

## ORIGINAL RESEARCH

Predicting species occurrences with habitat network models Damian O. Ortiz-Rodríguez<sup>1,2,3,4</sup>  | Antoine Guisan<sup>3,5</sup>  | Rolf Holderegger<sup>1,4</sup> | Maarten J. van Strien<sup>2</sup> <sup>1</sup>WSL Swiss Federal Research Institute, Birmensdorf, Switzerland<sup>2</sup>Planning of Landscape and Urban Systems (PLUS), Institute for Spatial and Landscape Planning, ETH Zurich, Zürich, Switzerland<sup>3</sup>Department of Ecology and Evolution, University of Lausanne, Lausanne, Switzerland<sup>4</sup>Department of Environmental Systems Science, ETH Zurich, Zürich, Switzerland<sup>5</sup>Institute of Earth Surface Dynamics, University of Lausanne, Lausanne, Switzerland**Correspondence**

Damian O. Ortiz-Rodríguez, Planning of Landscape and Urban Systems, ETH Zurich, Stefano-Franscini-Platz 5, CH-8093 Zurich, Switzerland.

Email: damiano@ethz.ch

**Funding information**

Schweizerischer Nationalfonds zur Förderung der Wissenschaftlichen Forschung, Grant/Award Number: CR30I3\_159250

**Abstract**

1. Biodiversity conservation requires modeling tools capable of predicting the presence or absence (i.e., occurrence-state) of species in habitat patches. Local habitat characteristics of a patch ( $lh$ ), the cost of traversing the landscape matrix between patches (weighted connectivity [ $wc$ ]), and the position of the patch in the habitat network topology ( $nt$ ) all influence occurrence-state. Existing models are data demanding or consider only local habitat characteristics. We address these shortcomings and present a network-based modeling approach, which aims to predict species occurrence-state in habitat patches using readily available presence-only records.
2. For the tree frog *Hyla arborea* in the Swiss Plateau, we delineated habitat network nodes from an ensemble habitat suitability model and used different cost surfaces to generate the edges of three networks: one limited only by dispersal distance (Uniform), another incorporating traffic, and a third based on inverse habitat suitability. For each network, we calculated explanatory variables representing the three categories ( $lh$ ,  $wc$ , and  $nt$ ). The response variable, occurrence-state, was parametrized by a sampling intensity procedure assessing observations of comparable species over a threshold of patch visits. The explanatory variables from the three networks and an additional non-topological model were related to the response variable with boosted regression trees.
3. The habitat network models had a similar fit; they all outperformed the non-topological model. Habitat suitability index ( $lh$ ) was the most important predictor in all networks, followed by third-order neighborhood ( $nt$ ). Patch size ( $lh$ ) was unimportant in all three networks.
4. We found that topological variables of habitat networks are relevant for the prediction of species occurrence-state, a step-forward from models considering only local habitat characteristics. For any habitat patch, occurrence-state is most prominently influenced by its habitat suitability and then by the number of patches in a wide neighborhood. Our approach is generic and can be applied to multiple species in different habitats.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2019 The Authors. *Ecology and Evolution* published by John Wiley & Sons Ltd.

## KEYWORDS

connectivity, cost surface, habitat network, habitat suitability, network topology, species occurrence

## 1 | INTRODUCTION

Knowledge about the spatial distribution of species is a key element for any conservation effort. To gain insights on the presence or absence of a species at specific locations (occurrence-state; from occurrence in Kéry & Schaub, 2012), it is necessary but not sufficient to consider the conditions that make the specific sites suitable for a species, that is, to define patches of suitable habitats (sensu Guisan & Zimmermann, 2000). One also needs to take into account habitat connectivity, which is the way the suitable habitat patches are accessible and thus connected to each other, or “the degree to which the landscape facilitates or impedes” the movement of species (Taylor, Fahrig, Henein, & Merriam, 1993). The consideration of connectivity is important, as a habitat patch in which the environmental conditions are suitable for a certain species can actually be unoccupied due to the inability or low probability of the species to reach the patch (Barve et al., 2011). In such a case, the occurrence-state of the known suitable habitat patch would be 0, as occurrence-state is a property of habitat patches with two alternative states: presence (1) or absence (0). To better capture the factors influencing the occurrence-state of a species, and to be able to make predictions about this state, it is necessary to develop new modeling approaches that do not only consider the local conditions in a habitat patch, but also the connectivity between patches. This was the main goal of the present study.

Habitat patches and their connectivity can be represented in a network-theoretical framework. Since the work of Bunn, Urban, and Keitt (2000), spatially explicit habitat network models have been in common use (e.g., Dufлот, Avon, Roche, & Bergès, 2018; Saura & Pascual-Hortal, 2007; Urban, Minor, Treml, & Schick, 2009). In such networks, nodes usually represent habitat patches potentially inhabited by a species, and edges commonly represent potential movement among them. In many habitat networks, edges are modeled with cost surfaces (i.e., raster maps in which each cell has a value of resistance to movement) from which likely movement routes can be derived (Adriaensen et al., 2003; McRae, 2006). In other cases, edges are modeled with straight-line transects (Jordán, Magura, Tóthmérész, Vasas, & Ködöböcz, 2007; van Strien et al., 2014). The specific arrangement of nodes and edges is the network topology (Kauffman, 1993; Urban et al., 2009), which can be analyzed at different scales, ranging from the immediate vicinity of a patch to the whole network (Baranyi, Saura, Podani, & Jordán, 2011). Following this logic, the presence of a species in a certain habitat patch is influenced by three different key categories of factors, which can be summarized with the following conceptual equation:

