

Using artificial neural networks to assess wolf distribution patterns in Portugal

Carmen Bessa-Gomes^{1,2} and Francisco Petrucci-Fonseca^{1,3}

¹Grupo Lobo, Departamento de Zoologia e Antropologia, Edifício C2, Campo Grande, 1749–016 Lisboa, Portugal

²Departamento de Matemática, Instituto Superior de Agronomia, Tapada da Ajuda, 1349–017 Lisboa, Portugal

³Centro de Biologia Ambiental, Edifício C2, Campo Grande, 1749–016 Lisboa, Portugal

(Received 30 July 2002; accepted 18 December 2002)

Abstract

Although the conservation of endangered species often implies the definition of priority areas for conservation, detailed information on their distribution patterns is seldom available over a large geographic range. The present paper explores the performance of an alternative data analysis approach, artificial neural networks, for assessing distribution patterns of endangered mammals when data are scarce and noisy. This approach was applied to identify wolf occupancy in Portugal based on information on wolf depredation and inquiries. Artificial neural networks were able to discriminate successfully between areas sporadically used by vagrant individuals and areas occupied by resident wolves, with a low estimated prediction error (around 1%). Only 33% of the wolf range is regularly occupied, being fragmented in five nuclei. These nuclei are surrounded by more disturbed marginal areas, with a source–sink dynamic between nucleus and marginal areas. Selection of priority areas for wolf conservation in the context of the Natura 2000 Network increased twofold the proportion of wolf regular range within protected areas. The good performance of artificial neural networks in assessing wolf distribution patterns suggests that this approach may be applied to other species where detailed records of distribution are limited.

INTRODUCTION

Conservation biology has frequently been described as a crisis discipline (Soulé, 1986), whose emergency character stems not only from the impending biodiversity crisis, but also from the limited availability of resources and knowledge necessary to avert it. More often than not, wildlife managers are forced to envisage conservation plans for endangered species based on limited biological information. In particular, although fundamental for the establishment of conservation strategies, detailed information on a population spatial structure is seldom available over large geographic ranges. This is typified in large carnivores, where the available information is frequently scarce and noisy, or very localized in space and/or time (Schonewald-Cox, Azari & Blumer, 1991), and conventional data analysis fails to identify the areas of occupancy, offering little information for managers to design conservation strategies. Thus, there is an urgent need to develop alternative analytical approaches

which can maximize the information extracted from available data.

The conflict between data availability and conservation needs is well illustrated by the selection of priority areas for wolf (*Canis lupus*) conservation in Portugal. The wolf distribution in the Iberian Peninsula seriously declined during the twentieth century (Fig. 1). In Portugal, wolves have progressively disappeared from littoral, south and central regions of the country (Petrucci-Fonseca, 1990; Bessa-Gomes, 1996). In addition, several studies suggest that the remaining distribution is fragmented into several main nuclei (Petrucci-Fonseca, 1990; Moreira, 1992; Álvares, 1995; Carreira, 1996). Wolf population decline in Portugal is associated primarily with the scarcity (and often absence) of natural prey, and to conflict with local people because of livestock depredation. In fact, wolves in Portugal depend largely on livestock, which may constitute up to 80% of their diet (Petrucci-Fonseca, 1990; Moreira, 1992; Álvares, 1995; Carreira, 1996).

It was only in 1990 that the wolf was classified as an endangered species in Portugal, and was given complete protection (SNPRCN, 1990). Since then, the Instituto de Conservação da Natureza, a governmental agency for nature conservancy, compensates losses due to wolf depredation in order to minimise the conflict with people

All correspondence to: Carmen Bessa-Gomes, Laboratoire d'écologie (CNRS-UMR 7625) Université Pierre et Marie Curie, 7 quai St Bernard, 75005 Paris, France. Tel.: ++33 1 44 27 52 04; Fax: ++33 1 44 27 35 16; E-mail: cbessa@snv.jussieu.fr.

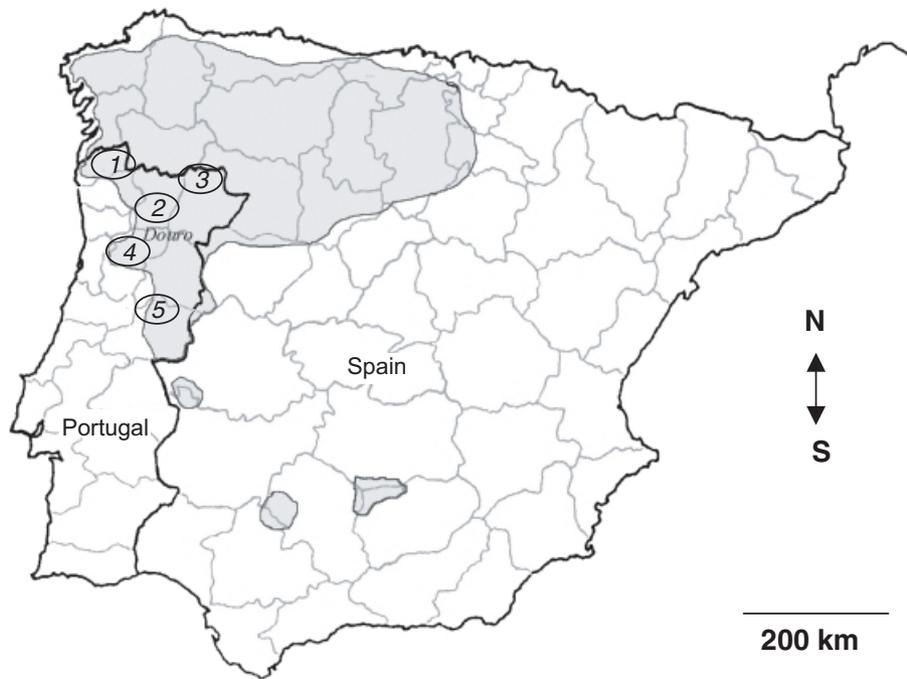


Fig. 1. Wolf distribution in the Iberian Peninsula in the 1980s (after Blanco *et al.*, 1990 and Petrucci-Fonseca, 1990). Location of ecological studies: (1) Álvares, 1995, (2) Carreira, 1996, (3) Moreira, 1992, (4) A. Alexandre and A. T. Cândido, pers. comm., (5) A. T. Cândido, pers. comm.

because of livestock damages. Nevertheless, by the mid-1990s, less than 15% of their distribution range was within protected areas (Bessa-Gomes, 1996).

