Deliverable D1.3

Version: FINAL
Date: 31/10/2012
Author: NRM
Dissemination status: PU
Document reference: Deliverable_D1.3M27

Implementation of “AquaMaps” and “European Fish Index” tools on the portal
(M27)

STATUS: FINAL VERSION

Project acronym: BIOFRESH
Project name: Biodiversity of Freshwater Ecosystems: Status, Trends, Pressures, and Conservation Priorities
Call and Contract: FP7-ENV-2008-1
Grant agreement no.: 226874
Project Duration: 01/11/2009 – 30.04.2014 (54 months)
Co-ordinator: Leibniz-Institute of Freshwater Ecology and Inland Fisheries at Forschungsverbund Berlin e.V., Germany

Partners:
RBINS, Royal Belgian Institute of Natural Sciences, Belgium
BOKU, Universität für Bodenkultur Wien, Austria
ICLARM, International Center for Living Aquatic Resources Management, Malaysia
IRD, Institut de Recherche pour le Développement, France
UDE, Universität Duisburg-Essen, Germany
IUCN, International Union for Conservation of Nature, Switzerland
UOXF.AC, Oxford University, UK
UB, Universitat de Barcelona, Spain
UFZ, Helmholtz Zentrum für Umweltforschung, Germany
UCL, University College of London, UK
UCBL, Université Claude Bernard - Lyon 1, France
UPS, Université Paul Sabatier- Toulouse 3, France
ECOLOGIC, Ecologic GmbH Institut für Internationale und Europäische Umweltpolitik, Germany
EC-ERC, Commission of the European Communities - Directorate General Joint Research Centre, Italy
UD, University of Debrecin, Hungary
NRM, Naturhistoriska riksmuseet, Sweden
FIN, FishBase Information and Research Group, Inc.

BIOFRESH

Biodiversity of Freshwater Ecosystems: Status, Trends, Pressures, and Conservation Priorities

Project no. 226874

Large scale collaborative project

Deliverable number	D1.3
Deliverable name	Implementation of “AquaMaps” and “European Fish Index” tools on the portal
WP no.	WP1
Lead Beneficiary (full name and Acronym)	NRM
Nature	DB/Written report
delivery date from Annex I (proj. month)	M27
Delivered	yes
Actual forecast delivery date	2012-10-31
Comments	Document delivery delayed but results presented at the 3rd Project Meeting (March 2012).

Project funded by the European Commission within the Seventh Framework Programme Dissemination Level		
PU	Public	✓
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

This project is supported by funding from the specific programme 'Cooperation', theme 'Environment (including Climate Change)' under the 7th Research Framework Programme of the European Union.

Name of the Authors	Name of the Partner	Logo of the Partner
Sven O Kullander	NRM	Naturhistoriska riksmuseet
Aaike De Wever	RBINS	museum

In case the report consists of the delivery of materials (guidelines, manuscripts, etc)

Delivery name	Delivery file name	From Partner	To Partner

Introduction

The integration of the web-based analytical tools “AquaMaps” and “European Fish Index” are key components of Task 1.5 Installation of analytical tools and predictive models. The EFI+-tool (Improvement and Spatial extension of the European Fish Index) was integrated on a specific ‘tools page’ on the data portal. This report gives some relevant background information on these tools and their integration in the portal. This report is the first out of two reports on the Freshwater AquaMaps. The current report focuses on the creation of a stable statistical platform for mapping freshwater organisms including a stable, reliable, and tested collection of environmental layers and a stable system for taking known distribution areas into account (Phase 1) and the integration of some example maps into the portal. A second report scheduled for month 44 will report on the bulk production and validation of models (Phase 2) and their further integration in the data portal through the webservices from the AquaMaps website.

AquaMaps

Freshwater AquaMaps is an approach to generating model-based, large-scale predictions of freshwater species and is based on a methodology, which was originally developed for marine mammals. Models for the freshwater AquaMaps are constructed from estimates of the environmental tolerance of a given species and occurrence data available through GBIF. Maps show the colour-coded relative likelihood of a species to occur in a global grid of half-degree latitude/longitude cell dimensions, which corresponds to about 50 km near the equator. Predictions are generated by matching habitat usage of species, in the form of environmental envelopes, against local environmental conditions to determine the relative suitability of specific geographic areas for a given species. Knowledge of species' distributions within river basins available as shapefiles, is also used to exclude potentially suitable habitats in which the species is not known to occur.

The objective of the implementation of AquaMaps in BioFresh is to **provide a robust visual interface to predicted and actual distribution of primarily European freshwater organisms**. Since map parameters can be manipulated by the user, e.g., testing for elevated temperatures, or modifying habitat optima, it is **also a tool for generating scenario maps according to different hypotheses**. AquaMaps can also deliver species richness maps for selected organisms. By mapping the suitable habitat it can also be used for predicting areas where invasive species may become established.

NRM work on the AquaMaps component of BioFresh proceeded in **two phases**: (1) establishing a stable, reliable, and tested collection of environmental layers for a variety of aquatic organisms, and a stable system of areas for delimitation of known areas of distribution (Pfaffstetter basins used as shape files to construct range maps). (2) Phase two consists of adding models of species in a bulk fashion, but with continuous evaluation by a GIS specialist familiar with freshwater species distribution.

Achievements. Twenty-seven global environmental parameters were tested for their relevance in freshwater fish modelling. Of these parameters relating to elevation, temperature, precipitation and moisture were most relevant, but the degree of contribution of each parameter varied between species.

Model calibration was done using 10 model species drawn from fish, mammals, and amphibians, and three sets of data. The first was the ecological relevant set of six to eight parameters for each taxon (termed *original*). The second set was selected based on a statistical approach in two steps (*VFIM*), and the third was using the full set of 27 parameters (*ALL27*). For the VFIM, first the 27 parameters were tested for correlation using correlation matrix and the variance inflation factor (VIF). VIF is a measure of variable multicollinearity, and have been used in ecological modeling to verify non-correlations of variables. Correlated variables may induce a bias in model estimations and should therefore be treated with caution. A VIF between 4 and 10 has been used as cutoff for eliminating variables, since higher levels of VIF are considered to indicate multicollinearity. To simply use VIF as a cutoff has been questioned, since the VIF value depends on the properties of the model. Here, we selected the value of 10 as cutoff and removed variables with higher values only after confirming high correlation via the correlation matrix. When checking for VIF, the variable with the highest value for VIF was removed and this process was repeated until no variable with a VIF larger than 10 are apparent in the parameter set.

The second step to select parameters for VFIM was based on boosted regression trees (BRT). BRTs are a boosted version of regression trees. Classification and regression trees are considered as a very good method to use in data mining and investigating the importance of different variables, e.g., ecological variables in a modeling set up. It has been used in ecological studies and also in spatial modeling predicting species presence versus absence. The BRT approach gives a value of relative importance for each variable used in the model. BRTs were constructed for the ten model species separately, using presence/absence data. Since only presence data are available from GBIF, we used the species native distributions maps downloaded from IUCN as to define the presence area of a species. We then selected 500 random cells within the IUCN map area to represent the presence cells, and 500 random cells outside the IUCN map area to represent absence cells. If the species was restricted to, or absent from, <500 cells, all cells were used as presence and/or absence data, respectively. From the BRT analyses, the six to eight parameters with highest relative importance were selected and used for map predictions. The statistical analyses are conducted in R.

To compare the outcome of the parameter sets we calculated the overall accuracy of the models (in percent), as the number of correct predictions, i.e., predicted presence in a cell within the IUCN map area and predicted absence in a cell outside the IUCN map area, divided by the total number of cells. The proportion of correctly predicted presence cells of the total number of presence cells was also calculated for each species. In AquaMaps, the predictions gave a probability of species occurrence for each cell, and in this study we investigate two levels defining presence of a species; the first at $P > 0.05$ and the second at $P > 0.50$.

To validate the model further we used one fish species (*Abramis brama*) parameter values and occurrence data from Sweden, France and the UK, in total 697 cells. These areas were selected since the national program of survey fishing in freshwater report all their observations to GBIF in Sweden, France and the UK (Swedish National Board of Fisheries). We assume that the occurrence reflects the presence and absence data, and hence assume that if a species exists, or have been observed, it is also reported to GBIF. This above mentioned assumptions might however include some bias of the actual species distribution, but as a base to evaluate the model it should be sufficiently reliable. In this exercise, we use only Sweden and France to extract data from cells that contain an occurrence point of the species, i.e., to create the good cells. We then run predictions on the UK. In this way we have an independent presence/absence data (UK) on which the prediction using two different sets of variables are compared.

In the model we include Pfafstetter basins to limit native distributions. Independent survey data, and literature range maps, were used to validate the model outputs showing that AquaMaps predictions of ten species representing fish, amphibian, and mammal taxa were well in accordance with the species native distributions described in the literature. The accuracy of the model predictions were 83–85% depending on selection and number of predictor parameters used in the model. Predictions of suitability of a given species across Europe indicated that many of the species are likely to survive or establish outside of their current distribution range. We also show that the choice of environmental parameters used in the model effect the accuracy of the predictions and that the optimal predictions not necessarily are achieved with statistically selected species-specific parameters

Outcomes (1) A manuscript in progress was produced to explain the procedures for Freshwater Aquamaps modelling, parameter testing, and validation of model distributions. This manuscript is appended. The contributors request that this manuscript does not appear in public domain until it has been accepted by a journal.

(2) Environmental parameter layers and drainage basins based on the Pfaffstetter system have been established for European freshwater fishes, amphibians, and mammals. Parameters were tested by calculating their relative contribution to the predicted distribution patterns for different organisms. Predicted ranges have been tested against IUCN shapefile boundaries based on known distributions. Based on this work, a streamlined procedure for producing modelled maps for the BioFresh portal is available.

(3) A symposium on AquaMaps modelling was organised by NRM, held in Stockholm 5 September 2011, with nine speakers presenting different aspects of modelling with AquaMaps. BioFresh was represented by Johan

Östergren (NRM) with the presentation *New homes for Fire Salamander, Catfish, and Racoon Dog? – Insights from AquaMaps predictions.* The abstracts are appended.

(4) The European AquaMaps are currently available for a dozen of freshwater species and will be produced for a wider range of species in the near future. These maps can be consulted at both <http://aquamaps.org/AM_Europe/search.php> and <http://data.freshwaterbiodiversity.eu/search/listAquamaps>

European Fish Index (EFI+)

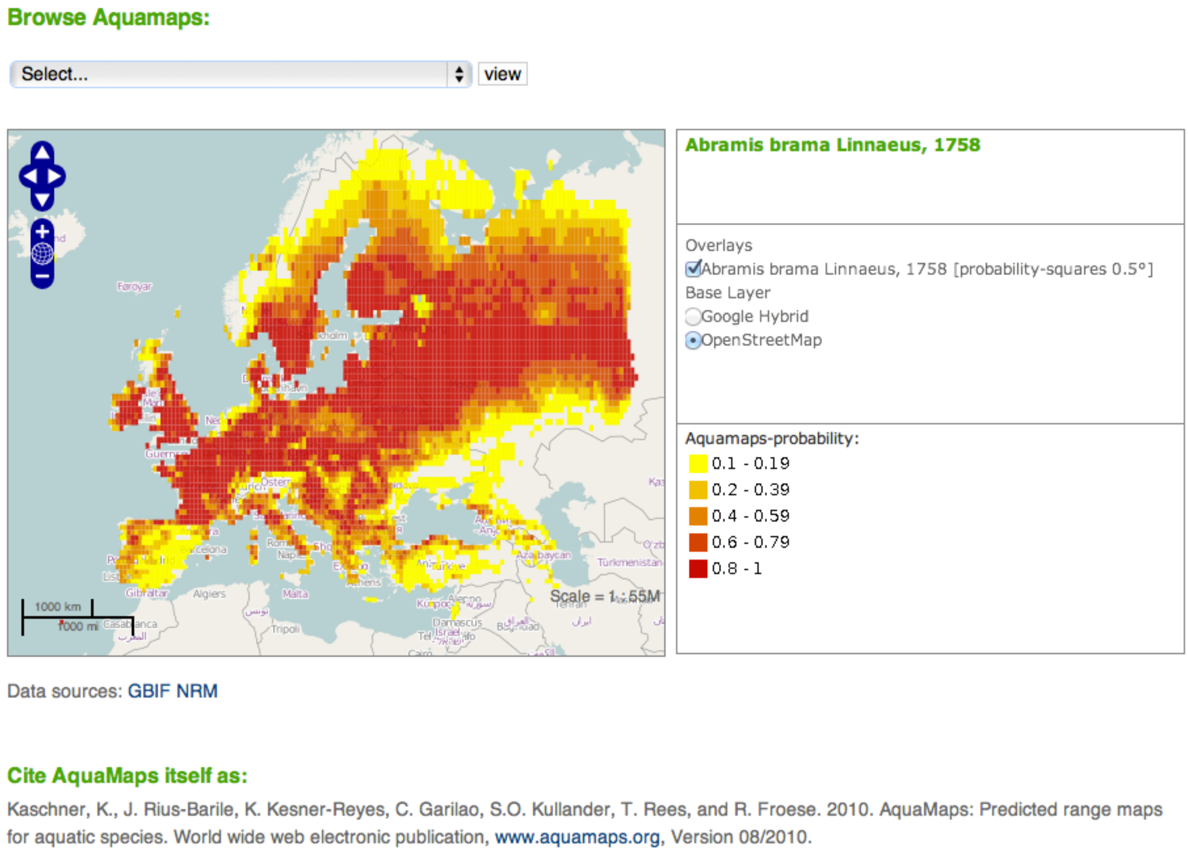
The EFI+-tool (Improvement and Spatial extension of the European Fish Index) allows assessing the ecological status of rivers in accordance with the EU Water Framework Directive. The tool is based on the European Fish Index (EFI) developed within the FAME and EFI+ project as a standardised fish-based assessment method applicable across a wide range of European rivers. The EFI employs a number of environmental descriptors (see http://efi-plus.boku.ac.at/software/insert_data_manual.php, data input matrix) to predict biological reference conditions and quantifies the deviation from reference conditions on a statistical basis.

Integration in the portal

AquaMaps

European freshwater AquaMaps were generated for a dozen of species. A scientist hired at NRM modified the parameter selection based on practical tests and made a C version of AquaMaps showing maps in ArcView and statistics in R, and FIN constructed the web interface map data for a large number of species (fish, mammals, and amphibians).

The European freshwater AquaMaps are currently integrated in the BioFresh data portal based on the by mapping the original output data. This allows the user to browse the different species, visualise them and zoom in on a specific area (Fig. 1). In the future, these maps will be integrated in the BioFresh atlas interface, which is currently under construction, and will provide access to a wider variety of species and a number of interactive options as specified under “Future developments”.



Browse Aquamaps:

Abramis brama Linnaeus, 1758

Overlays
☑ Abramis brama Linnaeus, 1758 [probability-squares 0.5°]
Base Layer
○ Google Hybrid
● OpenStreetMap

Aquamaps-probability:
🟨 0.1 - 0.19
🟨 0.2 - 0.39
🟧 0.4 - 0.59
🟥 0.6 - 0.79
🟥 0.8 - 1

Data sources: GBIF NRM

Cite AquaMaps itself as:

Kaschner, K., J. Rius-Barile, K. Kesner-Reyes, C. Garilao, S.O. Kullander, T. Rees, and R. Froese. 2010. AquaMaps: Predicted range maps for aquatic species. World wide web electronic publication, www.aquamaps.org, Version 08/2010.

Fig. 1: Screenshot of the AquaMaps implementation

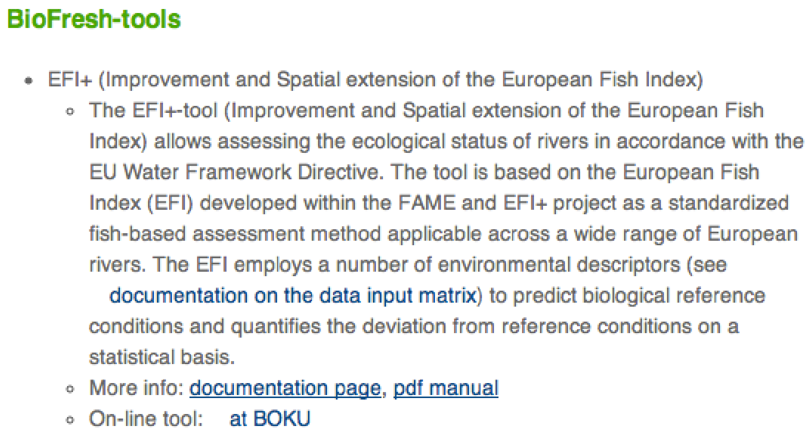
Future Freshwater AquaMaps developments

The start of the work on AquaMaps was delayed by the difficulty to recruit a scientist with the appropriate qualifications, initially foreseen for early 2010, it was realized only in January 2011. Phase 1 was concluded up to the end of 2011, when the scientist in charge unexpectedly left the position. Phase 2 activities necessarily were underprioritized for the need of a stable platform for freshwater modeling. A replacement is available only from 1 December 2012, with the explicit mission of populating BioFresh with AquaMaps-modelled maps. Meanwhile, NRM has also produced the necessary map files, environmental layers, and a test species list for Australia as a step towards having a global AquaMaps presentation instead of a continent-based as it is at present.

The integration of the freshwater AquaMaps in the BioFresh portal will be further improved. The intention is to use the web service developed for GBIF, and link on to the AquaMaps website where users can view or create maps *ad libitum*. This web service is currently available, and will be integrated during the development of a more advanced mapping interface and geoportal for the BioFresh data portal.

European Fish Index (EFI+)

This tool is integrated on a specific 'tools page' on the data portal (Fig. 2) along with a listing of other relevant models and tools for freshwater biodiversity researchers.



BioFresh-tools

- EFI+ (Improvement and Spatial extension of the European Fish Index)
    - The EFI+-tool (Improvement and Spatial extension of the European Fish Index) allows assessing the ecological status of rivers in accordance with the EU Water Framework Directive. The tool is based on the European Fish Index (EFI) developed within the FAME and EFI+ project as a standardized fish-based assessment method applicable across a wide range of European rivers. The EFI employs a number of environmental descriptors (see documentation on the data input matrix) to predict biological reference conditions and quantifies the deviation from reference conditions on a statistical basis.
    - More info: documentation page, pdf manual
    - On-line tool: at BOKU

Fig. 2: Screenshot of the BioFresh tools page linking to the EFI+-tool

Predicting freshwater-dependent species distributions in Europe

Johan Östergren[1,2], Sven Kullander[1], Kathleen-Kesner Reyes[3], Kristin Kaschner[4] and Rainer Froese[5]

[1] Department of Vertebrate Zoology, The Swedish Museum of Natural History, SE-104 05 Stockholm, Sweden

[2] Department of Aquatic Resources, Institute of Freshwater Research, Swedish University of Agricultural Sciences, Stångholmsvägen 2, SE-178 93 Drottningholm, Sweden

[3] FIN, Philippines

[4] Evolutionary Biology & Ecology Lab, Institute of Biology I (Zoology), Albert-Ludwigs-University Freiburg, Germany

[5] Leibniz-Institute of Marine Sciences, Düsternbrooker Weg 20, D-24105 Kiel, Germany

Keywords: spatial modeling, envelope model, GIS-modeling, range maps

Corresponding author: Johan Östergren, Phone: +46703461429 fax: +468######, email: johan.ostergren@nrm.se

Abstract

The distribution of species dependent on a freshwater environment is modeled on a large scale using a freshwater version of the model AquaMaps. Predictions are based on global occurrence data and generated by matching habitat usage of species, termed environmental envelopes, against local environmental conditions to determine the relative suitability of specific geographic areas for a given species. In the freshwater environment there is a significant challenge to the production of range maps due to large biases in the amount and location of occurrence data, and due to historical events affecting the connectivity and dispersal of species. In the model we therefore include Pfafstetter basins to limit native distributions. Independent survey data, and literature range maps, were used to validate the model outputs showing that AquaMaps predictions of ten species representing fish, amphibian, and mammal taxa were well in accordance with the species native distributions described in the literature. The accuracy of the model predictions were 83–85% depending on selection and number of predictor parameters used in the model. Predictions of suitability of a given species across Europe indicated that many of the species are likely to survive or establish outside of their current distribution range. We also show that the choice of environmental parameters used in the model effect the accuracy of the predictions and that the optimal predictions not necessarily are achieved with statistically selected species-specific parameters.

Introduction

Knowledge on spatial and temporal species distribution is crucial for conservation and management of a species. Predicting species distribution in space and time is therefore very important. Despite recent advances in spatial- and niche-modeling, and the increasing availability of comprehensive online databases of environmental and species occurrence data, predicting distribution of species at a global scale remains challenging. Availability of global data sets, potential bias in occurrence data, and adequate scale and resolution for the models used, needs to be considered. The freshwater environment is particularly challenging to model at a global scale, since the habitat may change significantly in short distances, and species special habitat requirements may limit species distributions at a close distance. Also, migration and dispersal of species in freshwater are affected by historical events as the ice ages, geomorphology, e.g., through migration barriers, and also by human activities. Humans have been transferring animals, e.g., fish pass migration barriers, to achieve additional areas with a species that could serve as an additional food resource (Spens 2008). Humans have also introduced migration barriers as dams for milling, water power production etc, which together with other anthropogenic effects (e.g. pollution, habitat degradation) have affected the distribution of many species and freshwater biodiversity (NRC 1996; Nilsson et al 2005; Geist 2011). A good species distribution model that uses presence of species, as occurrence points reported to some database, optimally would use comprehensive presence and absence data from surveys. Many country-based authorities conduct these kinds of surveys in freshwater, however, the availability of this information is still limited at a global scale. Another challenge is the selection of appropriate data sets of environmental parameters for the model. Beaumont et al. (2005) found that increasing the number of predictor variables caused a decrease in the size of areas of predicted presence. There is a risk of over-fitting the model when using too many predictive parameters. However, too few parameters will probably under-fit the model. The biological relevance of the parameters used also needs to be considered (Synes and Osborne 2011).

