Ecological niche separation of two sympatric sibling wood ant species: a habitat reciprocal approach.

1 Summary

Formica lugubris and *Formica paralugubris* are two sibling wood ant species that, in contrast with classical rules of the biological view, seem to occupy the same ecological niche, in alpine forest ecosystems. In order to complete recent research on these species and better understand their sympatry, occurrence data are collected through a random-stratified sampling in the Swiss Jura Mountains. A first attempt is made in modelling *F. lugubris* and *F. paralugubris* habitats by the mean of General Linear Models (GLM) and Geographic Information Systems (GIS). One general and two specific models were fitted on climatic, topographic and environmental spatially explicit predictors. This approach provided evidences for distinct distribution patterns of the two species, *F. lugubris* occurring more frequently in locations at woodland borders and *F. paralugubris* in deep forest. Moreover, the two species display different response to the seasonal number of frost days and to the sites' topographic exposure.

Key words: Formica lugubris, Formica paralugubris, habitat suitability models, niche differences, predictive distribution maps, sympatry.

2 Introduction

Much past and current research in wild life biology focuses on describing and modelling habitats of various species. In the last 30 years, statistical performances in the field of ecological modelling greatly improved and tools such as General Linear and Additive Models (GLMs, GAMs) provided a considerable contribution to that development. Nowadays, extensive and high-resolution environmental data are increasingly available, and their implementation in a Geographic Information System (GIS) coupled with the use of modern regression approaches has proven particularly useful in the modelling of the spatial distribution of organisms (see Franklin 1995, Guisan & Zimmermann 2000, Scott *et al.* 2002).

GIS-based statistical modelling of plant and animal relationships with environmental conditions has frequently been associated with efforts in the protection of threatened species (Gerrard *et al.* 2001, Broennimann 2003), management of problematic species (Radeloff, Pidgeon & Hostert 1999, Patthey 2003), biodiversity assessment and preservation of highly valuable habitats (Store and Jokimäki 2003, Heikkinen and Neuvonen 1997, Lehmann *et al.* 2002). It is also a precious tool allowing researchers to get insights in predicting potential impacts of climate-change phenomena on the distribution of species and communities (see for ex. Guisan and Theurillat 2001, Bakkenes *et al.* 2002, Rebetez 2002).

Meanwhile, even if our methods and results might reveal useful for future conservation works, the basic aim of the present study is to discuss on some hypotheses concerning the ecology of two sympatric, occasionally syntopic, ant species: *Formica lugubris* and *Formica paralugubris*.

These ants belong to the *Formica rufa* group (so-called red wood ants), a well-studied but phylogenetically complex and still controversial taxon (Maeder & Cherix 2001). Until 1996 (Seifert 1996) *F. paralugubris* had never been considered as a species in its own, but rather as a *F. lugubris* "type B" (Kutter 1967 for queen's hairiness differences, Pamilo, Chautems & Cherix (1992) for genetic differences in allozymes, Rosengren *et al.* 1994 according to behavioural tests for cocoon transport). The description of *F. paralugubris* nov. spec. based on comparative studies of external morphology (Seifert 1996), brought forth new questions, specifically in relation to the reasons allowing two sibling species to dominate a vast environment (coniferous alpine forests) when inhabiting same geographic areas and apparently identical habitats. Some important differences regarding the social structure and the reproduction strategies have recently been pointed out (for a review see Cherix *et al.* 2004, *in press*).

In the present work, the study of *F. lugubris* and *F. paralugubris* and their relationships with environment is approached from the perspective of spatial ecology. A general and two specific models were calibrated on occurrence data provided by a random-stratified sampling, which was performed on a 819 km^2 study area, the part of the Jura Mountains belonging to the Canton Vaud. We concentrate the sampling efforts in this region, since *F. lugubris* and *F. paralugubris* are here particularly abundant.

Niche differences between species are investigated through model comparisons.

To our knowledge, despite the great amount of information that has been gathered by numerous studies, this is the first attempt to build GIS-based habitat suitability models for wood ant species.

3 Methods

3.1 Studied species

Formica lugubris Zett. and *Formica paralugubris* Seifert are two boreo-alpine sympatric species belonging to the *Formica rufa* group (so called red wood-ants), who are characterized by large mound-building nests, dwelled by very populous societies. The central role played by wood ants in maintaining the biological equilibrium of the forest ecosystem is well known (Cherix 1991); consequently, these insects are protected by law in many countries (for instance by the Swiss federal law for nature and landscape protection, since 1966). Colonies[1] depend on forest, principally coniferous woods, for food supply (arthropods, aphids' honeydew), for nest building materials such as fir-tree's and pine's needles and probably to benefit of a certain amount of shade.

F. lugubris. and *F. paralugubris* are mainly distributed across Central and Northern Europe, ranging from 600 m up to 2200 m (Gösswald *et al.* 1965, Gösswald 1989; *in* Cherix *et al.* 2003, for *F. lugubris* sensu lato[2]). Since its description as a separate species, *F. paralugubris* was found in the Swiss Jura Mountains (type locality), in the French Jura, in the Swiss and Austrian Alps (Seifert 1996), and, according to Maeder *et al.* (*in prep.*), in the Pyrenees, the French and Italian Alps as well.

In the Jura Mountains, *F. lugubris* is more widely distributed, although *F. paralugubris* locally reaches higher densities (Freitag 2002). This pattern reflects differences in social structures and reproduction strategies (Cherix et al 2004, *in press*). Indeed, *F. lugubris* colonies are mostly monogynous and monocalic. Queens need a nuptial flight before mating; therefore they generally disperse far from the native nest to found a new colony. On the contrary, *F. paralugubris* generally shows high polygyny rates, that are caused by the displaying of an alternative sexual behaviour consisting in mating inside or at the surface of the mother nest, where sexuals shed their wings and stay in. After a year or more some queens may found new nests in the neighbourhoods by a process called budding, leading to highly polycalic colonies (i.e. super-colonies). Therefore, generally speaking, whereas *F. lugubris* has a dispersal opportunistic strategy, *F. paralugubris* opt for local habitat dominance, based on competitive power.

In the Swiss National Park two populations of *F. lugubris* with distinct reproductive strategies have been recently described (Bernasconi 2002). One shows the classical dispersal strategy whereas the

[1] The terms nest, colony and population will be used here according to the following definitions. Nest: group of individuals (queens and workers) interacting cooperatively and living in the same anthill (or "mound"). Colony: functional and cooperative ensemble of individuals, which can be constituted of one or more nests. Population = ensemble of colonies inhabiting the same geographical area.

[2] *F. lugubris* sensu lato = *F. lugubris* and *F. paralugubris* pooled togheter (corresponding to data before 1996).

other behaves, similarly to *F. paralugubris* but at a lower degree, following a local and polycalic pattern. Nevertheless, this alternative form of *F. lugubris* seems to be present in the Alpine region only. In the Jura Mountains, the only few observed cases of high densities of *F. lugubris* don't display any visible connection (paths) among nests, which is the necessary condition for polycalism (Maeder pers. comm.). The absence of a common social structure in our study area is a very important condition as our goal is to differentiate the two species as much as possible.

3.2 Sampling design

Field samples have been selected following a random-stratified strategy, an approach that has been shown to be particularly appropriate in modelling species distribution (Hirzel & Guisan 2002, Maggini, Guisan & Cherix 2002). Its greatest advantage is that the different environmental situations occurring in the study area tend to be equally sampled. From the target species point of view, this is a fundamental requirement in order to accurately discriminate propitious from adverse ecological circumstances.

Stratification was based on four environmental factors: elevation, slope aspect, slope angle and location relatively to the forest. They are considered as being *a priori* potentially very influent on *Formica* ants distribution (Devenoges 1999, Borges 2001, Maggini *et al.* 2002).

Two main classes have been defined for each factor. Considering all the possible combinations of these environmental situations, the studied area has been split in 16 strata. An equal number of replicates have then been randomly sampled into each one of them, according to the concept of ecological representativity.

For elevation a first class covers the range from 800 m to 1200 m, a second one the range above 1200 m (lowlands have not been considered because the target species are rarely found below 800 m). This 1200 meters threshold was chosen according with the design set up by Freitag (2002). Elevation is strongly correlated with climatic conditions to which various ant species are differently adapted, in order to avoid or moderate competition with other species. Slope and aspect are key-factors since they determine solar radiation that ants can exploit for heating their nest and thus provide optimal conditions for brood's development. Sites having a South-East exposition (between 45° and 225°, "favourable") and a North-West exposition (between 225° and 360°, 0° and 45°, "unfavourable") were distinguished. Slope is either weak (1° to 20°) or steep (25° to 45°). Flat sites were not sampled since they have no aspect, intermediate slopes (20° to 25°) were excluded in order to emphasize species' response to this gradient and extreme slopes (>45°) were omitted as well because one considers that they physically prevent the building of wood ant mounds. Finally sites can be located either into or in

the border of a forest, which is supposed to affect solar radiation incidence, food availability and the potential presence of foundation sites (slave ants' nests of the genus *Serviformica*).

All GIS manipulations were performed in the ArcInfo® software (ESRI Inc., Redlands, CA, USA).

Grid maps for elevation, aspect and slope all derive from the Swiss digital elevation model (DEM) at a resolution of 25 x 25 m (DHM25, OFS[3], 2002). As the surveyed sites' surface is one hectare, these maps have been aggregated to obtain pixels of 100 x 100 m. To each new pixel, the value of a given class has been attributed only in the case where all the 16 contained pixels of the higher resolution belonged to this same class; every time that one or more cells at a 25 m resolution had a different value compared to the others (or even "nodata"), the result pixel became a nodata. This strategy allows avoiding the sampling of inhomogeneous sites (for example having only part of their surface below or above the chosen threshold of 1200 m). Every 100 m pixel was than saved as a point in a new vectorial map. Sites' position relatively to the forest was then determined on the basis of a vectorial map of forest cover of the studied area: points falling into a two sided 30 m buffer around woodland edges were put in the "forest border" class, those falling in forest-free areas were ignored and all the others formed the "deep forest" class.