The implementation of the Natura 2000 Network constituted an important opportunity to increase wolf protection in Portugal because wolves were included in the Annex II of the Habitat Directive of the European Community (92/43/EEC). Consequently, it is mandatory to identify and designate as Special Areas for Conservation sites that are important for wolf conservation. These Special Areas will make up the European Network of protected sites, Natura 2000 Network (hereafter N2000).

The establishment of priority areas for wolf conservation in Portugal requires information about wolf occupancy at a national scale. However, previous analysis using conventional techniques have been inadequate, owing to the scarcity of detailed ecological studies in areas of sporadic wolf presence (i.e. the training data set was unbalanced). As a consequence, if we ignored this inadequacy and used a conventional approach, a large proportion of the areas of sporadic wolf presence in the training data set were incorrectly classified (one fourth for logistic regression or half for discriminant analysis (Bessa-Gomes, 1996)).

In this paper we explore the performance of artificial neural networks as an alternative data analysis approach for classifying wolf presence. Data considered were collected from 1990 to 1994, and are based on information on wolf depredation on domestic animals and oral inquiries. The ability of artificial neural networks to detect patterns in multivariate data suggests that they are ideally

suitable to classification problems (Weeks & Gaston, 1997). Additionally, by making no assumptions about data distribution and by being less sensitive to noise, they may constitute a powerful data analysis tool.

ARTIFICIAL NEURAL NETWORKS AS A MODERN REGRESSION TECHNIQUE

The use of conventional linear models often proves inadequate when relationships between variables are non-linear, which is often the case in ecology. In order to overcome this problem some variables may be transformed yet, despite these manipulations, the results often remain disappointing, owing to poor predictive power (Lek *et al.*, 1995). Additionally, these models frequently assume normality and homogeneity of the variance/covariance matrices, assumptions that rarely (or never) hold true.

Artificial neural networks (hereafter ANN) were developed to model, although very crudely, the behaviour of the brain. They are parallel processing systems with the ability to 'learn' through examples and generalize from inferred patterns (Fausett, 1994). They can be described as non-linear models constituting powerful pattern recognition and data analysis tools that make no assumptions about the data (Ripley, 1994; Lek *et al.*, 1996; Weeks & Gaston, 1997). In spite of their potential the application of ANN to ecological models is still incipient. Nevertheless, there is a growing awareness of their possibilities, as can be attested by the organization of meetings on the subject in recent years (e.g. Lek & Guégan, 1999; Recknagel, 2001).

ANN have proved to be more appropriate than conventional techniques such as multivariate linear models or discriminant factorial analysis in several ecological studies (Lek *et al.*, 1996; Mastrotrillo *et al.*, 1997; Paruelo & Tomasel, 1997; but see Manel *et al.*, 1999a; Manel, Dias & Ormerod, 1999b). Some previous applications of ANN to ecological modelling include vegetation mapping using remote sensing (Lein, 1995; Paruelo & Tomasel, 1997), estimation of phytoplankton production (Scardi, 1996; Recknagel *et al.*, 1997; Scardi & Harding Jr., 1999; Scardi, 2001), estimation of food consumption by fish populations (Lek *et al.*, 1995), prediction of productivity/biomass of benthic organisms (Brey, Jarre-Teichmann & Borlich, 1996), and estimation of fish abundance (Baran *et al.*, 1996; Aoki & Komatsu, 1997; Brosse *et al.*, 1999; Lae, Lek & Moreau, 1999).

In spite of these successful applications, the use of ANNs to study terrestrial vertebrates, and particularly in conservation, is still relatively rare. However, ANNs have recently been used to model (1) species distribution as a function of habitat suitability (Himalayan river birds (Manel *et al.*, 1999a, b), New Zealand fur seals (Bradshaw *et al.*, 2002)); (2) species abundance (Northern bobwhite (Lusk, Guthery & DeMaso, 2001)); (3) individual recognition based on vocalizations (corncrake (Terry & McGregor, 2002)); (4) wildlife damage risk assessment (wild boar (Spitz & Lek, 1999), flamingos (Tourenq *et al.*, 1999; Tourenq *et al.*, 2001)).

ANN MODELS FOR CLASSIFYING WOLF PRESENCE

In this study, we aimed at developing an ANN model that would 'learn' to classify wolf presence given patterns of information in the 'Project *Signatus*' database (Petrucci-Fonseca, 1990). We have chosen to use a Feed-Forward Neural Network model (hereafter FFNN), because it provides a flexible way to generalize linear regression functions (Fausett, 1994; Ripley, 1994; Venables & Ripley, 1994; Lek & Guégan, 1999). A brief overview of feed-forward neural structure and training can be found in Lek & Guégan (1999).

The FFNN was implemented using the S Plus package 'nnet' developed by Venables & Ripley (1994), which is available from the 'StatLib-Software and extensions for the S (*Splus*) language' (<http://lib.stat.cmu.edu/S/>). This package can also be found for R on 'The Comprehensive R Archive Network' (<http://lib.stat.cmu.edu/R/CRAN/>), and the R system itself can be downloaded, free of charge, from this site.