One approach to modeling species distributions at a global scale is the AquaMaps model (Ready et al 2010). The AquaMaps is an approach to generating model-based, large-scale predictions of species distribution and is based on a methodology, which was originally developed for marine mammals (Kashner et al 2006). Predictions are based on occurrence data and estimates of the environmental tolerance of a given species at a resolution of 0.5◦ latitude×0.5◦ longitude. AquaMaps is designed to be a transparent and easy-to-understand model, which can be used online by researchers and stakeholders. The model also allows the incorporation of expert knowledge about habitat usage. Ready et al (2010) compared the marine AquaMaps with several other ecological niche models (GARP, GLM, GAM, & MAXENT) and concluded that AquaMaps outputs compare well to methods tested. Its transparency, ability to incorporate expert knowledge, and online access with a user friendly interface, makes it a valuable tool increasing practical use in the context of decision making and planning processes.

This study describes a version of AquaMaps that is calibrated for the freshwater environment. The model use occurrence data from the Global Biodiversity Information Facility GBIF, and environmental parameters suitable for niche modeling of freshwater-dependent species. For the freshwater environment, to our knowledge, there is no global data set for, e.g., pH, water temperatures, ice-cover, depths, etc. available. We therefore use soil data, topography, and climatic variables. Predictions from AquaMaps for ten selected freshwater-dependent fish, amphibian, and mammal species are presented and compared. We validate the model using the species distribution maps from literature and IUCN, as well as electro-fishing surveys.

The aim with this study is to describe and discuss species distribution maps generated by a freshwater version of the model AquaMaps. We also compare two methods of selecting environmental data to use in the modeling, one manually selected data set based on ecological theory and a second where selection is based on a statistical approach.

Methods

Occurrence data and environmental parameters

Global occurrence records were extracted from the Global Biodiversity Information Facility (GBIF), a comprehensive, online database of species occurrence records from museum collections, surveys and other sources (GBIF, 2011). In this study, we extracted data from 10 species representing three taxa; four fishes, three mammals, and three amphibians, to illustrate the model and for model validation. The selected species represent a broad range of life histories and habitat preference. For each taxon, we chose species occurring in Europe with a relatively restricted as well as a wide native range. Fish species were adequately represented in fisheries surveys (e.g. National Board of Fisheries, Sweden) which served as test data in the validation of the model. Raw occurrence data (all accumulated occurrence data per species) from GBIF were extracted for all species. Occurrence points were cleaned from outliers and points outside of the species native range and spatially aggregated into grid cells at a resolution of 0.5◦ latitude×0.5◦ longitude. Outliers that were largely separated spatially were checked individually and removed if the source indicated the wrong species scientific name. The native range was defined by the literature (Kottelat & Freyhof 2007; Arnold & Burton 1978; Steward 1969) and by the species distribution maps available online on the IUCN red list of threatened species http://www.iucnredlist.org/.

The choice of environmental parameters for species distribution modeling may significantly affect the outcome of the model, and ecological relevance as well as statistical methods may reduce the uncertainty of model predictions (Synes and Osborne 2011). Here, we use a set of 27 variables that are comprehensive, publicly accessible global raster data sets and key environmental variables based on long-term average conditions. From this data set we select six to eight parameters using two methods. First, the original AquaMaps parameters are selected based on ecological theory (Austin, 2007), using ecologically relevant parameters for the taxa fish, amphibians, and mammals, respectively. A second approach was a statistical choice of parameters to use for all species individually. The 27 parameters used are elevation, average air temperature in June, mean annual precipitation, net primary production, soil pH, soil moisture, soil carbon, compound topographic index (CTI), and additional 19 climatic variables which are also considered to be biologically meaningful, and are often used in ecological niche modeling (e.g., BIOCLIM, the Genetic Algorithm Rule-set Procedure GARP, maximum entropy Maxent). Global datasets of the 27 parameters were downloaded from internet, prepared at 0.5◦ resolution covering Europe (8593 cells). All geospatial data analysis was performed using ArcMap v.10 (Environmental Systems Research Institute). Data is generally available at greater resolutions than the 0.5◦ resolution used here, and was converted to such by calculating mean, minimum and maximum values, and used as appropriate for mean, minimum and maximum layers. Sources and an explanation to all parameters are presented in Appendix A.1.

Bounding boxes

In freshwater, species distributions are in many cases restricted by the lack of connectivity or by historic events (Spens 2008). This is naturally more common among fish species since they are dependent on an open water environment to relocate or invade new areas. Thus, a species may be absent in an area with suitable environment or habitat simply because it has not been able to migrate there. In this version of AquaMaps, we use Pfafstetter basins to spatially limit predictions of native distributions and in this way account for connectivity issues and historic events. The Pfafstetter system is a system for watershed classification, and was developed by Otto Pfafstetter (Pfafstetter 1989; Verdin & Verdin 1999). The Pfafstetter system is hierarchical assignment of watershed IDs based on topology of land surface (digital elevation model or DEM and river networks). The watershed IDs are used in an area-to-area navigation program for automatic identification of watersheds upstream and downstream of a given watershed. There are six basin levels where L1 is the continental scale watersheds and L2 – L6 are finer tessellations/sub-watersheds. Here, the Pfafstetter basins are used to create 'bounding boxes'. The bounding boxes are based on the literature and IUCN maps simple by overlaying distribution maps of the specific species over the Pfafstetter basin layer in ArcMap v10. We used several levels in a modified layer of Pfafstetter basins in order to get a sufficiently detailed yet reasonable resolution matching native maps. The modification was done in the following steps: Pfafstetter basins were downloaded from USGS Hydro 1K (http://eros.usgs.gov/#/Find_Data/Products_and_Data_Available/gtopo30/hydro). L1 divided Europe into ten major basins, e.g., Danube, Don, Dniepr, and watersheds draining into the Caspian Sea (Fig 1). Since the L1 resolution was too wide for our modeling approach, we selected L2. This level gave sufficient detail for several areas, for example watersheds draining into the Black Sea (Fig 1). However, most of Western Europe was part of one basin and the watersheds draining into the Caspian Sea was not separated sufficiently (Fig 1). In these areas, it would not have been possible to create native ranges for species restricted by some of Europes' main watersheds, e.g., Oder, Elbe, Rhine, Seine, and Loire. Therefore, these areas were divided into L3 (Fig 2). The Pfafstetter basin 911 (L3) covered the entire northern coast of the Mediterranean Sea and was further divided into L4 (Fig 3). This gave the possibility to predict the native distributions of all species that are restricted to a small area in the northern Mediterranean (see examples of restricted species in Kottetlat & Freyhof 2007 and Steward 1969). Finally, the Caspian Sea basin (L2 = 2; Fig 1) was also divided into L3 and islands were treated as unique basins instead of part of an existing Pfafstetter basin since islands have a natural lack of connectivity through the open sea (Fig 3). The islands were originally part of Pfafstetter basins, for example Ireland and the UK were part of basin 913 together with the river Seine drainage (Fig 2). In total, Europe was divided into 131 basins with an average area of 130 180 km^2 (SD: 245 209 km^2).

Model description

A detailed description of the marine version of AquaMaps is provided in Ready et al (2010) and also in Kashner et al (2010). Here, we provide a model overview and details on the parts that differ in the freshwater version of AquaMaps used in this study.

Briefly described, the model creates environmental tolerance or suitability curves (envelopes) for each environmental parameter. These envelopes are based on the parameter values at the locations where the species is found to occur within the species native range and have a trapezoidal shape (Fig. 1 of Ready et al 2010). The trapezoidal shape derives from a set of rules based on the absolute and preferred minimum, and the absolute and preferred maximum for each parameter (for details see Ready et al 2010). The envelope is rescaled to give a suitability value (or probability) between 0 and 1, for all parameter values. Predictions are then made by comparing the parameter values in all cells with the respective environmental envelope, which then give the parameter-specific likelihood (or probability value) for a species to occur in each cell. Thus, the likelihood for a species to occur in each cell is zero for values ≤absolute minimum and ≥absolute maximum. Species-specific relative habitat suitability for each grid cell is then computed as the geometric mean of all individual predictor probabilities, assuming an equal weighting for all. This is then presented as color-coded probability (between 0 and 1) for a species to occur in one cell.

For the model predictions, we define cells with a species occurrence record that are located within the bounding box of the species in question as 'good cells', and these are used for creating the environmental envelopes. We predict and create two types of distribution maps for the ten modeled species; a *native map* and a *suitable habitat map*. The native map is restricted with the bounding box. The suitable habitat map has no geographic restrictions and covers the entire Europe and parts of western Asia, according to the limitations by all Pfafstetter basins together (Fig 1). For all species we use three sets of predictor parameters; one parameter set based on ecological relevance (termed *original*), a second species-specific statistically derived parameter set (termed *VIFM*), and finally the whole set of 27 parameters (termed *ALL27*).

All permutations of the AquaMaps methodology were implemented in a desktop version of the software, copies of which are available from the lead author. Note that the online version of AquaMaps (http://www.aquamaps.org) is only available in the form of the basic 'expert quaMaps' method which incorporates independent information about depth preferences and maximum *latitudinal and longitudinal boundaries that are available from FishBase. Envelope settings for all other environmental parameters have only been reviewed on a case-by-case basis. Given the same input settings to AquaMaps, the outputs will always be the same. This in contrast to some other methods such as GARP, which produce slightly different outputs with the same input information.* (Ready et al., 2010).

Calibration and validation
Model calibration was done with the use of three sets of data. The first was the ecological relevant set of six to eight parameters for each taxon (termed *original*). The second set was selected based on a statistical approach in two steps (*VFIM*), and the third was using the full set of 27 parameters (*ALL27*). For the VFIM, first the 27 parameters were tested for correlation using correlation matrix and the variance inflation factor (VIF). VIF is a measure of variable multicollinearity, and have been used in ecological modeling to verify non-correlations of variables (e.g. Singh et al 2010). Correlated variables may induce a bias in model estimations and should therefore be treated with care (Graham, 2003). A VIF between 4 and 10 has been used as cutoff for eliminating variables, since higher levels of VIF are considered to indicate multicollinearity (Graham, 2003). To simply use VIF as a cutoff has been questioned, since the VIF value depends on the properties of the model (number of variables, sample size etc., see O'Brien 2007 for discussion). Here, we selected the value of 10 as cutoff and removed variables with higher values only after confirming high correlation via the correlation matrix. When checking for VIF, the variable with the highest value for VIF was removed and this process was repeated until no variable with a VIF larger than 10 are apparent in the parameter set.

The second step to select parameters for VFIM was based on boosted regression trees (BRT). BRTs are a boosted version of regression trees. Classification and regression trees (CART; Breiman 1983) are considered as a very good method to use in data mining and investigating the importance of different variables, e.g., ecological variables in a modeling set up. It has been used in ecological studies (De'ath and Fabricius 2000; Kolar and Lodge 2002) and also in spatial modeling predicting species presence versus absence (Spens et al 200?; Elith et al 2008). In this study, we follow the methods described by Elith et al (2008). The BRT approach gives a value of relative importance for each variable used in the model. BRTs were constructed for the ten model species separately, using presence/absence data. Since only presence data are available from GBIF, we used the species native distributions maps downloaded from IUCN as to define the presence area of a species. We then selected 500 random cells within the IUCN map area to represent the presence cells, and 500 random cells outside the IUCN map area to represent absence cells. If the species was restricted to, or absent from, <500 cells, all cells were used as presence and/or absence data, respectively. From the BRT analyses, the six to eight parameters with highest relative importance were selected and used for map predictions. The statistical analyses are conducted in R (www.r-project.org).

To compare the outcome of the parameter sets we calculated the overall accuracy of the models (in percent), as the number of correct predictions, i.e., predicted presence in a cell within the IUCN map area and predicted absence in a cell outside the IUCN map area, divided by the total number of cells. The proportion of correctly predicted presence cells of the total number of presence cells was also calculated for each species. In AquaMaps, the predictions gave a probability of species occurrence for each cell, and in this study we investigate two levels defining presence of a species; the first at $P > 0.05$ and the second at $P > 0.50$. In addition, plots of ROC values and accuracy versus cutoff level are provided in Appendix A.2.

To validate the model further we used one fish species, parameter values and occurrence data from Sweden, France and the UK, in total 697 cells. These areas are selected since the national program of survey fishing in freshwater report all their observations to GBIF in Sweden, France and the UK (Swedish National Board of Fisheries). We assume that the occurrence reflects the presence and absence data, and hence assume that if a species exists, or have been observed, it is also reported to GBIF. This above mentioned assumptions might however include some bias of the actual species distribution, but as a base to evaluate the model it should be sufficiently reliable. In this exercise, we use only Sweden and France to extract data from cells that contain an occurrence point of the species, i.e., to create the good cells. We then run predictions on the UK. In this way we have an independent presence/absence data (UK) on which the prediction using two different sets of variables are compared.

Results

Occurrence and environmental data

The occurrence point-data downloaded from GBIF used in the study was not evenly distributed within the native range of the model species, with the exception of species with a native range restricted to the western part of Europe. Overall, occurrence point data was missing in Eastern Europe. Countries as Russia, Ukraine, and Belarus together had < 10 occurrence points of all modeled fish and mammals within the native distributions described in the literature. This of course had an influence on the predictive potential of the model and illustrates one of the largest challenges with spatial modeling using occurrence databases, i.e., lack of or biased data.

The average number of occurrence cells (i.e. good cells) for our ten species used for modeling was 329 cells (Min–Max: $n = 32–548$; Table 1). The native range was on average 2292 cells (Min–Max: $n = 72–4831$) and the relative ratio of good cells (n good cells/ n cells within native range) varied between 1 and 87% (Table 1).

The VIF values of all 27 parameters, the original and parameters used for VIFM sets are given in table 2. All VIFM sets, i.e., the species-specific statistically derived sets, were unique across species (Table 3). No species had the same set of parameters as another species. One parameter was apparent in all sets, *net primary production* NPP, and had also the highest relative importance with an average of 21% (Table 3). After NPP, six climatic parameters followed, with *isothermally* and *mean temperature of driest quarter* as the most important. *Compound topographic index* CTI and *soil moisture* were the two least important parameters (Table 3).

Map predictions

Given the lack of occurrence data, freshwater AquaMaps performed well in predicting the distributions of the modeled species with an average accuracy (suitable habitat map) of 83% for the original, VIFM and ALL27 parameter sets at the $P > 0.05$ level, and an accuracy of 85, 84, and 79% for the three data sets, respectively at $P > 0.50$ (Table 4). Predictions of a species with as few as 32 good cells (*Silurus glanis*) gave an accuracy of 70–72% the original and VIFM, despite that the relative ratio of good cells only was 1% (Table 4). The highest accuracy were achieved for species restricted to the western part of Europe; *Lissotriton helveticus*, *Pleurodeles waltl*, and *Galemys pyreneaicus*. These three species had the smallest native range and the highest relative ratio of good cells (Table 1). There was a positive relationship between relative ratio of good cells and accuracy (original model linear regression: $p < 0.05$; Adjusted R-squared: 0.41–0.67 for $P > 0.05$ and $P > 0.50$, respectively), but no relationship between number of good cells and accuracy (original model linear regression: $p > 0.05$; Adjusted R-squared: 0.00–0.06 for $P > 0.05$ and $P > 0.50$, respectively). In spite of the relatively high accuracy, distributions of wide spread species were poorly predicted in the eastern part of Europe, e.g., *Silurus glanis* (Fig. 4–#).

The predicted presence of a species (i.e. presence cells within the native range) ranged between 34 (*Silurus glanis*) to 96% (*Lissotriton helveticus*) at the $P > 0.05$ level the original and VIFM models. There was a positive relationship between relative ratio of good cells and presence at the $P > 0.50$ level (original model linear regression: $p < 0.05$; Adjusted R-squared: 0.43), but not at the $P > 0.05$ level. There was no relationship between number of good cells and presence (original model linear regression: $p > 0.05$; Adjusted R-squared: 0.05–0.15 for $P > 0.05$ and $P > 0.50$, respectively).

The use of the bounding box significantly improved the predictions of the native maps (Fig. 4–#). The bounding box was particularly helpful to restrict distributions on islands, e.g., *Lissotriton helveticus* on Ireland. Also within the native range AquaMaps could identify areas not suitable for the species, e.g., the high altitude in the Alps. The native maps were also more similar between the models than the all suitable maps (Fig. 4–#).

In seven out of ten suitable habitat maps, VIFM predictions where more restricted, i.e., fewer cells had suitable habitat according to the model predictions, than the maps predicted with the original model. For the two species that had been introduced to outside of their native range (*Silurus glanis* and *Chondrostoma nasus*), this revealed one important difference between the models. VIFM restricted the suitability to the native range and predicted that the area to where it had been introduced not was suitable for this species. Thus, the VIFM predicted unsuitable environmental conditions in areas where the species evidently survived and established. The original model, however, predicted correctly that the areas where the species had been introduced were suitable for the species.

The suitable habitat maps predicted with all 27 parameters gave an accuracy that was similar or lower than the models original and VFIM (Table 4), but the predicted presence was overall low with an average of 28% compared to 56% at the $P > 0.50$ level for the original model (51 and 77% at $P > 0.05$; Table 4). For *Silurus glanis* the predicted presence at the $P > 0.5$ level was only 4%. This also illustrates the importance of considering the number of predictors in the model.

Discussion

The freshwater AquaMaps model presented here was developed for predictions of species distribution at a global scale, using presence only data and available global environmental data. As such, the distribution maps predicted by the model performed well. However, there were relatively poor predictions for species with a wide range, in the eastern part of Europe. This is most likely due to the lack of occurrence points in this area. Without point data that adequately represent the species environmental preference, the model will not be able to predict the species distribution correctly. Several countries has not yet digitalized their occurrence information from surveys and museum collections, or not yet submitted the data to the GBIF database used in this study (Freyhof 2011, unpublished information). Once the information in GBIF is complete, the model predictions will improve. The species with the highest ratios of presence cells with occurrence points of all cells within the native distribution gave the best results in terms of accuracy of the suitable habitat maps. This illustrates the importance of occurrence data, however, the AquaMaps model was able to predict relatively well even at a ratio of 1%, achieving an accuracy of 70–72%.

We used bounding boxes based on Pfafstetter basins (Pfafstetter 1989) to limit the native maps. This significantly helped the distribution maps to match with the native map in the literature. This was not surprising since the bounding boxes were based on the native maps, but the good conformity with the native maps from IUCN shows that this can be a suitable approach when creating distribution maps for freshwater-dependent species. Also, the model correctly predicted high probability of occurrence within the entire bounding box for most species, and the suitable habitat maps were also similar to the native distributions. The suitable habitat map may be used to investigate the potential distribution outside of the species native range. From the suitable habitat maps presented here it is worth noting that many species may be expected to survive and maybe establish also outside of the native distribution range. Some examples are the species *Pleurodeles waltl*, *Lissotriton helveticus*, and *Triturus cristatus*, that all seem to have suitable habitats in parts of Italy, Greece, and Turkey. The two first ones could also potentially establish on Ireland. This is interesting and raises the question on why they are not established in these areas. One explanation could be the lack of connectivity, or a high cost of colonization migration. One obvious reason for the absence of *Lissotriton helveticus* and *Triturus cristatus* on Ireland and is that this species has not been able to pass the sea that separates Ireland from the UK and the rest of Europe. According to Stewart (1969) this two species were able to colonize the UK before it separated from the European main land. Other species might have had problems to pass large mountains as the Pyrenees and the Alps, e.g., *Triturus cristatus*, *Lissotriton helveticus*, and *Abramis brama,* which hindered them to colonize Spain and/or Italy.

For all species, the current distribution has been affected by historical events, such as the ice-ages. The ice cover in Europe has naturally limited species to refuge in central and southern Europe. The re-colonization that has taken place then depended on the geographic position of the refuge (Steward 1969). Other potential factors that might affect the distribution of species, not encountered for here, are inter- and intraspecific competition, and predation pressure from other species.

One interesting observation was that the two ways of selecting explanatory parameters, or predictors, did not give any overall difference in prediction accuracy. Thus, original parameters selected based on ecological theory (Austin 2007), gave similar predictions as our statistical approach with individually selected 'best' parameter sets VIFM. The VIFM gave a more restricted distribution map over the majority of species. For the two fish species that had been introduced to areas outside of their native range, VIFM indicated a too restricted prediction. According to the model, *Silurus glanis* and *Chondrostoma nasus* did not have suitable habitat in areas where they had been introduced. Since they evidently survived and established in these areas, this illustrates that one should be aware of the potential of a model to over-fit or over-restrict predictions, and one should be careful to use too many and only statistically derived predictors. Our predicted maps using all 27 parameters illustrate the importance of considering the number of predictors further. The predicted maps were very restricted compared to the other two maps, and the ALL27 map predicted species presence poor and much lower than the other models. Beaumont et al. (2008), predicted butterfly distributions using 35 parameters from BIOCLIM, and showed that the area of predicted presence decreased with number of parameters used in the model with a risk of over-fitting as a consequence. In our study, using the 6–8 parameters seems reasonable and using all 27 seems to over-fit the model predictions.

MISSING

Conclusions

The freshwater AquaMaps model presented here performs well in predicting both native and suitable habitat distribution maps of freshwater-dependent species. This means that the AquaMaps model, originally developed for marine mammals and fish, as a transparent and user friendly online mapping tool, also can be used in terrestrial species distribution modeling. The approach of using Pfafstetter basins to limit the native distributions works well, and we also show with the suitable habitat map that potential distribution outside of the native range can be predicted with high accuracy, with as few occurrence points as 32 cells. Further, the use of a parameter set selected based on ecological relevance performed equally well as a parameter set selected based on a statistical approach, while the full set of 27 predictors gave too narrow maps, basically just reproducing a point map.

Acknowledgement

The study was conducted with funding from the BioFresh project, grant agreement number 226874.

References

Arnold, A. N. & Burton, J.A. 1978 A field guide to the reptiles and amphibians of Britain and Europe. 264 p. William Collins Sons & Co. Ltd. London.
Austin, M. 2007. Species distribution models and ecological theory: A critical assessment and some possible new approaches. Ecological Modelling, 200: 1-19.
Beaumont, L.J., Hughes, L. & Poulsen, M. (2005) Predicting species distributions: use of climatic parameters in BIOCLIM and its impact on predictions of species' current and future distributions. Ecological Modelling, 186, 250–269.
Breiman, L., Friedland, J.H., Olshen, R.A. & Stone, C.G. 1984. Classification and Regression Trees. Belmont,California, USA: Wadsworth International Group.
De'ath, G. & Fabricius, K.E. 2000. Classification and regression trees: A powerful yet simple technique for ecological data analysis. Ecology 81: 3178-3192.

