

Ecological niche models for four native cockroach species (Blattaria: Blattellidae: Ectobiinae) in the Netherlands

Pasquale Ciliberti¹, Herman de Jong^{1,4}, Marieke A. Schouten², Pita A. Verweij², Wijnand R.B. Heitmans^{1,3}

¹ Zoological Museum Amsterdam, Section Entomology, Plantage Middenlaan 64, 1018 DH Amsterdam, the Netherlands

² Copernicus Institute, Utrecht University, Group Science, Technology and Society, Heidelberglaan 2, 3584 CS Utrecht, the Netherlands

³ Institute of Biology, Leiden University, Kaiserstraat 63, 2300 RA Leiden, the Netherlands

⁴E-mail: H.deJong@uva.nl

Key words: *Capraiellus panzeri*, cockroaches, ecological niche modelling, *Ectobius lapponicus*, *Ectobius pallidus*, *Ectobius sylvestris*, Maxent, model stability, the Netherlands

Abstract

This paper aims at modelling the spatial distribution of the cockroach species *Capraiellus panzeri*, *Ectobius lapponicus*, *Ectobius pallidus* and *Ectobius sylvestris* within the Netherlands and comparing the habitat preferences of these species. Maxent was used to calculate habitat suitability and to identify environmental variables underlying the differences in observed distribution patterns. A sub-sampling procedure was employed to test model stability. Models were evaluated by calculating the Area Under the Curve (AUC). The analyses show that except for the coastal dune area, the western part of the Netherlands is unsuitable for the species. Suitability predictions for *C. panzeri*, *E. lapponicus* and *E. sylvestris* are very similar, with suitable areas concentrated in the eastern and the north-eastern parts of the country and along the western coast. The prediction model for *E. pallidus* is somewhat more restricted, especially in the northern part of the country. Soil type, land cover and altitudinal range are most important in predicting the distribution of all species. A correspondence analysis was performed to identify the association between the species distribution and the most influential environmental variables. Correspondence analysis indicated that the species distributions are comparably associated with soil type and land cover while species appear to have different preferences with respect to altitudinal range.

Contents

Introduction	149
Materials and methods	150
Study area and datasets	150
Modelling approach	151
Model evaluation and stability	151
Correspondence analysis	153
Results	153
Modelling results	153
Evaluation and stability of the models	153

Correspondence analysis	155
Discussion	155
Acknowledgements	156
References	156

Introduction

To gain understanding of the spatial distribution patterns of species is a general challenge of ecology. Species distribution patterns are influenced by the environmental requirements of species and inter-specific relations like parasitism, competition and predation (Costa *et al.*, 2008). Furthermore, historical factors such as geographical barriers or species properties such as dispersal ability are also responsible for the spatial distribution of species.

In practice, studies of species range size are often hampered by problems such as differences in sampling effort and temporal variation (Gaston, 1996). This can cause the observed patterns to be geographically, taxonomically and temporally highly skewed or incomplete. Today, several approaches exist to cope with insufficiently sampled datasets. Ecological Niche Modelling is a relatively new approach that generates predictions of potential distributions by relating observed points of occurrence to environmental variables (Guisan and Zimmerman, 2000; Guisan and Thuiller, 2005). An advantage of this approach is that predictions of potential distributions can be easily compared between species with relatively little sampling effort. A disadvantage may be that biotic interactions are difficult to insert in the model (Davis *et al.*, 1998; Hampe, 2004).

Here, we used ecological niche modelling to study the spatial distribution of four species of Ectobiinae cockroaches, namely *Capraiellus panzeri* Stephens, 1835, *Ectobius lapponicus* Linnaeus, 1758, *Ectobius pallidus* Olivier, 1789, and *Ectobius sylvestris* Poda, 1761 within the Netherlands. Maxent was used to calculate habitat suitability and to identify individual environmental factors underlying differences in observed distribution patterns. In order to test the stability of model output we used a sub-sampling procedure. It is known, in fact, that ecological niche modelling is sensitive to the number and identity of the occurrence points (Pearson *et al.*, 2007). Furthermore, we attempted to discern the differences in response to the environmental variables by performing a correspondence analysis.