$$\psi_i = f(lh_i, wc_i, nt_i) \quad (1)$$

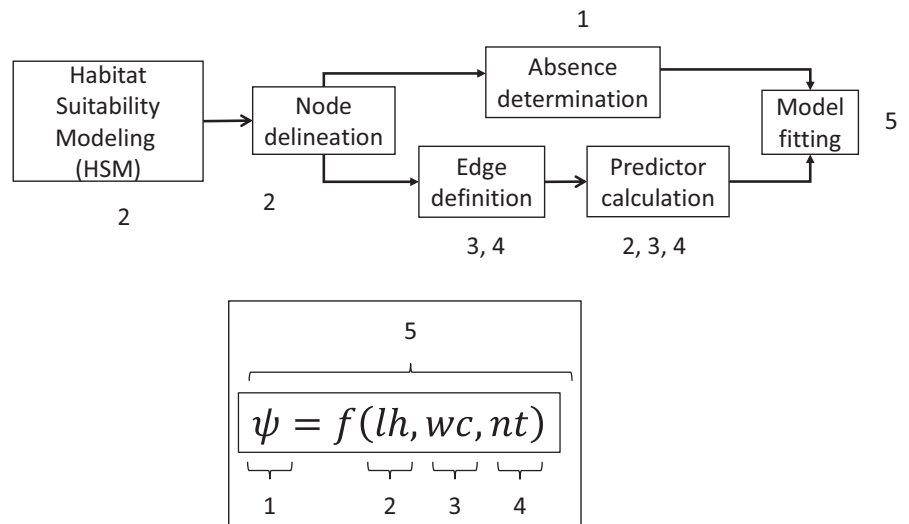
where  $\psi_i$  is the occurrence-state of a species (whether it is present or absent) in a habitat patch  $i$ ,  $lh_i$  refers to the local habitat characteristics of such patch,  $wc_i$  is the weighted connectivity of the patch to surrounding patches, and  $nt_i$  is the place of the patch (node) in the network topology.

Local habitat characteristics ( $lh_i$ ) are defined by the properties of suitability and size of a habitat patch. Patch size is an important factor in metapopulation biology (Hanski, 1992), and its relevance is widely acknowledged in studies dealing with occurrence and distribution of species (Hodgson, Moilanen, Wintle, & Thomas, 2011; Saura & Pascual-Hortal, 2007). The suitability of a patch is determined by the environmental requirements (i.e., environmental niche) of species. These requirements can be assessed with habitat suitability modeling (HSM), which aims to predict the distribution of species across a study area based on mapped environmental factors (Guisan & Zimmermann, 2000; Thuiller & Münkemüller, 2010).

Habitat connectivity depends on the movement ability and behavior of a species, reflected in species-specific maximum dispersal distances (Jenkins et al., 2007). It also depends on factors that facilitate or inhibit the movement of a species through the landscape between neighboring suitable patches (Prevedello & Vieira, 2009). The weighted connectivity ( $wc$ ) component of conceptual Equation (1) includes those variables that explicitly incorporate the probability of traversing the landscape matrix (the latter determining the “weights”) into their calculation. The  $wc$  factors give rise to the emergent large-scale structure of a network. The network topology ( $nt$ ) refers to this large-scale structure (Albert & Barabási, 2000). For a given node  $i$ ,  $nt_i$  refers to variables that describe its neighborhood, position, and importance in the whole network, independent of any weights specific to a certain environmental or species-specific context. The context-independent nature of  $nt$  variables makes them ideal to compare habitat networks of different species in different environments, as well as to compare habitat networks with other kinds of natural networks (Watts & Strogatz, 1998).

Determining the occurrence-state of a habitat patch is difficult for non-sessile species (MacKenzie et al., 2002). Although it can be performed by site occupancy models (Kéry & Schaub, 2012), these can only be used in situations where sites have been sampled in a regular and systematic way (MacKenzie et al., 2006). Another difficulty is the empirical estimation of connectivity between habitat patches, which is usually performed by means of mark-recapture, radio tracking, GPS sensors, or genetic methods (Kool, Moilanen, & Treml, 2013; Straka, Paule, Ionescu, Štofík, & Adamec, 2012). Due to their high costs and labor intensity, these methods are usually not implemented over large spatial scales, for several species, or by institutions under economic hardship. In summary, the determination of both the occurrence-state of

**FIGURE 1** Workflow of steps used in this study. The numbers indicate the relation of each step to Equation (1) (inset below)



habitat patches and of the connectivity among them is based on data that are relatively expensive, laborious, and time-consuming to obtain. In contrast, spatially explicit data on species observations are readily available in many countries, such as the data aggregated in the GBIF international database ([www.gbif.org](http://www.gbif.org)) or in the Swiss InfoSpecies database ([www.infospecies.ch](http://www.infospecies.ch)). These data consist of confirmed presences, but usually do neither contain any absences nor information on whether all potential habitat patches were surveyed. Given these biases, it is a challenge to parameterize habitat networks with such incomplete data. Nevertheless, given the high prevalence of such data, it is worthwhile to explore the possibilities of using it to parametrize habitat network models aiming to predict species occurrence. By aggregating observation data from groups of comparable species (Anderson, 2003) to determine a habitat patch's sampling intensity, we expect that likely absences for a focal species can be estimated.

In this study, we developed a habitat network modeling approach to predict species occurrences in habitat patches following conceptual Equation (1). We aimed to develop a generic method that (a) includes insights about the topology of the habitat networks and (b) makes use of readily available presence-only records. We expected that the incorporation of network topological variables would increase the explanatory power of models as compared to nontopological ones, addressing the omission of connectivity factors incurred by traditional models capable of predicting species occurrences (such as HSM and resource selection models; Boyce, Vernier, Nielsen, & Schmiegelow, 2002). We anticipate that the approach can be applied to a multitude of species in different environments at minimal cost. We exemplify our approach with the European tree frog (*Hyla arborea* L.) in the Swiss Plateau.

We followed a multistep procedure with two modeling stages (Figure 1). First, we used HSM to delineate suitable patches, that is, the nodes of the network. We then defined the edges based on least-cost calculations on different cost surfaces, which incorporated different environmental, biological, and human influences on the landscape, generating three different networks. From these

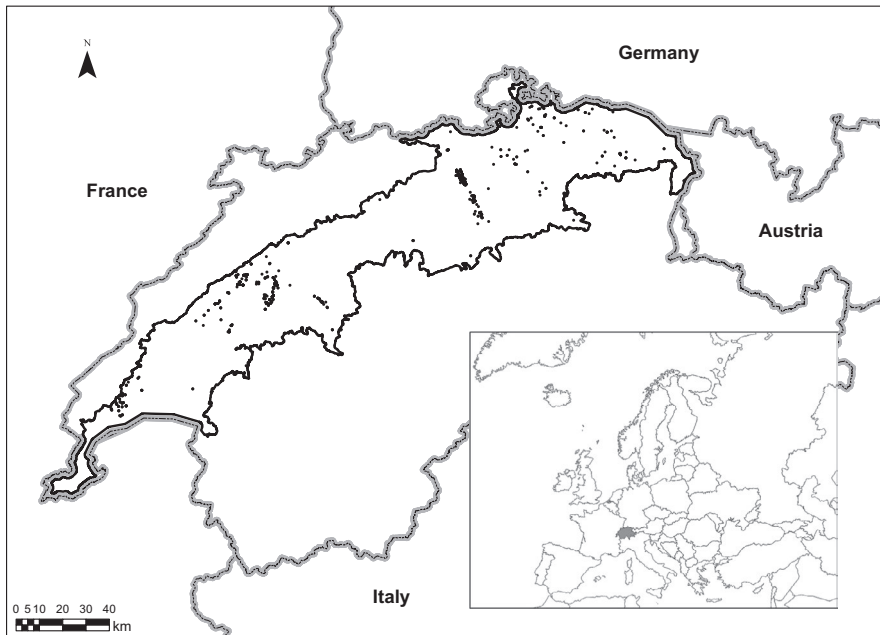
networks, we calculated several variables quantifying the three categories of factors (i.e.,  $lh$ ,  $wc$ , and  $nt$ ) in Equation (1), which were used as explanatory variables in models that related them to the response variable occurrence-state. We then compared the fit of models with and without the  $wc$  and  $nt$  variables. In order to calculate occurrence-state, we developed an approach inspired by Anderson (2003) that uses comparative sampling intensity to define absences of the focal species in habitat patches. Finally, by means of boosted regression trees (BRTs), we tested the explanatory power of predictor variables related to the three factors of Equation (1) on occurrence-state.

