn et al. 1998). Widespread species tend, on average, to be locally more abundant than rare restricted species at large spatial scales. A positive interspecific (in space) and intraspecific (in time) relationship arises because common species are abundant at some sites, whereas rare species are rare at all sites. Unfortunately, little is known about parasitic and infectious diseases. Restricted parasite endemics, e.g. Chagas disease, may probably differ in its eco-

logical interactions with hosts, and/or in its population genetics to a more similar widespread disease. Such differences might result from evolved adaptations to the condition of rarity, or, on the contrary, not all species might be equally armed to become rare. Extensive data sets are now available on parasitic and infectious diseases to be analysed in this way. Combining different information about genetics, physiology, life history, epidemiology etc., may allow us to address some unresolved issues in public health. We are developing in our laboratory such methodological approaches to integrate a vast amount of data (see Rapport et al. 1998) on some core research projects, i.e. Chagas disease, trypanosomiasis, paludism and liver-fluke.

Changes in attribute data entered into models are very easy to deal with. In some ways, it is possible to modify a variable, or a group of values within a variable, allowing comparison of trends before and after modification, and to observe effects on prediction of disease occurrences across countries. For instance, this would be possible with population density, death rate, GNP data or fertility rate values to detect declines in site occupancy for a disease, synchronous declines for a group of co-occurring diseases, or on the contrary dispersion of some others. In addition, climate change, deforestation and desertification scenarios, which are highly probable in the near future, may provide an immediate response on some spatial disease occupancies. Lattice illustrations using Kohonen (1984) mapping may be ideal for representing human disease distributions in space and time, for detecting regions or areas at risk of new invasions by infective agents (work in progress). We feel that such analysis methods are not yet particularly well developed, and they are completely general to answer some urgent problems the earth is facing within the very near future, such as emerging and resurging diseases (Gratz 1999)!

Acknowledgements

This work was funded by I.R.D. (J.F.G.) and C.N.R.S. (F.T., T.d.M., S.L. and F.R.). We thank the staff of C.E.P.M. at Montpellier (France), two anonymous referees for comments on an earlier draft of this work, and the thousands of field epidemiologists who have carried out the disease census-taking.

References

Alexander HM (1987) Pollinator limitation in a population of *Silene alba* infected by the anter-smut fungus *Ustilago violacea*. J Ecol 75:771–780

Anderson RM, May RM (1991) Infectious diseases of humans: Dynamics and control. Oxford University Press, Oxford

Ashford RW (1991) The human parasite fauna: towards an analysis and interpretation. Ann Trop Med Parasitol 85:189–198

Blackburn TM, Gaston KJ, Greenwood JJD, Gregory RD (1998) The anatomy of the interspecific abundance-range size relationship for the British avifauna: II. Temporal dynamics. Ecology Lett 1:47–55

Brooke M de L, Davies NB, Noble DG (1998) Rapid decline of host defences in response to reduced cuckoo parasitism: Behavioural flexibility of reed warblers in a changing world. Proc R Soc Lond B 265:1277–1282

Brown JH (1995) Macroecology. University of Chicago Press, Chicago

Cavalli-Sforza LL (1997) Genes, peoples, and languages. Proc Natl Acad Sci USA 94:7719–7724

Combes C (1995) Interactions durables: Ecologie et evolution du parasitisme. Masson Ed., Paris

Craig MH, Snow RW, le Sueur D (1999) A climate-based distribution model of malaria transmission in Sub-Saharan Africa. Parasitology Today 15:105–111

Edwards M, Morse DR (1995) The potential for computer-aided identification in biodiversity research. Tree 10:153–158