Although the ectobiinid cockroaches belong to different phylogenetic lineages (Failla and Messina, 1978; Bohn, 1989), they are biologically very similar species: they have a mainly vegetarian diet, consuming plant litter material and soil particles. The occurrence of cockroaches is not related to any particular plant species, although the younger nymphs, and to a lesser degree the adult females, display a preference for Scots Pine (*Pinus sylvestris* L.) during the summer and the first weeks of autumn. In the Netherlands, Ectobiinae are known to inhabit higher sandy soils and heathlands, including the inland wooded heathlands and open woodlands as well as the coastal and inland dunes (the Kempen lands) and river dunes of the Meuse basin. All four species are present on one or more West Frisian islands. The habitat of the species studied can be best described as very warm and sunny, but also concealed. Typical cockroach localities are sunny edges of woodland with heather bundles and grass tussocks, inclines along wood-paths and roads, slanted light woodlands, wooded heathlands and dune landscapes with scattered bundles of oak trees and Scots Pine. Cockroaches strongly avoid windy habitats and other kinds of bare or exposed localities, such as sand-drifts and unplanted dunes eroded by the wind. Wet or boggy heathlands as well as extraordinary dry heathlands and sand dunes represent no (permanent) suitable habitats. Cockroaches will also leave a habitat when it becomes overgrown with weeds or tree branches that shade their residence. Human activities such as agriculture have a negative impact on these species. Knowing that the four species have such similar life history traits and overall preferences, we were interested in assessing whether the species show differences in their spatial distributions.

Materials and methods

Study area and datasets

The geographical area for which the distribution models are built is the country of the Netherlands. For the purpose of this study we divided the country into grid cells of 250 m x 250 m.

We selected four native cockroach species of continental north-western Europe. The Ectobiinae data originate from the database of Wijnand R.B. Heitmans (WRBH) which contains records from national and regional natural history museum collections as well as some private collections. Since 1996 large amounts of new records from field biologist as well as material from ecological field studies and hundreds of systematic hand samplings by WRBH from all over the Netherlands have been added. Records range from the 1880s to present. Nearly all implemented data (both nymphs and adult cockroaches) were (re-)identified to species level by WRBH. Records lacking coordinates were georeferenced using the *Grote Topografische Atlas van Nederland* (Comprehensive Topographic Atlas of the Netherlands, 1989).

Data on environmental variables (Table 1) were obtained from the ECOGRID project of the Institute for Biodiversity and Ecosystem Dynamics, University of Amsterdam. The environmental variables comprise soil type, land cover, altitudinal range, duration of drainage and climatic variables. Soil type is divided in 6 categories (Table 2).

Data from the Corine Land Cover 2000 project of the European Environmental Agency were used to construct a land cover map containing 30 categories. A comprehensive elevation map of the Netherlands based on LiDAR technology was used (AHN2). Differences between the upper and lower surface model extracted

Table 1. Environmental variables and corresponding units

Variable	Unit
Altitudinal range	ratio
Duration of drainage	days
Land cover	30 categories
Mean temperature	°C
Precipitation sum	mm/y
Principal component 1	NA
Principal component 2	NA
Principal component 3	NA
Soil type	6 categories

NA = not applicable.

Table 2. Categories of the variables soil type and land cover

SOIL TYPE	
2=dunes	5=rivers
3=hills	6=sand soils
4=bogland	7=sea-clay soils
LAND COVER	
1=discontinuous urban fabric	16=broad-leaved forests
2=industrial or commercial units	17=coniferous forests
3= road and rail networks and associated land	18=mixed forests
4= port areas	19=natural grassland
5=airports	20=moors and heathland
6=mineral extraction sites	21=transitional woods and shrubs
7=dump sites	22=beaches
8= construction sites	23=inland marshes
9= green urban areas	24=peatbogs
10=sport and leisure facilities	25=salt-marshes
11=non-irrigated arable land	26=intertidal flats
12=permanently irrigated arable land	27=water courses
13=pastures	28=water bodies
14=complex cultivation patterns	29=estuaries
15=land mainly occupied by agriculture	30=sea and ocean