Elith, J, J. R. Leathwick and T. Hastie. 2008. A working guide to boosted regression trees. Journal of Animal Ecology. 77: 802–813.
GBIF, 2011. Global Biodiversity Information Facility. World Wide Web electronic publication http://www.gbif.org/. Accessed September 2011.
Geist, J. Integrative freshwater ecology and biodiversity conservation. Ecological Indicators. 11 (2011) 1507–1516. doi:10.1016/j.ecolind.2011.04.002
Graham, M. (2003) Confronting multicollinearity in ecological multiple regression. Ecology, 84, 2809–2815.
Kaschner, K., J. S. Ready, E. Agbayani, J. Rius, K. Kesner-Reyes, C. Garilao, P. D. Eastwood, A. B. South, S. O. Kullander, T. Rees, C. H. Close, R. Watson, D. Pauly, and R. Froese. 2010. AquaMaps: Predicted range maps for aquatic species. World wide web electronic publication, www.aquamaps.org, Version 08/2010.
Kolar, C.S. & Lodge, D.M. 2002. Ecological predictions and risk assessment for alien fishes in North America. Science 298: 1233-1236.
O'Brien, R. M. 2007. A Caution Regarding Rules of Thumb for Variance Inflation Factors. Quality & Quantity. 41:673–690.
Kottelat, M., and Freyhof, J. 2007. Handbook of European freshwater fishes, Publications Kottelat, Cornol.
Pfafstetter, Otto., 1989. Classification of hydrographic basins: coding methodology, unpublished manuscript, Departamento Nacional de Obras de Saneamento, August 18, 1989, Rio de Janeiro; available from J.P. Verdin, U.S. Geological Survey, EROS Data Center, Sioux Falls, South Dakota 57198 USA
Ready, J.S, K. Kaschner, A. B. South, P. D. Eastwood, T. Rees, J. Rius, E. Agbayani, S. O. Kullander and R. Froese. 2010. Predicting the distribution of marine organisms at the global scale. Ecological modelling. 221: 467–478.
Singh N.J., I.A. Grachev, A.B. Bekenov, E.J. Milner-Gulland. 2010. Saiga antelope calving site selection is increasingly driven by human disturbance. Biological Conservation 143: 1770–1779.
Spens, J. 2008. Using GIS to Predict Landscape-scale Establishment and Extinctions of Lacustrine Salmonids. In Department of Wildlife, Fish, and Environmental Studies, p. 44. Swedish University of Agricultural Sciences, Umeå.
Spens, J., Englund, G., and Lundqvist, H. 2007. Network connectivity and dispersal barriers: using geographical information system (GIS) tools to predict landscape scale distribution of a key predator (Esox lucius) among lakes. Journal of Applied Ecology, 44: 1127-1137.
Steward, J. W. 1969. The Tailed Amphibians of Europe. David & Charles. Ltd. Newton Abbot.
Verdin, K.L. and J.P. Verdin, 1999, A topological system for delineation and codification of the Earth's river basins," Journal of Hydrology, vol. 218, nos. 1-2, pp. 1-12
Synes, N. W., and Osborne, P. E. 2011. Choice of predictor variables as a source of uncertainty in continental-scale species distribution modelling under climate change. Global Ecology and Biogeography, 20: 904-914.
Verdin, Kristine L. 1997. "A System for Topologically Coding Global Drainage Basins and Stream Networks". 1997 Esri International GIS User Conference Proceedings: (6/14/01)

http://proceedings.Esri.com/library/userconf/proc97/proc97/to350/pap311/p311.htm
http://www.aquamaps.org/
http://www.fishbase.org/
http://www.gbif.org/
http://www.iucnredlist.org/

Table 1. Fish, amphibian, and mammal species occurrence data used for model training and model testing. Common name taken from IUCN.

Species name	Abbreviation	Common name	Taxon	N native cells (NC)	N good cells (GC)	GC/NC (%)
Galemys pyrenaicus	*G.pyr*	Pyrenean Desman	Mammal	72	59	82
Pleurodeles waltl	*P.wal*	Iberian Ribbed Newt	Amphibian	136	118	87
Lissotriton helveticus	*L.hel*	Palmate Newt	Amphibian	526	390	74
Chondrostoma nasus	*C.nas*	Nase	Fish	1140	44	4
Neomys anomalus	*N.ano*	Southern Water Shrew	Mammal	1528	188	12
Triturus cristatus	*T.cri*	Northern Crested Newt	Amphibian	2328	516	22
Silurus glanis	*S.gla*	Wels Catfish	Fish	3204	32	1
Abramis brama	*A.bra*	Bream	Fish	4483	548	12
Lota lota	*L.lot*	Burbot	Fish	4673	464	10
Lutra lutra	*L.lut*	Eurasian Otter	Mammal	4831	933	19
Average				*2292*	*329*	*32*

Table 2. Variance influence factor (VIF) for parameters used in the models for all 27 (ALL), the 13 with VIF < 10 after systematically removing VIF > 10, and the parameters used in the original models for fishes, amphibians, and mammals, respectively.

Parameter	ALL	VIF<10	Original Fish	Original Amphibian	Original Mammal
CTI_Max	1	1.3			
Elevation	4	2.1	1.2	1.3	1.4
NPP	6	4.0	2.8	2.9	2.9
PrecipAnMe	7299		5.2	5.2	5.2
SoilCarbon	4	3.1		3.0	3.0
SoilMoistu	9	7.3	5.8	6.1	6.1
SoilpH	6	4.0	2.5	3.3	3.3
TempMonthM	62		3.4	7.3	8.0
Bio1	2004				
Bio2	44				
Bio3	42	6.6			
Bio4	1492				
Bio5	$8 \cdot 10^{10}$				
Bio6	$1 \cdot 10^{11}$			2.4	
Bio7	$6 \cdot 10^{10}$	3.9			
Bio8	4	2.3			
Bio9	22	7.3			
Bio10	3054				
Bio11	7965				2.7
Bio12	7270				
Bio13	111				
Bio14	118	5.9			
Bio15	12	5.9			
Bio16	220				
Bio17	189				
Bio18	52				
Bio19	33	3.7			

Table 3. Relative variable importance in percent for the VIFM model of all species. For abbreviations see table 1.

Variable	*L.hel*	*T.cri*	*P.wal*	*N.ano*	*L.lut*	*G.pyr*	*L.lot*	*C.nas*	*A.bra*	*S.gla*
CTI_Max							5.4			
Elevation				7.3	2.0	17.7	5.3		5.8	
NPP	10.8	39.6	1.9	21.5	48.1	9.2	5.9	47.2	11.2	11.6
SoilCarbon		5.4	11.1	3.9		3.9			10.2	16.9
SoilMoistu	2.2	5.1			1.7					
SoilpH		5.9								8.1
Bio3	9.1		33.3	7.5	26.3	15.0		12.9		7.6
Bio7	35.6	3.0		14.3		4.1		6.6		
Bio8		11.9	1.7	7.8	14.6		8.5	13.6	16.8	24.4
Bio9	3.3	8.1	7.6	6.7			44.7	3.3	23.9	9.4
Bio14	30.8	10.8	14.7	14.4	3.0	3.2	11.4	6.7		
Bio15	1.4		25.0		1.3	4.8			6.5	
Bio19	4.5		1.3		1.2	35.5				

Table 4. Accuracy and presence messures of the predicted suitable habitat maps. For species abbreviations see table 1.

Species	Accuracy ($P>0.05$) %			Accuracy ($P>0.50$) %			Presence ($P>0.05$) %			Presence ($P>0.50$) %		
	Original	VIFM	All27	Original	VIFM	All27	Original	VIFM	All27	Original	VIFM	All27
L.hel	88	93	97	96	98	97	96	94	82	77	78	47
T.cri	77	72	82	81	79	80	77	77	55	50	47	29
P.wal	94	98	99	98	98	99	84	82	70	65	63	46
N.ano	85	84	87	88	88	85	83	72	48	51	43	21
L. lut	85	78	74	77	74	61	91	92	56	63	67	31
G.pyr	89	91	97	97	97	99	89	82	69	69	64	46
L.lot	80	84	66	76	73	64	46	34	11	34	20	4
C.nas	89	91	69	88	89	59	78	92	51	60	74	23
A.bra	75	70	89	76	74	88	47	42	18	29	27	8
S.gla	72	70	71	70	67	57	81	83	49	63	54	22
Average	*83*	*83*	*83*	*85*	*84*	*79*	*77*	*75*	*51*	*56*	*54*	*28*

Figure 1.

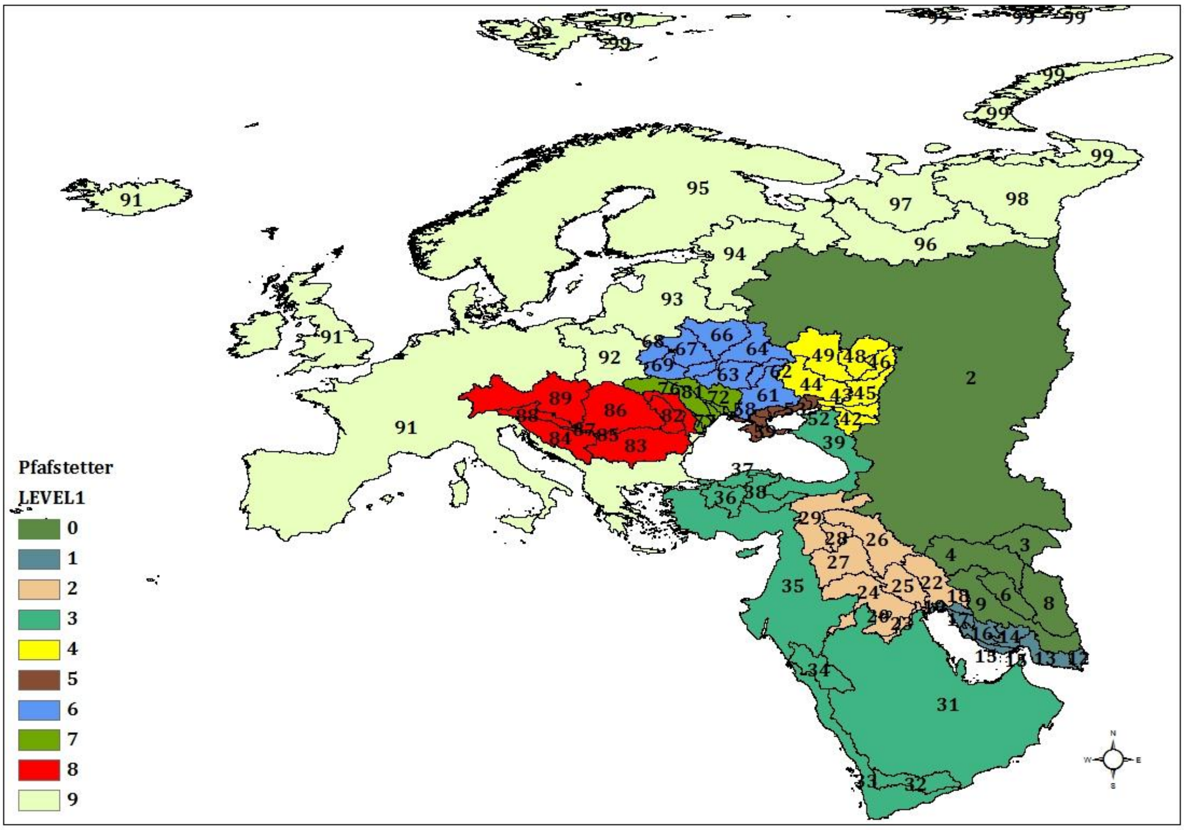


Figure 1. Pfafstetter basins of Europe and Western Asia downloaded from USGS. Level 1 is indicated with colors showing ten basins (0–9), and level 2 is indicated with numbers. Examples of Pfafstetter basins corresponding to river drainages are Danube (red), Dniepr (blue), and Don (yellow).

Figure 2.

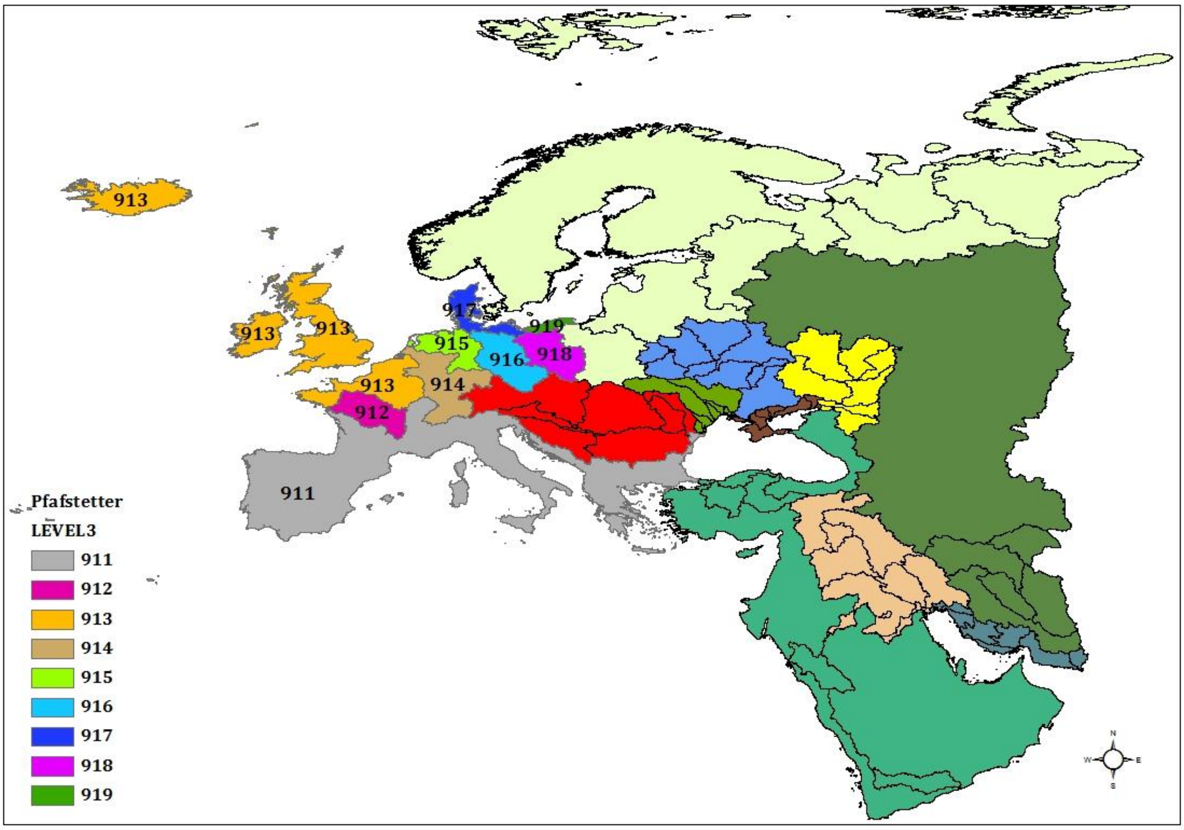


Figure 2. Pfafstetter basins of Europe and Western Asia where basin 91 (L2) is further divided into level 3, indicated by colors and corresponding numbers.

Figure 3.

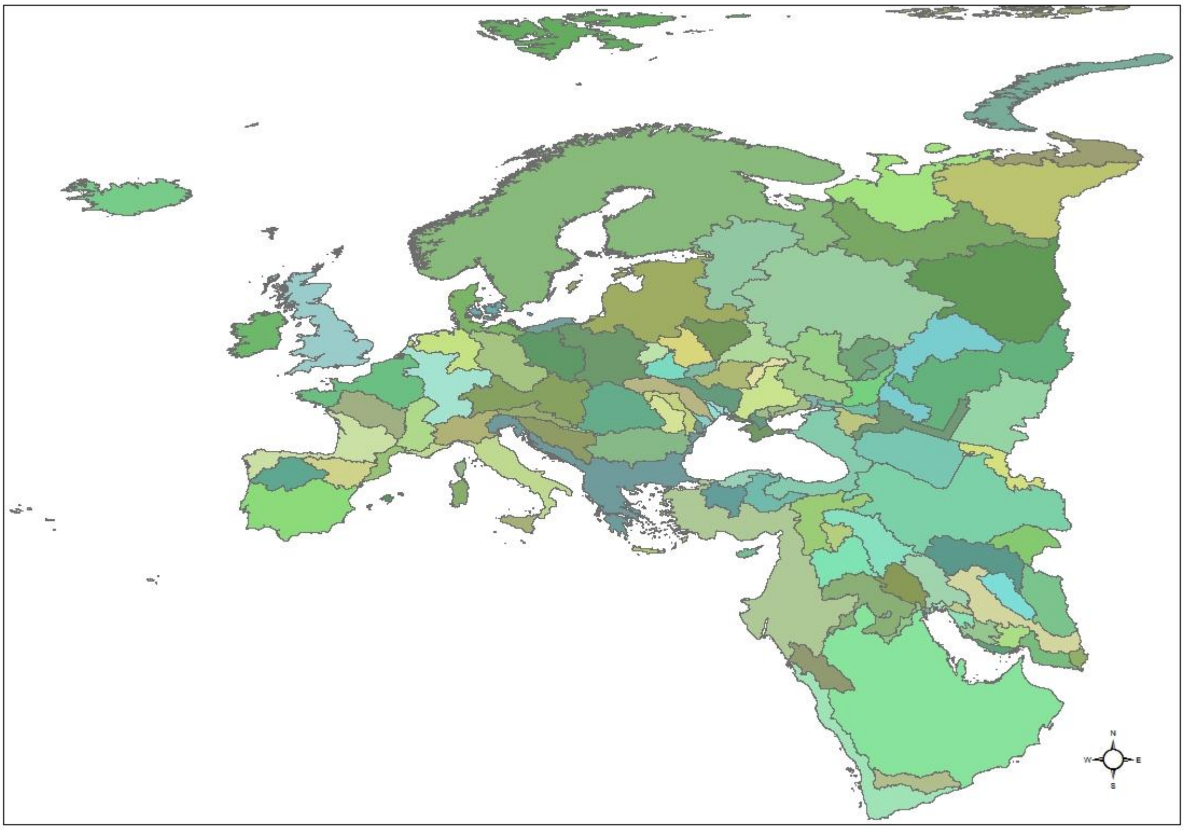
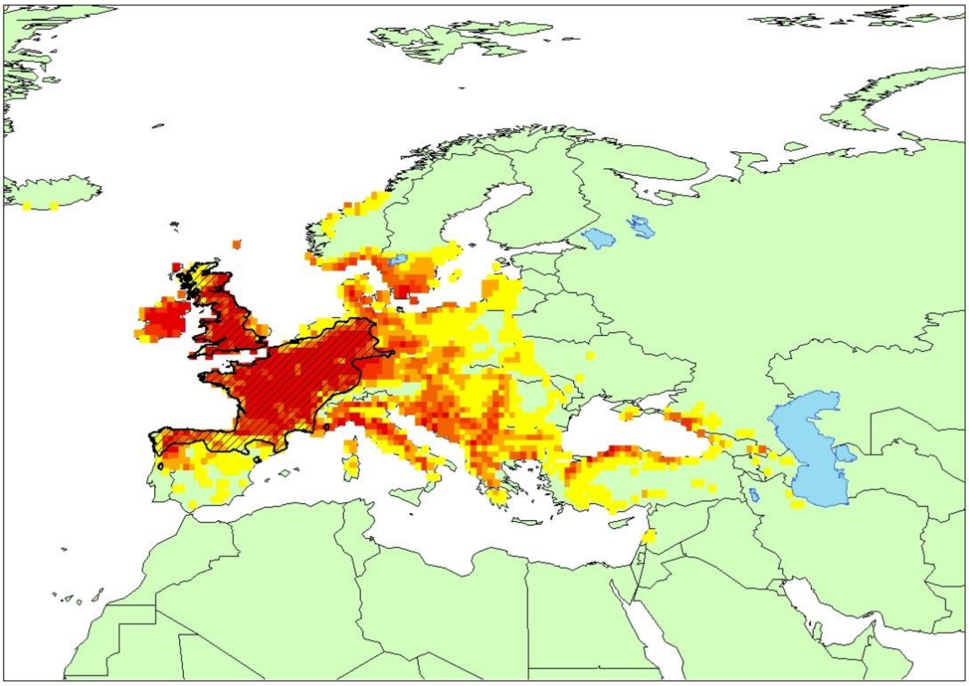


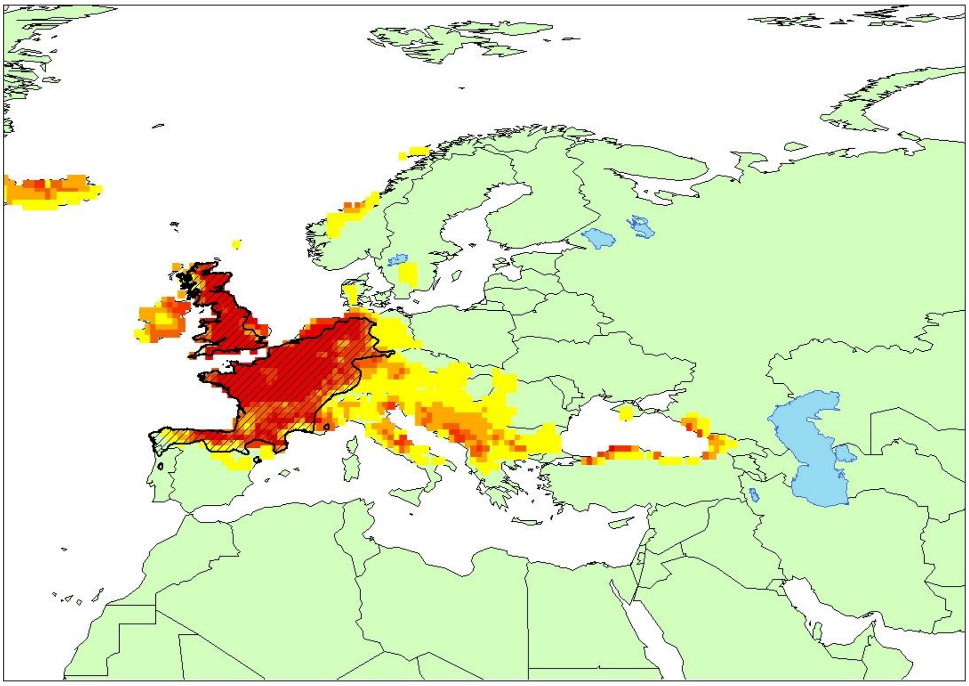
Figure 3. Final modification of Pfafstetter basins as used in the model comprising 131 basins indicated by colors. Three basin levels are used and an additional level created by islands. See text for details.