Model structure

The FFNNs are highly structured networks of discrete processing elements, arranged in layers. Besides the input layer and output layer, which roughly correspond to the predictor and response variables of conventional statistical models, most FFNNs have an intermediate layer, named the hidden layer, that influences its predictive power and ability to generalize (Fausett, 1994). The present model

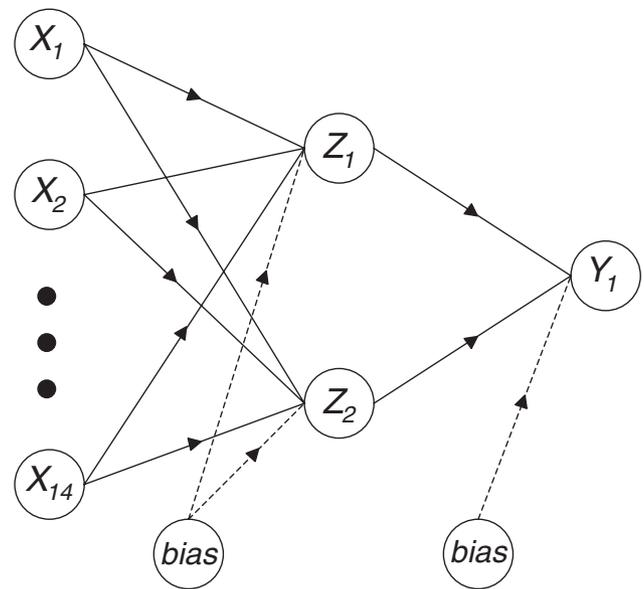


Fig. 2. Structure of the artificial neural network model. The input neurons are denoted X_i , and the output neuron is denoted Y_1 . The neurons in the hidden layer are denoted by Z_1 and Z_2 . Each connection between neurons of adjacent layers is characterised by a weight W_{ij} , with i being the index of the neuron at its origin and j the index of its terminal neuron. The 'bias' is represented by a dashed line.

aimed at discriminating between areas where wolf presence is limited to vagrant individuals and areas of occupancy (i.e. comprising resident wolves/packs), using the information in the Project *Signatus* database, and had two elements (neurons) in the hidden layer (Fig. 2).

Predictor variables: the 'Project Signatus' database

Predictor variables assessed were based on the 'Project *Signatus*' database (Petrucci-Fonseca, 1990). This database gathers a wide range of information on wolf occurrence, as well as information on environmental variables obtained from governmental sources, such as human and road density (INE, 1991a,b, 1993). The information is referenced to a UTM 10×10 km² grid, and the database management is made using a relational database and a raster geographic information system (*Microsoft Access* and *Idrisi for Windows*).

The present study used two main sources of information, namely (1) oral inquiries made to nature conservancy and game wardens, naturalists, shepherds, hunters and farmers, and (2) records of wolf attacks on domestic animals made by the Instituto de Conservação da Natureza in order to provide financial compensation.

Oral inquiries were carried out by biologists working in association with the 'Grupo Lobo', a NGO dedicated to wolf conservation in Portugal. The area surveyed during oral inquiries extends over the wolf distribution in the 1980s and the neighbouring areas of Serra da Estrela, Castelo Branco and Alto Alentejo, comprising 413

10 × 10 km grid cells. Data were checked by interviewing independent informants. As for the records of wolf attacks, every damage claim was examined by wardens in order to ensure its authenticity.

For each 10 × 10 km grid cell, we assessed the information related to wolf depredation on domestic animals, direct observations of wolves, and observation of wolf corpses per year, and summarized it in 11 variables (Table 1). Additionally three categorical variables were included as potential surrogates to regional differences on the availability of natural prey, husbandry practices, hunters behaviour and other factors (Table 1).

Information on wolf depredation on livestock was subdivided into seven variables (Table 1). Besides the total number of attacks, which may be a relative measure of wolf presence (Blanco, Cuesta & Reig, 1990; Blanco, Reig & Cuesta, 1992), we have also accounted for the proportion of attacks in each season and for their regularity. The seasonal variability of attacks may be linked to the dynamics of the packs (Álvares, 1995; Carreira, 1996). The inclusion of attack regularity accounts for the fact that dispersing and vagrant individuals may occasionally provoke large damages.

The information from oral enquiries was summarized in four variables. Besides the maximum number of wolves observed together in a given year, we also considered the maximum number of wolves observed together in October/November, as this variable has been taken to be an indicator of reproduction (Álvares, 1995). The third variable is the observation of juveniles, because pups tend to be rather conspicuous (Blanco *et al.*, 1992), and are often seen, trapped, hunted or road killed. Finally, we have also accounted for mortality in the previous or current year, because this may also affect the dynamics of packs (Álvares, 1995; Haight, Mladnoff & Wydeven, 1998).

Table 1. Predictor variables used in this study. Variables 1 to 7 are related to data on wolf attacks on domestic animals, namely the annual total per 10 × 10 grid cell, regularity (false if 95% of the attacks took place in a 4 weeks interval) and seasonality. Variables 8 and 9 are related to direct observation of wolves. The detection of juveniles may be through direct observation, or indirect cues such as observation of females with signs of lactation. Mortality refers to reports of dead wolves, mainly due to human actions. If attacks are very concentrated in time, i.e. if one single season represents more than 95% of the attacks, we considered that attacks were not regular.

