A PREDICTIVE DISTRIBUTION MODEL FOR *GRAPHODERUS BILINEATUS* IN THE NETHERLANDS (COLEOPTERA: DYTISCIDAE)

Henk Sierdsema & Jan Cuppen

In 2004 and 2005 the distribution and habitat of the dytiscid water beetle *Graphoderus bilineatus* in the Netherlands were investigated. All old localities for this species were reinvestigated and habitat characteristics were gathered. In order to designate areas where *G. bilineatus* potentially occurs, a predictive distribution model was created. The best performing model was based on the occurrence of ten plant species known to be associated with *G. bilineatus,* electronic conductivity of the water and the distribution of fen peat soils without a top layer of sand or clay. The model indicates that the species is likely to be more widespread in the north of country than is currently known. Surveying the appointed kilometre squares with a high probability of occurrence of *G. bilineatus* can be used to validate the model.

INTRODUCTION

On request of the province of Zuid-Holland and the Ministry of Agriculture, Nature and Food quality the current distribution and habitat of the Natura2000-species *Graphoderus bilineatus* (Degeer, 1774) was investigated in 2004 and 2005. The main goal of the project was to re-investigate known localities of *G. bilineatus*. Of the 85 surveyed sites *G. bilineatus* was present on 38 sites (Cuppen et al. 2006). To compile a complete overview of the distribution one would ideally sample all kilometre squares in the whole country. To reduce the effort we have created a distribution model, predicting the kilometre squares in which the species is most likely to occur.

Elodea canadensis
Hottonia palustris
Hydrocharis morsus-ranae
Lemna trisulca
Nuphar lutea
Nymphaea alba
Potamogeton acutifoliius
Potamogeton obtusifolius
Stratiotes aloides
Utricularia vulgaris

Table 1. Plant species associated with the occurrence of *Graphoderus bilineatus*.
Tabel 1. Plantensoorten die geassocieerd worden met het voorkomen van *Graphoderus bilineatus*.

METHODS

The distribution data were collected in 2004-2005 (Cuppen et al. 2006). Both presence and absence were recorded. Sites that were regarded as insufficiently investigated to obtain trustworthy information on absence were discarded from the dataset. This resulted in a dataset of 73 sites with 38 positive observations. From all sites information on a number of explanatory variables were recorded. See 'Sampling procedure' in Cuppen et al. (2006) for a description of these variables. In addition, the presence of plant species per kilometre square that are associated with the occurrence of *G. bilineatus* was used (table 1) (data provided by FLORON). The total number of these plant species per kilometre square (fig. 1) was used as a predictive variable. The analysis of the data was performed within the statistical package 'Genstat' (Rothamsted-Experimental-Station 2003). First descriptive analyses like correlations and histograms were carried out. Following the descriptive analyses a number of logistic regression models

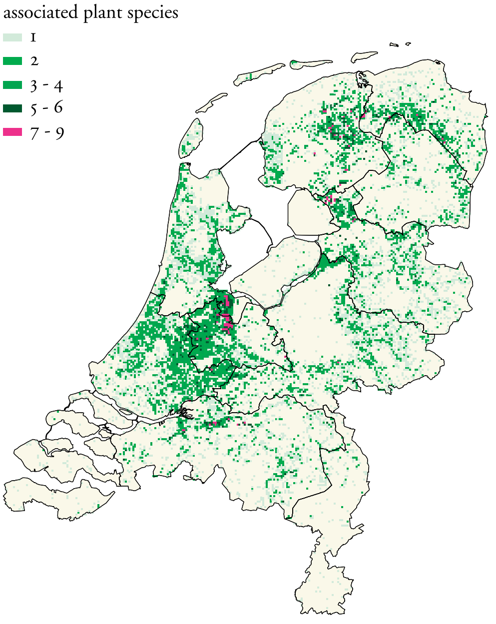

Figure 1. Number of plant species per kilometre square associated with *Graphoderus bilineatus* (data Florbase FLORON). The presence of plant species is biased towards observer effort: absence of species may be caused by lack by sufficient field work.
Figuur 1. Aantal met *Graphoderus bilineatus* geassocieerde plantensoorten per kilometerhok (data Florbase FLORON). Het aantal soorten wordt mede beïnvloed door waarnemersinvloeden: de afwezigheid van soorten kan het gevolg zijn van onvoldoende veldonderzoek.