Figures 4–# Maps showing species-specific relative habitat suitability for each grid cell as color-coded probability between 0 (yellow) and 1(red) for a species to occur in one cell. IUCN native (black) and introduced (blue) area maps are shown.

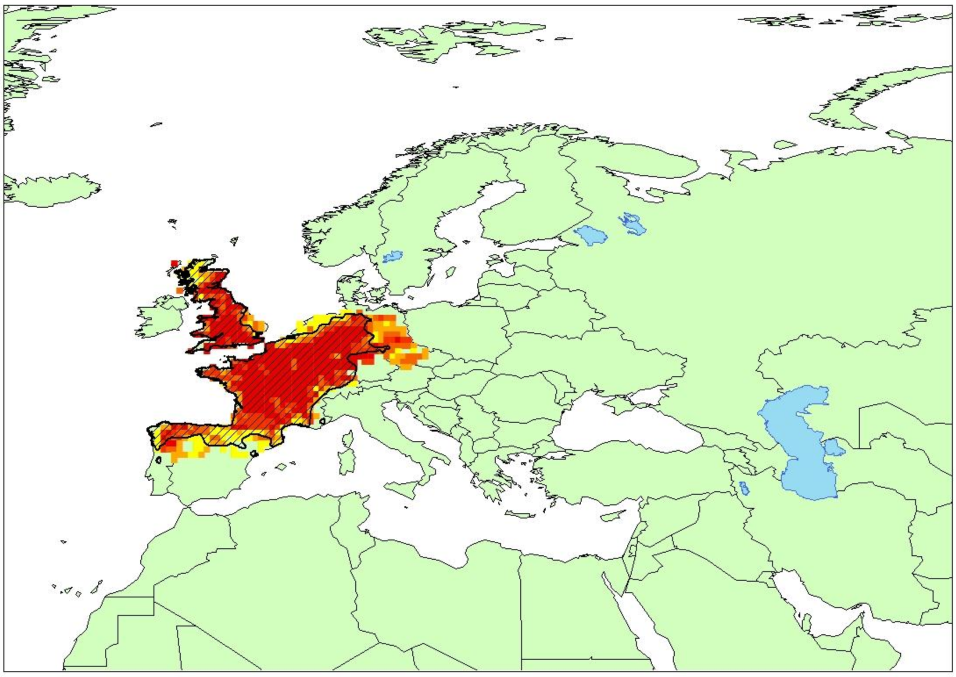
Lissotriton helveticus, suitable habitat, original (8 parameter)



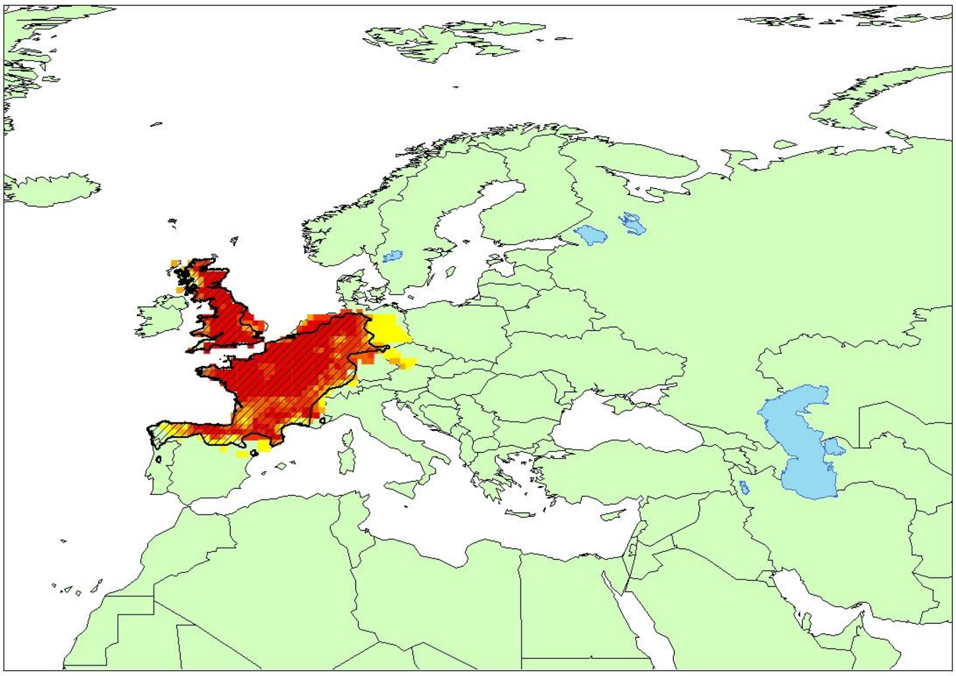
Lissotriton helveticus, suitable habitat, VIFM (8 parameter)



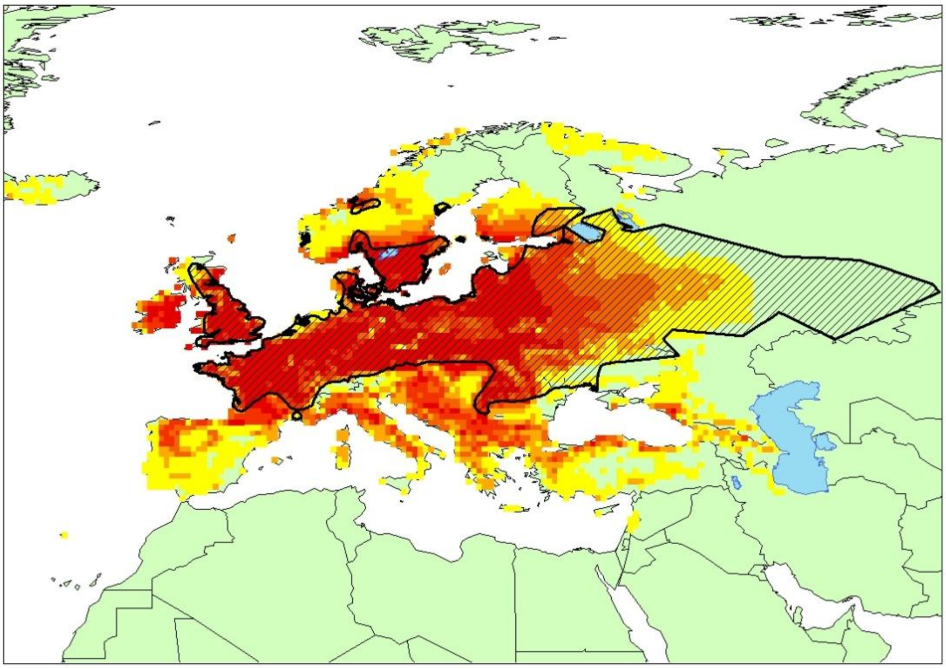
Lissotriton helveticus, native, original (8 parameter)



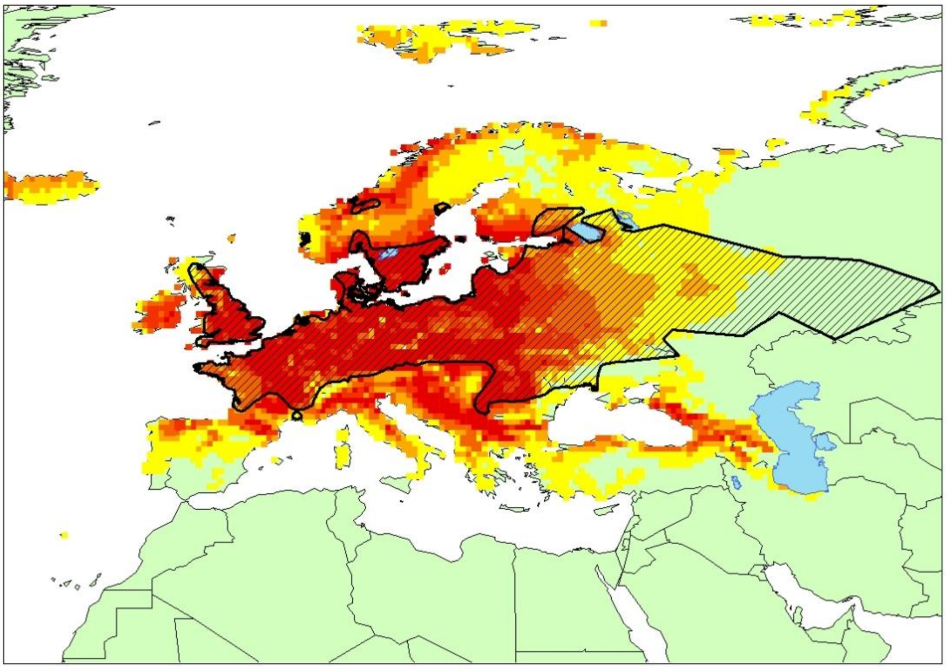
Lissotriton helveticus, native, VIFM (8 parameter)



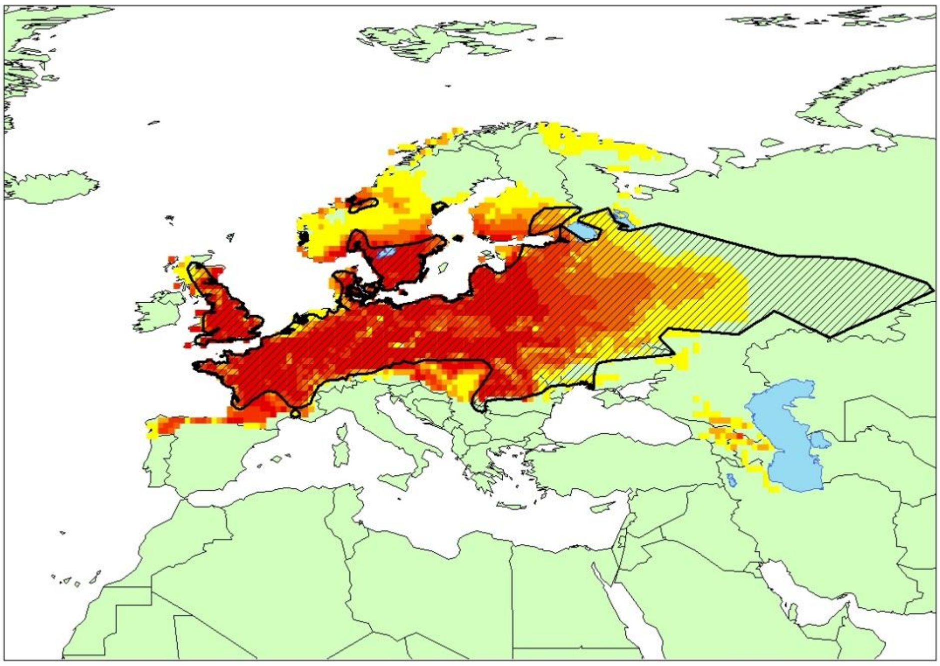
Triturus cristatus, suitable habitat, original (8 parameter)



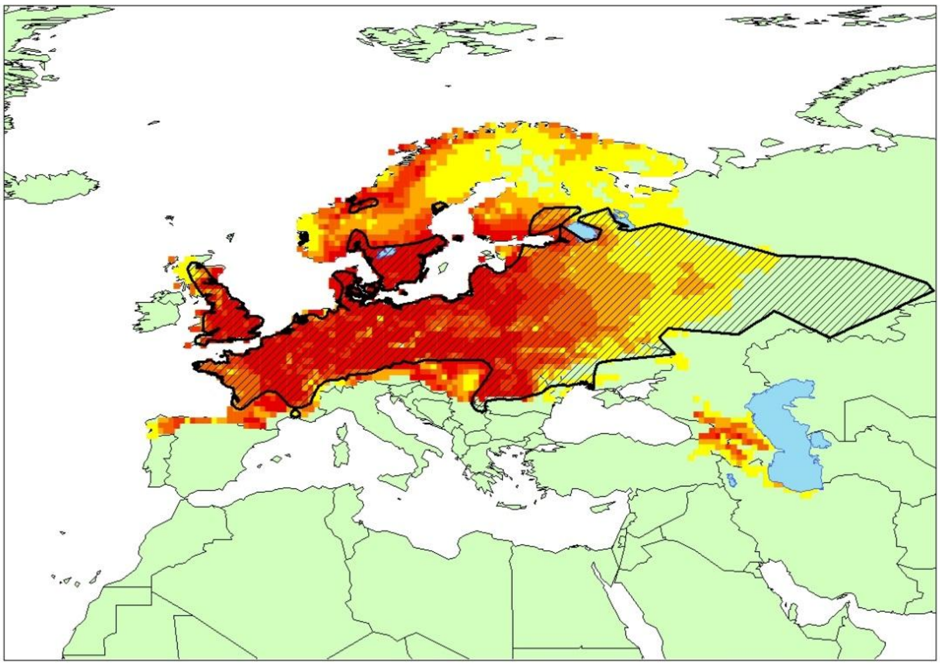
Triturus cristatus, suitable habitat, VIFM (8 parameter)



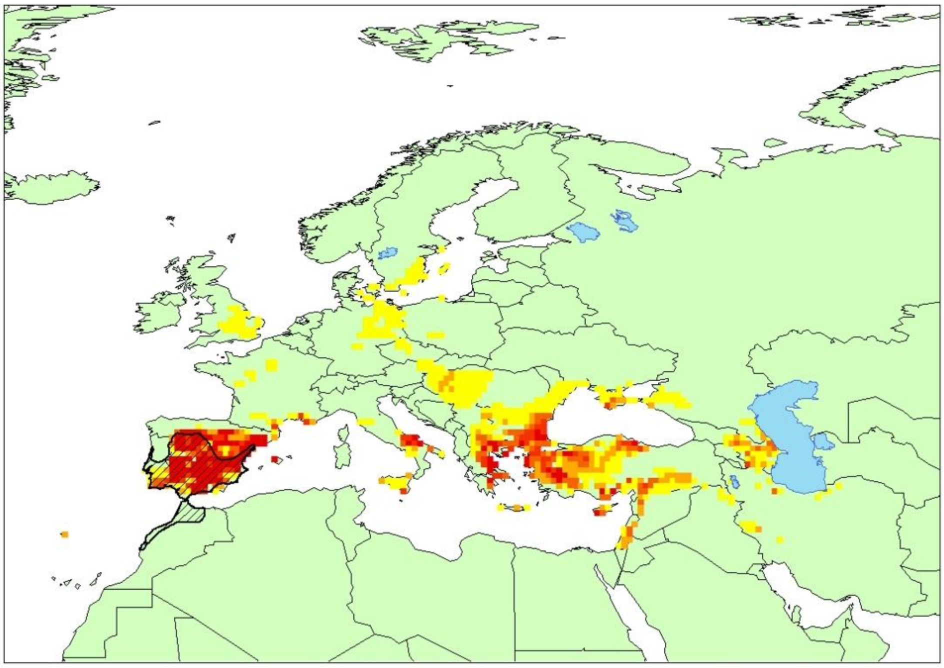
Triturus cristatus, native, original (8 parameter)



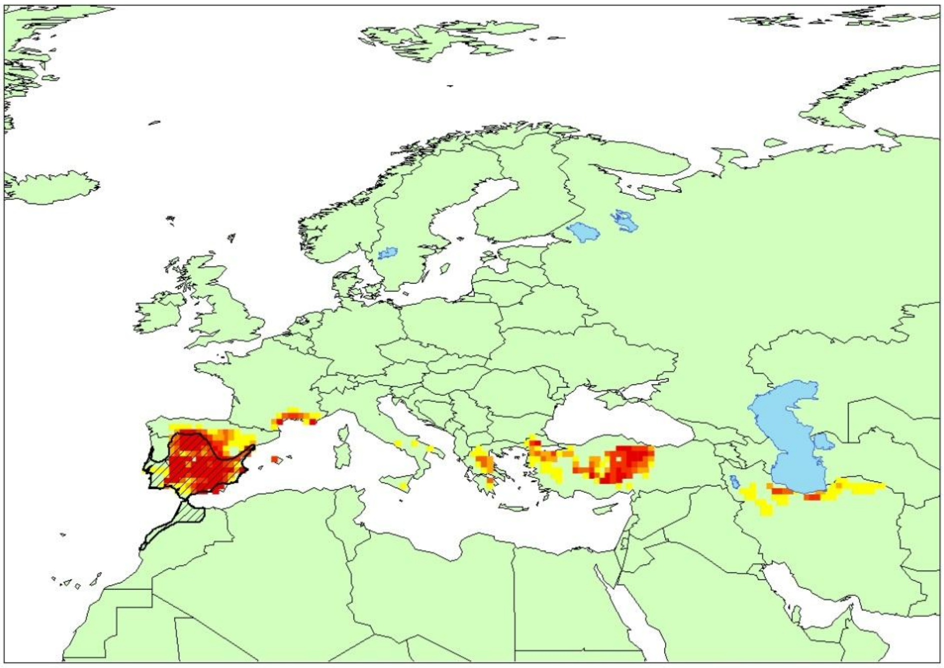
Triturus cristatus, native, VIFM (8 parameter)



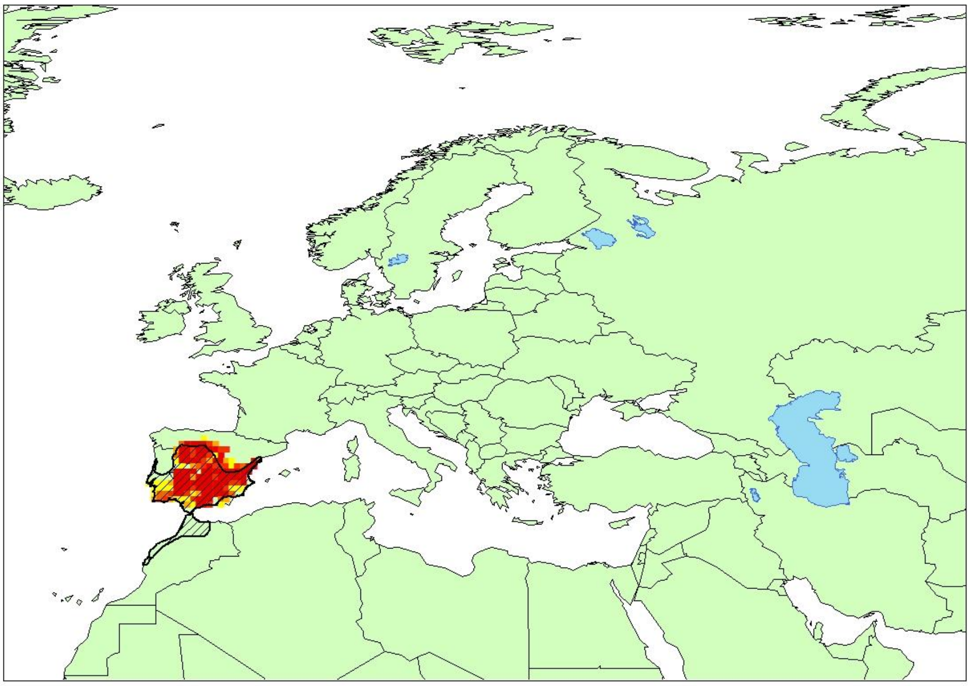
Pleurodeles waltl, suitable habitat, original (8 parameter)



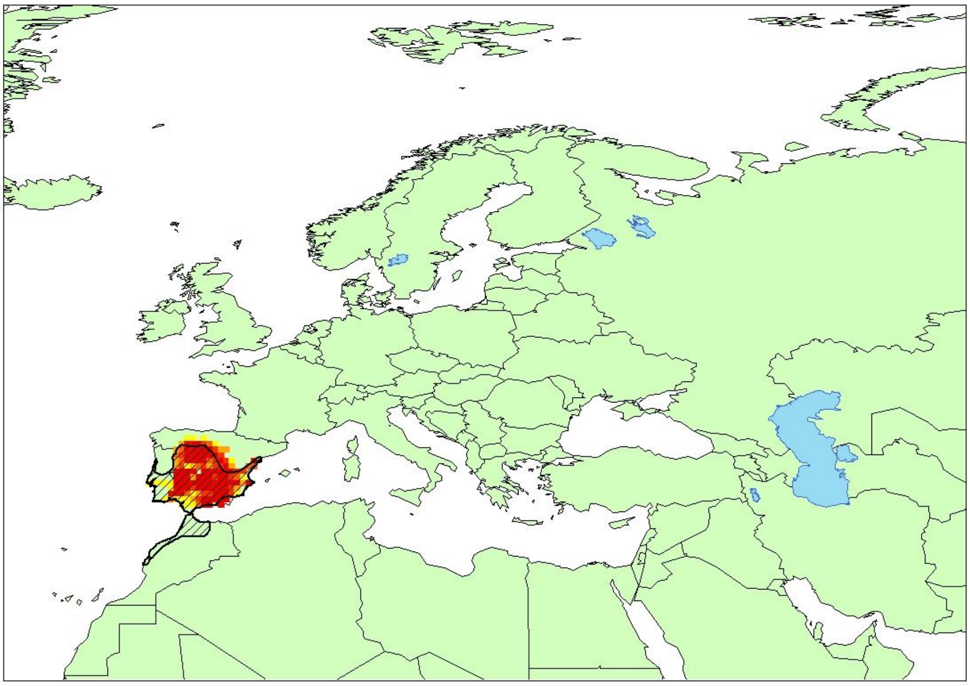
Pleurodeles waltl, suitable habitat, VIFM (8 parameter)