## 2 | METHODS

### 2.1 | Study area, focal species and presence records

Our study area consisted of the Swiss Plateau, a densely populated region (426 inhabitants/km<sup>2</sup>; Müller-Jentsch, 2012) of 11,168 km<sup>2</sup>, where strong increases in landscape fragmentation and urban sprawl have recently occurred (Roth, Schwick, & Spichtig, 2010). The area is dominated by human land use, with a patchy distribution of settlements, agricultural land, and forests. The exact shape of the study area (Figure 2) was defined by the boundaries of the Swiss Plateau from the official map of the biogeographical regions of Switzerland (OFEV, 2011) minus a 2 km (i.e., commonly reported amphibian dispersal distances; Smith & Green, 2005) negative buffer away from the international borders of Switzerland to prevent border effects. We chose the European tree frog (*H. arborea* L.) as our focal species, as it is a neither abundant nor rare habitat specialist, vulnerable to environmental disturbances and restricted to well-defined natural features, which are areas close to sunny forest edges and bushy landscape elements surrounding vegetation-poor ponds, in which it spawns (Clauzel, Girardet, & Foltete, 2013).

Our dataset on species occurrences consisted of geopositioned records of 13 amphibian species (Appendix A1) that were sampled from water bodies (mainly ponds, but also shallow lakeshores) between 2006 and 2015 across the Swiss Plateau, provided by



**FIGURE 2** Location of the study area (Swiss Plateau; black line) in Switzerland (gray line, solid gray in inset) and presence records of *Hyla arborea* (black dots)

InfoSpecies-KARCH ([www.infospecies.ch](http://www.infospecies.ch)). The records in this dataset originated from a variety of sources and observers and are limited to sites visited and reported on during the above-mentioned period. Some sites were visited only once, while others were visited annually or many times. It was also not clear whether observers reported all species encountered at a visit, an arbitrary subset of species or a single species. Absences of particular species from particular sites were thus not explicitly reported. Such data limitations are frequently encountered in national and international observation databases. Records on amphibian occurrences were aggregated at a 1 ha resolution. In total, the dataset consisted of 2,354 locations with at least one amphibian species presence in one or more years. Out of these, 291 contained the focal species *H. arborea* (Figure 2).

## 2.2 | Habitat suitability modelling

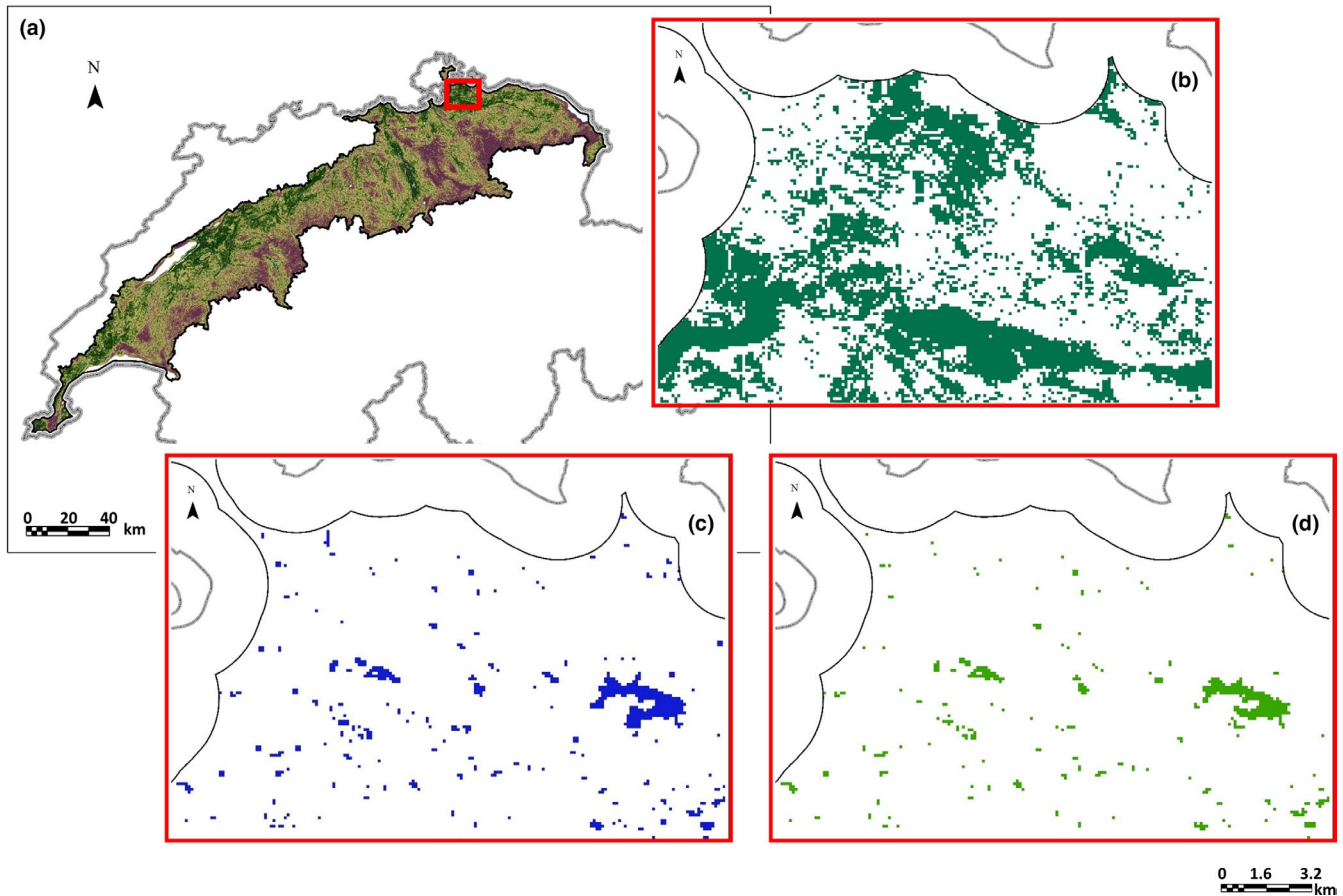
For HSM, we compiled a dataset of 25 environmental predictor variables based on previous studies describing the environmental preferences of pond-based amphibians in general and *H. arborea* in particular (Pellet, Hoehn, & Perrin, 2004; Van Buskirk, 2005; Zanini, Pellet, & Schmidt, 2009), as well as additional variables quantifying human influence on ecosystems. Our final HSM predictors fell under three basic categories: human influence, natural landscape features, and climate variables. All predictor variables were converted to a resolution of 1 ha. Circular moving windows with a 2 km radius (common dispersal distance of amphibians) were used for calculating many of the predictors. We eliminated collinear predictor variables based on pairwise Pearson's correlation coefficients (Gillham, 2001) with a reference threshold of 0.75 and based on variance inflation factors (VIF) with a threshold of 0.9, using the packages USDM (Naimi, Hamm, Groen, Skidmore, & Toxopeus, 2014) and stats in R 3.3 (R Development Core Team, 2016). The removed variables were mean annual precipitation, total noise at daytime,