Variables

- (1) Wolf attacks on domestic animals: annual total
 - (2) Regular attacks (true or false)
 - (3) Proportion of attacks per season: spring
 - (4) Proportion of attacks per season: summer
 - (5) Proportion of attacks per season: autumn
 - (6) Proportion of attacks per season: winter
 - (7) Proportion of attacks that occurred during wolves' breeding season
 - (8) Maximum number of individuals observed together
 - (9) Maximum number of individuals observed together in October/November
 - (10) Juvenile observation (true or false)
 - (11) Mortality in the current or in the previous year (true or false)
-

Categorical variables (surrogates to regional differences)

- (12) South of river Douro (true or false)
 - (13) Interior (true or false)
 - (14) Inside a protected area (true or false)
-

Response variables

In the present analysis we assumed that, given the existence of information on wolf occurrence in a given area, its presence per grid cell/year could be classified as either sporadic/marginal (i.e. used by vagrant individuals or on territory boundary), or resident when occupied by resident individuals/packs. Therefore, our response variable is binary (resident or not), and the FFNN model has only one output neuron that is either equal to 1 or to zero. In addition, we have assumed that in the absence of information, wolves could not be considered as being residents. This assumption is supported by the fact that all of the sporadic or resident areas identified by local studies corresponded to information in the database. Moreover, given the importance of wolf depredation on domestic animals and the human density over its distribution range, it is unlikely that resident wolves would go undetected by oral enquiries and damage records.

Model training

Training was performed using the back-propagation algorithm (Fausett, 1994; Venables & Ripley, 1994). The training data set was selected using studies on wolf ecology that allowed the precise identification of wolf presence at a local scale (Moreira, 1992 (2 years); Álvares, 1995 (2 years); Carreira, 1996 (2 years); A. Alexandre and A. T. Cândido pers. comm. (1 year)). Together, these studies cover the wolf distribution range in Portugal (Fig. 1), including populations both within protected areas as well as outside, north and south of the river Douro. They have characterized wolf occurrence at a local scale using a wide range of approaches, including observation of indirect indices (tracks, footprints, scats, howling playbacks), and capture and radio-tracking.

Areas occupied by resident wolves correspond to areas that are comprised within individual wolves' or packs' territories. The areas classified as sporadic/marginal presence correspond to territory margins (i.e. less than 20% of the grid cell is within wolf or pack territories), and to areas sporadically used by vagrant individuals or during dispersal. These results were then referenced to the 10 × 10 km grid cells and year, comprising 12 level 1 and 44 level 2 training input vectors, giving a total of 56 training input vectors.

During training, the FFNN model 'learned' to classify correctly the training data set and the apparent error was null, indicating a successful training process.

Model validation

Given the small size of the data set available for model fitting, it was not possible to validate our models with independent data. To overcome this problem, the model was validated using the bootstrap technique (Efron & Tibshirani, 1993). The bootstrap analyses were implemented in S Plus and, as for the FFNN package, the bootstrap package (Efron & Tibshirani, 1993) was available from the 'StatLib—Software and extensions for the S (*Splus*) language' (<http://lib.stat.cmu.edu/S/>).

The application of bootstrap to estimate the prediction error consisted in sampling with replacement 500 times the training data set and fitting the model using the bootstrap samples. We have chosen to use two bootstrap measures of prediction error that correct for the bias due to the lack of independence between the training and the test data set: the bootstrap bias corrected estimate and the 0.632 bootstrap estimate (Efron & Tibshirani, 1993).

Model validation using the bootstrap technique supported the apparent good fit of the model, and bootstrap estimations of prediction error were around 1% for both estimators (bias corrected prediction error = 1.13%, 0.632 bootstrap prediction error = 0.72%).

WOLF DISTRIBUTION PATTERNS IN PORTUGAL

We have used the FFNN model to classify wolf presence per grid cell based on the data in the *Signatus* database. The present study focused on the period 1990–1994, summing a total of 3625 registers. According to these results, less than half of the wolf range in Portugal holds resident wolves each year (Table 2). This pattern is not homogeneous over the occurrence range. Thus, for each year of the study period, approximately one third of the range south of the Douro holds resident wolves, whereas north of the Douro more than half of the range holds residents (Table 2).

Classifications are not constant over time and the same grid cell may be classified as holding resident wolves one year and be unoccupied in the following year, and vice versa. Moreover, some grid cells are frequently classified as holding resident wolves, whereas others have a low frequency of wolf occupancy (Fig. 3(a)), or may

often change classification during the study period (Fig. 3(b)).

Based on these observations, we have calculated both the ‘colonization’ and the ‘extinction’ rate. The ‘colonization’ rate is the proportion of regularly occupied grid cells per year that were not occupied or only sporadically used in the previous year. Likewise, the ‘extinction’ rate is the proportion of regularly occupied grid cells per year that were only sporadically or not used in the next year. Based on these indices, we observed a high turnover rate on grid cell occupancy (Table 3), so that more than one third of the range is colonized or goes extinct every year. This turnover rate depends on whether the grid cell is located south or north of the Douro, being much higher in the south.

Population fragmentation and efficiency of protected areas

As mentioned in the introduction, one of our major aims has been to identify areas of occupancy in order to support the selection of Special Conservation Areas in the context of N2000. Priority areas for wolf conservation should ensure the protection of stable core areas, which are assumed to include potential breeding areas. If we consider grid cells whose frequency of occupancy is either 4 or 5 years (1990–94) as being the putative core areas of the wolf population in Portugal, only 30% of the wolf range matches this criterion (Table 4). In addition, during the study period only around 30% of the core areas were inside protected areas. However, the establishment of N2000 increased this proportion twofold (Table 4).

The putative core areas denote the fragmentation of the Portuguese wolf population into five main nuclei:

Table 2. Proportion of grid cells within wolf range in 1990–94 where wolf presence was classified as ‘sporadic’, or ‘resident’, per year. The class ‘absent’ accounts for the proportion of grid cells within wolf range for which there is no information in each year. Within parenthesis is the proportion of grid cells comprised in protected areas. Results are presented separately for regions north or south of the river Douro.