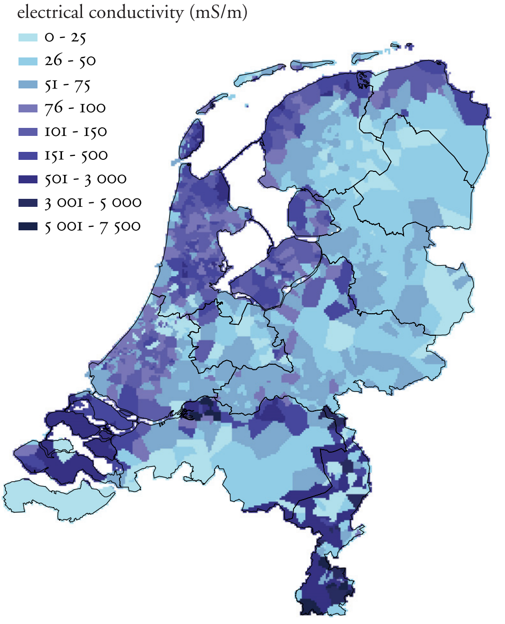

Figure 2. Map of electrical conductivity interpolated from sample points in the Limnodata Neerlandica database. Note the pronounced difference between Limburg in the far southeast and the neighbouring province of Noord-Brabant. This is caused by difference in scaling used in those provinces: the Limburg-numbers are probably 10 times too high as compared to Noord-Brabant.
Figuur 2. Kaart van het elektrisch geleidingsvermogen geïnterpoleerd van meetpunten in de Limnodata Neerlandica database. Let op de scherpe overgang tussen Limburg en Noord-Brabant. Deze wordt waarschijnlijk veroorzaakt doordat in Limburgse waarden in de database 10 maal te hoog zijn in vergelijking met Noord-Brabant.

(Guisan & Zimmermann 2000, Sierdsema et al. 2005) were built that describe the chance of occurrence (probability) in relation to one or more variables that showed high correlations with observations of *G. bilineatus*. Predictive variables were first tested in their original form ($y = constant + x$) to look for linear relationships. In order to search for optima a quadratic form was added to the model ($y = constant + x + x2$) (Jongman et al. 1987). Finally, variables were modelled as third degree splines in order to investigate more complex relationships.

Based on the best performing logistic regression model a predictive distribution map was created. For this map nationwide information on electrical conductivity (EC) was provided by STOWA (www.limnodata.nl, Limnodata Neerlandica). However, these are sample points and therefore not directly suitable for predictions for each kilometre square in the Netherlands. Therefore the dataset was interpolated to a grid with the 1-Nearest Neighbour technique (Cressie 1991) (fig. 2). With the map of the number of

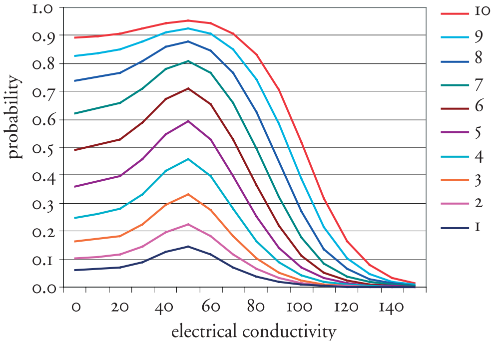

Figure 3. The chance of occurrence (probability) of *G. bilineatus* in relation to electrical conductivity (EC) and the number of associative plant species per kilometre square, ranging from 1 to 10, per EC-value.
Figuur 3. De kans op voorkomen (probability) van *G. bilineatus* in relatie tot het elektrisch geleidingsvermogen (EGV) (electrical conductivity) and het aantal associatieve plantensoorten, variërend van 1 tot 10, per EGV-waarde.