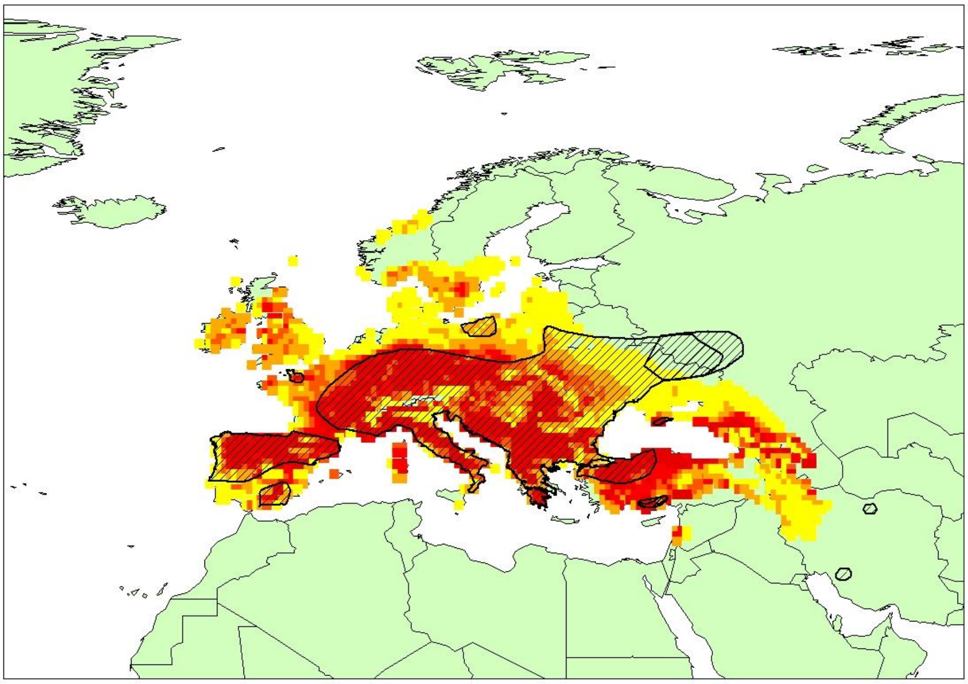
Pleurodeles waltl, native, original (8 parameter)



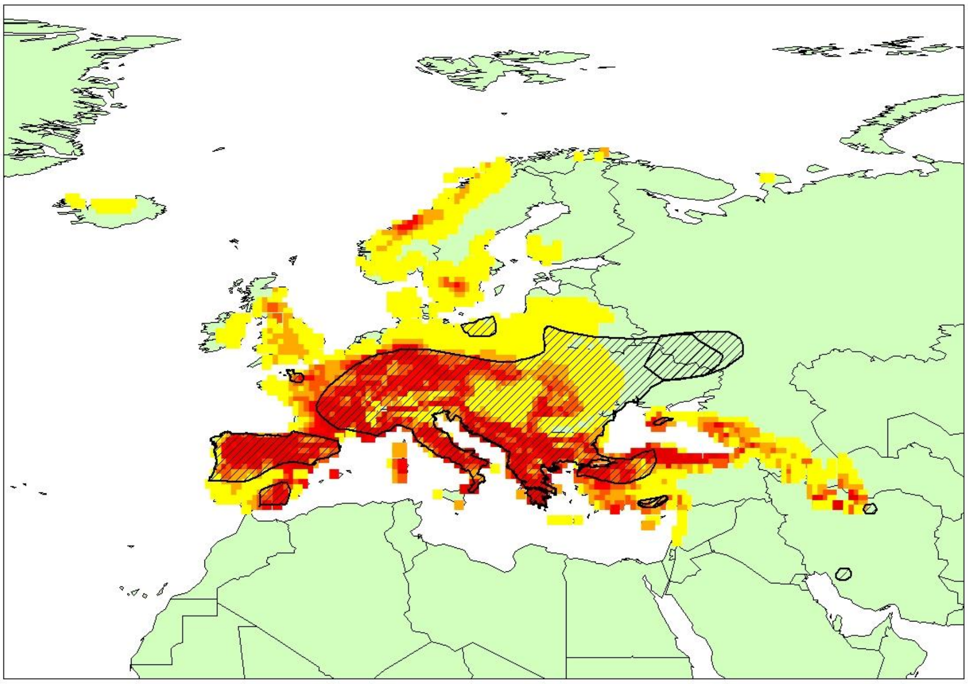
Pleurodeles waltl, native, VIF2 (8 parameter)



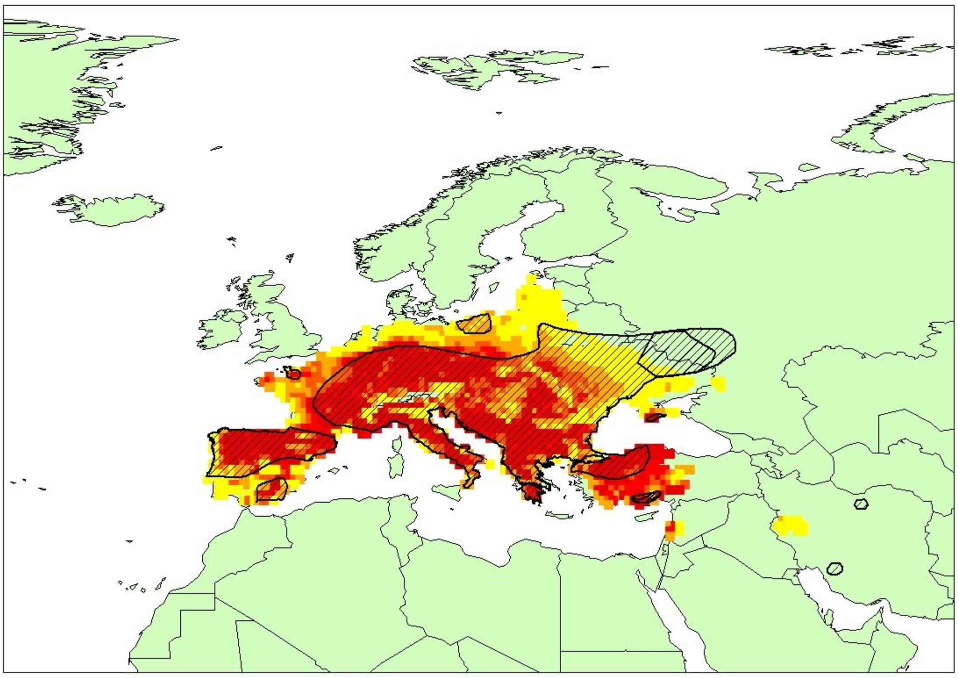
Neomys anomalus, suitable habitat, original (8 parameter)



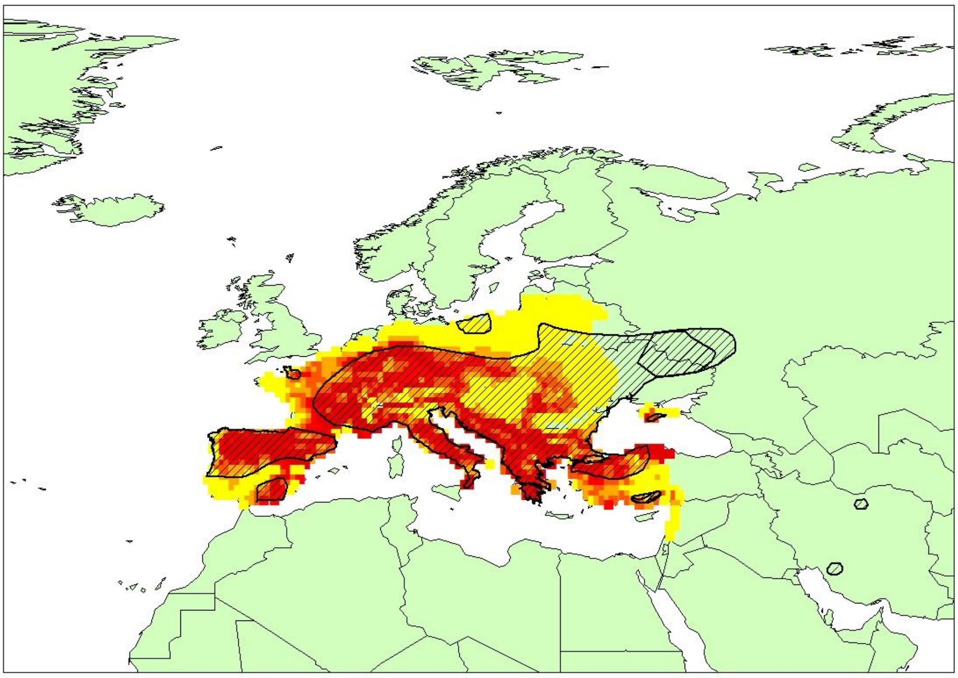
Neomys anomalus, suitable habitat, VIFM (8 parameter)



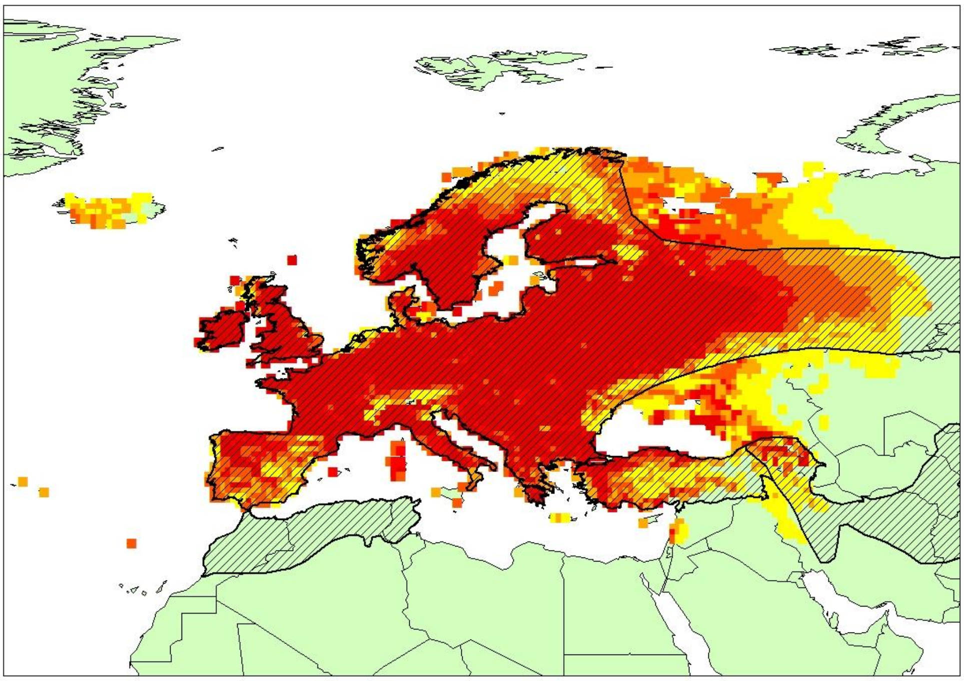
Neomys anomalus, native, original (8 parameter)



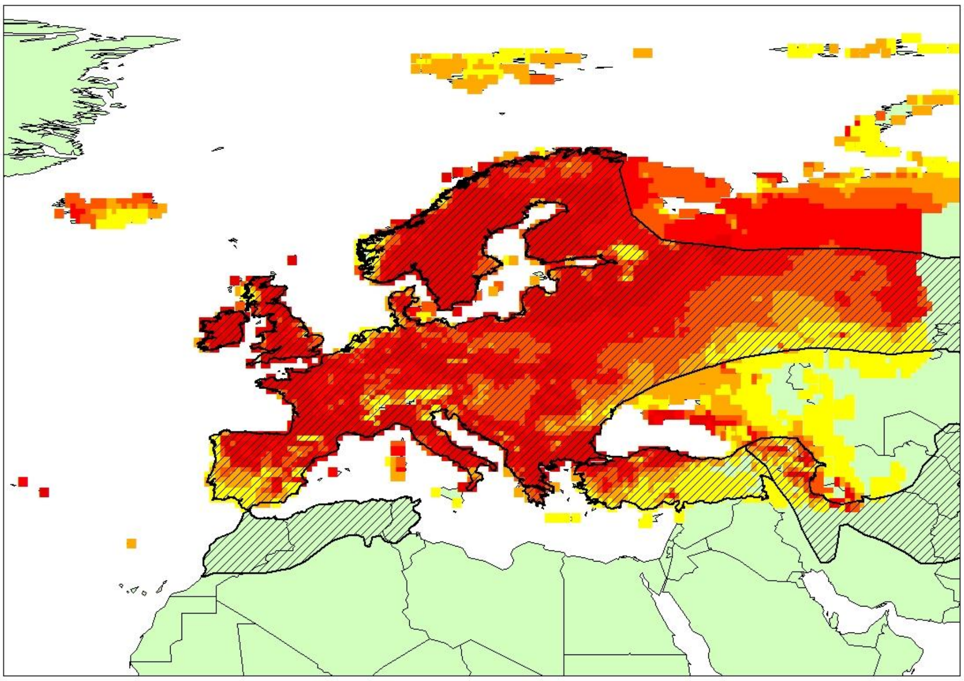
Neomys anomalus, native, VIFM (8 parameter)



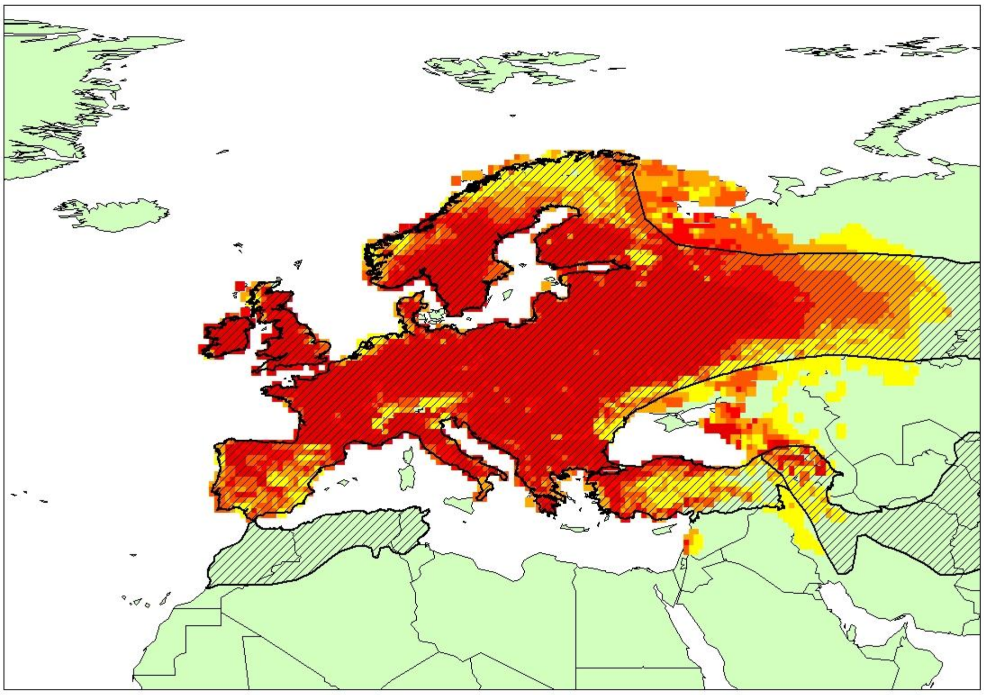
Lutra lutra, suitable habitat, original (8 parameter)



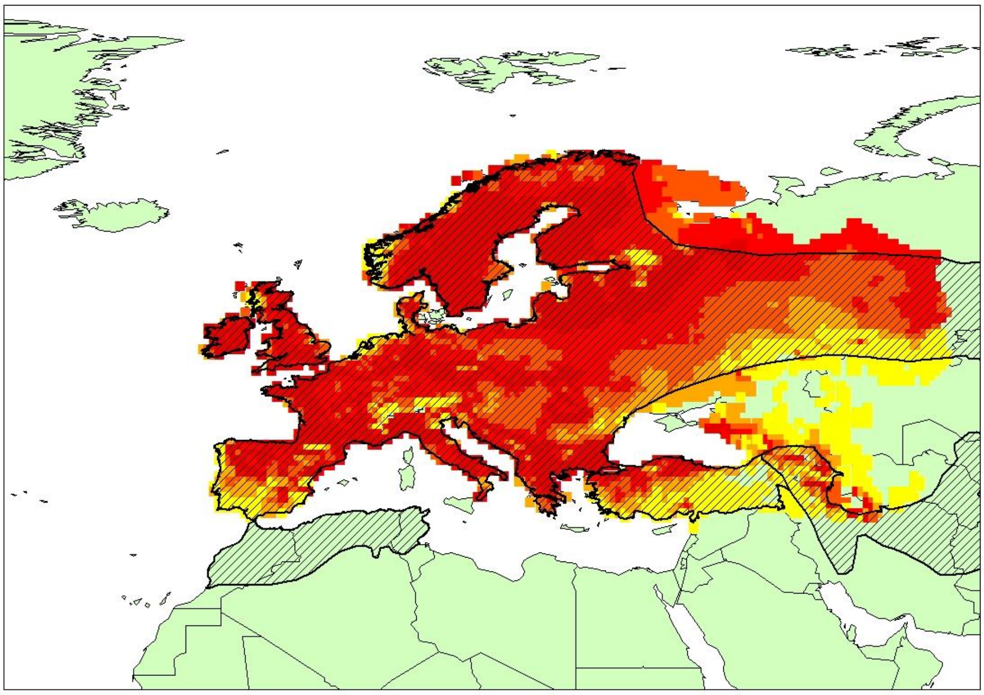
Lutra lutra, suitable habitat, VIFM (8 parameter)



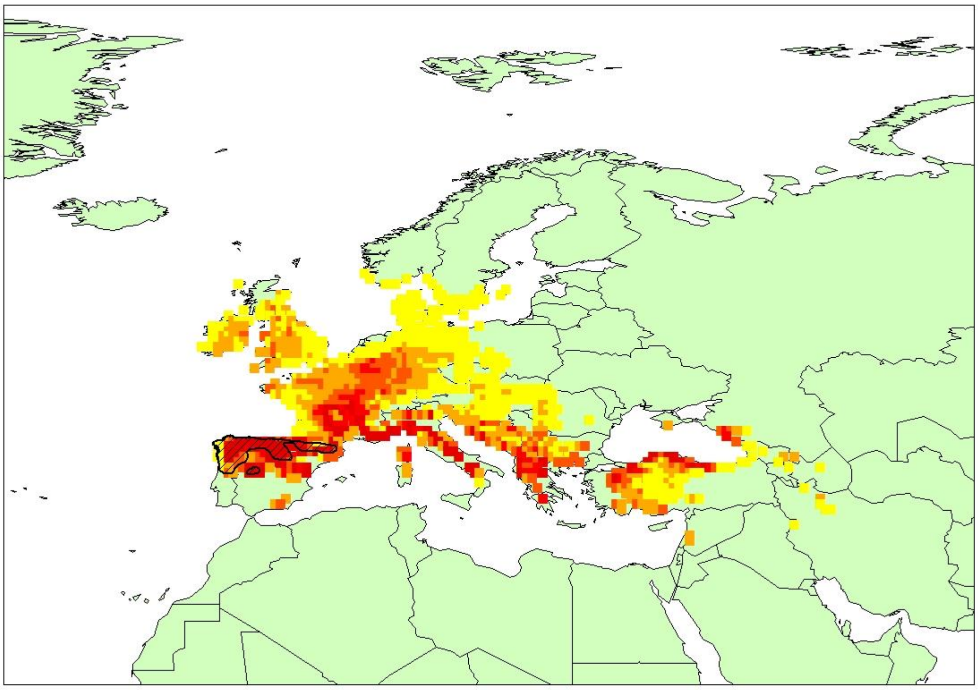
Lutra lutra, native, original (8 parameter)



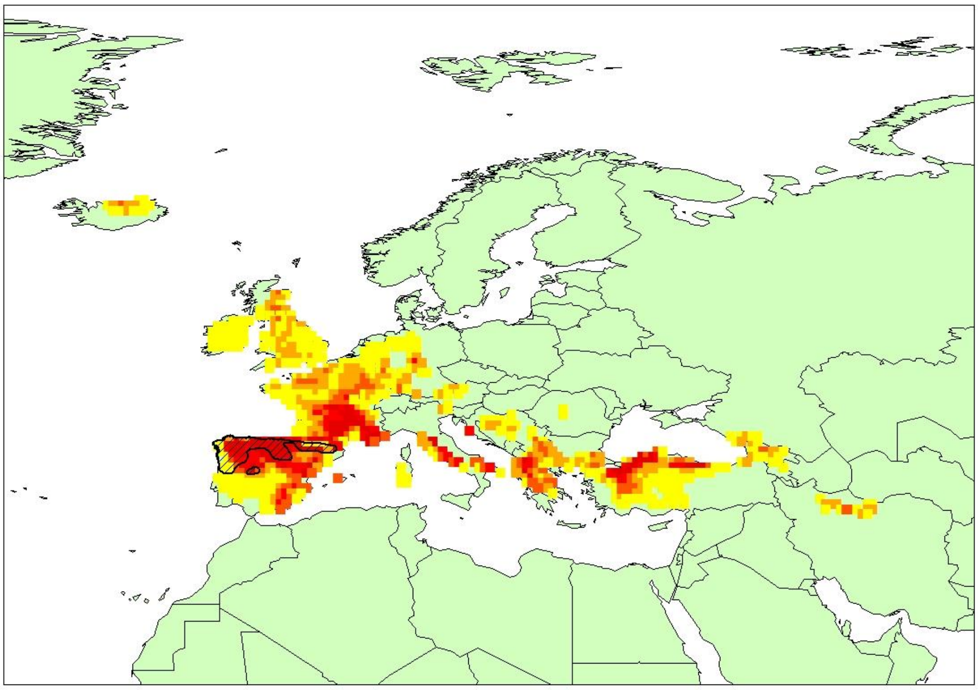
Lutra lutra, native, VIFM (8 parameter)