3.3 Field sampling

Field sampling was carried out between September and October 2003. Sampling units were localised with 1:25000 maps and a GPS (Global Positioning System, Garmin®) and then visited following a standard zigzag path. When *F. lugubris* and/or *F. paralugubris* were found into the site's perimeter, the number of nests and their spatial coordinates, provided by the GPS, were recorded (average precision of 7 meters, in case of sufficient satellite cover). Abandoned or small seasonal nests were not taken into account for abundance assessments. For each nest, ten ant workers were collected for identification purposes. Some sampling unit had to be skipped because forest stands were damaged, either by human activities or by natural perturbations. Furthermore, spatial autocorrelation was avoided by taking into account only sites that were at least 200 m apart from one another; the mean (maximum?) range of action of a wood ant's colony being approximately 100 m, this distance was considered to be sufficient (D. Cherix, pers. comm.).
Collected data were merged with an already existing data set, which had been obtained with an identical sampling design during a recently conducted inventory of wood ants in the Canton Vaud (Freitag 2002).

[3] Swiss Federal Office Statistical Office.

3.4 Species identification

The worker-based distinction between *F. lugubris* and *F. paralugubris* is a quite difficult task, that cannot be achieved in the field. A 70x binocular equipped with a graduate lent was used for external morphology measures on collected workers. Following a simplified version of the identification key proposed by Seifert (1996), an index allowing species differentiation was estimated for each nest. When this index falls into a given intermediate interval of values, identification remains unsure.

3.5 Ecogeographical predictors

As a basic assumption for building a static model (i.e. that does not take into account a temporal dimension of spatial distribution), it is considered that the species whose distribution is being modelled is in equilibrium, or at least pseudoequilibrium, with the surrounding environment that sustains it (Guisan & Zimmermann 2000). This means that the species is theoretically not spreading, regressing or switching habitat within the study area, if not at a geological time-scale.

According to Harrell, Lee & Mark (1996), and in order to enhance predictions' accuracy, the set of explanatory variables in the final models should be restricted to no more than n/10 predictors, where n is the number of observations. A choice was therefore made (Table 1) among the available environmental predictors, in respect of various criteria, such as the existing knowledge on wood ants' biology, the correlation rates among predictors and the results of many exploratory uni- and multivariate GLMs. Collinearity among predictors is particularly important to avoid, since it is known to create problems in stepwise procedures for model selection (Brauner & Shacham 1998, in Guisan, Edwards & Hastie 2002). Table 2 shows the correlation matrix (Spearman correlation coefficients) of our selected predictors.

The forest cover layer, from which the binary variable FORLIS (deep forest *vs* forest borders) was drawn from, is extracted from VECTOR25[4]. ASPSIN, measuring the "eastness" of a location, was derived from the Swiss DEM (the Digital Elevation Model of Switzerland)[5]. DRESLIS200, the average coniferous density in a 200 meters moving window, was constructed on the basis of a forest "mixity" grid layer (MIXFOR[5]). Finally, climatic layers were provided by N.E. Zimmermann et F. Kienast at WSL[6].

[4] Layer provided by the Swiss Federal Office for Topography, 3084 Wabern.
[5] Layer provided by the Swiss Federal Statistical Office, Neuchâtel.
[6] Swiss Federal Office for Forest, Snow and Landscape research.

[TABLE 1 HERE]

[TABLE 2 HERE]

DDEG300, more than simply elevation or average temperature gradients, is a biologically meaningful predictor. In fact, it is related to plants physiological activity (Bolliger, Kienast & Zimmermann 2000), 3°C being the temperature below which plants activity stops (average threshold for the Swiss flora). The relevance of taking into account plant-related predictors in our study context comes from the strict trophic dependence of wood ants on the honeydews that aphids produce from trees sap. Moreover, both DDEG300 and SFROYY, which refers to the period of vegetation growth, are considered more pertinent predictors for ants than equivalent all-year-long measures. Low temperatures and frost occurrence during winter probably very poorly influence wood ants, since these social insects remain largely inactive during the cold season, sheltered by their mounds.

Solar radiation (direct, SDIRYY, or diffuse, SDIFYY) is a very proximal[7] parameter, particularly appropriate for modelling wood ants distribution, since it greatly influences many microhabitat parameters such as soil temperature or humidity on the ground and is directly responsible for nest heating.

DRESLIS is actually an interaction (i.e. the multiplication) between a binary variable (1=forest edge location, 0=deep forest location) and a quantitative variable (% of surface covered by coniferous woods). It was employed in model calibration instead of the two basic variables for two reasons: (i) it condenses in one term the information of both FORLIS and DRES200 and (ii) the interaction of these two variables proved to fit more explanatory and significant regressions in comparison with separate effects (synergetic interaction).

In the study of ants one can obviously not work on occurrence data of *individuals*. The actual unit that is responding to natural conditions is the nest as a whole. The successful establishment of a colony is detected by the presence of a nest and nest number evaluates its breadth. Moreover, as suggested by Maggini *et al.* (2002), Freitag (2002) and also according to our own unpublished results, the utilization of a classical presence-absence approach has proven to be poorly satisfactory for ant distribution modelling purposes.

Our analyses were therefore based on nest number per sampling unit, accepting the following postulates:

- the more suitable a site the more likely a colony of *F. paralugubris* is to

[7] An environmental gradient is defined "proximal" (as opposed to "distal") when its effects have a direct and causal impact on a given organism (Austin 2002).

grow and multiply its nests;

- similarly, a site that allows the establishment of several monocalic colonies of *F. lugubris* is, , considered as being particularly suitable.

3.6 Statistical modelling

3.6.1 Generalized linear models

Generalised Linear Models (GLM; McCullagh & Nelder 1989) were calibrated and evaluated (see chapter 2.5.3) by using the statistical software RAqua (Version 1.8.1, Copyright 2003, The R Development Core Team.). GLMs are mathematical extensions of multiple linear regressions that allow assuming non-Gaussian distributions of data. They are able to handle binomial, Poisson, negative binomial or gamma distributions (for the response variable Y), which often better fit the non-normal error and non-constant variance structures of ecological data (Guisan *et al.* 2002).

The general formula of a GLM is:

$$f(Y) = LP + \varepsilon = (\alpha + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_n x_n) + \varepsilon$$

where Y is the response variable (values that we try to predict), $x_1, x_2, ..., x_n$ are the n predictor variables, $\beta_1, \beta_2, ..., \beta_n$ are the regression coefficients and ε is the vector of the regression residuals. Linearity is ensured by the link function $f(Y)$, which sets the relationship between the response variable Y and the linear predictor LP.

Since our models were calibrated on abundance data, a Poisson distribution was choosen for the response variable, because it generally better fits discrete count data. The link function is logarithmic.

As in the case of least-squares multiple regression, polynomial terms such as x_i^2 or x_i^3 can be included in the set of predictors to account for unimodal respectively skewed or bimodal responses (Guisan *et al.* 2002).

3.6.2 Analytical framework

Four models have been calibrated: FDB, L/P, PARA, LUGU.

A stepwise regression routine with AIC (Akaike's Information Criterion) was chosen as model selection tool. AIC is a parsimony criterion allowing the selection of a minimal number of predictors for a maximal amount of explained deviance. In their final form, models only include those terms, linear or quadratic, which satisfy tree main criteria: (i) to be selected by the stepwise procedure

(direction = "both"), (ii) to be significant at the 0.05 confidence level from a Student test of deviance[8] reduction, and (iii) to cause a reduction of at least 2% of the original (null) deviance when leaved aside. This third requirement should limit the possibility that a selected predictor explains a part of the deviance only by chance.

The FDB model considers the two studied species pooled together and is build upon standardized-abundance data. "Real" abundances cannot be used in this case, because the number of mounds found in a sample unit strictly depends not only on the site's habitat suitability but also, and especially, on species' social structure. One need to consider that *F. lugubris*, being primarily mono- or oligocalic, will generally not display very high nest densities like *F. paralugubris*, although habitat might be very propitious.
We therefore took the median of each species' nest abundance distribution as the criterion through which the nest number data can be translated into relative abundances allowing them to be pooled together, by applying the following transformations:

$$\mathbf{y_{rel\ (paralugubris)} = y / m_{paralugubris}}$$

$$\mathbf{y_{rel\ (lugubris)} = y / m_{lugubris}}$$

$$\mathbf{y_{rel(wood\ ants)} = round[(y_{rel\ (paralugubris)} + y_{rel\ (lugubris)})*10]}$$

where y_{rel} = relative nest abundance, y = nest number, m = median of the nest abundance distribution (characteristic of each species). "Wood ants" relative nest abundances were multiplied ten times and rounded to integer before model calibration, consenting to limit information lost as well as a better fitting by a Poisson probability distribution.
FDB is a general "red wood ants" model that should underline the principal common environmental factors influencing these species.

modèle binomial: presence lugu = 1, presence para = 0

tests de Wilcoxon: comparaison , au sein de chaque variable, de la moyenne des valeurs des sites occupés para *vs* lugu##

modèles spécifiques para et lugu, GLMs poisson qui tiennent compte des abondances##

[8] Deviance is the corresponding concept of variance in the case of multiple regressions where coefficients are estimated by a Maximum-Likelihood algorithm instead of the classical Least Squares

The building of separate (specific) models and the description of hypothetical niche differences is another delicate task. First of all, the models PARA and LUGU are built with the whole data set. Now, if one considers that *F. lugubris* and *F. paralugubris* are two sympatric and very similar species, it is natural to suppose that they probably share a substantial part of their fundamental niche (sensu Hutchinson 1978). Such scenario implies that, because of interspecific competition, where a colony of *F. lugubris* is established, the chances of finding *F. paralugubris* as well are much lowered (and *vice versa*). For that reason, two other models (PARA and LUGU) were constructed, using for each species a modified data set not taking into account absence sites corresponding to a presence of the sibling species. These points probably correspond, from the habitat modelling point of view, to "false absences" in the sense that the missing of one species is imputable to the presence of the other (competitor) rather than to a low habitat quality. If PARA and LUGU will provide worse results than PARA and LUGU, we'll assume this as a confirmation that a process of competitive exclusion is likely to take place, and only the two latter models will be considered for species' niche comparison.

3.6.3 Predictive distribution maps and models evaluation

The four models were fitted with spatially explicit predictors and were therefore easily implemented in a GIS to derive potential habitat distribution maps. The response variable, which is visually expressed by colour patterns, is the predicted species abundance and can be interpreted as a measure of habitat suitability.

Because of the mentioned biological differences among *F. lugubris* and *F. paralugubris*, a direct comparison of respective abundance predictions would lead to wrong deductions. Spatial predictions will therefore be represented by unequal classes (see Results) rather than by continuous values. As a result, maps are much easier to compare.

Finally, default zero (nodata) values were assigned to all those pixels being at least 30 m far from a forest (i.e. meadows), since those habitats are not expected to host F. *lugubris* nor *F. paralugubris*.