from this elevation map resulted in the variable altitudinal range. Duration of drainage was obtained from Rijkswaterstaat (the implementation organisation of the Dutch Ministry of Transport, Public Works and Water Management). Climatic data was obtained from the WorldClim 1.4 database (WorldClim, 2008). The climatic data of WorldClim is a global set of interpolated bioclimatic variables (Hijmans *et al.*, 2005). The data consist of monthly total precipitation, and monthly mean, minimum and maximum temperature, and 19 derived bioclimatic variables. The climatic variables of ECOGRID use four principal components to reduce the information contained into the 19 bioclimatic variables into three variables named Principal component 1, Principal component 2 and Principal component 3. The other two climatic variables used in this study are the mean temperature and annual precipitation sum which were re-sampled to the Dutch coordinates.

Modelling approach

Recent studies (Elith *et al.*, 2006; Hernandez *et al.*, 2006; Phillips *et al.*, 2004; 2006) show that the modelling program Maxent outperforms many of the alternative computerized approaches for Ecological Niche Modelling. Maxent is a maximum likelihood method. The basic principle of Maxent is to predict a target distribution from occurrence localities over a finite geographical space. In this finite space there are real valued environmental variables, which are empirically measured. The model is looking for a distribution in

which the expectations of the empirical variables are the same as the expectations of the theoretical distribution that Maxent outputs. There are obviously many distributions that satisfy this condition. Maxent outputs the maximum entropy distribution or the distribution closest to uniform subject to the condition that the empirical expectations are the same as the theoretical expectations. For more details on the properties and the theory behind Maxent we refer the reader to Phillips *et al.* (2004, 2006). We used Maxent (version 3.0 beta) in this study with the following parameters: regularization multiplier 1, random test percentage 30 and auto features. The models were built using 146 points for training and 63 points for test for *C. panzeri*, 150 training points and 66 test points for *E. lapponicus*, 93 training points and 41 test points for *E. pallidus* and 340 training points and 146 test points for *E. sylvestris*. The Maxent jackknife test was used to estimate the variability of the samples and allows an indication of how much better the established distribution fits the sample points than a uniform distribution would.

Model evaluation and stability

Evaluation of the generated model is based on the Area Under the Curve (AUC). The AUC is computed using a Receiver Operating Characteristic (ROC) curve. In an ROC curve the y-axis shows the *sensitivity* of a certain classifier, the x-axis *1-specificity*. In general, the sensitivity, which is also known as the true positive rate or recall, is estimated by the fraction of positive

instances that are correctly classified as positive relative to the total number of positive instances. *Specificity* is the fraction of true negative instances relative to the total number of false positives plus true negatives.

In niche modelling, *sensitivity* represents the presence points which are correctly predicted as present, while *specificity* relates to the absence points which are correctly classified as absent. The AUC is determined by