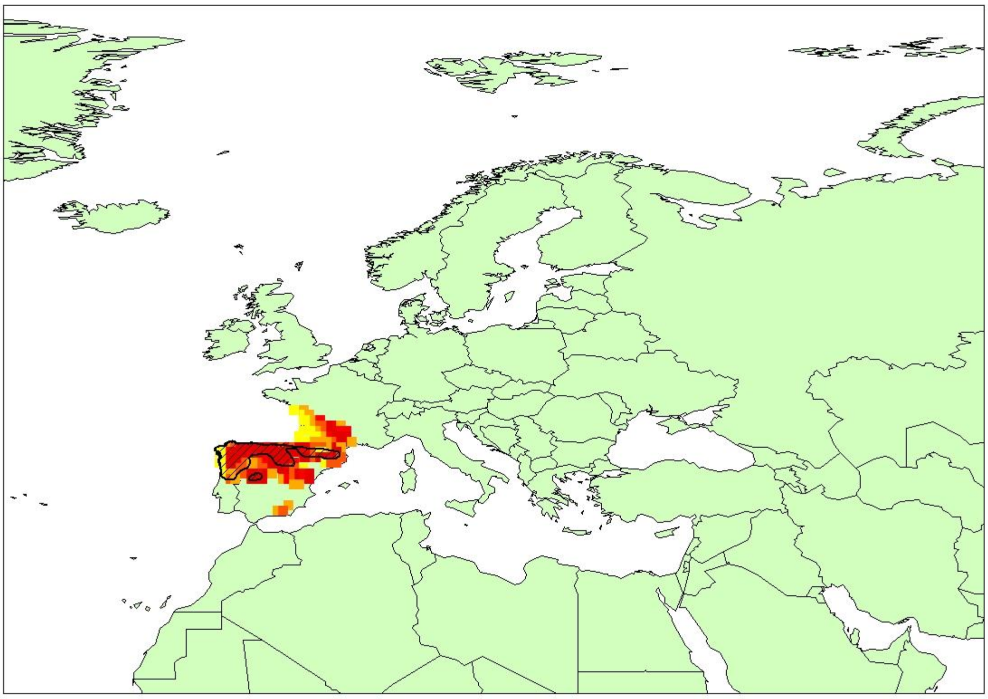
Galemys pyreneaicus, suitable habitat, original (8 parameter)



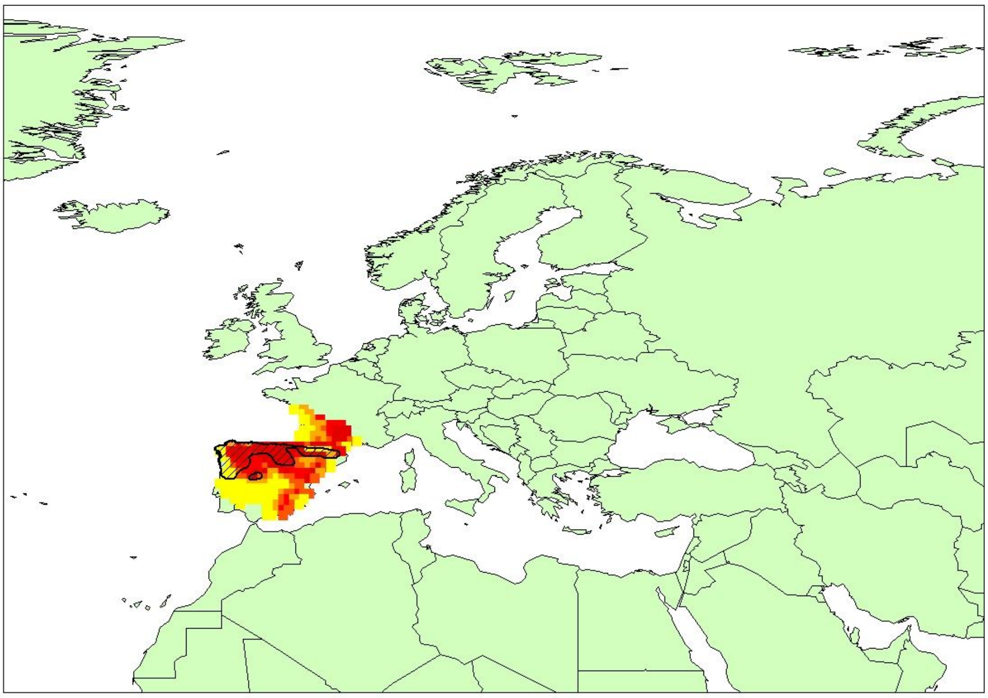
Galemys pyreneaicus, suitable habitat, VIFM (8 parameter)



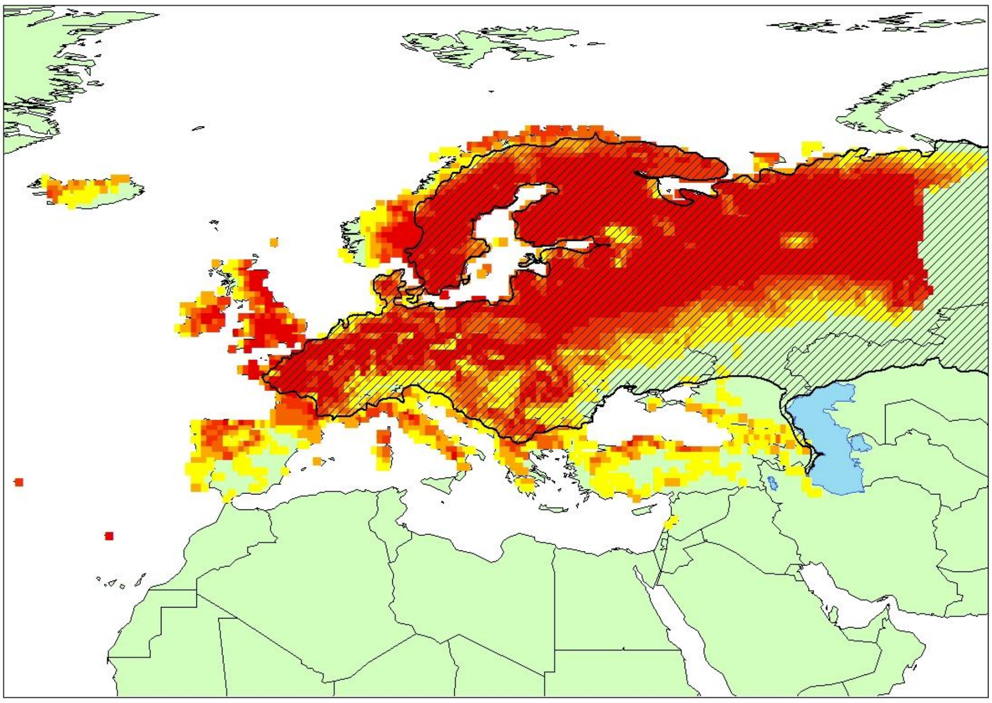
Galemys pyreneaicus, native, original (8 parameter)



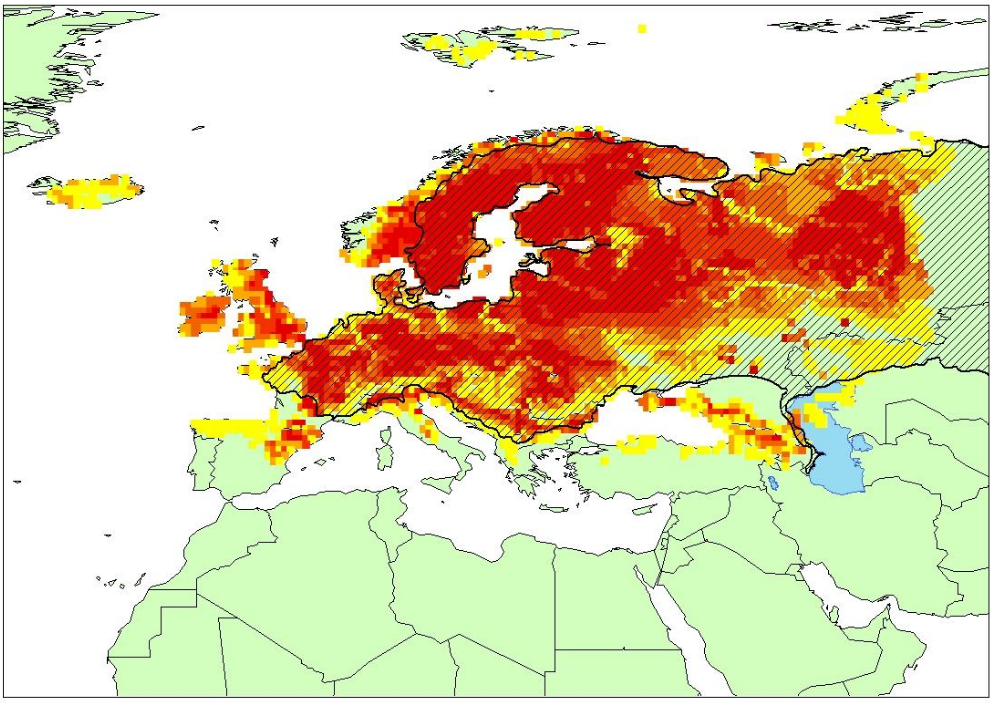
Galemys pyreneaicus, native, VIFM (8 parameter)



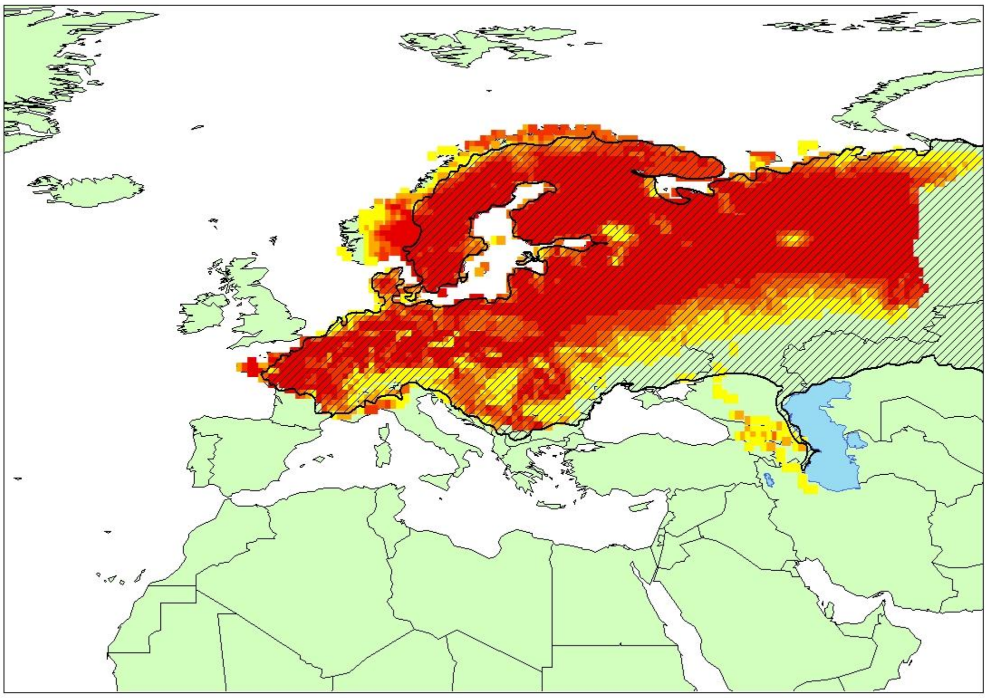
Lota lota, suitable habitat, original (6 parameter)



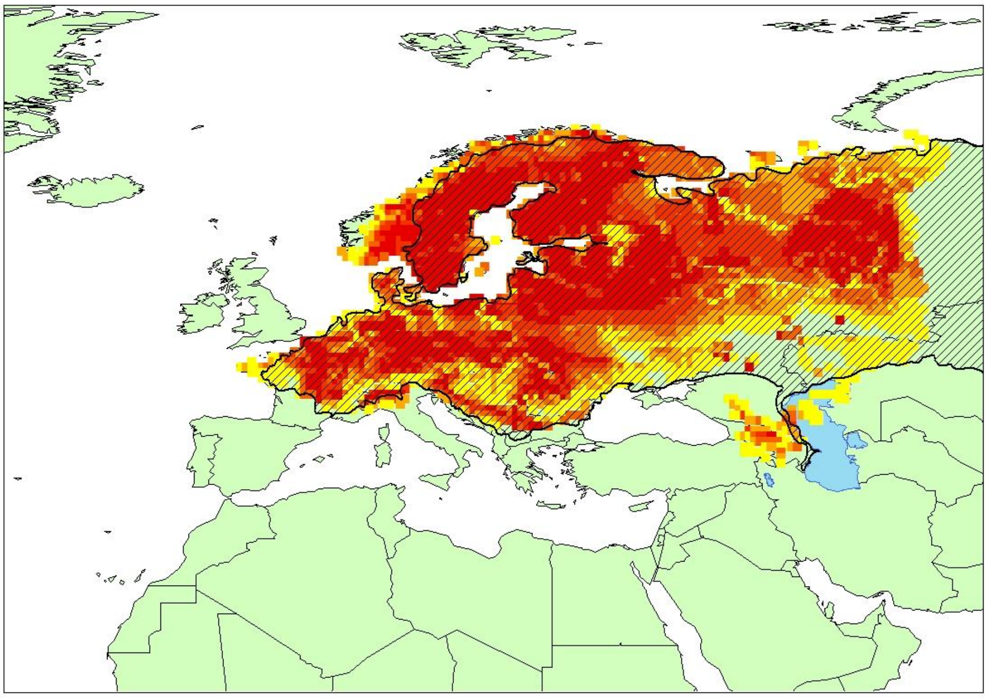
Lota lota, suitable habitat, VIFM (6 parameter)



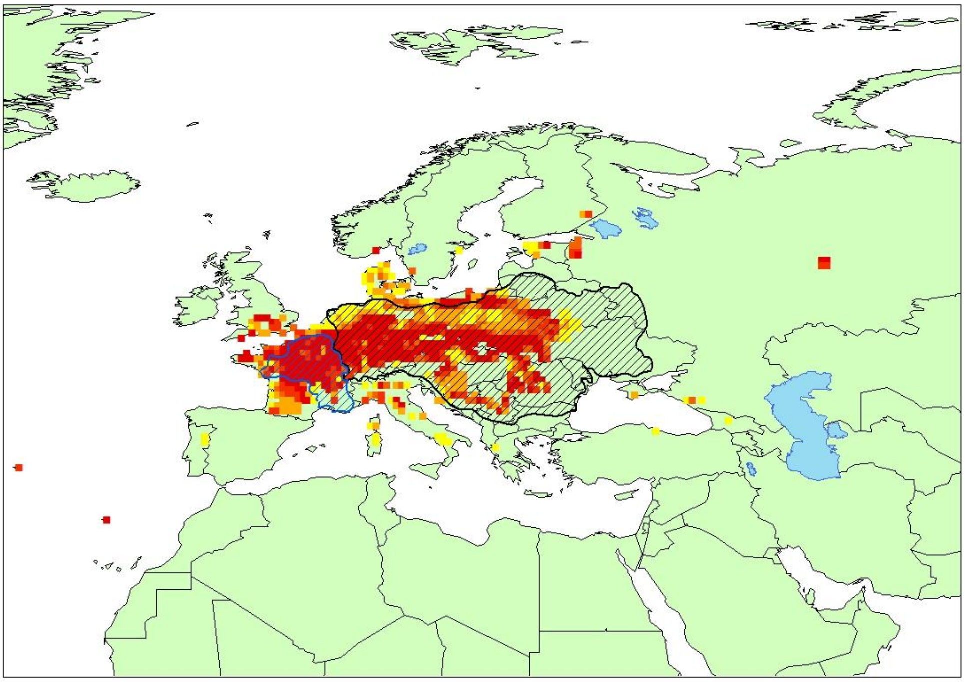
Lota lota, native, original (6 parameter)



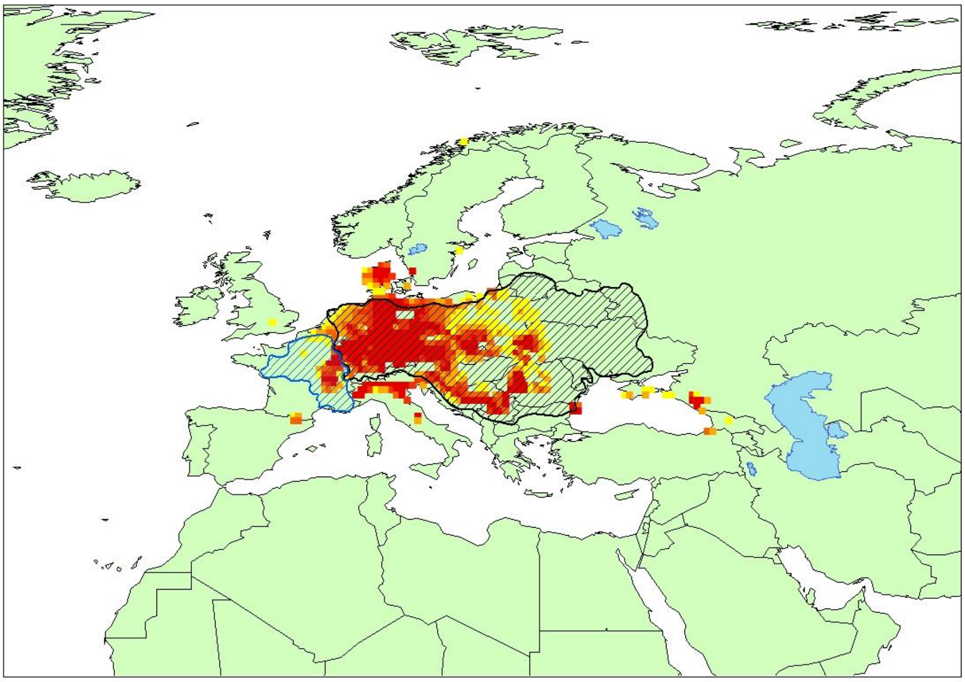
Lota lota, native, VIFM (6 parameter)



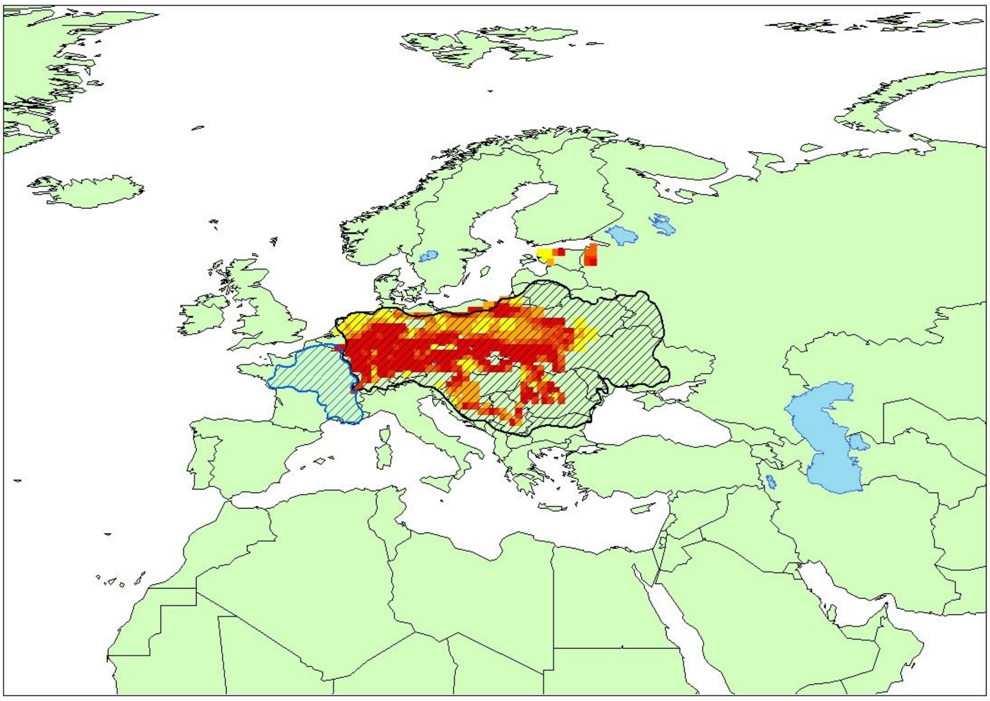
Chondrostoma nasus, suitable habitat, original (6 parameter)



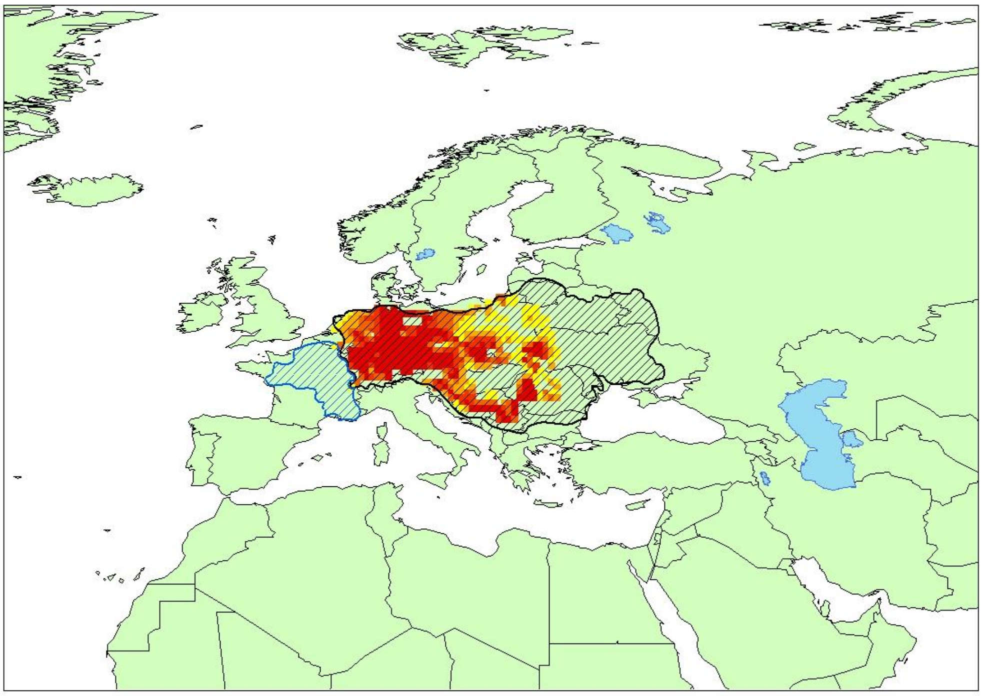
Chondrostoma nasus, suitable habitat, VIFM (6 parameter)