recreation intensity, highway density, and density of roads. This led to a final selection of 20 predictor variables for HSM (Appendix A2 lists the HSM predictors; Appendix A3 gives a short description and the sources of the data). All data processing was carried out with ArcGIS 10.4.1. (ESRI, 2016) in Python 2.7 (Python Software Foundation, 1995).

In order to delineate potential habitat patches of *H. arborea*, we generated an ensemble habitat suitability model (HSM) in which the 291 presences of *H. arborea* constituted the response variable. To prevent pseudoreplication, we included only one record of *H. arborea* per sampling site, even if the species was observed in multiple years. We generated 10,000 pseudoabsences as recommended by Barbet-Massin, Jiguet, Albert, and Thuiller (2012), with one round of pseudoabsence selection. We developed an ensemble using the R-package Biomod2 (Thuiller, Georges, Engler, & Breiner, 2016), which does multiple runs of different models, projects the models spatially, and generates consensus projections between the different models. In this study, we used the mean ensemble of a generalized linear (GLM), a random forest (RF), and a maximum entropy (MaxEnt) model. The models were evaluated with ROC AUC, with a quality threshold of  $AUC \geq 0.7$  (Bulluck, Fleishman, Betrus, & Blair, 2006). To binarize the continuous habitat suitability maps, the applied criterion was the point in the ROC curve that minimizes the difference between sensitivity and specificity. We used default settings unless otherwise specified.

## 2.3 | Node delineation

HSM resulted in a map indicating where the environmental conditions were potentially suitable for *H. arborea*. In order to delimit suitable habitat patches for this species, we intersected the binary results of the ensemble HSM with those areas in which *H. arborea* can reproduce, namely water bodies in the Swiss Plateau. Water



**FIGURE 3** Continuous (a) and binary (b; close-up of red area in a) suitability maps yielded by the ensemble habitat suitability model for *Hyla arborea* in the Swiss Plateau. Discrete habitat patches (d; same close-up as in b) were produced by the application of a mask (c; same close-up as in b)

bodies were defined by merging several spatial datasets: lakeshores (Swisstopo, 2016), mires (OFEV, 2010), amphibian spawning sites (BAFU, 2016), and all locations with at least one occurrence of at least one amphibian species in the period 2006–2015. The merged layer constituted a mask that was overlaid with the binarized HSm (Guisan et al., 2006; Figure 3). A habitat patch was considered unique if it was not connected to any other patch under a Moore neighborhood criterion (i.e., considering all eight neighbors of a raster cell). For each patch, we determined its size (ha) and its mean habitat suitability, which were later used as explanatory variables for the occurrence-state modeling (see below). The identified habitat patches constituted the nodes of the habitat network.

## 2.4 | Edge definition

Between pairs of nodes, we defined edges based on least-cost calculations (Etherington, 2016). We developed an algorithm that takes as input a binary raster of habitat patches and a cost surface. The algorithm determines the least cumulative cost between patches and draws an edge between patches if the total cost is below a certain threshold. Translation of dispersal probabilities into dispersal costs

and vice versa was performed following the  $p2p$  function of the R-package PopGenReport (Adamack & Gruber, 2014):

$$\text{cost} = \log(\text{prob}) / \log(p) * d0 \quad (2)$$

in which  $d0$  is the dispersal distance of a proportion  $p$  of individuals,  $\text{prob}$  is the probability of dispersal between patches, and  $\text{cost}$  is the cost-value associated with a certain probability  $\text{prob}$ . We set  $p = .5$ , so that  $d0$  equaled the median dispersal distance. We set the dispersal probability threshold beyond which no edges were drawn to 0.0001. Subsequently,  $d0$  was set to 200 m so that no edges were formed over cost distances of 2,658. When cost distances are just Euclidian distances, 2,658 m is slightly above the reported average maximum dispersal distance of *H. arborea* (Clauzel et al., 2013).

Making use of three different cost surfaces, we created three different networks: a Uniform, a Traffic, and an Inverse Habitat Suitability network. In the Uniform network, the cost distance was equal to the Euclidean distance among habitat patches. In the Traffic network, the default cost-value of a raster cell was 1, and for all the raster cells that intersected with a road (excluding tunnels), the traffic intensity on the respective road was converted to a cost-value. To do this, we calculated the probability that an animal successfully

crosses a road according to van Langevelde and Jaarsma (2009, equation A1.4). These authors include speed and physical dimensions of vehicles and animals, as well as traffic volumes on intersecting roads, to calculate probabilities of successful road crossing. The parameter settings for *H. arborea* were taken from Van Strien and Grêt-Regamey (2016). Subsequently, the calculated probabilities were transformed to costs using Equation (2). The traffic values per road segment were those calculated by the 2010 version of the Swiss national passenger transport model (ARE, 2010). For the Inverse Habitat Suitability network (HabSuit), we assumed that the probabilities of dispersing through the most unsuitable and suitable terrain were 0 and 1, respectively. Therefore, the continuous habitat suitability raster from HSM was divided by the maximum suitability and the inverse value was then taken as cost-value (sensu Ziótkowska, Ostapowicz, Radeloff, & Kueemmerle, 2014). With this approach, not only the network topology differed between the cost surfaces, but also the weight of individual edges in the network. Edge weights were calculated by transforming the least-cost values to dispersal probabilities following Equation (2); hence, costlier paths have lower dispersal probabilities. Edge calculations were performed using the Python packages numpy (Oliphant, 2006), arcpy (ESRI, 2016), and igraph (Csardi & Nepusz, 2006).