The predictive power of each model is evaluated performing Spearman correlation tests on observed *vs* predicted abundances in checked stations.

3.6.4 Comparing models

The two models (PARA, LUGU OR PARA, LUGU) giving the better results in terms of explained deviance and predictive accuracy (Spearman correlation coefficient) would be considered to analyse the hypothetical niche differences among species. The statistical exploration of differences between two

models calibrated on different datasets and for different species is quite an unusual analysis context. Because of the difficulty to find an existing statistical framework allowing this kind of comparison, a particular design was set up.

First of all, as a qualitative parameter, the explanatory potential of each predictive variable is evaluated by measuring their average explained deviance (in % of the null deviance). In the case of partially correlated (non-orthogonal) predictors, the sequential order in which terms are entered in an analysis of variance (ANOVA) influences the manner in which model's explained deviance is distributed among the different predictors. Therefore, each model was ran n! times (n being the number of predictors) in order to have a value of explained deviance of each predictor for all their possible permutations. This strategy allows, for each term of the model formula, the calculation of an average explained deviance (± standard error), which could be used to evaluate its respective importance in the different models.

The interspecific comparison is based on a reciprocal analysis consisting in applying the set of predictors selected for the model of *F. lugubris* on the dataset of *F. paralugubris* (model PARA'), and *viceversa* (model LUGU'). Indeed, two models are much easier to compare when they share exactly the same predictive variables.

At this stage, for each different variable, five possible situations can be distinguished:

1. it is selected for both specific models, but shows different types of response curves;
2. it is only selected for one species and its regression coefficient is non significant (*p-value*>0.05) when applied to the other species' dataset;
3. it is only selected for one species, but the other one also respond to it, although to a lesser extent, when models are reciprocally applied.
4. it is selected for both species, which show similar responses to it, but with different intensities;
5. it is selected for both species, which show a very similar response to it.

Situations 1. and 2. clearly indicate different ways to react to environmental conditions. On the other hand, when the two species' spatial distributions show the same response type to a given ecological factor, an eventual difference needs to be stressed by the comparison of average percentage of explained deviance (3..4., 5.).

4 Results

4.1 Field sampling

As often, when confronted to the field reality, the final number of sampled sites per stratum was not perfectly balanced (Table 3). Indeed, it was difficult to find intact hectares for some environmental situations (e.g. low elevation + South-East exposition + steep slope + forest border) that are infrequent in the study area. Nevertheless, we consider our sampling sufficiently balanced to ensure that models output actually simulate the species response to environmental gradients rather than simply reflecting the distribution of the sampling itself (Maggini *et al.* 2002).

[TABLE 3 HERE]

58 out of 144 sampling sites were occupied, by *F. lugubris* (39), *F. paralugubris* (14) or, in a few cases, by both species (5) (Fig. 1). *F. lugubris* is much frequently encountered in the Jura Mountains, but as it is shown by the histogram in Fig. 2, *F. paralugubris* often reaches higher local abundances (huge colonies, with densities up to 16 nests/ha).

[FIGURE 1 HERE]

[FIGURE 2 HERE]

4.2 Fitted models

4.2.1 FDB, a model for F. lugubris and F. paralugubris pooled together

Five predictors have been selected to fit a general "wood ants" model (FDB, see Table 4 for model parameters and Fig. 3 for the predicted distribution map). In spite of the fact that this model is surely influenced by the larger number of *F. lugubris* occurrences (44 *vs* 19 for *F. paralugubris*), it should nonetheless draw attention to factors determining both *F. lugubris* and *F. paralugubris* distribution patterns.

[TABLE 4 HERE]

[FIGURE 3 HERE]

FDB accounts for about 36% of the null deviance (adjusted-D^2) and the correlation between observed and fitted abundances has a Spearman coefficient exceeding 0.48.

A strong positive effect of the predictor DRESLIS is evident (11.4% of explained deviance with a very significant coefficient). The location at forest borders, especially in the case of high coniferous densities, is therefore a more favourable emplacement than inner forest locations.

The average number of annual frost days during vegetation growing season (SFROYY), for which only the quadratic term was retained, seems to have a negative impact on wood ants occurrence.

For direct solar radiations, both linear and quadratic terms were retained; the global response curve adjusted by the model (Fig. 4) clearly shows that wood ants are more likely to occur on sites benefiting of a good exposure to solar radiations. As previously said, this is a very proximal variable; it is therefore not surprising that it accounts for more of one third of the total explained variance. It would be difficult to say if the curve decreasing at very high radiation levels corresponds to a regression artefact or if it actually matches the biological reality.

Finally, the degree-days of vegetation growing season (with a threshold of 3°C, DDEG300) is represented in FDB by its quadratic term only. The regression coefficient has a negative value, which probably mainly reflects a general preference for higher elevations, a result that confirms the findings of Freitag (2002).

[FIGURE 4 HERE]

Résultats tests de Wilcoxon

[TABLE 5 HERE]

Specific models

The stepwise procedure for models calibrated on *F. paralugubris* and *F. lugubris* entire datasets, PARA and LUGU, led to two six-terms models, which account respectively for 33.3% and 35.4% of the null deviance (adjusted-D^2). The value of Spearman correlation coefficient is 0.36 for PARA and 0.43 for LUGU.

Two other specific models, PARA and LUGU (Tables 6 and 7), were fitted on modified datasets, where only those absences that are shared by the two species (*i.e.* no wood ants found in a given sampling unit) were taken into account. The dataset was restricted from 144 to 105 records for *F. paralugubris* and from 144 to 120 for *F. lugubris*. *A priori*, a considerable amount of information is therefore lost and a reduction of models' predictive capacity would be expected. Actually, predictions of PARA and LUGU have an equivalent, if not superior accuracy,. Spearman correlation coefficients have a value of 0.36 for PARA and of 0.37 for PARA and keep an approximate value of 0.43 in the case of *F. lugubris*' models. In addition, explained deviance rates increased in both circumstances, precisely from 33.3%

for PARA to 44.4% for PARA and from 35.4% for LUGU to 38.3% for LUGU (adjusted-D^2). It seems therefore justified to let aside this subset of absence data when fitting specific models.
Further analysis and discussions will be exclusively based on models LUGU and PARA.

[TABLE 6 HERE]

[TABLE 7 HERE]

Predictor contributions are much easier to get through graphical plotting (Fig. 5). *F. lugubris* develops colonies more frequently at coniferous forest boundaries (a) and displays a strong preference for sunny sites (c).
As previously touched upon, DDEG300 is a predictive variable that largely expresses the shortening of warm season along the altitudinal gradient. The apparent and somehow unexpected negative impact of high amounts of degree-days (b) actually masks the preference of *F. lugubris* for high elevations.
The response to the number of frost days' gradient is the sum of the linear plus the quadratic contributions (d). Until an approximate threshold of 15 frost days the predicted abundance of *F. lugubris* is positively affected; beyond about 30 frost day, however, conditions become unfavourable.

[FIGURE 5 HERE]

[FIGURE 6 HERE]

Plots in Fig. 6 show that *F. paralugubris* prefers a good exposure to direct solar radiations (d) and locations at high elevations (c, see the above interpretation for *F. lugubris*).
The model adjusts the species' response to the number of frost days by a negative and linear regression with a y-intercept at zero (null contribution on predicted values). This parameter therefore never has a positive impact on the predicted abundance of *F. paralugubris*. ### **LA COURBE A CHANGE (MODELE1): QUADRATIQUE!###**
Finally, the response to the topographic position is an unimodal curve centred on zero. TOPO expresses the relative topographic exposure (ridge, slope, toe slope, etc) at various spatial scales hierarchically integrated into a single grid. Generally speaking, values near zero correspond to sites for which elevation is approximately equivalent to the average elevation of the surrounding terrain. It can be either regular slopes or flat areas.

Note that neither PRECYY (sum of average monthly precipitations), ASPSIN (Eastness) nor SDIFYY (average monthly mean of diffuse solar radiations) were selected for any of the models; These

predictors are not able to significantly reduce the residual deviance of modelled points distribution further on.

4.3 Potential habitat distribution maps

Maps describing the potential habitat distribution for target species are provided by the implementation of the GLM formulas in a GIS (ArcInfo®). Environmental conditions in each 25 square meters of the Jura Vaudois region are established by the overlapping of predictor layers. This spatially explicit information is used by the GLM, calibrated on occurrence data from sampling units, to attribute a predicted value to each pixel. The modelled response variables consist here in *F. lugubris* (Fig. 5a) and *F. paralugubris* (Fig. 5b) nest abundances. A quicker reading of nest abundance distribution patterns, which should express suitable *vs* unsuitable habitat spatial distribution, is allowed by the clustering of predicted values in four general categories: very low, low, high and very high densities.

The results presented in Fig. 1 were taken into account for the choice of density classes' intervals. To be precise, the "very low" class was defined as the interval of predicted values ranging between 0 and 1 nests/ha. The "very high" class covers the interval including all high-density values for which a given species has only few observations (above 3 nests/ha for *F. lugubris* and above 10 nests/ha for *F. paralugubris*. Between these two extreme-ranges, two equal intermediate classes were created (1-2 and 2-3 nests/ha for *F. lugubris*, 1-5 and 5-10 nests/ha for *F. paralugubris*). This classification is of course at least partially subjective but, to our sense, biologically reasonable, since it is based on observed specific abundance occurrences. Such strategy provides a coherent framework ensuring continuity among potential distribution maps, in spite of the pronounced social structure differences among the compared species. Being aware of the limits of our models predictive accuracy, we won't go into any finer inquiry of predicted nest abundances.

[FIGURE 5 HERE]

In Fig. 5a and 5b, a zoom on a detail of the study area shows prediction patterns at a higher resolution. A curious complementarity of best habitat locations is frequently encountered.

4.4 Comparing F. lugubris and F. paralugubris

The reciprocal analysis consisting in fitting a GLM for *F. lugubris* using the selected set of predictors of *F. paralugubris*, LUGU', and conversely, PARA', gave the results summarised in Table 8 and Table 9.

[TABLE 8 HERE]

[TABLE 9 HERE]

Terms that correspond to valuable predictive variables for one species but prove themselves non significant when used to fit a model for the other species, are considered as potential main niche differences between *F. lugubris* and *F. paralugubris*. This scenario is encountered twice. First, the topographic position (TOPO) is only significant for *F. paralugubris*. Second, the quadratic term of the number of frost-days (SFROYY) only gives a significant contribution to the deviance reduction when fitting the distribution of *F. lugubris*. This latter result confirms that the two species own a different response towards the mentioned predictor (quadratic-unimodal *vs* a linear).