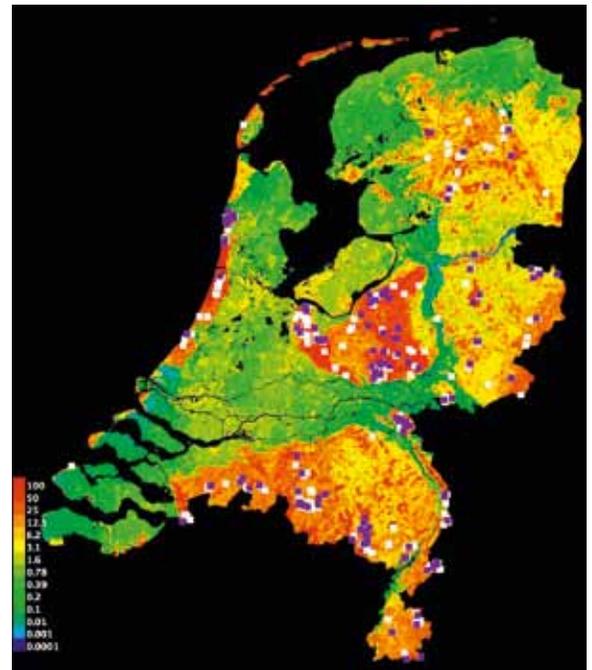
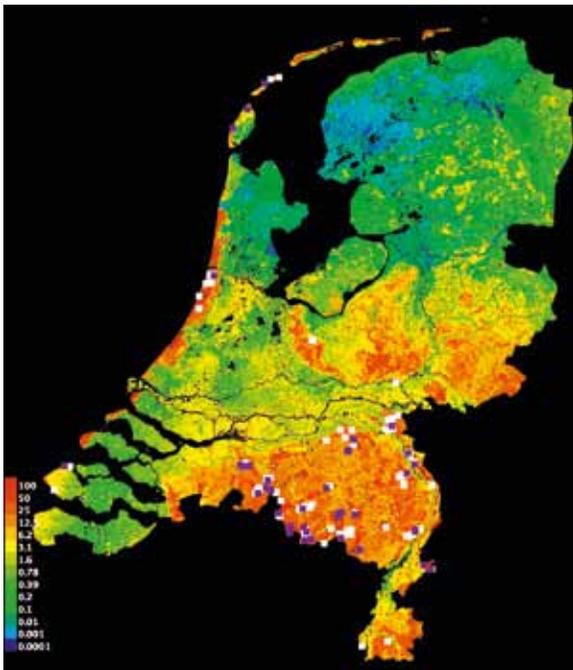
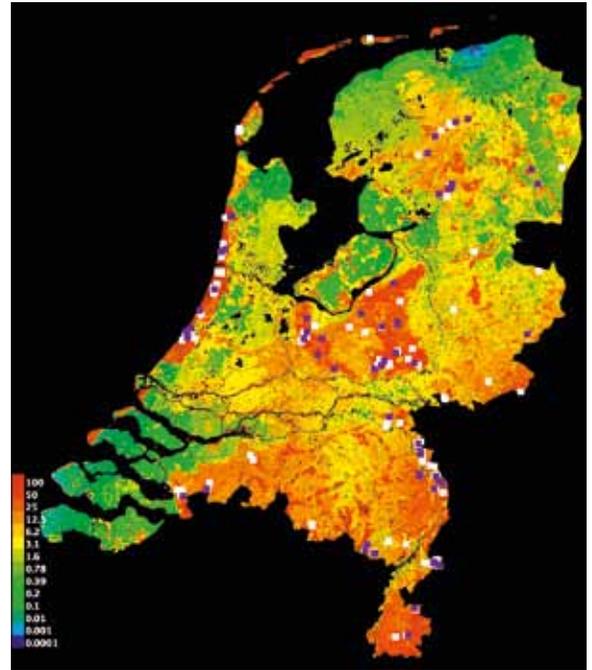
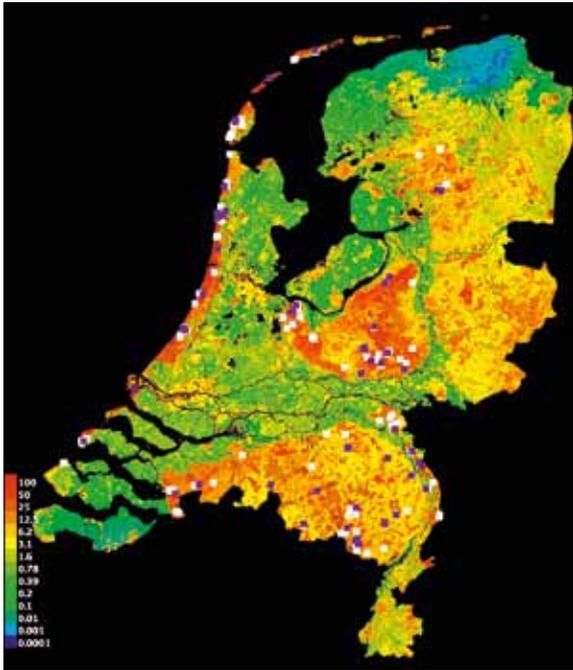


Fig. 1. Predicted suitability maps of *C. panzeri*, *E. lapponicus*, *E. pallidus* and *E. sylvestris* (from left to right, from above to below) for the models with all occurrence localities included. The red end of the spectrum represents the more suitable conditions and the blue end of the spectrum the less suitable ones. White cells represent the sample points used to build the model, violet cells represent test localities.

joining the points in the graph with a line (Fawcett, 2004; Phillips *et al.*, 2004, 2006). Maxent uses only presence data, and to build ROC curves absence points are needed. To circumvent this problem, Maxent randomly creates absence points, usually 10,000, which are then used as negative instances.