## 2.5 | Calculation of explanatory variables for occurrence-state modeling

We prepared a set of explanatory variables for network model assessment, which quantified the three types of factors from Equation (1) as patch (node) properties. To address the aspect of network topology ( $nt_i$ ) at different scales, we calculated for each habitat patch the degree, third-order neighborhood, and betweenness centrality. The degree is the number of connections (edges) a specific node  $i$  has to other nodes (Jordán & Scheuring, 2004). The third-order neighborhood measures the number of nodes (patches) that can be reached in maximally three topological steps through the network (Csardi & Nepusz, 2006). To measure the influence of topology at the whole-network scale, we used betweenness centrality, which measures how many connections between all node pairs in the network pass-through node  $i$  (Freeman, 1978). While Baranyi et al. (2011) define it as a meso-scale measure, it is actually calculated considering all other nodes in the network, so it is an appropriate proxy to check how the whole-network structure affects a node-specific property.

To account for the weighted connectivity of patches ( $wc_i$ ), the calculated variable was the strength, which is the sum of the weights of all the edges connecting a node to others (Barrat, Barthélemy, Pastor-Satorras, & Vespignani, 2004). It is thus also considered a topological variable. We also calculated the habitat availability, which is a hybrid variable incorporating aspects of  $nt$  and  $lh$ . This measure calculates a weighted sum of all patch sizes that can be reached from a focal patch  $i$ . The weights are calculated as the maximum product probability between two patches. Habitat availability is similar to the probability of connectivity index of Saura and Pascual-Hortal (2007), with the main difference that it is calculated for each node

separately (not summed over all nodes) and not divided by the total habitat area. In order to achieve efficient computation times, we limited the process to consider patches only up to second-order neighborhood, after having observed only negligible change for higher neighborhood order values.

In addition to those network topological variables, we also included the size ( $ha$ ) and the mean habitat suitability (habitat suitability index; HSI) values of the patches to evaluate the influence of local habitat characteristics ( $lh_i$ ). The habitat suitability values per patch were obtained by calculating the mean habitat suitability of all pixels that made up a discrete patch. The size was an attribute generated when defining the discrete patches.

## 2.6 | Determination of absences of *H. arborea*

To define the absence values of the binary response variable occurrence-state, we used an adapted version of the approach used by Anderson (2003), based on comparative sampling intensity. For each habitat patch, we assumed that a presence record of any amphibian species in a particular year represented a confirmed visit to the patch in that year by an observer familiar with amphibians. Furthermore, we assumed that if a patch was visited a considerable number of years and a certain focal species was not reported, it is likely that such species was indeed not present in that patch during these years. We calculated the total number of times a patch had been visited ( $V_t$ ) by aggregating all observations of the 13 amphibian species in the data set (Appendix A1). Multiple amphibian species recorded in one particular year for a certain location were counted as only one visit. As we had 10 years of observations with a temporal resolution of 1 year, the maximum number of visits ( $V_t$ ) was 10. For each patch, we also calculated the number of times *H. arborea* was observed ( $V_h$ ), and for each  $V_t$  value, we calculated the mean  $V_h$ . We assigned a likely absence (occurrence-state = 0) to all those patches with a  $V_t$  value that corresponded to a mean  $V_h \geq 1$  (i.e., on average *H. arborea* was found at least once over the years) and in which *H. arborea* had not been recorded. Patches that neither contained a confirmed presence (from the original occurrence data on *H. arborea*) nor a likely absence had an unknown occurrence-state and were therefore excluded from subsequent analyses.

## 2.7 | Occurrence-state network model fitting

In order to test what kind of variables were most important for explaining the occurrence-state of a species in a habitat patch, we ran boosted regression trees (BRTs; Elith & Leathwick, 2017). This modeling technique has been proven useful for the analysis of complex ecological data. It can handle interactions among variables and non-parametric relationships, and integrates the calculation of variable importance (Elith, Leathwick, & Hastie, 2008). The three different habitat networks (Uniform, Traffic, and HabSuit) were used to build three separate models. In all of them, we used the seven explanatory variables described above. In addition, we built a model without

**TABLE 1** Summary of the three network-based models (Uniform, Traffic, and HabSuit) and the model without topological predictors fitted by BRT's

Model	Resistance surface used to define edges	Predictors included
Uniform	Cost-value equal to Euclidean distance among habitat patches; edge formation limited by dispersal distance only.	Degree Strength Third-order neighborhood Habitat availability Betweenness centrality Mean HSI Mean patch area
Traffic	Traffic intensity on intersecting roads converted to cost-value.	Degree Strength Third-order neighborhood Habitat availability Betweenness centrality Mean HSI Mean patch area
HabSuit	Cost-value defined by inverse of maximum-weighted habitat suitability index.	Degree Strength Third-order neighborhood Habitat availability Betweenness centrality Mean HSI Mean patch area
noTopo	None. Network topology not considered.	Mean HSI Mean patch area

topological variables (noTopo), for which the only explanatory variables were patch size and HSI. For all four models (Summarized in Table 1), the response variable was the occurrence-state of *H. arborea*.

BRT models are based on an aggregation of numerous classification trees (at least 1,000 as recommended by Elith et al., 2008). The learning rate (lr) and tree complexity (tc) influence the number of trees that are used in a final BRT model. As stochastic factors give rise to differences in the prediction each time the model is run, we used 100 runs with lr = 0.001, tc = 5, and a bagging fraction of 0.75 (following the general guidelines of Elith et al., 2008) with the gbm step function of the R-package dismo (Hijmans, Phillips, Leathwick, & Elith, 2017). This function searches for the number of trees that yields the lowest deviance. We evaluated the predictive power of the models by comparing the distribution of their cross-validated AUC values (cv-AUC) over the 100 runs. We assessed the different variables with the distributions of their importance scores over all 100 runs.