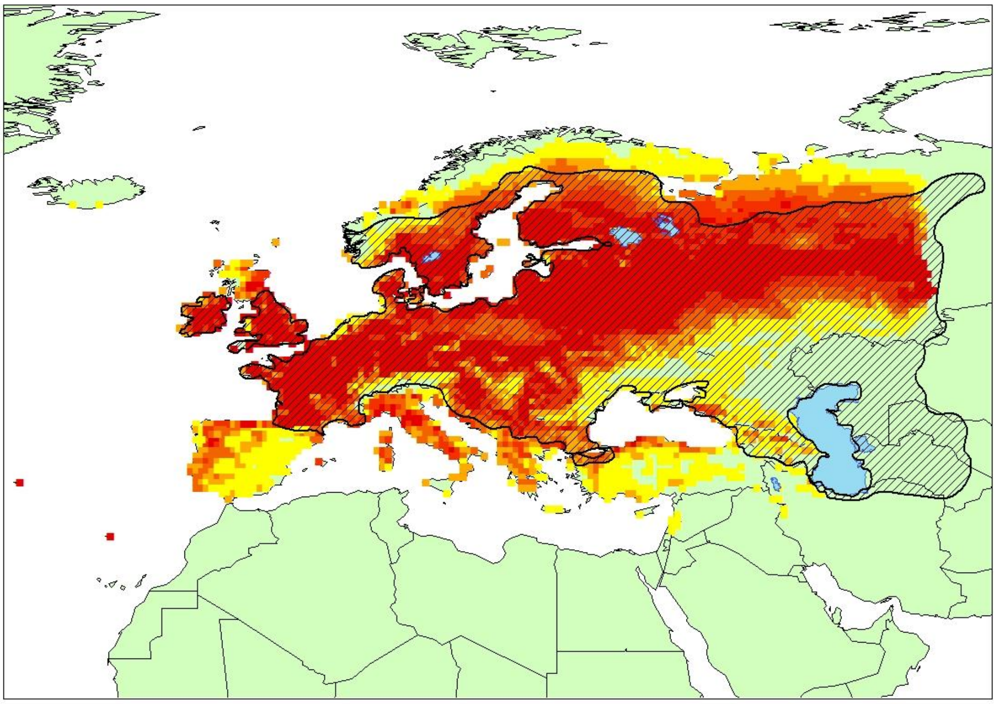
Chondrostoma nasus, native, original (6 parameter)



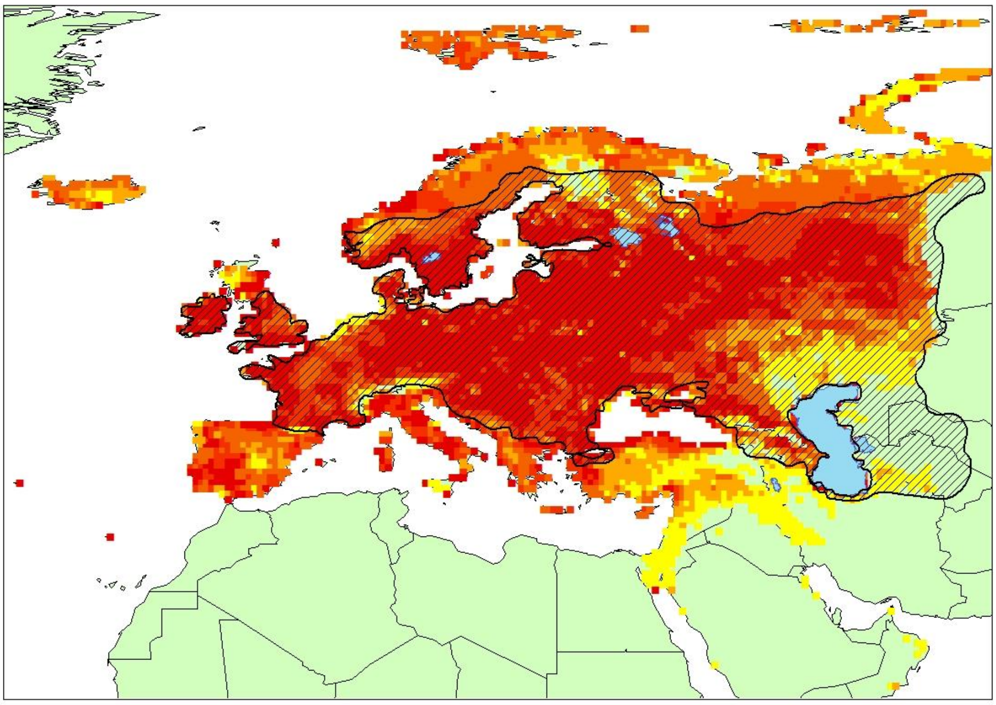
Chondrostoma nasus, native, VIFM (6 parameter)



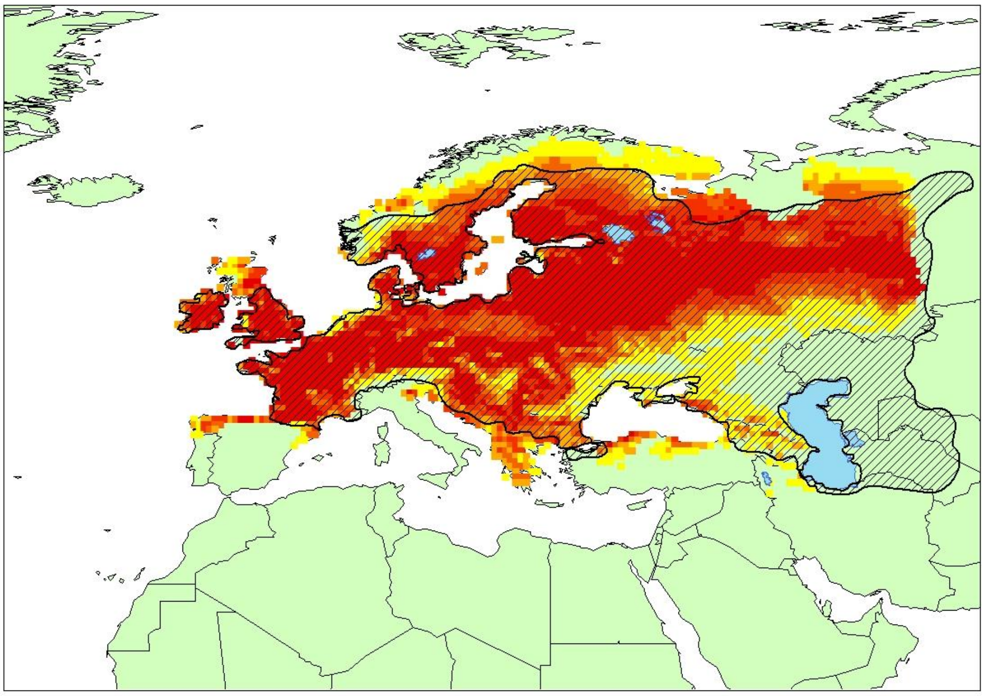
Abramis brama, suitable habitat, original (6 parameter)



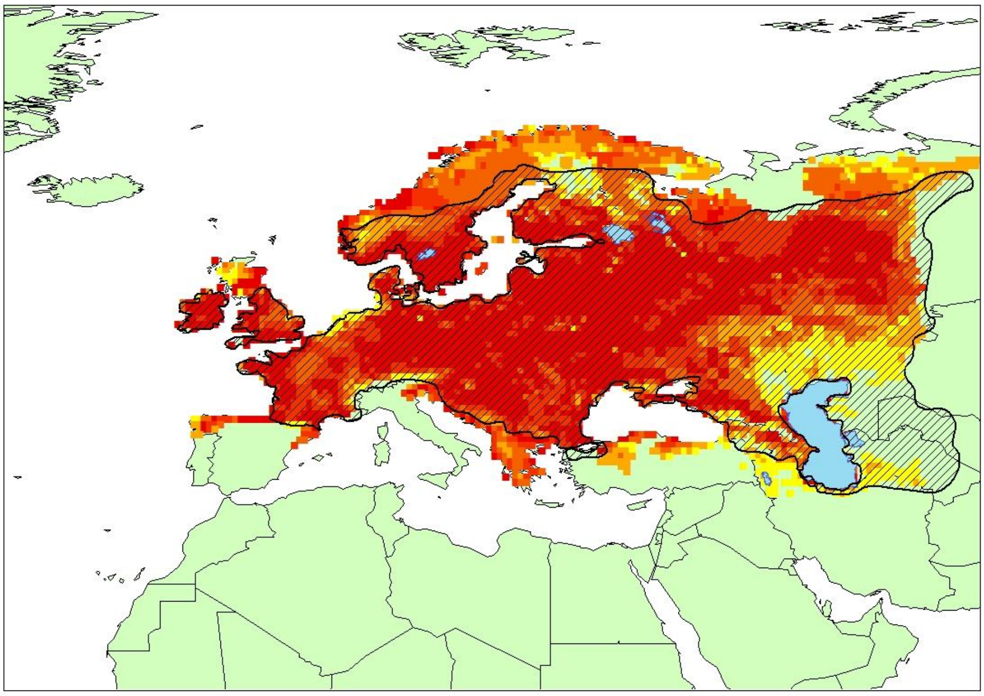
Abramis brama, suitable habitat, VIFM (6 parameter)



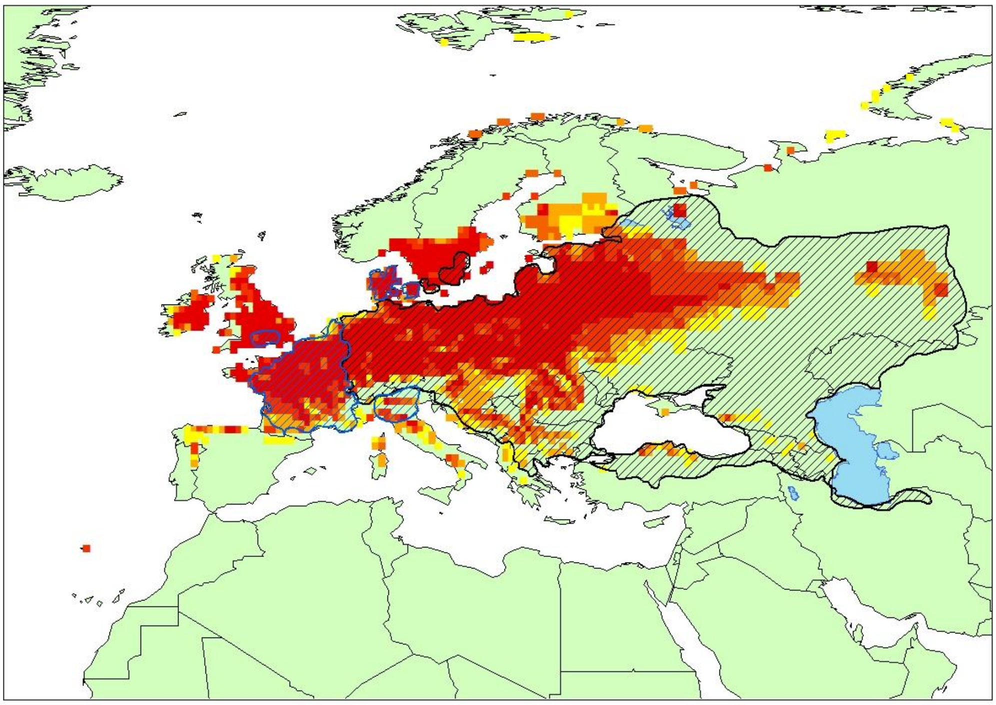
Abramis brama, native, original (6 parameter)



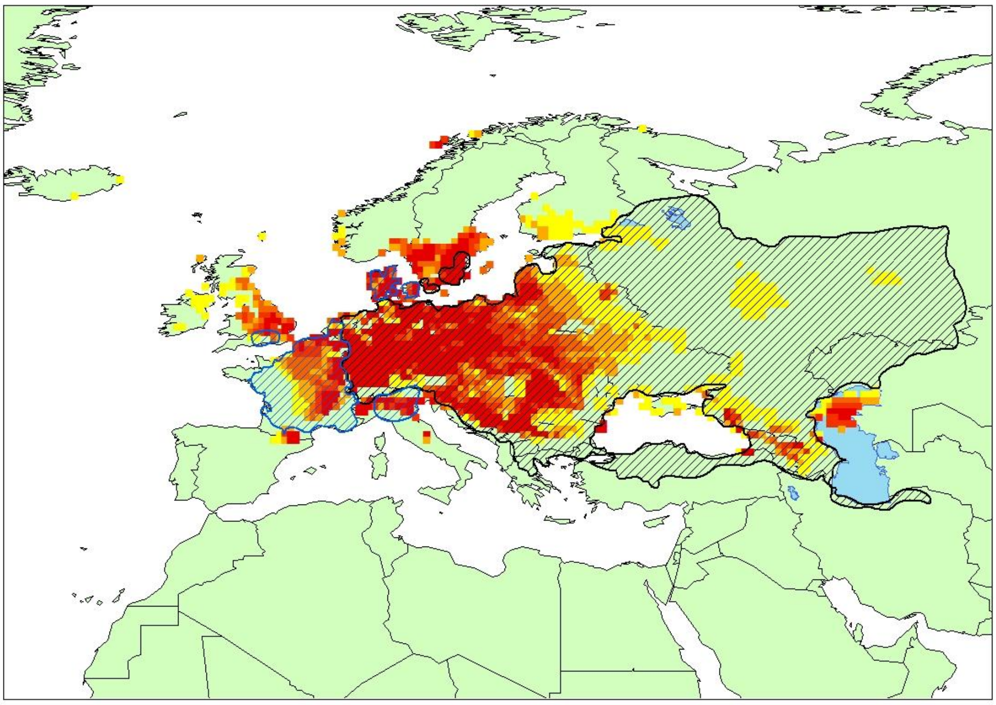
Abramis brama, native, VIFM (6 parameter)



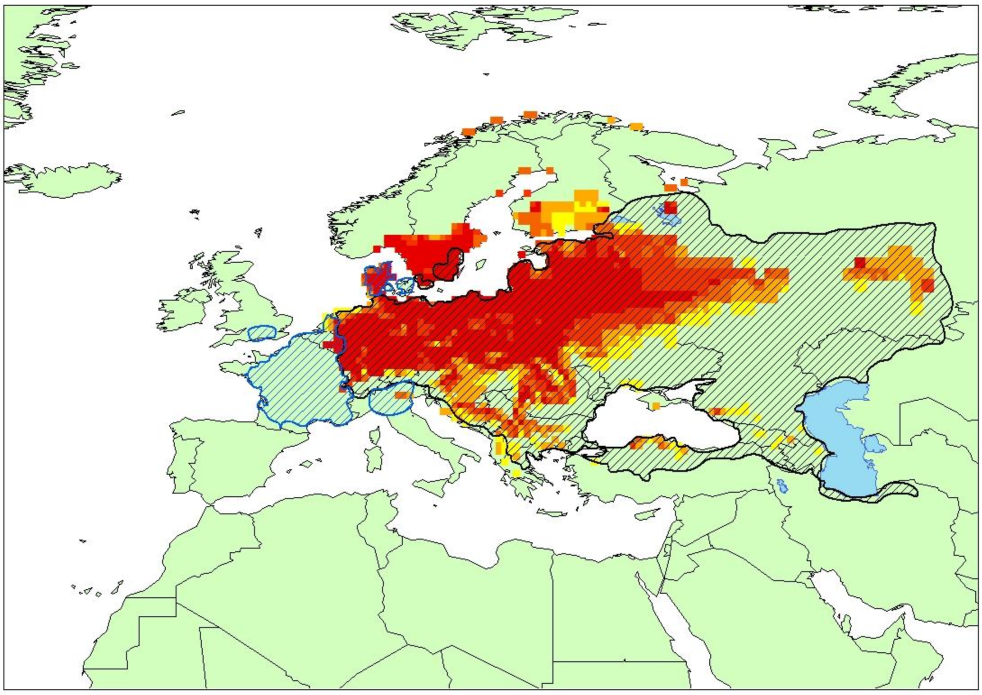
Silurus glanis, suitable habitat, original (6 parameter)



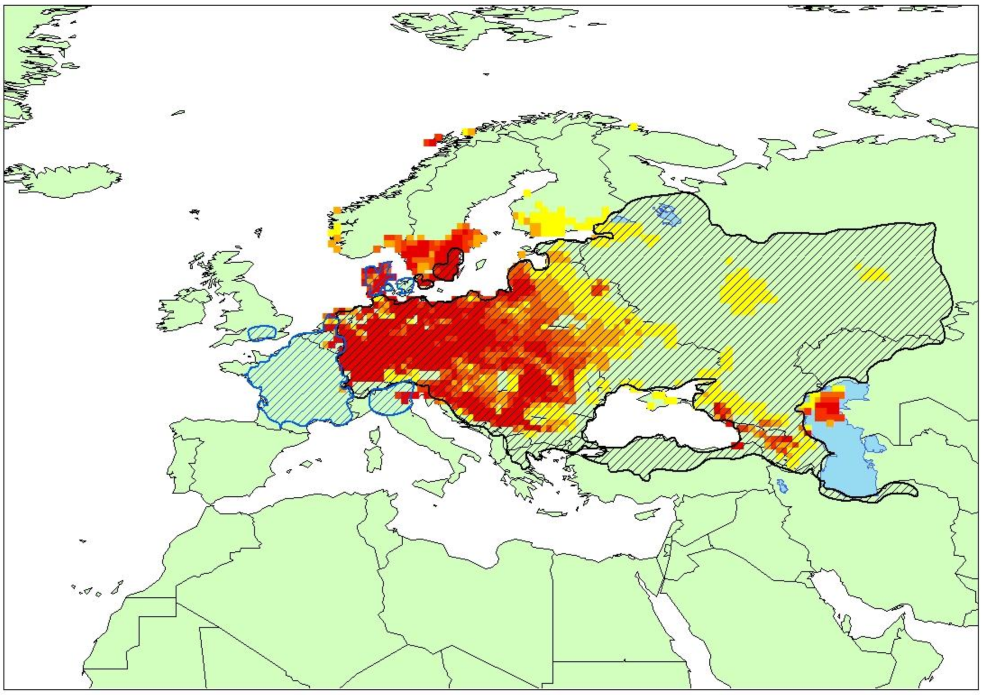
Silurus glanis, suitable habitat, VIFM (6 parameter)



Silurus glanis, native, original (6 parameter)

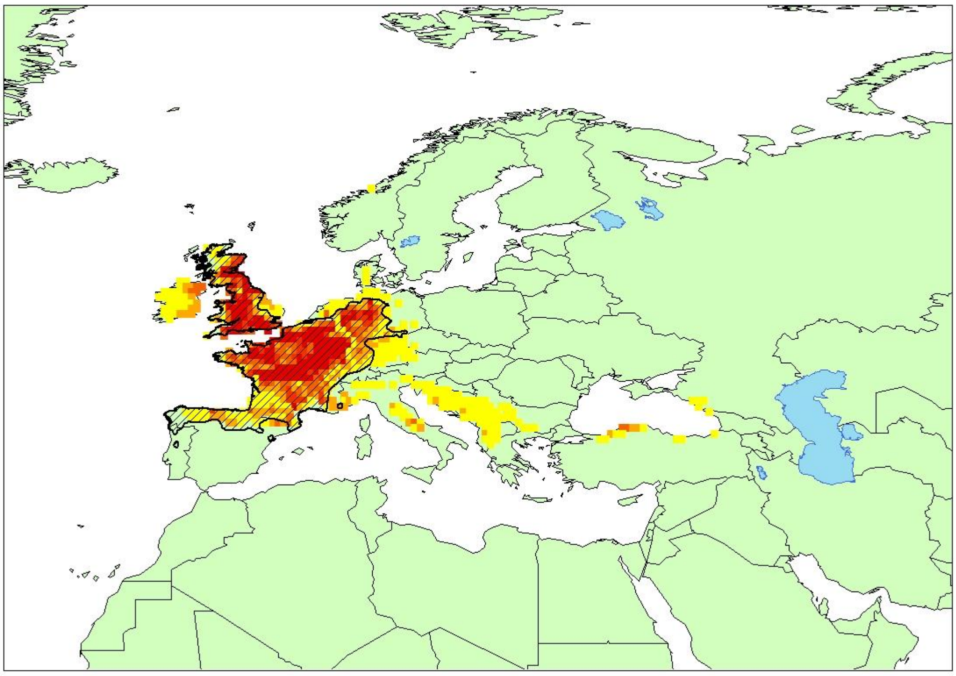


Silurus glanis, native, VIFM (6 parameter)