		1990	1991	1992	1993	1994
North	Absent	0.42 (0.11)	0.45 (0.06)	0.43 (0.12)	0.38 (0.03)	0.32 (0.06)
	Sporadic	0.08 (0.00)	0.03 (0.20)	0.05 (0.38)	0.08 (0.08)	0.13 (0.10)
	Resident	0.50 (0.29)	0.52 (0.30)	0.52 (0.23)	0.54 (0.32)	0.55 (0.29)
South	Absent	0.53 (0.07)	0.65 (0.09)	0.61 (0.08)	0.62 (0.10)	0.39 (0.13)
	Sporadic	0.12 (0.10)	0.05 (0.00)	0.05 (0.00)	0.08 (0.00)	0.07 (0.00)
	Resident	0.35 (0.03)	0.30 (0.00)	0.34 (0.04)	0.30 (0.00)	0.54 (0.02)
Total	Absent	0.46 (0.09)	0.52 (0.07)	0.50 (0.10)	0.46 (0.06)	0.34 (0.09)
	Sporadic	0.10 (0.05)	0.04 (0.11)	0.05 (0.25)	0.09 (0.05)	0.11 (0.07)
	Resident	0.44 (0.22)	0.44 (0.23)	0.45 (0.18)	0.45 (0.24)	0.54 (0.20)

Table 3. Proportion of transitions in site occupancy between consecutive years. The ‘extinction rate’ is the proportion of grid cells classified as ‘residents’ that are no longer classified in this class the next year and the ‘colonization rate’ is the proportion of grid cells classified as ‘residents’ that were not classified in this class the previous year. Within parenthesis is the proportion included in protected areas. Results are presented separately for regions north and south of the river Douro.

		1990–91	1991–92	1992–93	1993–94
North	Extinction	0.26 (0.05)	0.33 (0.25)	0.30 (0.00)	0.24 (0.08)
	Colonization	0.28 (0.13)	0.33 (0.00)	0.33 (0.31)	0.26 (0.00)
South	Extinction	0.52 (1.00)	0.44 (0.00)	0.54 (1.00)	0.28 (0.00)
	Colonization	0.44 (0.00)	0.50 (1.00)	0.48 (0.00)	0.60 (1.00)
Total	Extinction	0.33 (0.09)	0.36 (0.25)	0.37 (0.05)	0.26 (0.08)
	Colonization	0.32 (0.13)	0.38 (0.05)	0.37 (0.31)	0.38 (0.04)

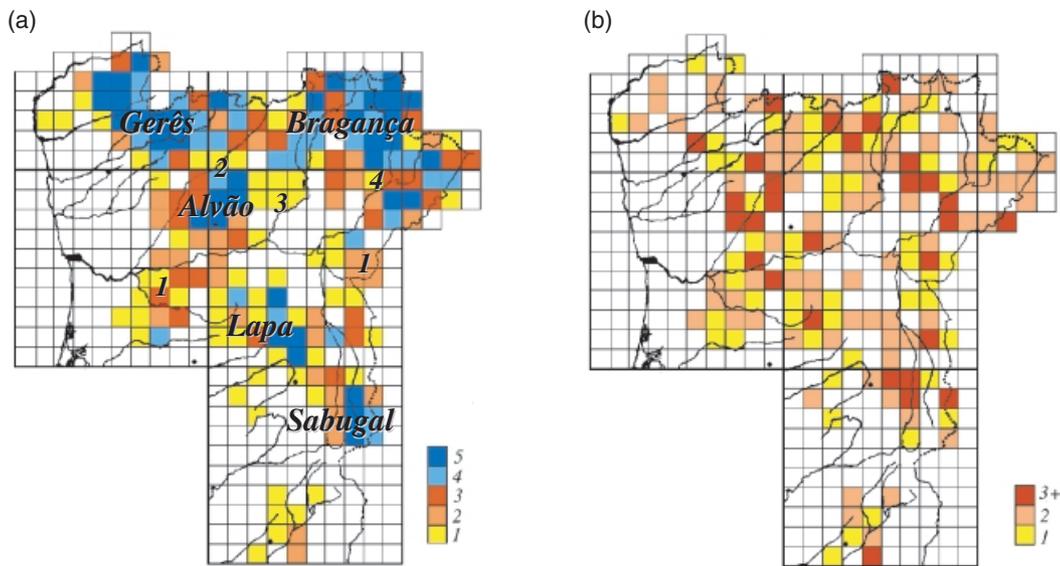


Fig. 3. Site occupancy by wolves after model 1. (a) Frequency of wolf occupancy, scored from 1 to 5, according to the total number of years of occupancy by resident wolves (colour index). The five population nuclei are identified as: *Gerês/Montalegre*; *Alvão*; *Bragança*; *Lapa/Arada*; *Sabugal/Figueira de Castelo Rodrigo*. Numbers 1 to 4 identify rivers associated with fragmentation: (1) Douro; (2) Tâmega; (3) Tua; (4) Sabor. (b) Frequency of transition between residency and sporadic use (and vice versa), scored as 1, 2 or 3 or more transitions (colour index).

Gerês/Montalegre; *Alvão*; *Bragança*; *Arada/Lapa*; *Sabugal/Figueira de Castelo Rodrigo* (Fig. 3(a)). The fragmentation lines seem to be associated with important river valleys, namely with the rivers Douro, Tâmega, Sabor and Tua (Fig. 3), and they correspond to regions of important human disturbance. In particular, human and road densities are, on average, significantly higher in marginal than in core areas (human density: 55.22 and 33.72 persons per km² respectively, $P = 0.014$, Mann–Whitney test; road density: 0.84 and 0.53 km/km² respectively, $P = 0.001$, Mann–Whitney test). Although these densities are much higher than similar values assessed for North American populations (Mech *et al.*, 1988), they are of the same magnitude of densities observed in other southern European populations, namely in Spain (Blanco *et al.*, 1990) and Italy (Corsi, Dupré & Boitani, 1999).