Efron B (1983) Estimating the error rate of a prediction rule: Improvement on cross-validation. J Am Stat Ass 78:316–330
Esch GW, Bush AO, Aho JM (1990) Parasite communities: Patterns and processes. Chapman and Hall Ltd., London
Ewald PW (1994) Evolution of infectious diseases. Oxford University Press, Oxford
Fielding AH, Bell JF (1997) A review of methods for assessment of prediction errors in conservation presence/absence models. Env Cons 24:38–49
Forbes MRL (1993) Parasitism and host reproductive effort. Oikos 67:444–450
Freeman JA, Skapura DM (1992) Neuronal networks: Algorithms, applications and programming techniques. Addison-Wesley Publishing Company, Reading
Friedman JH (1997) On bias, variance, o/1-loss and the curse-of-dimensionality. Data Mining and Knowledge Discovery 1:55–77
Gallant SI (1993) Neuronal network learning and expert systems. The MIT Press, Cambridge
Garine-Wichatitsky M de, Meeûs T de, Guégan JF, Renaud F (1999) Spatial and temporal distributions of parasites: Can wild and domestic ungulates avoid African tick larvae? Parasitology 119:455–466
Gaston KJ (1998) Species-range size distributions, extinction and transformation. Philos Trans R Soc Lond B 353:219–230
Gaston KJ, Blackburn TM, Gregory RD, Greenwood JJD (1998) The anatomy of the interspecific abundance-range size relationship for the British avifauna: I. Spatial patterns. Ecology Lett 1:38–46
Grafen A (1992) The uniqueness of the phylogenetic regression. J Theor Biol 156:405–423
Gratz NG (1999) Emerging and resurging vector-borne diseases. Annu Rev Entomol 44:51–75
Harvey PH, Pagel MD (1991) The comparative method in evolutionary biology. Oxford University Press, Oxford
Hochberg ME, Michalakis Y, de Meeüs T (1992) Parasitism as a constraint on the rate of life-history evolution. J Evol Biol 5:491–504
Immerman RS (1986) Sexually transmitted disease and human evolution: survival of the ugliest. Hum Ethol Newsletter 4:6–7
Jones H (1990) Population geography. Paul Chapman Publishing Ltd., London
Jongman RHG, ter Braak CJF, van Tongeren OFR (1995) Data analysis in community and landscape ecology. Cambridge University Press, Cambridge
Kohavi R (1995) A study of cross-validation and bootstrap for estimation and model selection. Proceedings of the 14th International Joint Conference on Artificial Intelligence, Morgan Kaufman Publishers Inc., pp 1137–1143
Kohonen T (1984) Self-organization and associative memory. Springer-Verlag, Berlin
Kris AC, Lively CM (1998) Experimental exposure of juvenile snails (*Potamopyrgus antipodarum*) to infection by trematode larvae (*Microphallus* sp.): Infectivity, fecundity compensation and growth. Oecologia 116:575–582
Kunin WM, Gaston KJ (1993) The biology of rarity: Patterns, causes and consequences. Tree 8:298–301
Lafferty KD (1993) The marine snail, *Cerithidea californica*, matures at smaller sizes where parasitism is high. Oikos 68:3–11
Lively CM (1987) Evidence from a New Zealand snail for the maintenance of sex by parasitism. Nature 328:519–521
Manel S, Dias JM, Ormerod SJ (1999) Comparing discriminant analysis, neuronal networks and logistic regression for predicting species distributions: a case study with a Himalayan river bird. Ecol Model 120:337–347
Martin D (1996) Depicting changing distributions through surface estimations. In: Longley P, Batty M (eds) Spatial analysis: Modelling in a GIS environment. GeoInformation International, Cambridge
Martins EP (1996) Phylogenies and the comparative method in animal behavior. Oxford University Press, Oxford
Martins EP, Hansen TF (1997) Phylogenies and the comparative method: A general approach to incorporating phylogenetic information. Am Nat 149:646–667
McCullagh P, Nelder JA (1989). Generalized linear models. Chapman and Hall Ltd., London
McNamara JM, Houston AI (1996) State-dependent life histories. Nature 380:215–221
Michalakis Y, Hochberg ME (1994) Parasitic effects on host life-history traits: A review of recent studies. Parasite 1:291–294
Møller AP (1997) Parasitism and the evolution of host life history. In: Clayton DH, Moore J (eds) Host-parasite evolution: General principles and avian models. Oxford University Press, Oxford
Møller AP (1998) Evidence of larger impact of parasites on hosts in the tropics: Investment in immune function within and outside the tropics. Oikos 82:265–270
Norusis MJ (1997) SPSS for Windows. Advanced Statistics 7.5., Chicago