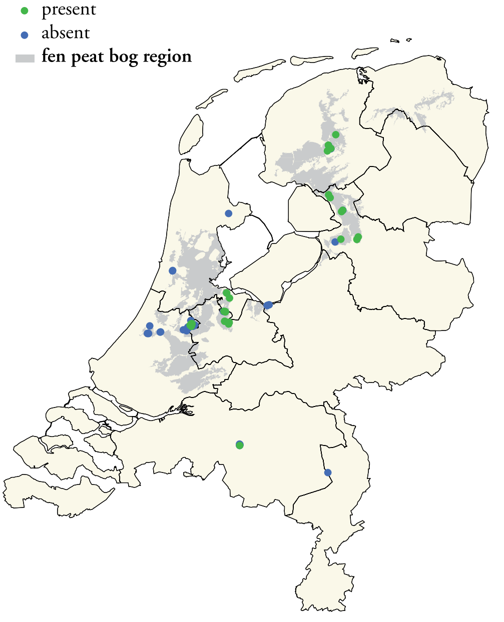

Figure 4. Observations of *Graphoderus bilineatus* and location of the Fen peat bog region.
Figuur 4. Waarnemingen van *Graphoderus bilineatus* en ligging van de Laagveenregio (present: aanwezig, absent: afwezig, fen peat bog region: Laagveenregio).

associated plant species and the interpolated map of electrical conductivity the probability of occurrence per kilometre square in the Netherlands was calculated. Since information on the associated plant species is not available for all kilometre squares in the Netherlands, a prediction solely based on electrical conductivity was made for squares without flora information.

The final step in the modelling process was the inclusion of soil information and the presence of surface water. A spatial comparison was made between the positive observations of *G. bilineatus* and the distribution of various types of peaty soils, their position in relation to the fen peat bog region in the Netherlands and amount of surface water.

RESULTS

The best performing model combines electrical conductivity (EC) and the number of associative plant species per kilometre square (table 2). This model explains 39.2% of the deviance. The chance of occurrence is highest with an electrical conductivity between 30-70 mS/m and 7-10 associated plant species within the square. Above an electrical conductivity of 100-120 mS/m the probability of occurrence sharply decreases (fig. 3). A model including only the number of associative plant species per kilometre square as predictive variables explained 30.8% of the deviance, whereas a model including only EC explained 25.4% of the deviance.

All but one of the recent observations of *G. bilineatus* are from the Fen peat bog region in the Netherlands (fig. 4). Moreover, almost all observations were conducted at locations where the top layer of the soil consists of peat bog or in lakes adjacent to this type of soil. Therefore the prediction map was constrained to squares where peaty soil with a top layer of peat is present. For the squares where no flora information was available and the predictions were only based on