Unlike the other river valleys, the river Douro is likely to constitute an almost impermeable barrier within Portugal, resulting in northern and southern nuclei belonging to different conservation realities.

Table 4. Proportion of wolf range classified as marginal or core areas (Wolf Range). Proportion of marginal and core areas within protected areas, both before (AP) and after (AP + N2000) the establishment of N2000, north and south of river Douro.

		Wolf range	AP	AP + N2000
North	Marginal	0.62	0.07	0.24
	Core	0.38	0.38	0.64
South	Marginal	0.83	0.07	0.25
	Core	0.17	0.00	0.43
Total	Marginal	0.69	0.07	0.25
	Core	0.31	0.31	0.60

North of the river Douro, the core areas comprise approximately 40% of the wolf range (Table 4). Prior to N2000, around 38% of the core areas north of river Douro were already within protected areas, and this proportion has increased after the N2000 (Table 4). In addition, the northern nuclei are probably linked to the Spanish wolf population (Blanco *et al.*, 1990). In fact, Portuguese and Spanish nuclei located north of the river Douro are likely to make up the large majority of the Iberian wolf population, which is currently estimated at around 2000 individuals (Blanco *et al.*, 1990; Blanco & Cortés, 2002). In particular, the *Bragança* nucleus is connected to north and east with the *Castilla-León* nucleus, which was considered as the area with the highest wolf density in Spain (Blanco *et al.*, 1990). Therefore, the *Castilla-León/Bragança* nucleus is probably one of the most important areas for wolves in the Iberian Peninsula.

South of the river Douro, the wolf range has been steadily declining and in 1990–94 it represented less than 80% of its range in the previous decade (1980–89). Additionally, the core areas comprise only 20% of the wolf range and, prior to the N2000, none of these areas was protected. Such a low degree of protection may be one of the reasons underlying the decline observed in this region. Although the N2000 has permitted the classification of part of the core areas, the proportion under protection is still inferior to north of the river Douro.

The nuclei south of the river Douro still pose a major conservation challenge, and it is essential to put forward measures to ensure their viability and to minimize their isolation. Although isolated from the wolf population in northern Portugal, these nuclei are probably connected to the wolf population in the neighbouring Spanish

provinces. These Spanish provinces have been recently recolonized (during the 1990s) as a result of dispersal across the Douro by wolves from northern Spanish provinces (Blanco & Cortés, 2002).

Marginal areas and occupancy turnover

The establishment of N2000 has also increased the proportion of marginal range included in protected areas by more than twofold (Table 4). Although this increase in protection is mostly a by-product of conservation plans focused on other species, the protection of marginal areas may serve as a buffer to the core areas, particularly by reducing mortality. Additionally, it may also increase the connectivity between nuclei.

Not surprisingly, the marginal areas are characterized by a high turnover rate on wolf presence, so that areas as regularly occupied in 1 year may be unoccupied the next year and vice versa. This result is probably related to pack dynamics, in that it may reflect the process of pack formation and extinction, after human related disturbance, namely the death of the pack leaders (Haight *et al.*, 1998).

During the study period, colonization seems to balance extinction, and although the limited time scale does not allow the detection of any temporal trend, the high turnover rate raises important questions in terms of population dynamics and, particularly, population viability. A high rate of colonization/extinction is often encountered at the limit of populations' distribution range, and is associated with a sink population dynamic (Cowlshaw & Dunbar, 2000). Therefore, marginal areas may act as population sinks, absorbing individuals that disperse from core areas but fail to establish because of human persecution or the difficulty of finding conspecifics.

Wolf conservation in Portugal requires an understanding of the processes underlying local extinction and colonization, which is likely to have a human dimension. Haight *et al.* (1998) investigated the impact of these two processes on wolf viability in situations where populations are disjointed yet persist in clusters dominated by disturbed sites using a simulation model. They estimated that for realistic values of human-induced mortality (adult mortality 0.35 and pup mortality ranging between 0.2 and 0.4), population viability will depend on the success of dispersing wolves in finding vacant sites and mates. Therefore, it is very important not only to quantify human-induced mortality, but also to assess how this will impact both pack and population dynamics. Furthermore, it is essential to study wolf dispersal further, in order to secure its role in maintenance of sink populations as well as the connection between the main nuclei.

CONCLUSION

Overall the ANNs proved to be useful tools, allowing the classification of wolf presence with relatively low prediction errors given the limited available information. In the present study ANNs increased our insight on wolf distribution patterns in Portugal from a simple distribution

range to a more complex and realistic understanding. The insights provided by our analysis allowed the identification of core areas that were taken into account when selecting priority areas for wolf conservation in Portugal in the context of N2000.

This approach may become an important tool to monitor wolf population status in Portugal, given limited resources and availability of information on wolf depredation. Its systematic use may allow managers to detect population trends and to assess turnover rates.

Can ANN be a useful tool for conservation biologists?

Although being well suited for classification problems, ANNs lack explanatory power, i.e. no information is given about how the identification is achieved or which variables were important in achieving it. Therefore, ANNs are of limited use for understanding the mechanisms underlying the phenomenon under analysis, although recent papers have explored computational approaches that may allow this limitation to be surpassed (Lek *et al.*, 1996; Olden & Jackson, 2002).

Nevertheless, ANNs are well suited for data analysis when the process *per se* is not under analysis. This is notably the case when the aim is to assess the occupancy by endangered species and detailed information is not available over large geographic and/or temporal ranges. The predictive power of ANNs will depend on the quality of the available data. Hence, ANNs are valuable when reliable data are available, but these data do not respect the assumptions of most classic statistics, such as normal distribution, homogeneity of variance/covariance matrices. Hence, ANNs can be applied whenever we have a reliable data set for the entire range of a taxa, and an independent training data set (i.e. local studies for which the response variable is known). This is often the case of oral inquiries, which are frequently used in assessing large carnivores distribution over wide geographic ranges (Easterbee, Hepburn & Jefferies, 1991; Blanco *et al.*, 1992; Rodrigues & Delibes, 1992; Gros, 1998; Stahl, 1998), without identifying occupancy areas within this range. The integration of broad-range assessment techniques with local studies through ANN may then allow us further to characterize species distribution, as long as local studies are independent and are representative of the entire range under study.