In order to test the stability of the prediction and selection of the variables that explain cockroach distributions we executed two sets of runs in which we sub-sampled the input data by randomly omitting one occurrence point or locality and ten occurrence points or localities respectively. Each model building procedure was repeated five times. In this way we tested whether the representation of the prediction and the identity of the most explanatory variable were affected by the employed set of occurrence points.

Correspondence analysis

A correspondence analysis was performed to check whether the association between the three most influential variables of the Maxent models and the predicted species distribution differs among the four species. In the correspondence analysis the variable altitudinal range was categorized as follows: difference in height from -5m to -1m was attributed a value -1, difference in height from +1 to +5m was attributed a value 1, difference in height from +5 to +10 m was attributed a value 2 and above +10m the attributed value was 3. The correspondence analysis was performed using SPSS version 16.00.

Results

Modelling results

The predicted suitability maps of the four species of cockroaches show a fair degree of overall similarity (Fig. 1). The western and the northern parts of the country are less suitable for the species while the eastern and southern areas of the Netherlands are the parts with the highest number of grid cells that would be suitable for these species. However, differences between the predictive suitability maps are also evident.

The predictive maps for *C. panzeri* and *E. lapponicus* show high suitability in the central part of the country and in the western coastal dunes including the west Frisian islands. The eastern part of the country is also predicted to be suitable while the northern

part and the region corresponding to the province Zeeland in the south-west are considered rather unsuitable for *C. panzeri* and *E. lapponicus*. The differences between these two species are most pronounced in the utmost south-eastern part of the country, the southern part of the province of Limburg. Here, conditions suit *E. lapponicus* better than *C. panzeri*. The predictive map of *E. pallidus* shows moderate suitability in the central eastern and south eastern part of the country and in the central western dunes. A remarkable difference with the other three species is that the northern part of the country, as a whole, is quite unsuitable for *E. pallidus*. *E. sylvestris* has the largest set of occurrence localities. Suitability is indicated in most of the eastern, south-eastern and the central parts of the country. In general the predictive map of *E. sylvestris* is very similar to that of *C. panzeri*.

The results of the jack knife plots show that the variables soil type and land cover have the highest impact on the predicted distributions of *C. panzeri*. The variables that best explain suitability for *E. pallidus*, *E. lapponicus* and *E. sylvestris* are soil type, land cover and altitudinal range (Table 3). Climatic variables are less important in explaining suitability.

Evaluation and stability of the models

The ROC curves indicate high accuracy of the generated models. The AUC for *C. panzeri* training data and test data are respectively 0.957 and 0.941 with 146 and 63 occurrence points for the training and test data respectively. The AUC for *E. lapponicus* training data is 0.939 while for test data the AUC is 0.913, with 150

Table 3. Results of the Jackknife analysis of the importance of the variables used for the models of the four species with all occurrence localities included. Regularized training gain is given for the variable alone. The higher the gain of a variable when it is used alone, the more useful it is in predicting the species' distribution. Species names abbreviated by initials.

	CP	EP	ES	EL
altitudinal range	0.218	0.614	0.736	0.680
duration of drainage	0.339	0.359	0.343	0.198
land cover	1.390	1.168	1.052	1.105
mean temperature	0.152	0.413	0.149	0.154
precipitation sum	0.049	0.098	0.175	0.093
principal component 1	0.192	0.069	0.106	0.095
principal component 2	0.091	0.200	0.026	0.238
principal component 3	0.133	0.239	0.110	0.295
soil type	0.805	0.481	0.574	0.581

sample points used for the training data and 66 for the test data. For the *E. pallidus* training data the AUC is 0.968 (based on 93 sample points) and for the test data 0.953 (based on 41 sample points). The model of *E. sylvestris* has an AUC of 0.940 for the training data (based on 340 sample points) and 0.927 for the test data (based on 146 sample points).