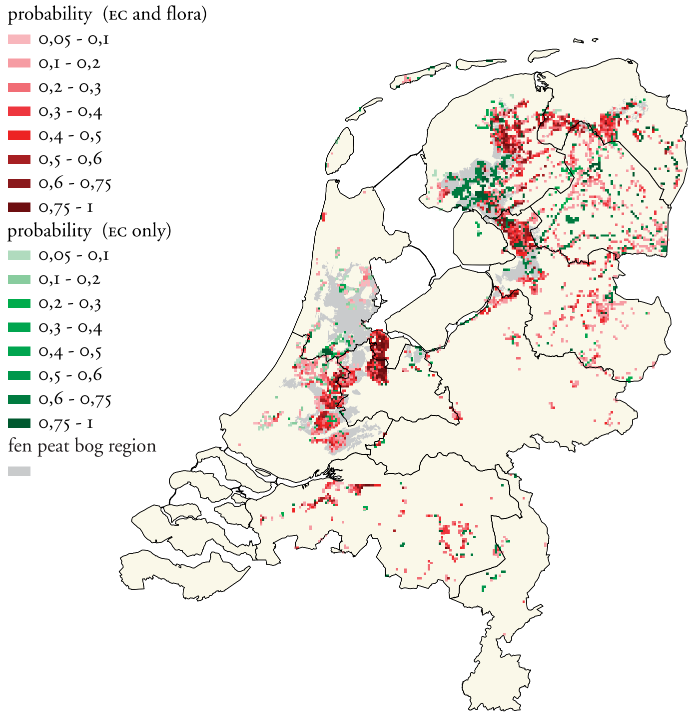

Figure 5. Predictive distribution map showing the chance of occurrence (probability) of *Graphoderus bilineatus* based on both electrical conductivity (EC) and the number of associative plant species ('flora') (in red) or only on electrical conductivity (in green). The map is clipped to squares where a peat soil with a top layer of peat bog is present. The fen peat bog region is shown in grey.
Figuur 5. Kansenkaart (probability) van *Graphoderus bilineatus* gebaseerd op zowel elektrisch geleidingsvermogen (EC) én het aantal associatieve plantensoorten (flora) (in rood) of alleen gebaseerd op elektrisch geleidingsvermogen (in groen). De verspreiding is beperkt tot kilometerhokken waar veengrond met een toplaag van veen aanwezig is. De Laagveenregio is weergegeven in grijs.

electrical conductivity, soil and surface water another criterion was added: the squares should have at least 2.5 ha of surface water.

The predictive map was evaluated by comparing the observations with the predicted chances of occurrence. This showed that than 90% of the observations are within squares with a probability of at least 0.5. Moreover, more than 80% of these observations are within squares with a predicted probability of more than 0.7. Two observations are outside the constrained area: one on peat bog with a clay top layer and one in a small lake in the south of country.

The final map showing the predicted distribution (fig. 5) was limited to those squares where the probability is more than 0.05. Especially in the north of the country a large number of squares are found where the prediction is only based on electrical conductivity, soil and surface water because of the lack of flora information.

DISCUSSION

Despite the fact that only 38 positive observations of *G. bilineatus* were available for the modelling procedure, a relatively well performing logistic regression model could be built. A combination of both the number of associative plant species and electrical conductivity (EC) performs significantly better than either one of these variables separately. Flora information alone performs better than EC alone. The outcome of this study is a predictive map that shows where we may expect *G. bilineatus* to occur in order to improve the protection of the species (Cuppen et al. 2006).

The quality of a regression model or a predictive distribution map is limited by the quality of the predictive variables used in the modelling process. Both EC and flora must be used with some caution. In case of the EC, there appears to be a difference in order of magnitude 10 between various data sources. For example there

Parameter	df	estimate	mean deviance	deviance ratio	approx prob
Constant		-1.83			0.141
EC (3rd degree spline)	3	-0.0107	8.4024	8.40	<0.001
Floraspecies	1	0.535	12.9821	12.98	<0.001
Residual	59		0.8554		
Explained Deviance: 39.21%					

Table 2. Summary of the best performing logistic regression model. df: degrees of freedom, estimate: parameter estimate of variable, approx prob: approximate significance of parameter or variable; EC: electrical conductivity, Floraspecies: number of associative plant species.
Tabel 2. Samenvatting van het beste logistische regressiemodel. df: aantal vrijheidsgraden, estimate: parameter schatting van de variabele, approx prob: geschatte significantie van parameter of variabele; EC: electrisch geleidingsvermogen, Floraspecies: aantal associatieve plantensoorten; Explained Deviance: percentage verklaarde variantie.