### 3 | RESULTS

#### 3.1 | Habitat suitability modeling and patch delineation

The ensemble HSm, a mean of nine models, yielded a continuous and a binarized habitat suitability map (Figure 3a,b). The mask (Figure 3c) applied on the binarized suitability map yielded the definitive delineation of habitat patches (Figure 3d). The total number of habitat patches was 1900.

#### 3.2 | Edge definition

The three different cost surfaces yielded different networks. The Traffic network comprised 254 components (groups of linked patches) and the Uniform network 134 components. The HabSuit network was the most sparsely connected, divided into 850 components (Figure 4).

#### 3.3 | Values of explanatory variables

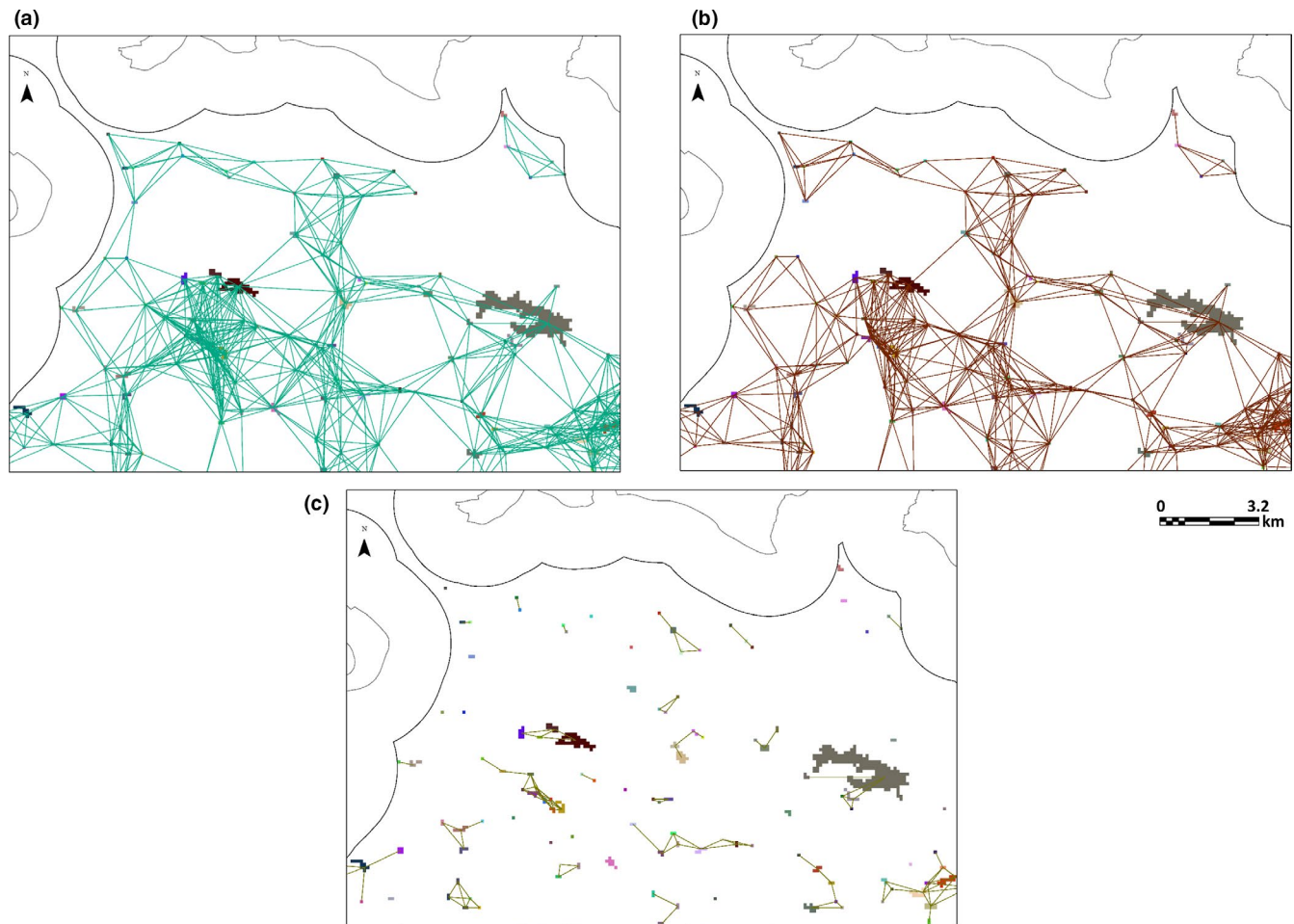
Mean values of all variables are shown in Table 2. For most network variables, the HabSuit network exhibited much smaller values (Table 2) than the Uniform and Traffic networks. For habitat availability, the differences were less pronounced.

#### 3.4 | Determination of absences

The mean number of times that *H. arborea* was sighted (mean  $V_h$ ) increased with the number of times a patch was visited ( $V_t$ ). At  $V_t \geq 6$ , mean  $V_h \geq 1.8125$ , indicating that on average there was more than one sighting of *H. arborea* when a site was visited six or more times (Figure 5). Therefore, we regarded every patch with  $V_t \geq 6$  and no confirmed *H. arborea* sighting as a likely absence. In doing so, we determined 46 likely absences of *H. arborea*, complementing the 209 confirmed presences.

#### 3.5 | Occurrence-state network model fitting

The boosted regression trees showed a similar predictive performance among the three network models, as indicated by the



**FIGURE 4** Habitat networks of *Hyla arborea* in the Swiss Plateau (same close-up as in Figure 3). The networks were based on three different cost surfaces: Uniform (a), Traffic (b), and Inverse Habitat Suitability (HabSuit; c)

**TABLE 2** Mean values of explanatory variables for the three different networks of *Hyla arborea* in the Swiss Plateau: Uniform, Traffic, and Inverse Habitat Suitability (HabSuit)

	Uniform	Traffic	HabSuit
Degree	11.32	8.86	4.02
Strength	0.58	0.49	0.22
Third-order neighborhood	41.63	30.39	11.65
Habitat availability	311,578.5	297,289.4	265,234.7
Betweenness centrality	1,029.20	911.85	34.19
Mean HSI	650.15	650.15	650.15
Mean patch area (ha)	18.97	18.97	18.97

distributions of their cv-AUC scores (Figure 6). The model without topological measures (noTopo) was outperformed by all three models with topological variables (Table 3).