The maps of the Maxent predictions obtained with the sub-sampled dataset are generally similar to the

maps of the models in which all the occurrence points are used (figures available on request). After sub-sampling, the variables that contribute most to the predictions (Table 4) are essentially the same as those for the models including all occurrence points (see above).

The variables land cover and soil type are the most explanatory in all models for *C. panzeri*. Land cover, altitudinal range and soil type are always explaining suitability for *E. lapponicus*. Only in two models for

Table 4. Number of times variables occur among the three most explanatory ones for the predicted distributions. The numbers 1, 2, 3 indicate the order of importance of the variables. Abbreviations: LC: land cover; ST: soil type; AR: altitudinal range. Note that for *C. panzeri* only the first two variables are considered because the other variables contribute very little to the models (see also Fig. 2).

Species	<i>C. panzeri</i>			<i>E. lapponicus</i>			<i>E. pallidus</i>			<i>E. sylvestris</i>		
	LC	ST	AR	LC	ST	AR	LC	ST	AR	LC	ST	AR
1	10	0	0	10	0	0	10	0	0	10	0	0
2	0	10	0	0	2	8	0	0	10	0	0	10
3			0	8	2	0	10	0	0	10	0	0

Table 5. AUC of the models based on sub-sampling occurrence localities. Abbreviations: a: model with one occurrence locality omitted; b: model with ten occurrence localities omitted.

	Training AUC				Test AUC			
	<i>C. panzeri</i>	<i>E. lapponicus</i>	<i>E. pallidus</i>	<i>E. sylvestris</i>	<i>C. panzeri</i>	<i>E. lapponicus</i>	<i>E. pallidus</i>	<i>E. sylvestris</i>
a	0.953	0.940	0.967	0.939	0.941	0.913	0.955	0.924
a	0.957	0.947	0.965	0.939	0.932	0.886	0.972	0.924
a	0.959	0.946	0.967	0.937	0.919	0.887	0.954	0.931
a	0.953	0.946	0.969	0.938	0.943	0.898	0.950	0.930
a	0.957	0.947	0.967	0.940	0.932	0.898	0.956	0.924
b	0.955	0.939	0.961	0.940	0.927	0.908	0.966	0.922
b	0.963	0.938	0.965	0.941	0.935	0.914	0.949	0.924
b	0.954	0.935	0.963	0.945	0.954	0.924	0.967	0.909
b	0.958	0.940	0.964	0.938	0.938	0.907	0.962	0.925
b	0.954	0.941	0.967	0.940	0.931	0.919	0.967	0.927

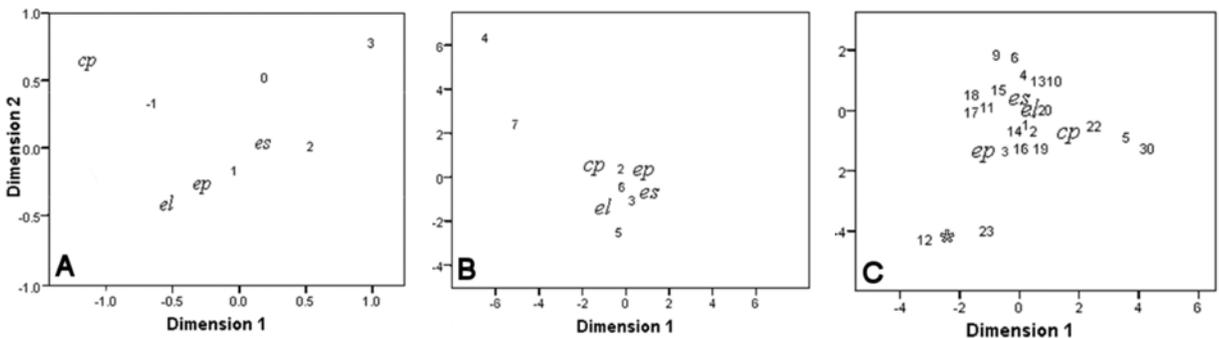


Fig. 2. Results of the correspondence analysis of cockroach species responses to the three most influential environmental variables determined by the Maxent analysis: A altitudinal range, B soiltype and C land cover. Abbreviations are: cp=*Capraieillus panzeri*, el=*Ectobius lapponicus*, ep=*Ectobius pallidus* and es=*Ectobius sylvestris*. See table 2 for categorical variables. In figure 2c, the star symbol represent variables 26 and 27.

the latter species the order of the second and third most explanatory variable is inverted resulting in a model where soil type is indicated as the second most explanatory variable and altitudinal range as the third. Also the models for *E. pallidus* specify the land cover as the most explanatory variable and altitudinal range and soil type as the second and third one. All models built for *E. sylvestris* indicate land cover, altitudinal range and soil type as the variables which most influence the predictions of suitability.