Moreover, the position relatively to the forest corrected by coniferous density (FORLIS), who is a very explicative predictor in the model LUGU, has a significant regression coefficient for PARA' as well; however, both explained deviance (3.04% *vs* 11.64% for LUGU) and significance level (p-value = 0.011 *vs* p-value = 6.84E-08 for LUGU) are in this case clearly lower, allowing to affirm that here is another unambiguous difference in *F. lugubris* and *F. paralugubris* best habitat conditions.
Direct solar radiations (SDIRYY, linear + quadratic terms) and degree-days (DDEG300, quadratic term) were selected in both PARA and LUGU models… **### LES COMMENTAIRES QUI SUIVAIENT CA NE SONT PLUS VALABLES POUR LES MODELES 1…###**

5 Discussion

As one can see from results of the model FDB and the reciprocal models LUGU' and PARA', some important relations with environmental factors are clearly shared by the two modelled species.
First of all, it appears essential for wood ants that locations they occupy should be well exposed to direct solar radiations. Moreover, high-elevation ranges are more likely to be colonised by these insects (negative regression curve on the number of degree-days above 3°C). This may reflect, at least to some extent, their adaptation towards conifers, trees that became dominant in cold environments. Actually, wood ants abundantly feed on aphid-extracted coniferous honeydews (Cherix 1981), build their mounds with needles and protect themselves against fungal and bacterial pathogens by gathering large quantities of solidified resin (Christe *et al.* 2003). Indeed, according to Freitag (2002), both species are mostly distributed above 1200 metres.
The positive effect of a location at forest boundaries is easily interpretable not only in terms of stronger solar radiation, but also by the fact that both *F. lugubris* and *F. paralugubris* need to start new colonies by temporary social parasitism on nests of the *Serviformica* ant genus (Kutter 1969, Hölldobler & Wilson 1990). This latter species are typically found in open habitats and along woodland borders.

To affirm that such fundamental similarities exist is not surprising at all; on the contrary it is quite a trivial, but not useless, statement. It could explain why model predictive accuracy cannot be improved at all by those absence points that match with a sibling species' occurrence (PARA, LUGU). From an evolutionary perspective, this partial niche overlap might have been even greater in the past, which probably implied a further differentiation between the two species in order to minimise competitive interactions.

The general model FDB has a greater predictive accuracy (rho_{FDB2} = 0.48) compared to separate specific models (rho_{LUGU} = 0.36, rho_{PARA} = 0.43), since more realistic regressions can be adjusted when using a larger sample.
However, this is true only for predictors that generate similar responses for both species, basically direct solar radiation and degree-days. Meanwhile, the percentage of unexplained (residual) deviance is greater than in the case of PARA or LUGU. This result can be interpreted as the expression of an actual unshared part of the respective ecological niches, that can obviously not be modelled when *F.lugubris* and *F. paralugubris* occurrence data are pooled together.

The fitting of specific models PARA, LUGU and their comparison through the reciprocal analysis in PARA' and LUGU' revealed that *F.lugubris* and *F. paralugubris* react in slightly different ways to three

of the considered variables: the position relatively to the forest (DRESLIS), the topographic position (TOPO) and the number of frost-days during the growing season (SFROYY). These ants seem therefore to have undergone a niche differentiation that involves distinct reactions to environmental conditions. This is confirmed by a visual comparison of model predictions (Fig. 5): distribution patterns are clearly different between the two species, although areas likely to host at least low nest densities match together to a great extent (i.e. sympatry).

First, DRESLIS is the term that accounts for the greatest percentage of explained deviance in the model LUGU (11.64%), whereas it is the less explanatory predictor when the same model is calibrated on *F. paralugubris* occurrence data, i.e. PARA' (only 3.04% of the null deviance).
Reasons for which *F.lugubris* would be much more tightly associated with forest borders than *F. paralugubris* can be found in their colonisation behaviours (dispersal *vs* local strategies). *F. paralugubris* tend to spread locally and occupy a suitable patch with numerous connected nests. On the other hand, *F.lugubris* chooses more long-distance dispersion of its queens. This behavioural divergence allows the first species to profoundly penetrate into the forest whilst it keeps the latter more dependent from *Serviformica* slave ants and thus at forest boundaries. Queens of weakly polygynous and monocalic populations of *F.lugubris* have been shown to have a great capacity for social temporary parasitism (Castella 2003).

Second, from a biological point of view, an interpretation of the response of *F. paralugubris* to the topographic position (TOPO) is rather risky. What can be said is that this species seems to display a preference for locations having an elevation being approximately equivalent to the average elevation of the surrounding terrain (values of TOPO near zero), i.e. not ridges nor slope toes. A partial explanation could be that, being a species that frequently dwells in deep forest, locations in slope toes or valley bottoms (negative values of TOPO) may be insufficiently exposed to solar radiations and therefore too cold and humid. This problem would be reduced for *F. lugubris*, as its mounds are more frequently located at the forest edge. For the apparent unsuitability of ridges and peaks (positive values of TOPO) we do not yet have any particular explanation.

The third main difference rising from models concerns the impact of the number of frost-days per growing season (SFROYY). Abundance of *F. paralugubris* appears to be negatively related with this climatic variable, indicating that the species poorly endures frost during the period of activity. By looking at the potential distribution map of this species (Fig. XXX) and at the meantime at the SFROYY layer (Annexe D), the key role played by this variable in driving the potential distribution of *F. paralugubris* becomes easily recognizable. The most suitable sites are mainly grouped in the Southern part of the Jura, where frosts outside the winter season are less recurrent. On the contrary, *F. lugubris*

displays an optimum corresponding to more or less 15 frost-days per growing season and apparently only suffers from this kind of climatic event when frequency exceeds about 30 days per growing season.

Open environments are known to involve a quicker heat loss at the ground during the night. Moreover, the soil is more frequently free from the snow cover, because of quicker melting in sites exposed to solar radiations. These factors are supposed to intensely expose anthills to frost. It could be that, because of living mostly in the zone of transition between forest and meadows, *F. lugubris* evolved towards an increased resistance against frost-induced injuries.

5.1 Some critics…

Although the above-discussed results may appear quite clear, they should nonetheless be taken with care. Many parameters at different stages of the modelling process remains in fact hardly controlled.
First of all, the morphological species determination of *F. lugubris* and *F. paralugubris* still remains very difficult and the influence on models of a few identification errors is potentially non-negligible.

In addition, still very little is known about the potential existence of some degree of hybridisation between *F. lugubris* and *F. paralugubris* (for evidences of hybridisation between other species of the *Formica rufa* group, namely *F. polyctena* and *F.rufa*, see Seifert 1991).

Spearman correlation tests provide, for all models, quite modest evaluations. We principally ascribe these results to the great complexity of underlying reasons determining nest abundance pattern in wood ants species. Particularly, to assume that nest abundance indirectly measures habitat suitability is probably a useful but weak postulate. Conclusions on *F. lugubris* only apply to the oligogynous-monocalic populations and not to the alternative polygynous-polycalic form, found in the Alps.
In fact, the true absence of the alternative form (i.e. polygynous-polycalic) of *F. lugubris* in the Jura Mountains cannot be completely assured. This eventuality could seriously affect predictive capacity of models fitted on abundance data. A sharp difference would exist between one polycalic colony, composed for example of four cooperative nests, and four competitive monocalic colonies.
On the other hand, the high polygyny and consequent polycalism of *F. paralugubris* is probably the outcome of a selection for individuals minimising the risks of dispersal and social parasitism (Rosengren, Sundström & Fortelius 1993). Nevertheless, it still remains difficult to evaluate the real significance of environmental trophic and climatic conditions for the phenomena of budding leading to polycalism. High nest densities could indeed be the consequence of a poor and rude habitat, as suggested by studies on a huge super-colony in the Jura (Cherix 1981).

Moreover, the data transformation into relative abundances when building a common predictive model for the two sibling species (FDB) was a necessary but nonetheless rather subjective way of proceeding. This highlights the problem of comparing the abundance spatial distribution of two ant species displaying different social structures (polycalism *vs* monocalism).

Finally, one should be aware that the generalisation of our models to geographical areas other than the Jura biogeographical region would probably be a hazardous process, as biotic interactions and stochastic effects are likely to differ from one region to another (Guisan & Theurillat 2000, *in* Guisan *et al.* 2002).

6 Conclusions

Despite the numerous simplifications and inevitable errors that are inherent to the process of modelling the spatial distribution of a species, as well as the lack of microclimatic-meaningful predictors, concrete findings about ecological preferences of *F.lugubris* and *F. paralugubris* were reached through this approach.

Admitting that our modelling results actually proceed from real biological facts, these would provide new substantial explanations through which the sympatry of the studied sibling ant species and their local spatial distribution patterns could be more easily understood.

In conclusion, without trying to define any cause-to-effect relationship, we can say that a coherent picture including behavioural, physiological and ecological findings is gradually emerging from recent research on wood ants. This knowledge has to be seen not only as a powerful mean for ant researchers self-enjoyment and intellectual satisfaction, but chiefly as a solid base on which efficient conservation strategies can be designed.

For instance, as *F. paralugubris* was only described in 1996, the actual Swiss Red List only considers *F. lugubris sensu lato* (Agosti and Cherix 1994). It would be useful to re-evaluate these species status as soon as possible. Habitat suitability models and potential distribution maps will be helpful for a better assessment of *F. lugubris* and *F. paralugubris* populations and of the potential threatens burdening on them.

We consider our first attempt in modelling the habitat of wood ant species by means of GIS and GLMs a valuable contribution for the conservation of these near threatened European insects.

7 Perspectives

First, it would be very useful to return on the field and collect individuals to perform the identification of some doubtful sample for a second time, possibly using a genetically based method. Further field work should also be carried out to enlarge and balance our sampling size (recorded *F. paralugubris* occurrences being nowadays still rather scarce).

Second, the climatic, topographic and environmental spatially explicit predictors available today are still rather inadequate for a work on insect habitat at the local scale. More accurate parameters need to be included in such distribution predictive models. We suggest two kinds of variables which, because of their influence on microclimatic conditions, could prove particularly efficient for ant habitat modelling purposes: (i) a predictor related with the spring melting of the snow cover, which is supposed to affect the starting date of ant's activity season; and (ii) more detailed and updated forest covers.