The use of ANNs should always be associated with model validation techniques, in order to evaluate their predictive power. Even if validation with an independent data set is not possible, other computer-intensive techniques such as jackknife or bootstrap resampling procedures can be used (Efron & Tibshirani, 1993).

Acknowledgements

This work was partially supported by the European Community Program LIFE 'Programa para o Conhecimento e Gestão do Património Natural' through

the Instituto de Conservação da Natureza (ICN). We thank all the members of the Project *Signatus* who collaborated in the fieldwork and database maintenance, in particular S. Alcindo, S. Alexandre, F. Álvares, A. T. Cândido, R. Carreira and A. E. Pires. We also thank A. St Aubyn and M. Neves for their advice on ANN and bootstrap and T. Boulinier, G. Chapron, B. N. K. Davis, S. Legendre, R. Merigii, M. Nicholson, O. Renault, E. Rocha, P. Rothery and two anonymous referees for comments on early versions of the manuscript, and J. Ewen for having reviewed our English.

REFERENCES

- Álvares, F. (1995). *Aspectos da distribuição e ecologia do lobo no Noroeste de Portugal: o caso do Parque Nacional da Peneda-Gerês*. Faculty of Sciences, University of Lisbon, Lisbon.
- Aoki, I. & Komatsu, T. (1997). Analysis and prediction of the fluctuation of sardine abundance using a neural network. *Oceanol. Acta* **20**: 81–88.
- Baran, P., Lek, S., Delacoste, M. & Belaud, A. (1996). Stochastic models that predict trout population density or biomass on a mesohabitat scale. *Hydrobiology* **337**: 1–9.
- Bessa-Gomes, C. (1996). *Contribuição para o estudo da distribuição do lobo em Portugal: caracterização da estrutura espacial da população no primeiro quinquénio da década de 90 e análise da adequação do habitat*. Departamento de Zoologia e Antropologia, Faculdade de Ciências da Universidade de Lisboa, Lisboa.
- Blanco, J. C. & Cortés, Y. (2002). *Ecología, censos, percepción y evolución del lobo en España: análisis de un conflicto*. Malaga: SECEM.
- Blanco, J. C., Cuesta, L. & Reig, S. (1990). El lobo en España: una visión global. In Blanco, J. C., Cuesta, L. & Reig, S. (Eds). *El lobo (Canis lupus) en España: situación, problemática y apontes sur su ecología*: 69–94. Madrid: Ministerio de Agricultura, Pesca y Alimentación, ICONA.
- Blanco, J. C., Reig, S. & Cuesta, L. (1992). Distribution, status and conservation problems of the wolf *Canis lupus* in Spain. *Biol. Conserv.* **60**: 73–80.
- Bradshaw, C. J. A., Davis, L. S., Purvis, M., Zhou, Q. & Benwell, G. L. (2002). Using artificial neural networks to model the suitability of coastline for breeding by New Zealand fur seals (*Arctocephalus forsteri*). *Ecol. Model.* **148**: 111–131.
- Brey, T., Jarre-Teichmann, A. & Borlich, O. (1996). Artificial neural network versus multiple linear regression: predicting P/B ratios from empirical data. *Marine Ecol. Progr. ser.* **140**: 251–256.
- Brosse, S., Guégan, J.-F., Tourenq, J.-N. & Lek, S. (1999). The use of artificial neural networks to assess fish abundance and spatial occupancy in the littoral zone of a mesotrophic lake. *Ecol. Model.* **120**: 299–311.
- Carreira, R. (1996). *Situação populacional e biologia alimentar do lobo na área de influência do Parque natural do Alvão*. Faculty of Sciences, University of Lisbon, Lisbon.
- Corsi, F., Dupré, E. & Boitani, L. (1999). A large scale model of wolf distribution in Italy for conservation planning. *Conserv. Biol.* **13**: 150–159.
- Cowlishaw, G. & Dunbar, R. I. M. (2000). *Primate conservation biology*. Chicago: Chicago University Press.
- Easterbee, N., Hepburn, L. V. & Jefferies, D. J. (1991). Survey of the status and distribution of the wildcat in Scotland, 1983–1987. Edinburgh: Nature Conservancy Council for Scotland.
- Efron, B. & Tibshirani, R. J. (1993). *An introduction to the bootstrap*. New York: Chapman and Hall, New York.
- Fausett, L. (1994). *Fundamentals of neural networks*. Englewood Cliffs, NJ: Prentice Hall International.
- Gros, P. M. (1998). Status of the cheetah *Acinonyx jubatus* in Kenya: a field-interview assessment. *Biol. Conserv.* **85**: 137–149.
- Haight, R. G., Mladnoff, D. J. & Wydeven, A. P. (1998). Modeling disjunct gray wolf populations in semi-wild landscapes. *Conserv. Biol.* **12**: 879–888.
- INE (1991a). *Anuário estatístico da região centro*. Lisbon: Instituto Nacional de Estatísticas.
- INE (1991b). *Anuário estatístico da região norte*. Lisbon: Instituto Nacional de Estatísticas.
- INE (1993). *Census 1992*. Lisbon: Instituto Nacional de Estatísticas.
- Lae, R., Lek, S. & Moreau, J. (1999). Predicting fish yield of African lakes using neural networks. *Ecol. Model.* **120**: 325–335.
- Lein, J. K. (1995). Mapping environmental carrying capacity using an artificial neural network: a first experiment. *Land Degrad. and Rehab.* **6**: 17–28.
- Lek, S., Belaud, A., Dimopoulos, I., Lauga, J. & Moreau, J. (1995). Improved estimation, using neural networks, of the food consumption of fish populations. *Marine Freshwater Res.* **46**: 1229–1236.
- Lek, S., Delacoste, M., Baran, P., Dimopoulos, I., Lauga, J. & Aulagnier, S. (1996). Application of neural networks to modelling nonlinear relationships in ecology. *Ecol. Model.* **90**: 39–52.
- Lek, S. & Guégan, J. F. (1999). Artificial neural networks as a tool in ecological modelling, an introduction. *Ecol. Model.* **120**: 65–73.
- Lusk, J. J., Guthery, F. S. & DeMaso, S. J. (2001). Northern bobwhite (*Colinus virginianus*) abundance in relation to yearly weather and long-term climate patterns. *Ecol. Model.* **146**: 3–15.
- Manel, S., Dias, J. M., Buckton, S. T. & Ormerod, S. J. (1999a). Alternative methods for predicting species distribution: an illustration with Himalayan river birds. *J. Appl. Ecol.* **36**: 734–747.
- Manel, S., Dias, J.-M. & Ormerod, S. J. (1999b). Comparing discriminant analysis, neural networks and logistic regression for predicting species distributions: a case study with a Himalayan river bird. *Ecol. Model.* **120**: 337–347.
- Mastrorillo, S., Lek, S., Dauba, F. & Belaud, A. (1997). The use of artificial neural networks to predict the presence of small-bodied fish in a river. *Freshwater Biol.* **38**: 237–246.
- Mech, L. D., Fritts, S. H., Rade, G. L. & Paul, W. J. (1988). Wolf distribution and road density in Minnesota. *Wildl. Soc. Bull.* **16**: 85–87.
- Moreira, L. M. (1992). *Contribuição para o estudo da ecologia do lobo (Canis lupus signatus Cabrera, 1907) no Parque Natural de Montesinho*. Faculty of Sciences, University of Lisbon, Lisbon.
- Olden, J. D. & Jackson, D. A. (2002). Illuminating the ‘black box’: a randomization approach for understanding variable contributions in artificial neural networks. *Ecol. Model.* **154**: 135–150.
- Paruelo, J. M. & Tomasel, F. (1997). Prediction of functional characteristics of ecosystems: a comparison of artificial neural networks and regression models. *Ecol. Model.* **98**: 173–186.
- Petrucci-Fonseca, F. (1990). *O lobo ibérico (Canis lupus signatus Cabrera, 1907) em Portugal*. Faculty of Sciences, University of Lisbon, Lisbon.
- Recknagel, F. (2001). Applications of machine learning to ecological modelling. *Ecol. Model.* **146**: 303–310.
- Recknagel, F., French, M., Harkonen, P. & Yabunaka, K.-I. (1997). Artificial neural network approach for modelling and prediction of algal blooms. *Ecol. Model.* **96**: 11–28.
- Ripley, B. D. (1994). Neural networks and related methods for classification. *J. R. Stat. Soc. (B)* **56**: 409–456.
- Rodriguez, A. & Delibes, M. (1992). Current range and status of the Iberian lynx *Felis pardina* Temminck 1824 in Spain. *Biol. Conserv.* **61**: 189–196.
- Scardi, M. (1996). Artificial neural networks as empirical models for estimating phytoplankton production. *Marine Ecol. Progr. Ser.* **139**: 289–299.
- Scardi, M. (2001). Advances in neural network modeling of phytoplankton primary production. *Ecol. Model.* **146**: 33–45.
- Scardi, M. & Harding Jr., L. W. (1999). Developing an empirical model of phytoplankton primary production: a neural network case study. *Ecol. Model.* **120**: 213–223.