The AUCs of the models, resulting from deleting one or ten sample points, respectively, indicate a good performance (Table 5).

Correspondence analysis

The correspondence analysis indicates that the species do not differ with respect to their response to the variables soil type and land cover (Fig. 2). The categories bogland and sea clay separate from the bulk of the points indicating a negative correspondence of the species studied with these two soil type categories, the same is applicable to the land cover categories permanently irrigated arable land, inland marshes and intertidal flats.

The effect of the variable altitudinal range seems to differ among the four species. *Ectobius lapponicus*, *E. pallidus* and *E. sylvestris* are grouped together, showing slightly different correspondence to the categories while *C. panzeri* is separated from the three species. However, it must be noticed that the impact of altitudinal range is of less importance for the prediction of the distribution of *C. panzeri*.

Discussion

This study provides insight into the spatial distribution of four biologically similar species of cockroaches at the geographical scale of 250 m x 250 m, it identifies the environmental factors that are most influential in predicting the habitat suitability of each species and compares the ecological niche preferences of the four species. The quality of the result was tested by a sub-sampling procedure.

Various studies have shown that the models produced by Ecological Niche Modelling are sensitive to the amount and identity of points used (Stockwell and Peterson, 2002; Pearson *et al.*, 2007). In our study, both the predicted distribution and the selection of the best explanatory variables produce the same pattern in

the analyses using all the occurrence points and in the analysis using sub-sampled points. Possibly, the relatively large amount of points used to train and test the model results in the stability of the predictions and the selection of the best explanatory variable.

The grid cells predicted as most suitable are mainly located in the south-east, east and the coastal western part of the Netherlands, and for *C. panzeri*, *E. lapponicus* and *E. sylvestris* also in the north-east of the country. These areas have in common that they are characterized by sandy soils and are relatively sparsely populated. They contrast with the western part of the country underlain by clay soils, which is less suitable for the four species and more intensely populated. The areas that are predicted as being suitable for *Capraielus panzeri*, *E. lapponicus* and *E. sylvestris* show a high level of spatial congruence. The National Park Hoge Veluwe, located in the centre of the country, has a prominent place in the predicted distribution of these species. *Ectobius pallidus* differs from the other species in having no areas predicted as suitable habitat in the north of the country.

The environmental variables land cover, soil type and altitudinal range appeared to be the most important variables explaining the spatial distributions of the species. These three types of environmental variables also appeared as important factors in explaining the species richness patterns of five different taxonomic groups in the Netherlands (Schouten *et al.*, 2009). In general, the four cockroach species seem to be associated to similar environmental conditions. However, subtle differences mainly with respect to the altitudinal range seem to exist. For example, the Maxent prediction for *C. panzeri* indicates altitudinal range as relatively unimportant, while it is of great importance for the prediction of the distribution of the other species.

Habitat selection at the macro-scale level has been shown to occur for seven species of cockroaches on La Réunion Island (Boyer and Rivault, 2006). Our study suggests that at least at the macro environmental scale a certain degree of habitat partitioning occurs between the four species of cockroaches which may allow them to occupy the same geographical space without competition. However, at the micro environmental level inter-specific competition may still occur. In conclusion, *C. panzeri*, *E. lapponicus*, *E. pallidus* and *E. sylvestris* appeared to have very similar ecological properties at the macro-scale level, while partitioning of habitat seemed to occur due to different ecological niche preferences according to altitudinal range.

Acknowledgements

We thank Dr. Tomislav Hengl of the ECOGRID-team of IBED, University of Amsterdam. His comments and suggestions highly improved the quality of the manuscript.