Finally, with regard to the statistical framework, a supplementary method could be applied on our data. In the present study, predictors that were retained for both *F. lugubris* and *F. paralugubris* models were evaluated on their coefficient's degree of significance and on average percentage of explained deviance to judge if one environmental factor has a greater impact on one species. A more detailed analysis of model results could be achieved by a bootstrap approach followed by statistical tests on each predictor's regression coefficient (average ± standard error, provided by the bootstrapping). Compared coefficients would come from same models applied to the two species (that is to say PARA *vs* LUGU' and LUGU *vs* PARA'). This technique would allow stressing out other possible model divergences, which, being less evident, could be missed through the methods used here.

8 References

AGOSTI D. AND CHERIX D. 1994. Liste rouge des fourmis menacées de Suisse. In: Office fédéral de l'environnement, des forêts et du paysage (OFEFP) (eds). Listes rouges des espèces menacées de Suisse, pp. 45–47. Berne, Switzerland.

AUSTIN, M.P. (2002). Spatial prediction of species distribution: an interface between ecological theory and statistical modeling. *Ecological Modelling*, 157, 101-118.

BAKKENES, M., J. R. M. ALKEMADE, F. IHLE, R. LEEMANS & J. B. LATOUR (2002). Assessing the effects of forecasted climate change on the diversity and distribution of European higher plants for 2050. *Global Change Biology*, 8, 390 - 407.

BERNASCONI, C. (2002). Structure sociale et génétique de deux populations alpines de *Formica lugubris*. *Diplôme de biologie de l'Université de Lausanne*, 29 p. + annexes.

BOLLIGER, J., KIENAST, F., ZIMMERMANN, N.E. (2000). Risks of global warming on montane and subalpine forests of Switzerland - a modeling case study. *Regional Environmental Change*, 1, 99-111.

BORGES, A. (2001). Sampling along the trails to model the distribution of wood ants in the Swiss National Park? *Diplôme Postgrade Systématique et Gestion de la biodiversité de l'Université de Lausanne*, 13 p.

BROENNIMAN, O. (2003). Modéliser la distribution potentielle des espèces végétales rares et menacées. *Travail de diplôme, Université de Lausanne*.

CASTELLA, G. (2003). Structure sociale et parasitisme social temporaire chez *Formica paralugubris. Travail de diplôme, Université de Lausanne*, 45p. + annexes.

CHERIX, D., MAEDER, A., BERNASCONI, C., CASTELLA, G., FREITAG, A. (2004). Red wood ants: new species and new questions! *Ins. Soc. Life, in press*.

CHERIX, D. (1981). Contribution à la biologie et à l'écologie de *F.lugubris* Zett. (Hymenoptera, Formicidae). Le problème des super-colonies. *Thèse (non publiée), Université de Lausanne*.

CHERIX, D. (1991). Red wood ants. *Ethology, Ecology & Evolution*, Special Issue 1, 165.

CHRISTE P., OPPLIGER A., BANCALÀ F., CASTELLA G. & CHAPUISAT M. (2003). Evidence for collective medication in ants. *Ecology Letters*, 6, 19-22.

DEVENOGES, A. (1999). Distribution des fourmis des bois (groupe *rufa*) au Parc national Suisse, PNS. Applications à l'aide d'un système d'information géographique, SIG. *Diplôme de biologie de l'Université de Neuchâtel*, 49 p. + annexes.

FRANKLIN, J. (1995). Predictive vegetation mapping of biospatial patterns in relation to environmental gradients. *Progress in Physical Geography*, 19, 474-499.

FREITAG, A. (2002). « FORMIS-2000 » Inventaire et protection des fourmis des bois du canton de Vaud. En collaboration avec Antoine Guisan et Daniel Cherix. *Musée cantonale de zoologie, Lausanne*. 36 p. + annexes.

GERRARD, R., P. STINE, R. CHURCH, GILPIN, M. (2001). Habitat evaluation using GIS, a case study applied to the San Joaquin kit fox. *Landscape and Urban Planning*, 52, 239-255.

GUISAN, A., EDWARDS, T.C., HASTIE, T. (2002). Generalized linear and generalized additive models in studies of species distribution: setting the scene. *Ecological Modelling*, 157, 89-100.

GUISAN, A. & J. - P. THEURILLAT (2001). Assessing alpine plant vulnerability to climate change: A modeling perspective. Integrated Assessment, 1, 307 - 320.

GUISAN, A., ZIMMERMANN, N.E. (2000). Predictive habitat distribution models in ecology. *Ecological Modeling*, 135, 147-186.

HARRELL, F.E., LEE, K.L., MARK, D.B. (1996). Multivariable prognostic models: Issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Statistics in Medicine*, 15, 361- 387.

HEIKKINEN, R.K. & NEUVONEN, S. (1997). Species richness of vascular plants in the subarctic landscape of northern Finland: modelling relationships to the environment. *Biodiversity and conservation*, 6, 1181-1201.

HIRZEL, A., GUISAN, A. (2002). Which is the optimal sampling strategy for habitat suitability modelling? *Ecological Modeling*, 157, 329-339.

HÖLLDOBLER B. & WILSON E. O. (1990). *The Ants.* Cambridge, Mass., Harvard University Press.

HUTCHINSON, G.E. (1978). An Introduction to Population Ecology. *Yale University Press*, New Haven, CT, 260 p.

JABERG, C., GUISAN, A. (2001). Modelling the distribution of bats in relation to landscape structure in a temperate mountain environment. *Journal of Applied Ecology*, 38, 1169-1181.

KUTTER H. (1969). Die sozialparasitischen Ameisen der Schweiz. *Neujahrsblatt. Naturforschenden Gesellschaft in Zürich*, 5-62.

LEHMANN, A., LEATHWICK, J.R., OVERTON, J.MCC. (2002). ASSESSING New Zeland fern diversity from spatial predictions of species assemblages. *Biodiversity and conservation*, 11, 2217-2238.

MAEDER, A., CHERIX, D. (2001). Problèmes poses par *Formica paraluguris*, une nouvelle espèce de fourmis des bois. *Actes Coll. Insectes Sociaux*, 14, 21-25.

MAGGINI, R., GUISAN, A., CHERIX, D. (2002). A stratified approach for modeling the distribution of a threatened ant species in the Swiss National Park. *Biodiversity and Conservation*, 11, 2117-2141.

PAMILO, P., CHAUTEMS, D., CHERIX, D. (1992). Genetics differentiation of disjunct populations of the ants *Formica aquilonia* and *Formica lugubris* in Europe. *Insectes Sociaux*, 39, 15-29.

PATTHEY, P. (2003). Habitat and corridor selection of an expanding red deer (*Cervus elaphus*) population. Thèse, Université de Lausanne.

RADELOFF, V., PIDGEON, A.M., HOSTERT, P. (1999). Habitat and population modelling of roe deer using an interactive geographic information system. *Ecological Modelling*, 114, 287-304.

REBETEZ M. (2002). La Suisse se rechauffe. *Presses Polytechniques et Universitaires Romandes*, Lausanne.

ROSENGREN R., SUNDSTRÖM L., FORTELIUS W. (1993). Monogyny and polygyny in *Formica* ants: the result of alternative dispersal tactics. In: *Queen number and Sociality in Insects* (L. Kellller, ed.), Oxford University Press. pp 308-333.

ROSENGREN, R., CHAUTEMS, D, CHERIX, D., FORTELIUS, W., KELLER, L. (1994). Separation of two sibling species of *Formica* L. ants by a behavioural choice test based on brood discrimination. *Memorabilia Zoologica*, 48, 237-249.

SCOTT, J.M., HEGLUND, P.J., HAUFLER, J.B., MORRISON, M., RAPHAEL, M.G., WALL, W.B., SAMSON, F. (EDS) (2002). Predicting Species occurrences: Issues of accuracy and scale. Island press, Covelo, California.

SEIFERT, B. (1991). The phenotypes of the *F. rufa* complex in East Germany. *Abh.Ber. Naturkundemus. Görlitz*, 65:1, 1-27.

SEIFERT, B. (1996). Formica paralugubris nov. spec. - a sympatric sibling species of Formica lugubris from the western Alps (Insecta: Hymenoptera: Formicoidea: Formicidae). *Reichenbachia Mus. Tierkd. Dresden* 31, 193-201.

STORE, R., JOKIMÄKI, J. (2003). A GIS-based multi-scale approach to habitat suitability modeling. *Ecological Modeling* 169, 1-15.

9 Acknowledgements

TABLES AND FIGURES

Table 1: Eco-geographic variables used for models calibration. Climatic variables consist in average values over the period between 1961 and 1990.

Variable	Units	Information type	Abbreviation
Number of frost days during plants' growing season.	days	climatic	SFROYY
Degree-days of growing season with a 3°c treshold	day*deg	climatic	DDEG300
Direct solar radiation (average monthly mean)	KJ/day	climatic	SDIRYY
Diffuse solar radiation (average monthly mean)	KJ/day	climatic	SDIFYY
Sum of average monthly precipitation	1/10mm/mth	climatic	PRECYY
East-West exposure gradient	sin(aspect angle)	topographic	ASPSIN
Topographic position	unitless	topographic	TOPO
Mean coniferous density	% of surface	environmental	DRES200
Position relatively to the forest (forest edge vs deep forest)	unitless (binary 1/0)	environmental	FORLIS
"Forest edge effect" pondered by coniferous density	% of surface	environmental	DRESLIS

Table 2: Values of Spearman correlation between the selected predictors. Correlations were estimated on the values of a sample of 10'000 pixels randomly chosen on the study area. The maximum correlation is found between SFROYY and DDEG300 (correlation coefficient = -0.55).

	TOPO	DDEG300	ASPSIN	SFROYY	SDIFYY	SDIRYY	DRES200
TOPO	1.00						
DDEG300	-0.43	1.00					
ASPSIN	-0.09	0.28	1.00				
SFROYY	0.18	-0.55	-0.08	1.00			
SDIFYY	0.11	-0.01	0.01	0.10	1.00		
SDIRYY	0.03	0.12	0.35	0.04	0.19	1.00	
DRES200	0.12	-0.04	0.05	-0.26	0.12	0.01	1.00

Table 3: The 144 sampling units that provided data for habitat modelling were as equally distributed as possible among the 16 strata of the sampling; data were collected during two field campaigns sharing the same sampling design. Strata: first digit, 1=low elevation 2=high elevation; second digit, 1=North-West slope aspect 2=South-East slope aspect; third digit, 1=gentle slope 2=steep slope; fouth digit, 1= deep forest 2= forest border.