HSI was the variable with the highest importance in all models. While the lowest mean HSI value in a suitable patch determined by the ensemble HSM was 382, the partial dependence plots for all models (Appendix A4) pointed toward a threshold HSI

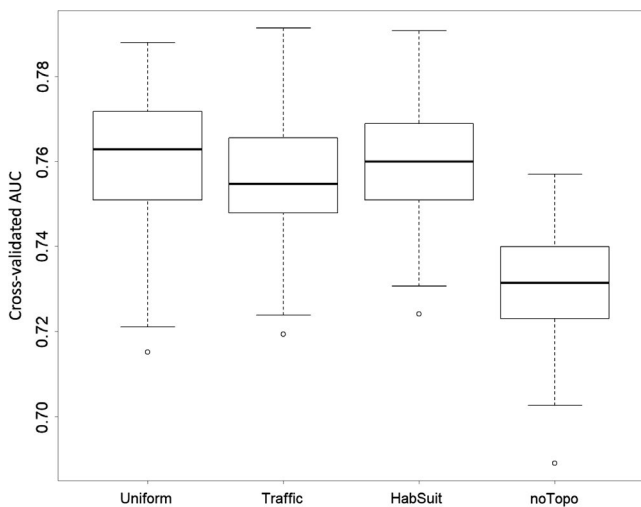
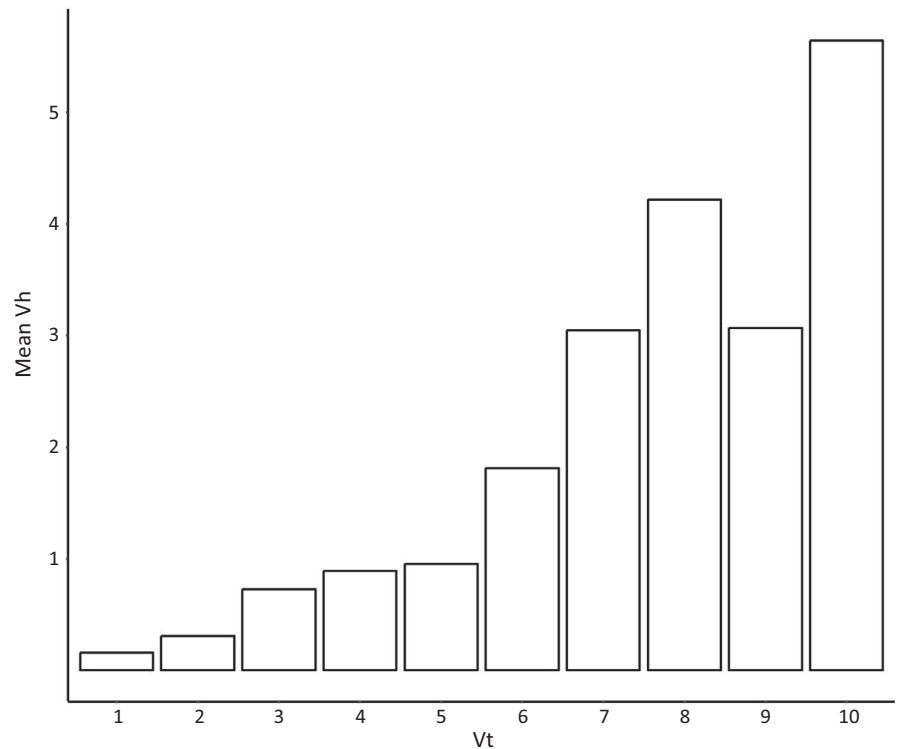
value above 500. In all the network models, third-order neighborhood was consistently the second most important variable (importance above 13%), followed by strength in the Uniform and Traffic models, and by habitat availability in the HabSuit model. Patch size was consistently among the least important variables in all models (Table 4).

## 4 | DISCUSSION

The goal of this study was to develop an approach to assess the occurrence-state of a species in habitat patches, one which would be inexpensive and practical by using widely available species presence data, and which would have an added predictive value by incorporating topological properties of the species' habitat networks as predictor variables. Our results support the expectation that topological variables of habitat networks are indeed relevant for explaining and predicting the occurrence-state of a species in habitat patches. This is showcased by the results on BRT model comparison, in which the model without topological variables (no-Topo) had the poorest performance in terms of its mean cv-AUC score (Figure 6).



**FIGURE 5** Relationship between the number of times a patch has been visited ( $V_t$ ) in the period 2006–2015 and the mean number of times that *Hyla arborea* has been spotted for any number of visits (mean  $V_h$ ) in the Swiss Plateau



**FIGURE 6** Distributions of cross-validated AUC scores over 100 runs for four models (Uniform, Traffic, and HabSuit networks, and noTopo) for *Hyla arborea* occurrence-state in the Swiss Plateau

**TABLE 3** Mean cross-validated AUC scores over 100 runs of four models (Uniform, Traffic, HabSuit, and noTopo) for *Hyla arborea* occurrence-state in the Swiss Plateau

	Uniform	Traffic	HabSuit	noTopo
Cross-validated AUC score	0.7610	0.7552	0.7597	0.7309

Following our other main objective, a novel aspect we present in this study is the derivation of likely absences from presence-only data. Given that we use unsystematically collected data with low temporal resolution, it was not possible to use traditional site

occupancy models (Kéry & Schaub, 2012). Instead, we developed a comparative sampling intensity approach, which yielded the likely absences necessary to model occurrence-state. A drawback of this approach was that we could only define likely absences for a small fraction of the habitat patches originally defined: our BRT models were built on 255 out of 1,900 patches, and the final response variable only included 46 absences. Nevertheless, even with this relatively small dataset, all of the network models had mean cv-AUC scores above the 0.75 AUC threshold of acceptability for good models (Elith, 2002).

The difference in predictive power between the three different network models was slight. While the Uniform network had a better mean predictive performance than the others did (Table 3), the difference was too small to warrant any conclusions regarding a most probable movement hypothesis for the focal species *H. arborea*. In order to study this aspect, it may be worthwhile to experiment with different combinations of cost factors or ways to define them (such as nonlinear cost increase; Dufлот et al., 2018). The application of the method to ecologically different taxa and landscapes of various sizes might reveal greater differences among network models. The contrast between different models as presented here could be used to identify the most likely movement hypothesis for a given target species and therefore indicate the optimal kinds of network models to use in other ecological contexts.