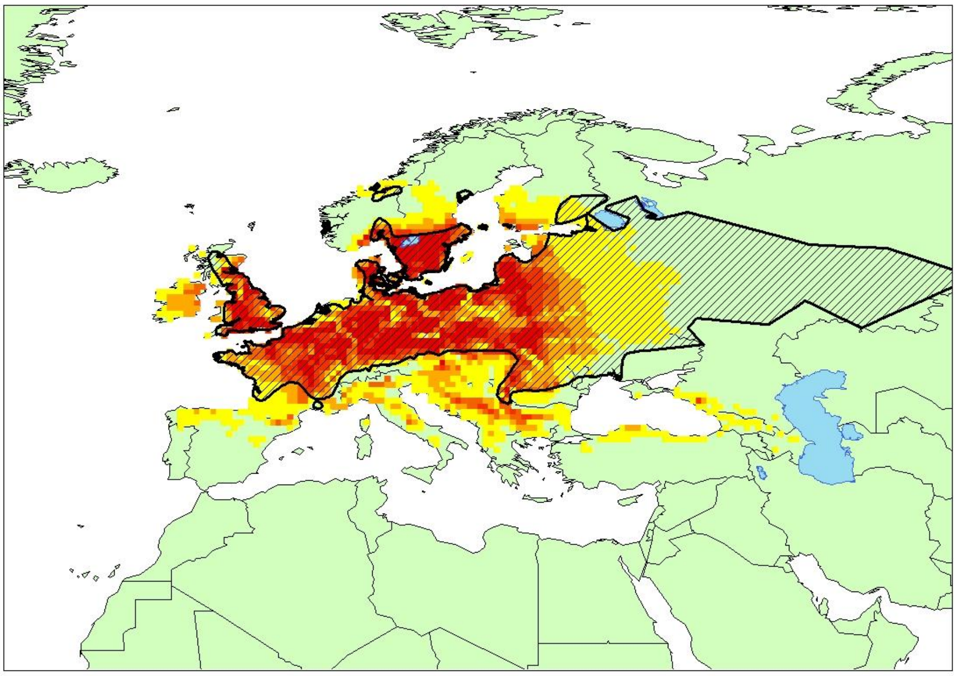


Figures ALL27 parameter maps

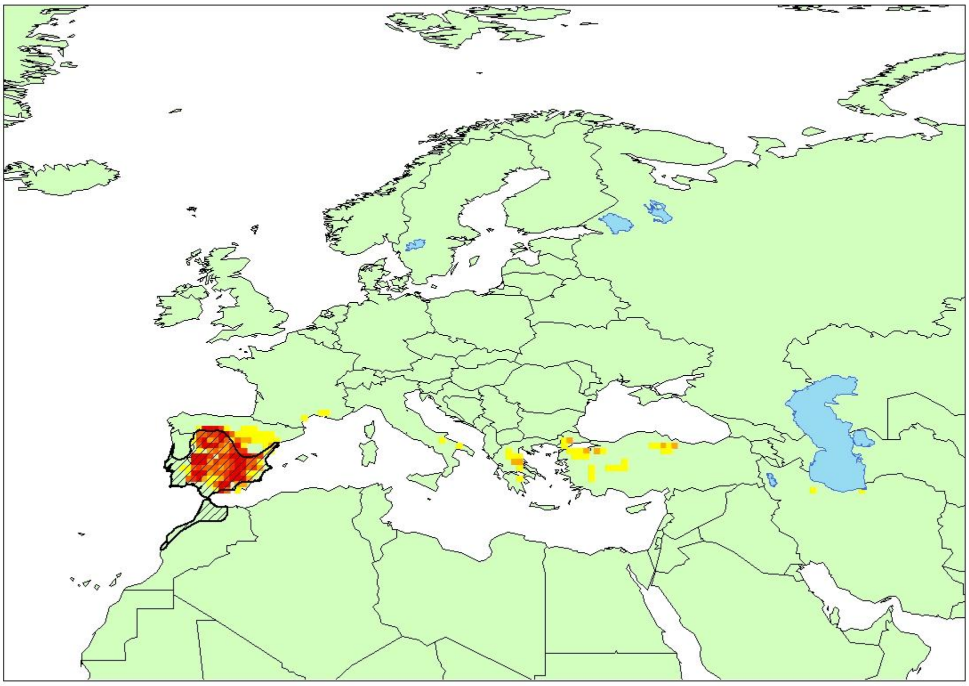
Lissotriton helveticus



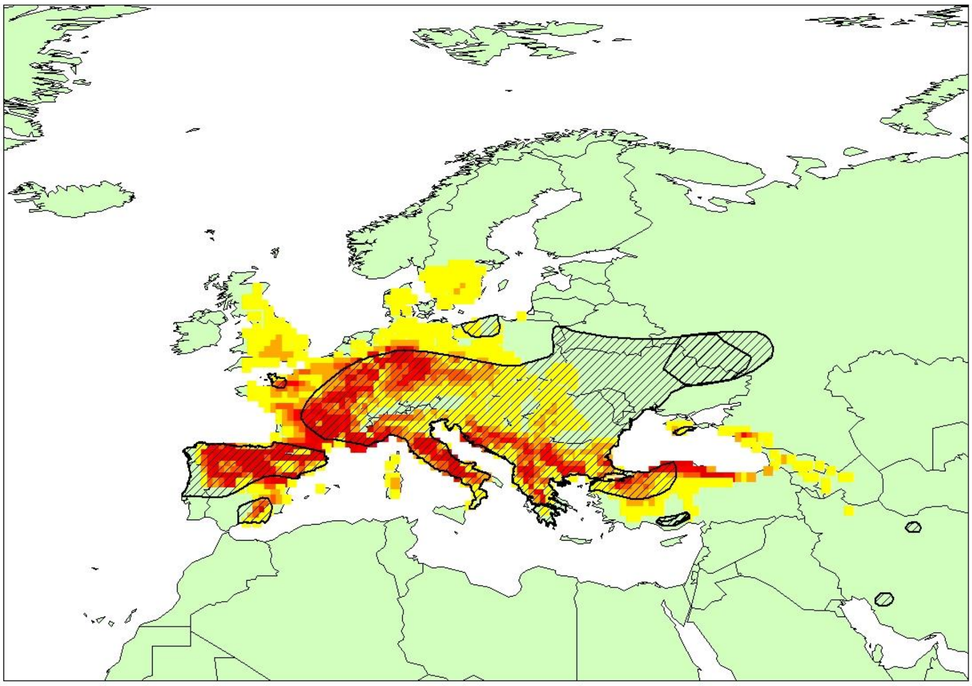
Triturus cristatus



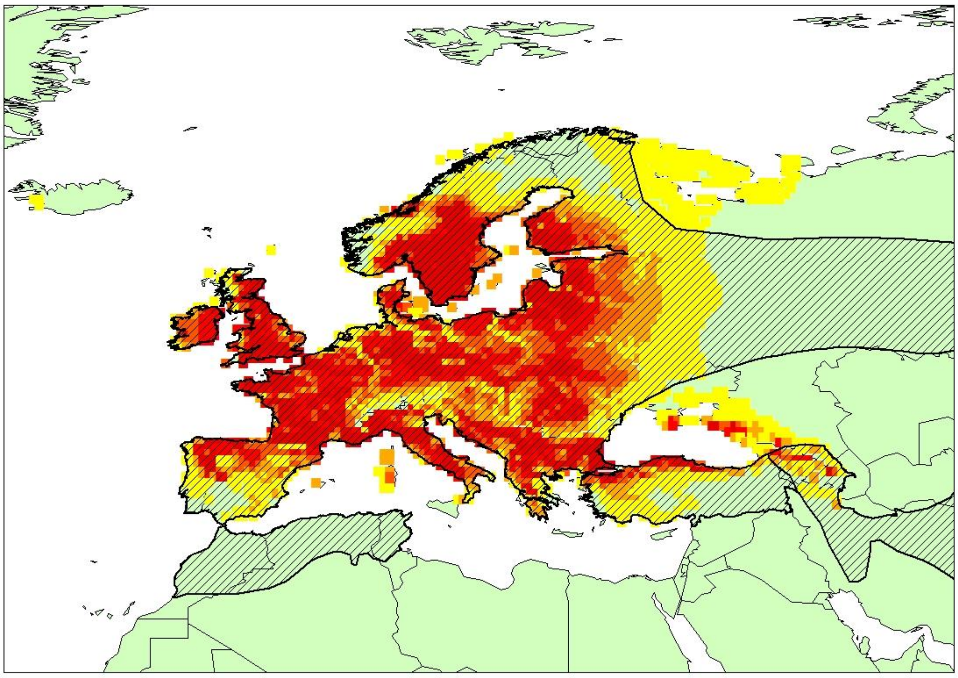
Pleurodeles waltl



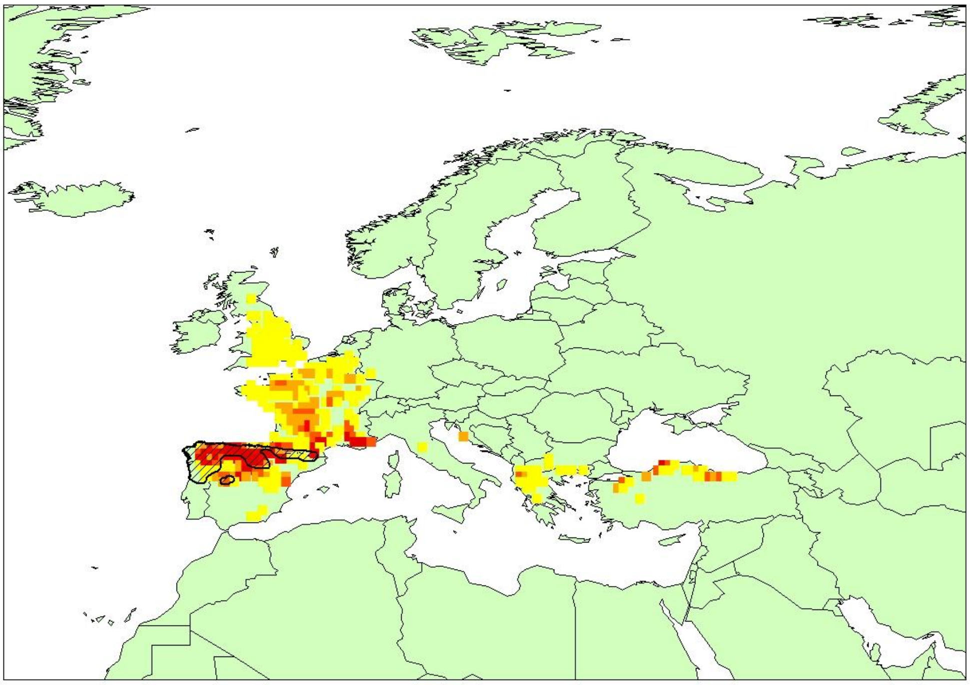
Neomys anomalus



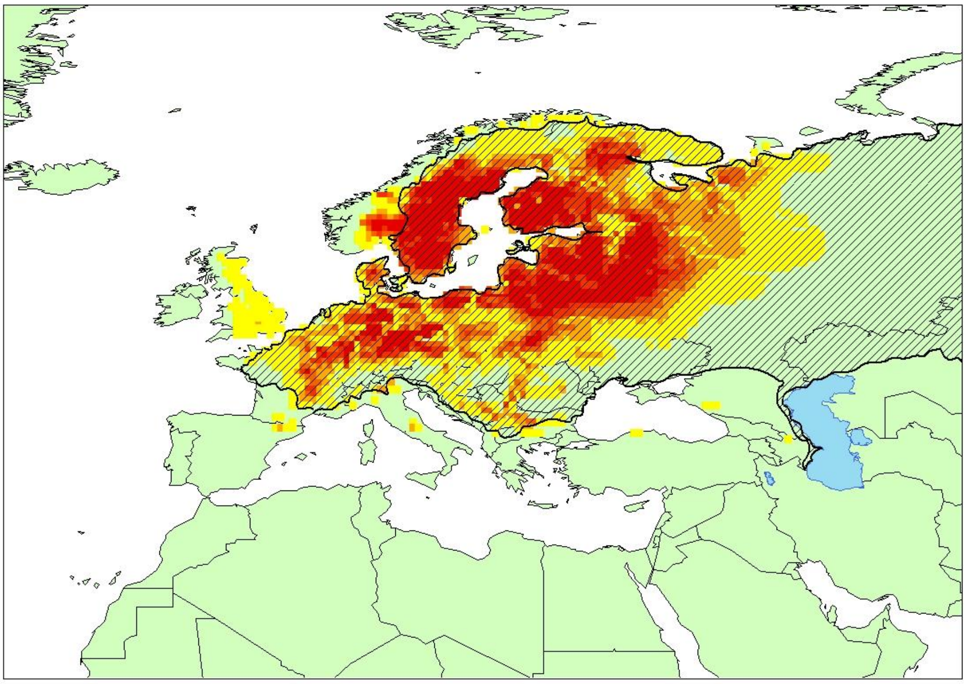
Lutra lutra



Galemys pyreneaicus



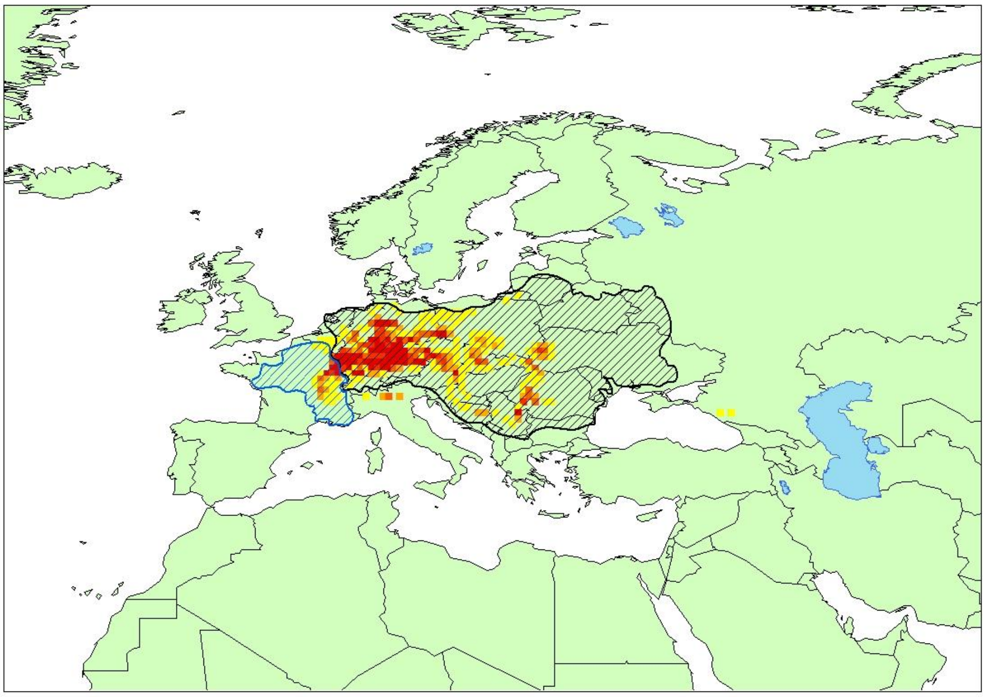
153 Lota lota



154

155

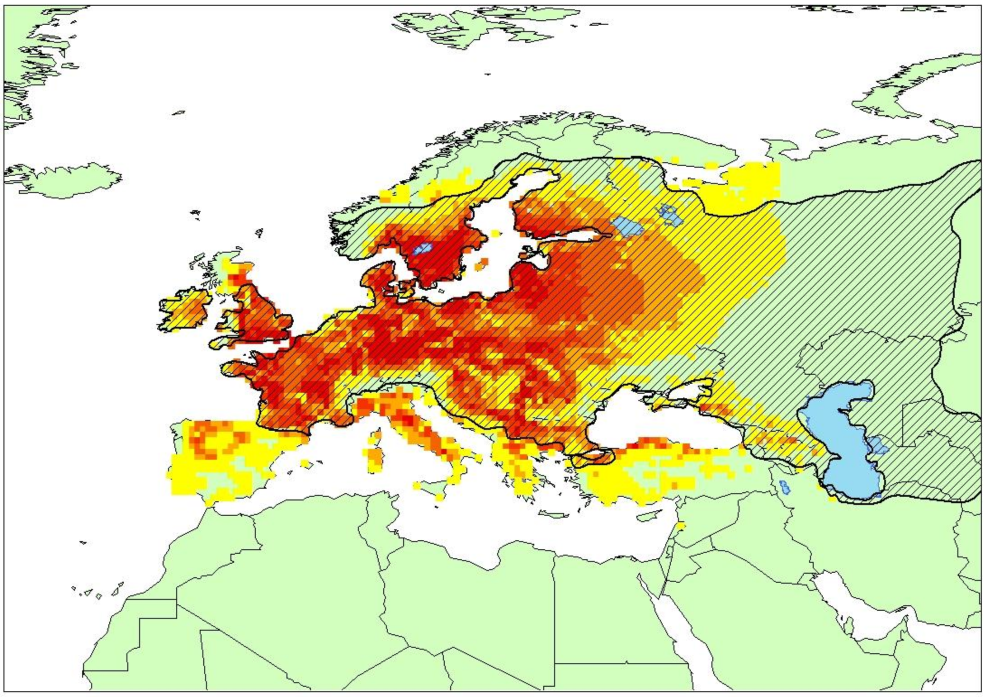
156 Chondrostoma nasus



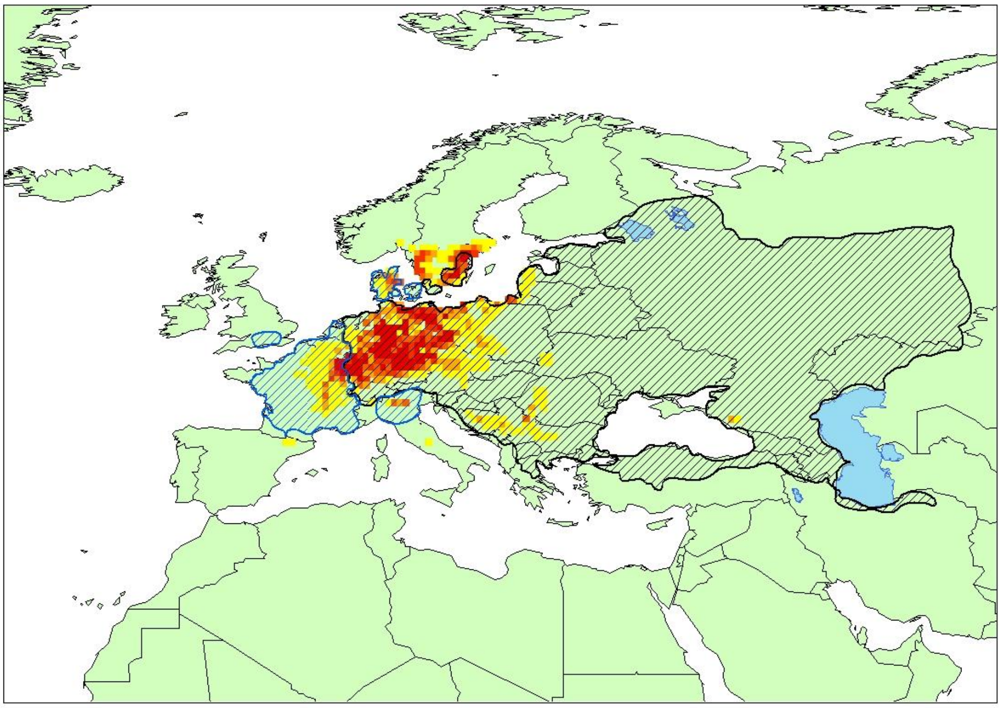
157

158

Abramis brama



Silurus glanis



Appendix 1

Data sources to freshwater AquaMaps

Pfafstetter basins: HYDRO1K, U.S. Geological Survey EROS Center
http://eros.usgs.gov/#/Find_Data/Products_and_Data_Available/gtopo30/hydro

Elevation: ETOPO2, National Geophysical Data Center
http://www.ngdc.noaa.gov/mgg/global/relief/ETOPO2/ETOPO2v2-2006/

Mean annual precipitation: CRU 0.5 Degree Dataset, New et al.(1998), Atlas of the Bioshpere
http://www.sage.wisc.edu/atlas/maps.php?datasetid=34&includerelatedlinks=1&dataset=34

Net primary production: Foley et al. (1996), Kucharik et al. (2000), Atlas of the Bioshpere
http://www.sage.wisc.edu/atlas/maps.php?datasetid=37&includerelatedlinks=1&dataset=37

Soil pH, soil moisture, and soil carbon: IGBP-DIS (1998), Atlas of the Bioshpere
http://www.sage.wisc.edu/atlas/maps.php?datasetid=20&includerelatedlinks=1&dataset=20
Willmott and Matsuura (2001), Atlas of the Bioshpere
http://www.sage.wisc.edu/atlas/maps.php?datasetid=23&includerelatedlinks=1&dataset=23
IGBP-DIS (1998), Atlas of the Bioshpere
http://www.sage.wisc.edu/atlas/maps.php?datasetid=21&includerelatedlinks=1&dataset=21

Worldclim data: 19 Bioclim variables (http://www.worldclim.org/bioclim).
BIOCLIM: Bioclimatic variables are derived from the monthly temperature and rainfall values in order to generate more biologically meaningful variables. These are often used in ecological niche modeling (e.g.. BIOCLIM. GARP). The bioclimatic variables represent annual trends (e.g.. mean annual temperature. annual precipitation) seasonality (e.g.. annual range in temperature and precipitation) and extreme or limiting environmental factors (e.g.. temperature of the coldest and warmest month. and precipitation of the wet and dry quarters). A quarter is a period of three months (1/4 of the year)

They are coded as follows:

BIO1 = Annual Mean Temperature
BIO2 = Mean Diurnal Range (Mean of monthly (max temp - min temp))
BIO3 = Isothermality (BIO2/BIO7) (* 100)
BIO4 = Temperature Seasonality (standard deviation *100)
BIO5 = Max Temperature of Warmest Month
BIO6 = Min Temperature of Coldest Month
BIO7 = Temperature Annual Range (BIO5-BIO6)
BIO8 = Mean Temperature of Wettest Quarter
BIO9 = Mean Temperature of Driest Quarter
BIO10 = Mean Temperature of Warmest Quarter
BIO11 = Mean Temperature of Coldest Quarter
BIO12 = Annual Precipitation
BIO13 = Precipitation of Wettest Month
BIO14 = Precipitation of Driest Month
BIO15 = Precipitation Seasonality (Coefficient of Variation)
BIO16 = Precipitation of Wettest Quarter
BIO17 = Precipitation of Driest Quarter
BIO18 = Precipitation of Warmest Quarter
BIO19 = Precipitation of Coldest Quarter

This scheme follows that of ANUCLIM, except that for temperature seasonality the standard deviation was used because a coefficient of variation does not make sense with temperatures between -1 and 1).