##A INTEGRER DS TEXTE##

Strata	Number of replicata
1111	9
1112	9
1121	9
1122	8
1211	10
1212	9
1221	9
1222	7
2111	10
2112	9
2121	9
2122	9
2211	10
2212	9
2221	9
2222	9
total	**144**

Table 4: Parameters of model FDB. Predictors selected by the stepwise procedure (see text for details on model selection criteria); regression coefficients with standard error and significance level of the t-test; Average explained deviance and standard error of each predictor were calculated on n! ANOVAs (n being the number of terms in the model formula, five in this case) each one corresponding to a different permutation of the entrance order of predictors; D2 = total percentage of null deviance explained by the model; adjusted-D2 = D2 corrected by a ratio of the number of predictors to the number of observations; rho = Spearman correlation coefficient between fitted and observed values.

Predictor	Regression coefficient	Standard error	p-value (coefficient)	Average % expl. deviance	Standard error
(INTERCEPT)	-12.282	1.721	***	-	-
DRESLIS	1.62E-02	1.29E-03	***	11.40	2.92
SFROYY^2	-1.24E-03	1.05E-04	***	8.87	3.20
SDIRYY	1.86E-03	2.11E-04	***	8.51	2.48
SDIRYY^2	-5.29E-08	6.43E-09	***	6.62	1.80
DDEG300^2	-6.18E-07	6.81E-08	***	2.36	2.38

Significance levels: ***<0.001, **<0.01, *<0.05, NS≥0.05.

D^2 = 0.380 ; adjusted-D^2 = 0.358 ; rho = 0.484.

Table 5: Parameters of model L/P. Binomial GLM. Response: occurrence of *F.lugubris* = 1, occurrence of *F.paralugubris* = 0. See Table 3 for more explanations.

Predictor	Regression coefficient	Standard error	p-value (coefficient)	Average % expl. deviance	Standard error
(INTERCEPT)	-1.071	1.006	**0.287**	-	-
SFROYY^2	-0.010	0.004	0.017	5.87	7.61
SFROYY	0.380	0.161	0.018	5.39	7.61

Significance levels: ***<0.001, **<0.01, *<0.05, NS≥0.05.

D^2 = 0.113 ; adjusted-D^2 = 0.077

Table 6: Parameters of model PARA. See Table 3 for explanations.

Predictor	Regression coefficient	Standard error	p-value (coefficient)	Average % expl. deviance	Standard error
(INTERCEPT)	2.742	0.913	***	-	-
TOPO^2	-4.91E-04	9.85E-05	***	10.79	1.48
DDEG300^2	-1.56E-06	2.42E-07	***	7.49	4.74
SFROYY	-0.222	0.050	***	6.45	3.75
SDIRYY	2.12E-04	4.04E-05	***	4.95	0.96
SFROYY^2	0.004	0.001	***	4.29	3.68
DRESLIS	0.013	0.004	***	2.11	0.65

Significance levels: ***<0.001, **<0.01, *<0.05, NS≥0.05.

D^2 = 0.361 ; adjusted-D^2 = 0.333 ; rho(Spearman) = 0.361.

Table 7: Parameters of model LUGU. See Table 3 for explanations.

Predictor	Regression coefficient	Standard error	p-value (coefficient)	Average % expl. deviance	Standard error
(INTERCEPT)	-14.429	4.518	***	-	-
DRESLIS	0.018	0.003	***	11.64	2.77
SFROYY^2	-0.006	0.001	***	8.12	2.25
SDIRYY	0.002	5.47E-04	***	6.16	2.00
SFROYY	0.209	0.045	***	5.40	1.76
SDIRYY^2	-4.58E-08	1.67E-08	***	5.02	1.86
DDEG300^2	-5.83E-07	1.85E-07	***	1.76	1.36

Significance levels: ***<0.001, **<0.01, *<0.05, NS≥0.05.

D^2 = 0.381 ; adjusted-D^2 = 0.354 ; rho(Spearman) = 0.426 .

Table 8: Parameters of model LUGU'. See Table 3 for explanations.

Predictor	Regression coefficient	Standard error	p-value (coefficient)	Average % expl. deviance	Standard error
(INTERCEPT)	-2.906	0.896	***	-	-
DRESLIS	0.021	0.003	***	12.30	2.17
SFROYY^2	-0.007	0.001	***	8.67	2.10
SDIRYY	1.87E-04	3.74E-05	***	6.81	0.86
SFROYY	0.222	0.046	***	5.67	1.61
DDEG300^2	-6.14E-07	1.78E-07	***	1.90	1.50
TOPO^2	2.25E-05	2.36E-05	NS	0.22	0.22

Significance levels: ***<0.001, **<0.01, *<0.05, NS≥0.05.

$D^2 = 0.36$; adjusted-$D^2 = 0.33$.

Table 9: Parameters of model PARA'. See Table 3 for explanations.

Predictor	Regression coefficient	Standard error	p-value (coefficient)	Average % expl. deviance	Standard error
(INTERCEPT)	-15.974	4.937	***	-	-
SFROYY	-0.234	0.044	***	9.90	4.88
DDEG300^2	-1.33E-06	2.18E-07	***	9.52	5.88
SDIRYY	0.003	6.10E-04	***	7.68	1.56
SFROYY^2	0.004	0.001	***	6.36	4.76
SDIRYY^2	-7.36E-08	1.84E-08	***	6.35	1.49
DRESLIS	0.011	0.004	***	3.04	1.04

Significance levels: ***<0.001, **<0.01, *<0.05, NS≥0.05.

$D^2 = 0.29$; adjusted-$D^2 = 0.26$.

Figure 1: Study area with sampling results…

Figure 2: Percentage of the occupied sites by each species for different abundance classes. It is quite easy to recognise the two different social structures on which F. lugubris and F. paralugubris usually organize their colonies (i.e. oligocaly vs polycaly).

Figure 3: Potential distribution map generated from the model fdb, which group F. lugubris and F. paralugubris together.

Figure 4: Response curve for direct solar radiations (SDIRYY) in the model FDB, where *F.lugubris* and *F. paralugubris* common habitat is modelled. x-axe=direct radiation amount, y-axes=contribution to the predicted abundance.

Figure 5: Response curves of *F. lugubris* to the environmental gradients selected to model its habitat. x-axes=environmental gradients, y-axes=contribution to the predicted abundances. DRESLIS=coniferous density at the edge of the forest, DDEG300=degree-days above 3°C during the vegetation growing season, SDIRYY=direct solar radiations, SFROYY=number of frost days during vegetation growing season.

Figure 7: Response curves of *F. paralugubris* to the environmental gradients selected to model its habitat. x-axes=environmental gradients, y-axes=contribution to the predicted abundances. TOPO=topographic position, DRESLIS=coniferous density at the edge of the forest, DDEG300=degree-days above 3°C during the vegetation growing season, SDIRYY=direct solar radiations, SFROYY=number of frost days during vegetation growing season.

Figure 8: Potential habitat distribution map for (a) *F. lugubris* and (b) *F. paralugubris.* The predicted nest density is a continuous variable, but to allow an easier interpretation and comparison of the distribution patterns spatial predictions are represented by abundance classes (upper range limits of each class not included). Green points display observed species occurrences, on which each model was calibrated.

Figure 8: Detail of the study area with potential habitat distribution maps of a) F. lugubris and b) F. paralugubris. When comparing maps at such small scale it is possible to recognize several patches of high habitat suitability for *F. lugubris* that correspond to low habitat suitability areas for *F. paralugubris*, and *vice-versa*.

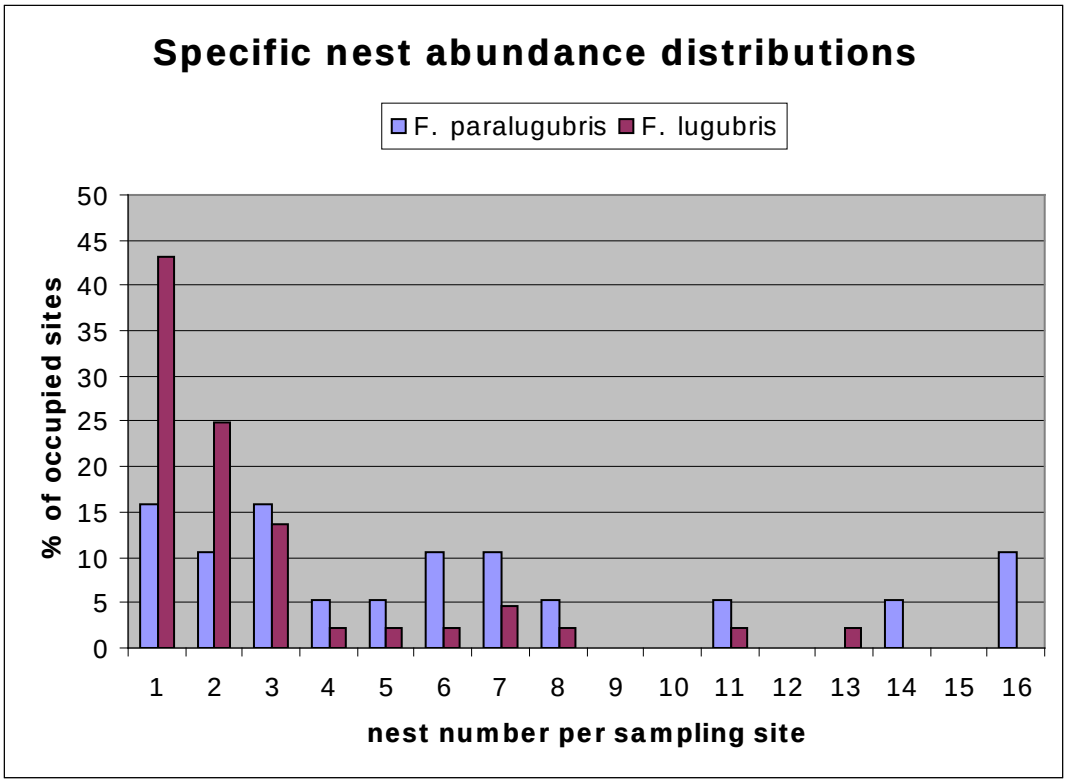

Specific nest abundance distributions
F. paralugubris
F. lugubris
% of occupied sites
50
45
40
35
30
25
20
15
10
5
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
nest number per sampling site