is a sharp decrease in EC at the border of the provinces of Limburg and Noord-Brabant (fig. 2) in the southeast of the country. This problem also arises in other regions like the Delta. We don't expect however, that these problems in the national EC data will have a major influence on the quality of the predicted distribution. All EC data for the regression model were collected by the same analyst and within the predicted region of occurrence of *G. bilineatus* the EC-map looks trust-worthy. The collection of the flora data is not standardized in most of the country. It is therefore difficult to judge whether the absence of associative plant species is indeed correct or the result of insufficient observation effort. This also applies for the total number of plant species. Especially in the north of country no species associated with *G. bilineatus* were recorded in Florbase in many kilometre squares (green squares in fig. 5), where they would be expected based on soil type, electrical conductivity and amount of water.
The final probability map is a result of expert knowledge on associated plants species, a regression model and comparison of observations with soil data. It is possible to generate a probability map based on a more extended regression model alone. This requires more observations and it is questionable whether this would improve the results. All recent observations but one are from the Fen peat bog region. Research on the distribution of the species will therefore probably be most successful in this region. The historical distribution shows that the species once also occurred in other parts of the country.

This study has shown that the combination of data from different taxa and involving different NGO's is a big step forward in the predictive modelling of the distribution of a rare species. The initiative of the University of Amsterdam and the Vereniging voor Onderzoek Flora en Fauna (VOFF) to create Ecogrid, a central portal for this type of data, will greatly improve the possibilities of trans-taxa distribution modelling.

ACKNOWLEDGEMENTS

The national data set on water quality and flora were provided by STOWA and FLORON respectively. The soil data were provided by WUR-Alterra. The fieldwork was funded by the province of Zuid-Holland and the Ministry of Agriculture, Nature and Food quality (LNV). The analysis has been made possible through funding from the WOT-IN program regarding the information supply for the Natura-2000 directives. We thank Judy Shamoun-Baranes of IBED-UVA for commenting on this manuscript.

REFERENCES

Cressie, N.A.C. 1991. Statistics for spatial data. – John Wiley & Sons, Inc., New York.

Cuppen, J.G.M., B. Koese & H. Sierdsema 2006. Distribution and biotope of *Graphoderus bilineatus* in the Netherlands (Coleoptera: Dytiscidae). – Nederlandse Faunistische Mededelingen 24.

Guisan, A. & N.E. Zimmermann 2000. Predictive habitat distribution models in ecology. – Ecological Modelling 135: 147-186.

Jongman, R.H.G., C.J.F. ter Braak & O.F.R. van Tongeren 1987. Data analysis in community and landscape ecology. – Pudoc, Wageningen.

Rothamsted-Experimental-Station 2003. Genstat 7. – Lawes Agricultural Trust.

Sierdsema, H., A. van Kleunen & C. van Swaay 2005. Van losse meldingen en steekproefgegevens naar verspreidingskaarten. – Vereniging Onderzoek Flora en Fauna, Nijmegen. [VOFF-rapport 2005/01]

SAMENVATTING

Een voorspellend verspreidingsmodel voor de gestreepte waterroofkever *Graphoderus bilineatus* in Nederland (Coleoptera: Dytiscidae)

Op verzoek van de Provincie Zuid-Holland en het Ministerie van LNV is het huidige voorkomen en de biotoopvoorkeur van de gestreepte waterroofkever *Graphoderus bilineatus* onderzocht. Om informatie te krijgen over de mogelijke verspreiding van deze Habitatrichtlijn-soort is een voorspellend verspreidingsmodel gemaakt. Het beste model is gebaseerd op de aanwezigheid van tien met *G. bilineatus* geassocieerde plantensoorten en het elektrisch geleidingsvermogen van het water. Met behulp van dit model en de sterke associatie van de soort met laagveenbodems zonder toplaag van zand of klei is de kans op voorkomen per kilometerhok in Nederland berekend. Dit laat onder meer zien dat *G. bilineatus* in Noord-Nederland waarschijnlijk veel wijder verspreid voorkomt dan tot op heden bekend is.

H. Sierdsema
SOVON Dutch Centre for Field Ornithology /
IBED-University of Amsterdam /
Alterra-Wageningen University and Research Centre
Rijksstraatweg 178
6573 DG Beek-Ubbergen
henk.sierdsema@sovon.nl

J.G.M. Cuppen
Aquatic Ecology and Water Quality Management
Group-Wageningen University and Research Centre
Ritzema Bosweg 32a
6703 AZ Wageningen
jan.cuppen@wur.nl