References

- Bohn H. 1989. Revision of the *sylvestris* group of *Ectobius* Stephens in Europe (Blattaria: Blattellidae). *Entomologica Scandinavica* 20: 317-342.
- Boyer S, Rivault C. 2006. Habitat selection and coexistence of invasive cockroach species (Dictyoptera) in sugar-cane fields on Réunion island. *Acta Oecologica* 29: 16-26.
- Comprehensive Topographic Atlas of the Netherlands/ Grote Topografische Atlas van Nederland. 1989. Volumes 1-4. Wolters-Noordhoff Atlasproducties. 2nd edition.
- Costa GC, Wolfe C, Shepard DB, Caldwell JP, Vitt JL. 2008. Detecting the influence of climatic variables on species distributions using GIS niche-based models along a steep longitudinal gradient. *Journal of Biogeography* 35: 637-646.
- Davis AJ, Jenkinson LS, Lawton JH, Shorrocks B, Wood S. 1998. Making mistakes when predicting shifts in species range in response to global warming. *Nature* 391: 783-786.
- Elith J, Graham CG, Anderson RP, Dudik M, Ferrier S, Guisan A, Hijmans RJ, Huettmann F, Leathwick JR, Lehman A, Li J, Lohmann LG, Loiselle BA, Manion G, Motitz C, Nakamura M, Nakazawa Y, Overton J, Peterson AT, Phillips SJ, Richardson K, Scachetti-Pereira R, Schapire RE, Soberon J, Williams S, Wisz MS, Zimmermann NE. 2006. Novel methods improve prediction of species' distributions from occurrence data. *Ecography* 29: 129-151.
- Failla MC, Messina A. 1978. Struttura della fossetta ghiandolare dei maschi delle specie italiane di *Ectobius* Steph. (Blattaria, Ectobiidae). *Animalia* 5: 357-394.
- Fawcett T. 2004. *ROC Graphs: notes and practical considerations for researchers*. Technical report HPL-2003 4: 1-38. home.comcast.net/~tom.fawcett/public_html/papers/ROC101.pdf.
- Gaston KJ. 1996. Species range size distributions: patterns, mechanisms and implications. *Trends in Ecology and Evolution* 11: 197-201.
- Guisan A, Zimmerman NE. 2000. Predictive habitat distribution models in ecology. *Ecological Modelling* 135: 147-186.
- Guisan A, Thuiller W. 2005. Predicting species distribution: offering more than simple habitat models. *Ecology Letters* 8: 993-1009.
- Hampe A. 2004. Bioclimate envelope models: what they detect and what they hide. *Global Ecology and Biogeography* 11: 469-471.
- Hernandez PA, Graham CH, Master LL, Albert D. 2006. The effect of sample size and species characteristics on performance of different species distribution modelling methods. *Ecography* 29: 773-785.
- Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A. 2005. Very high resolution interpolated climatic surfaces for global land areas. *International Journal of Climatology* 25: 1965-1978.
- Pearson RG, Raxworthy CJ, Nakamura M, Peterson AT. 2007. Predicting species distribution from small numbers of occurrence records: a test case using cryptic geckos in Madagascar. *Journal of Biogeography* 34: 102-117.
- Phillips SJ, Dudik M, Schapire RE. 2004. A maximum entropy approach to species distribution modeling. Pp. 655-662 in: Carla E Brodley, ed. *Machine Learning, Proceedings of the Twenty-First International Conference*.
- Phillips SJ, Anderson RP, Schapire RE. 2006. Maximum entropy modelling of species geographic distributions. *Ecological Modelling* 190: 231-259.
- Schouten MA, Verweij PA, Barendregt A, Kleukers RMJC, Kalkman VJ, de Ruiter PC. 2009. Determinants of species richness patterns in the Netherlands across multiple taxonomic groups. *Biodiversity and Conservation* 18: 203-217.
- Stockwell DRB, Peterson AT. 2002. Effect of sample size of species distribution models. *Ecological Modelling* 148: 1-13.
- WorldClim 2008. WorldClim Version 1.4, Museum of Vertebrate Zoology, University of California, Berkeley. <http://www.worldclim.org>.

Received: 25 November 2008

Accepted: 11 September 2009

Published online: 23 November 2009

Editor: J.W. Arntzen