Figure 1

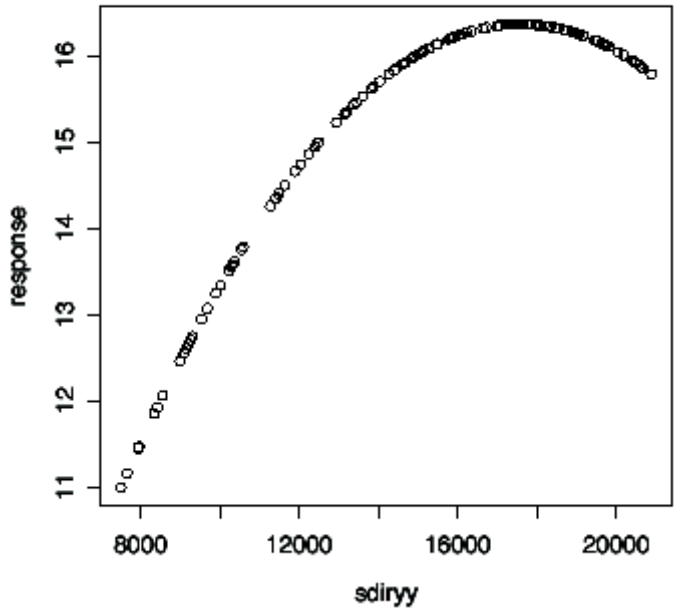


Figure 2

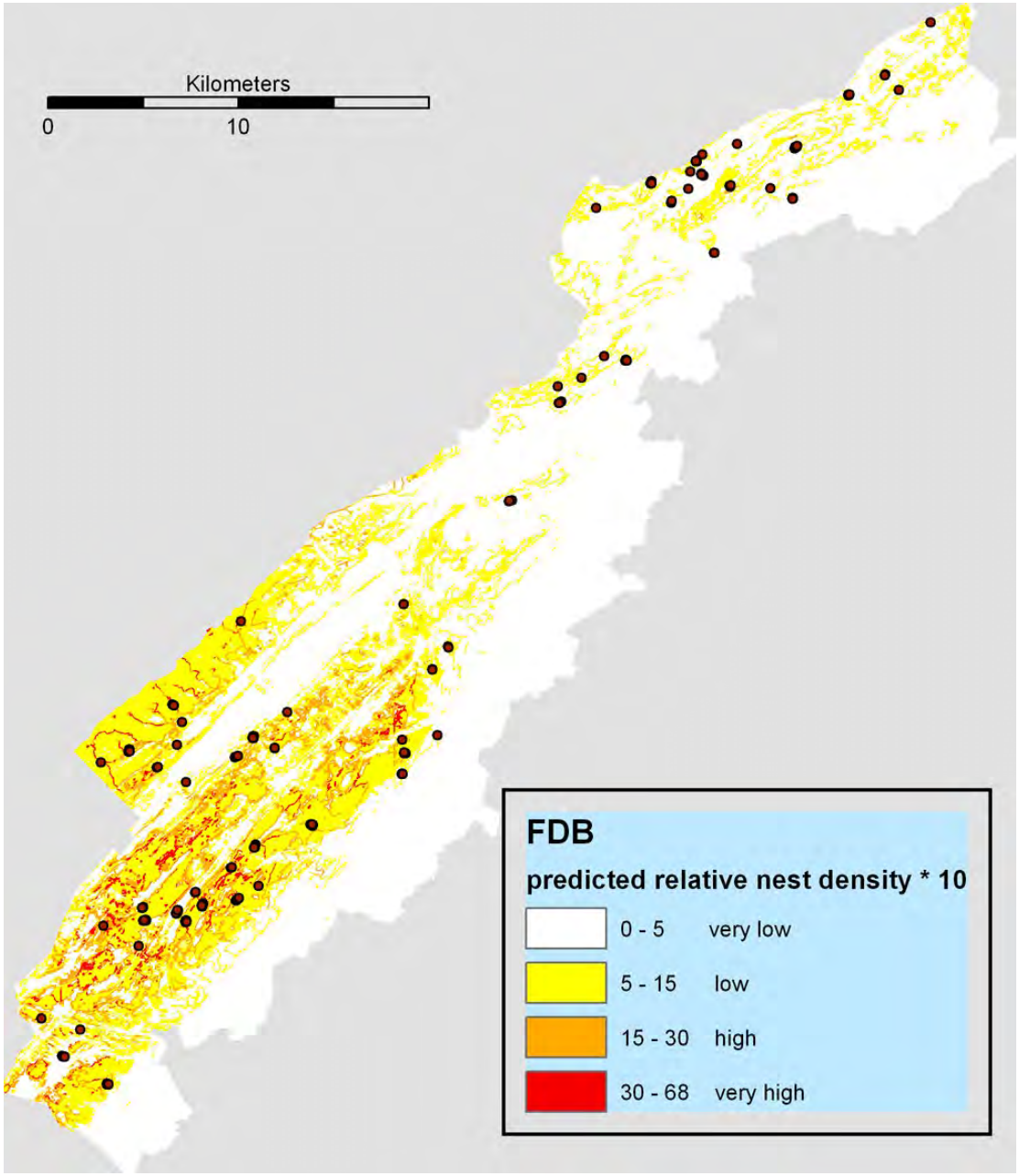

Kilometers
0
10
FDB
predicted relative nest density * 10
0 - 5 very low
5 - 15 low
15 - 30 high
30 - 68 very high

Figure 3

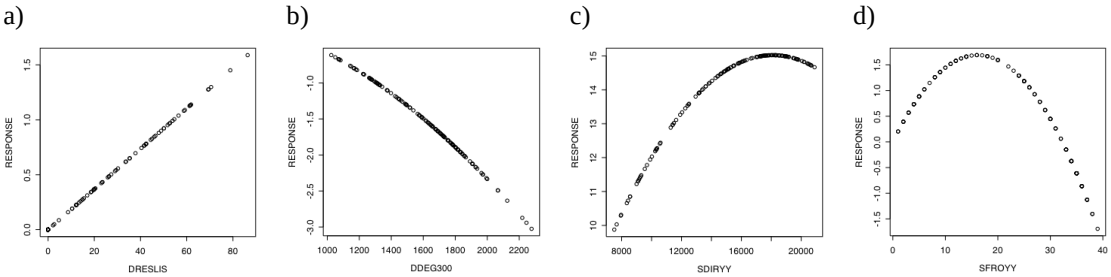


Figure 3

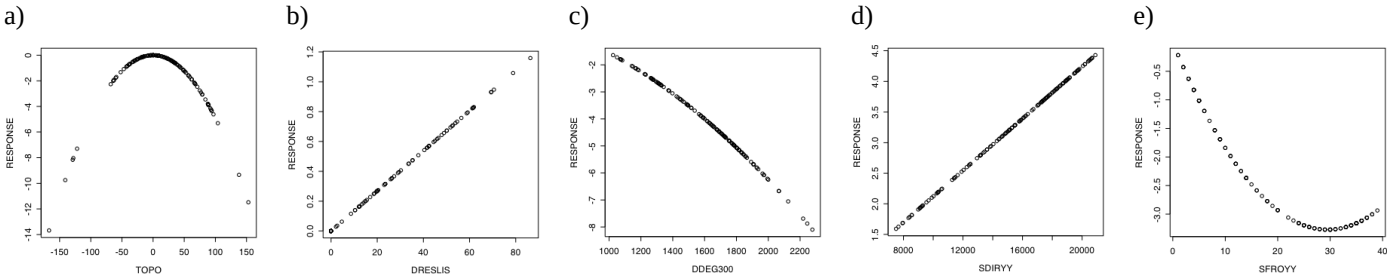
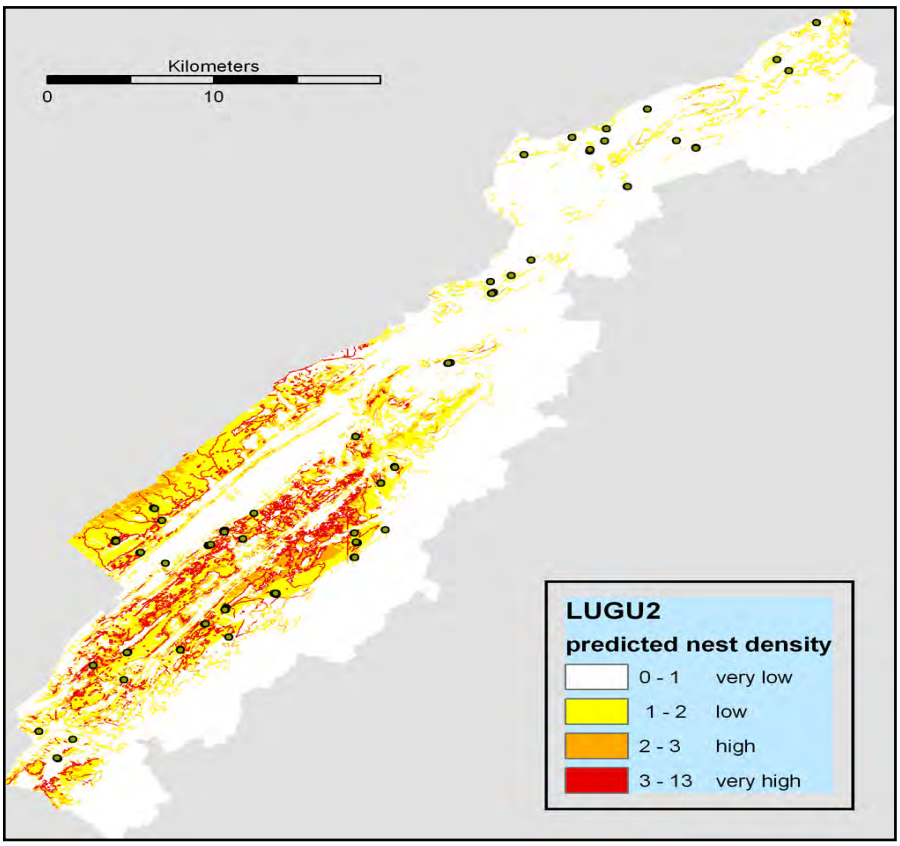


Figure 4

a)


Kilometers
0
10
LUGU2
predicted nest density
0 - 1 very low
1 - 2 low
2 - 3 high
3 - 13 very high

b)