- Schonewald-Cox, C., Azari, R. & Blume, S. (1991). Scale, variable density, and conservation planning for mammalian carnivores. *Conserv. Biol.* **5**: 491–495.
- SNPRCN (1990). *Livro vermelho dos vertebrados terrestres de Portugal*. Vol. I *Mamíferos, aves, répteis e anfíbios*. Lisbon: Serviço Nacional de Parques, Reservas e Conservação da Natureza.
- Soulé, M. (1986). *Conservation biology: the science of scarcity and diversity*. Sunderland, MA: Sinauer.
- Spitz, F. & Lek, S. (1999). Environmental impact prediction using neural network modelling. An example in wildlife damage. *J. Appl. Ecol.* **36**: 317–326.
- Stahl, P. (1998). *Evolution recente de la repartition du chat sauvage en France. Résumé du protocole d'enquête*. CNERA Prédateurs. ONC. Muséum d'Histoire Naturelle.
- Terry, A. M. R. & McGregor, P. K. M. (2002). Census and monitoring based on individually identifiable vocalizations: the role of neural networks. *Anim. Conserv.* **5**: 103–111.
- Tourenq, C., Aulagnier, S., Durieux, L., Lek, S., Mesleard, F., Johnson, A. & Martin, J.-L. (2001). Identifying rice fields at risk from damage by the greater flamingo. *J. Appl. Ecol.* **38**: 170–179.
- Tourenq, C., Aulagnier, S., Mesleard, F., Durieux, L., Johnson, A., Gonzalez, G. & Lek, S. (1999). Use of artificial neural networks for predicting rice crop damage by greater flamingos in the Camargue, France. *Ecol. Model.* **120**: 349–358.
- Venables, W. N. & Ripley, B. D. (1994). *Modern applied statistics with S-plus*. New York: Springer-Verlag.
- Weeks, P. J. D. & Gaston, K. J. (1997). Image analysis, neural networks, and the taxonomic impediment to biodiversity studies. *Biodivers. Conserv.* **6**: 263–274.