Oberdorff T, Hugueny B, Compin A, Belkesam D (1998) Non-interactive fish communities in coastal streams of the North-Western France. J Anim Ecol 67:472–484
Pedersen M (1995) Interactive and animated cartography. Prentice Hall Int., Upper Saddle River, N.J.
Petney TN, Andrews RH (1998) Multiparasite communities in animals and humans: Frequency, structure and pathogenic significance. Int J Parasitol 28:377–393
Rapport DJ, Costanza R, McMichael AJ (1998) Assessing ecosystem health. Tree 13:397–402
Reeson AF, Wilson K, Gunn A, Hails RS, Goulson D (1998) Baculovirus resistance in the noctuid *Spodoptera exempta* is phenotypically plastic and responds to population density. Proc R Soc Lond B 265:1787–1791
Rhodes CJ, Anderson RM (1996) Power laws governing epidemics in isolated populations. Nature 381:600–602
Rosenzweig ML (1995) Species diversity in space and time. Cambridge University Press, Cambridge
Rumelhart DE, Hinton GE, Williams RJ (1986) Learning representations by back-propagating errors. Nature 323:533–536
Schoener TW, Adler GH (1991) Greater resolution of distributional complementarities by controlling for habitat affinities: A study with Bahamian lizards and birds. Am. Nat. 137:669–692
Sheldon AL, Meffe GK (1995) Path analysis of collective properties and habitat relationships of fish assemblages in coastal plain streams. Can J Fish Aquat Sci 52:23–33
Shykoff JA, Bucheli E, Kaltz O (1996) Flower lifespan and disease risk. Nature 379:779
Sorci G, Clobert J, Michalakis Y (1996) Cost of reproduction and cost of parasitism in the common lizard, *Lacerta vivipara*. Oikos 76:121–130
Sorci G, Møller AP, Clobert J (1998) Plumage dichromatism of birds predicts introduction sucess in New Zealand. J Anim Ecol 67:263–269
Stearns SC (1992) The evolution of life-histories. Oxford University Press, Oxford
Sutherst RW (1998) Implications of global change and climate variability for vector-borne diseases: generic approaches to impact assessments. Int J Parasitol 28:935–945
Teriokhin AT (1998) Evolutionary optimal age of repair: Computer modelling of energy partition between current and future survival and reproduction. Evolutionary Ecology 12:291–307
Teriokhin AT, Budilova EV (2000) Evolutionary optimal networks for controlling energy allocation to reproduction and repair in men and women. (see Chapter 15 of this book)
Trexler JC, Travis J (1993) Nontraditional regression analyses. Ecology 74:1629–1637
Veltman CJ, Nee S, Crawley MJ (1996) Correlates of introduction success in exotic New Zealand birds. Am Nat 147:542–557
Whittaker RH (1975) Communities and ecosystems. Macmillan, New York
Zar JH (1996) Biostatistical analysis. Prentice Hall Int., Upper Saddle River, N.J.
Zweig MH, Campbell G (1993) Receiver-operating characteristic (ROC) plots: A fundamental tool in clinical medicine. Clin Chem 39:561–577

Guégan Jean-François, Thomas F., Meeüs T. de, Lek S., Renaud F. (2000).

The macroepidemiology of parasitic and infectious diseases : a comparative study using artificial neuronal nets and logistic regressions.

In : Lek S. (ed.), Guégan Jean-François (ed.). Artificial neuronal networks : application to ecology and evolution.

Berlin : Springer, 203-224.

(Environmental Science). Applications of Artificial Neural Networks to Ecological Modelling : Workshop, Toulouse (FRA), 1998/12.

ISBN 3-540-66921-3