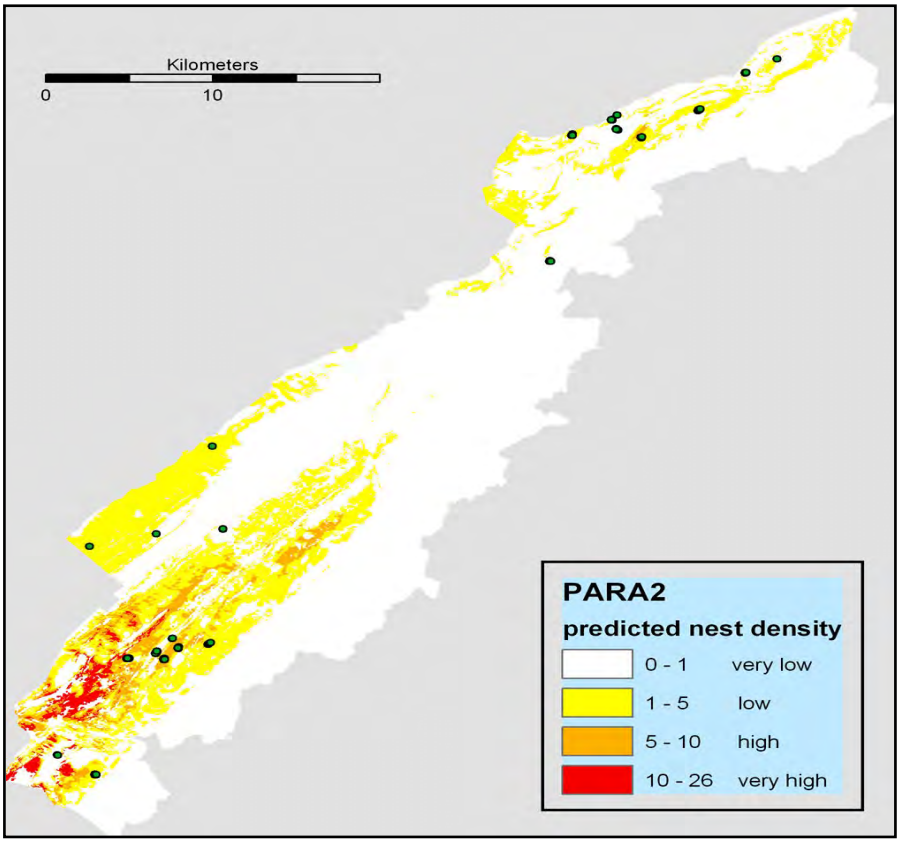

Kilometers
0
10
PARA2
predicted nest density
0 - 1 very low
1 - 5 low
5 - 10 high
10 - 26 very high

Figure 5

a) **b)**

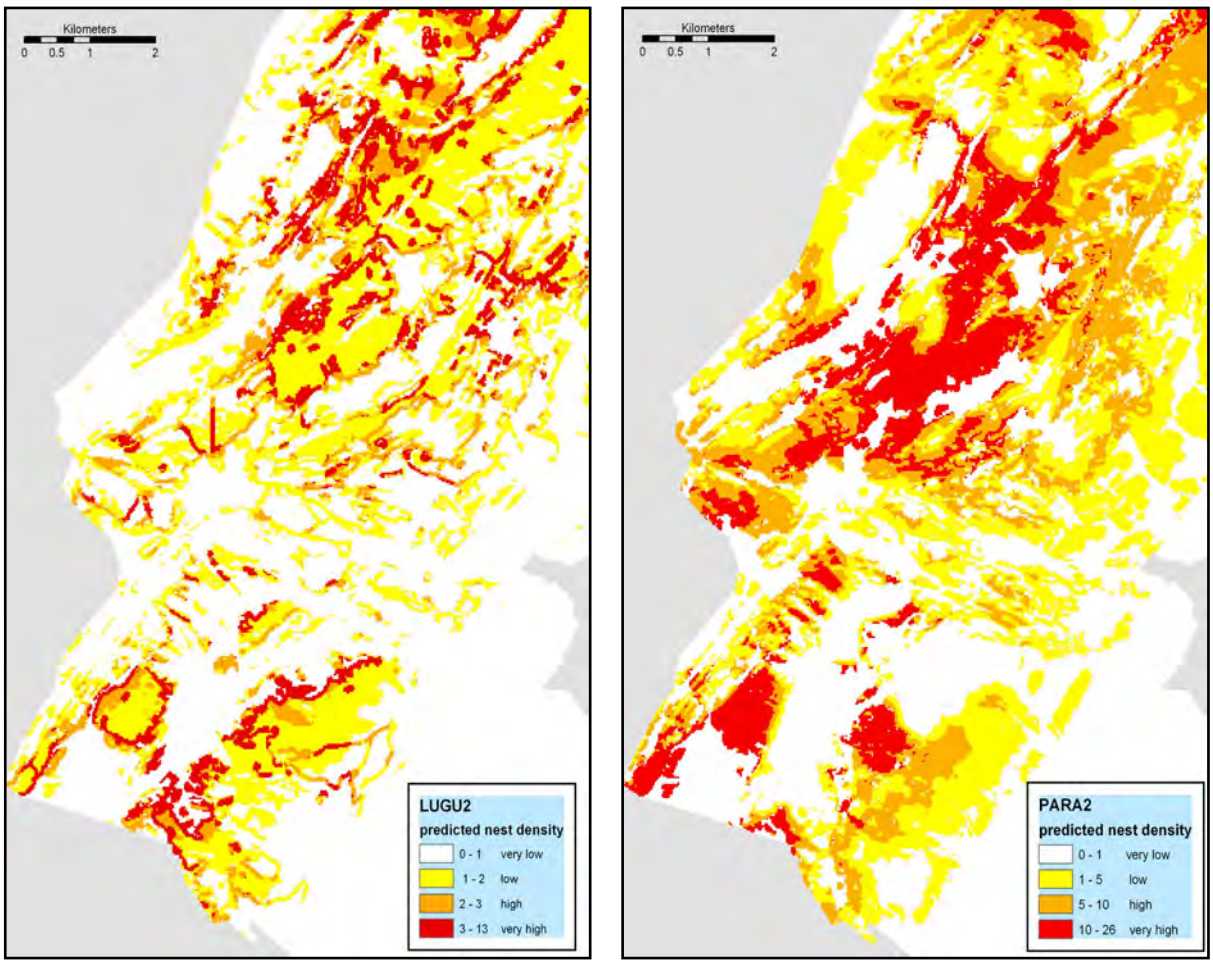
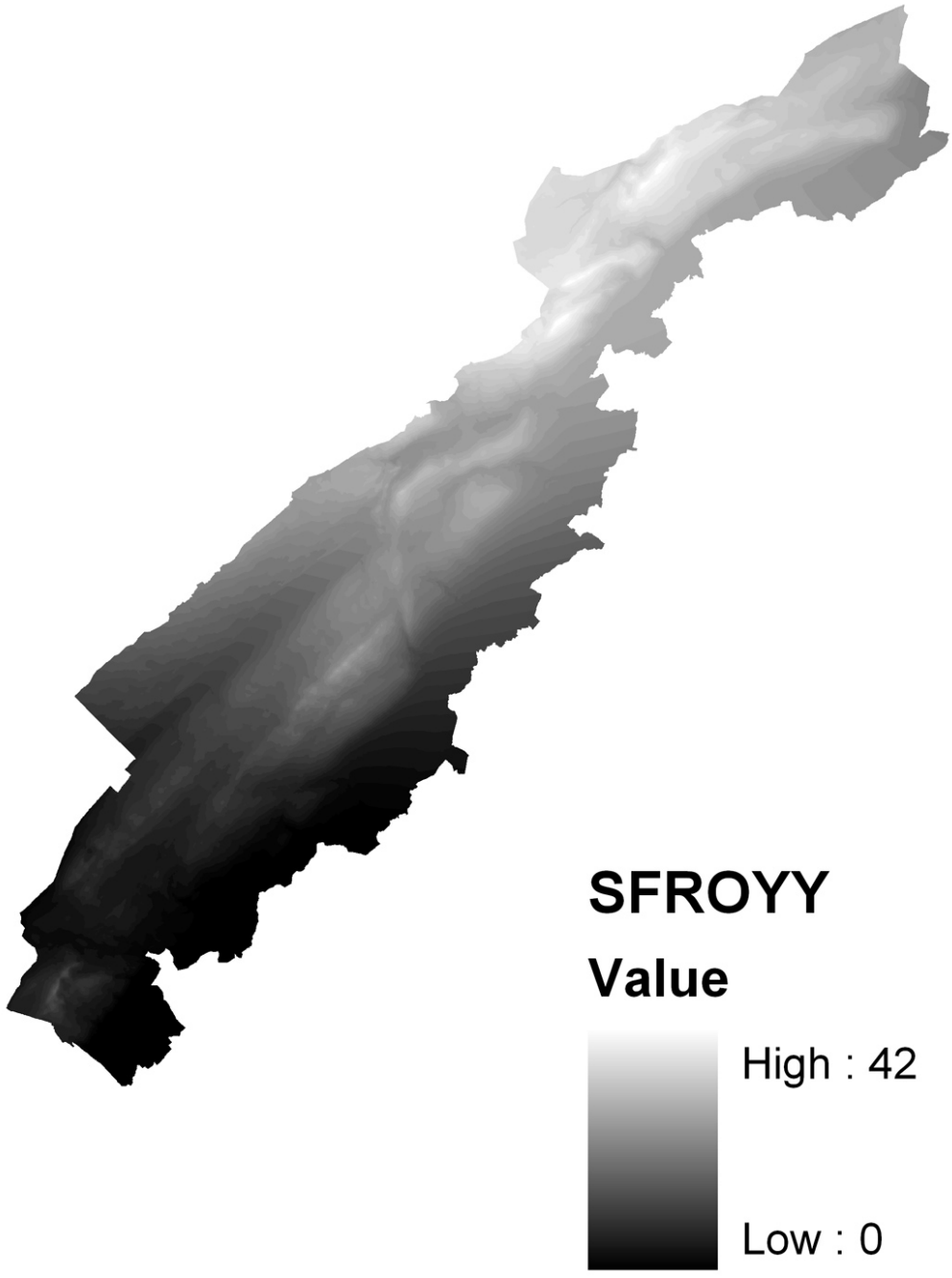

Kilometers
0 0.5 1 2
LUGU2
predicted nest density
0 - 1 very low
1 - 2 low
2 - 3 high
3 - 13 very high
Kilometers
0 0.5 1 2
PARA2
predicted nest density
0 - 1 very low
1 - 5 low
5 - 10 high
10 - 26 very high

Figure 6

ANNEXE D: Annual average number of frost days during vegetation growing season in the Jura Vaudois.



tests de Wilcoxon

Variable	p-value	
sdiryy	0.8222	NS
sfroyy	0.7988	NS
topo	0.3808	NS
ddeg300	0.2111	NS
dreslis	0.4492	NS