The explanatory variable with the highest importance across all three network models was the mean habitat suitability index per patch (HSI). This was partially expected, as the baseline requirement for occupancy of a patch is to be suitable habitat. However, the values of HSI related to presence or absence had a different threshold than the one from the initial habitat suitability model (see partial dependence plots,

**TABLE 4** Mean relative importance (in %) of seven predictors over 100 runs of boosted regression trees (BRTs) for four models (Uniform, Traffic, HabSuit, and noTopo) for occurrence-state of *Hyla arborea* in the Swiss Plateau

Explanatory variable	Mean relative importance			
	Uniform	Traffic	HabSuit	noTopo
Habitat Suitability Index (HSI)	43.10	44.96	53.67	84.34
Third-order neighborhood	13.05	19.49	19.80	–
Strength	12.86	11.25	5.87	–
Habitat availability	10.01	9.08	13.26	–
Degree	8.46	5.55	0.33	–
Betweenness centrality	7.69	3.77	0.79	–
Patch area	4.82	5.90	6.27	15.66

Appendix A4). This was probably due to the contrast of confirmed presences with an informed selection of likely absences instead of random pseudoabsences as in HSM. The approach to determine likely absences applied here could thus be used to fine-tune binarization thresholds of habitat suitability models in other applications.

The consistent importance of the third-order neighborhood variable shows that the number of patches at this wider scale influences the occurrence-state of a species more than in the immediate vicinity (as shown by the lower importance of degree). The centrality of a patch at the whole-network scale did not prove to be especially relevant for the occupancy-state of a species. Another variable that was not important in any of the models was patch area. This arrangement of variable performances shows the complementarity of the different factors of Equation (1) in determining species occurrence-state. Local habitat characteristics are important, but offer an incomplete assessment, which can be enhanced by the incorporation of topological variables.

Our multistep approach of occurrence-state assessment thus achieved the expectation of improving predictive performance with the integration of connectivity and network topological considerations. Methods that incorporate these considerations should lead to better-targeted conservation actions, with more satisfactory outcomes. Our approach is generic and can be applied to any other wildlife species. In addition to presence observations of the focal and related species, the main data requirements for its application are spatial datasets of variables deemed important for the habitat suitability (for patch delineation) or connectivity (for edge definition) of the focal species. Patch delineation, easier for habitat specialists such as *H. arborea* (Van Buskirk, 2011), can be helped by the implementation of masks (e.g., excluding roads) to reduce patch size for more generalist species.

Further testing of our approach will indicate how widely it can be implemented, and ground-truthing likely absences would be particularly interesting. Although confirming absences is a difficult endeavor, novel techniques like eDNA (Deiner et al., 2017) could help

with this task. An interesting expansion of our approach would be to incorporate temporal dynamics. With a higher temporal resolution (we only had 10 time points), one could get further insights into how occurrences in one time step influence those in the next, thereby incorporating perspectives from metapopulation theory (Hanski, 1998) or research on dynamical complex networks in other disciplines (e.g., Alvarez-Buylla et al., 2007; Sinatra, Wang, Deville, Song, & Barabási, 2016). Our approach also opens the possibility for multi-species analyses, comparing the networks of different taxa. With such an analysis, it would become possible to identify key spatial elements across multiple networks and areas of strategic importance for the conservation of groups of species (Foltête, 2019). Our approach is fully expandable, and we hope it can find use in conservation management in different contexts around the world, often in dire need of effective and inexpensive methods.

## ACKNOWLEDGMENTS

We thank InfoSpecies-KARCH, especially Benedikt Schmidt, for providing presence data on amphibians, Andreas Justen (UVEK-ARE) for help with traffic data, and Frank Breiner (Wetlands International) as well as Olivier Broenimann (ECOSPAT-UNIL) for technical and statistical advice and support. We also thank two anonymous referees and the associate editor for helpful comments on the manuscript. This study is part of the CHECNET project, financed by the Swiss National Science Foundation (Grant nr. CR3013\_159250).

## CONFLICT OF INTERESTS

None declared.

## AUTHOR CONTRIBUTIONS

All authors conceived the ideas and designed the methodology. DOR analyzed the data and wrote the manuscript. RH, AG, and MvS reviewed and commented critically on the manuscript. All authors gave final approval for publication.

## DATA AVAILABILITY STATEMENT

The scripts used to develop the method are available from the Dryad Digital Repository (<https://doi.org/10.5061/dryad.sc818d5>). The original data are either conservation-relevant or property of Swiss public institutions, only available by independent agreements with them; therefore, it is not included in the Dryad data package of this article. Upon request, the authors can give indications on how to request access to the data from the relevant institutions.

## ORCID

Damian O. Ortiz-Rodríguez  <https://orcid.org/0000-0003-2781-4146>

Maarten J. Strien  <https://orcid.org/0000-0002-4311-0926>

Antoine Guisan  <https://orcid.org/0000-0002-3998-4815>

## OPEN RESEARCH BADGES



This article has earned an Open Data Badge for making publicly available the digitally-shareable data necessary to reproduce the reported results. The data is available at <https://doi.org/10.5061/dryad.sc818d5>.

## REFERENCES

- Adamack, A. T., & Gruber, B. (2014). PopGenReport: Simplifying basic population genetic analyses in R. *Methods in Ecology and Evolution*, 5, 384–387. <https://doi.org/10.1111/2041-210x.12158>
- Adriaensen, F., Chardon, J. P., De Blust, G., Swinnen, E., Villalba, S., Gulinck, H., & Matthysen, E. (2003). The application of 'least-cost' modelling as a functional landscape model. *Landscape and Urban Planning*, 64, 233–247. [https://doi.org/10.1016/S0169-2046\(02\)00242-6](https://doi.org/10.1016/S0169-2046(02)00242-6)
- Albert, R., & Barabási, A.-L. (2000). Topology of evolving networks: Local events and universality. *Physical Review Letters*, 85, 5234. <https://doi.org/10.1103/PhysRevLett.85.5234>
- Alvarez-Buylla, E. R., Benítez, M., Dávila, E. B., Chaos, Á., Espinosa-Soto, C., & Padilla-Longoria, P. (2007). Gene regulatory network models for plant development. *Current Opinion in Plant Biology*, 10, 83–91. <https://doi.org/10.1016/j.pbi.2006.11.008>
- Anderson, R. P. (2003). Real vs. artefactual absences in species distributions: Tests for *Oryzomys albigularis* (Rodentia : Muridae) in Venezuela. *Journal of Biogeography*, 30, 591–605. <https://doi.org/10.1046/j.1365-2699.2003.00867.x>
- ARE (2010). *Nationales Personenverkehrsmodell des UVEK (NPVM)*. Berne, Switzerland: ARE.
- BAFU (2016). *Amphibienlaichgebiete*. Berne, Switzerland: BAFU.
- Baranyi, G., Saura, S., Podani, J., & Jordán, F. (2011). Contribution of habitat patches to network connectivity: Redundancy and uniqueness of topological indices. *Ecological Indicators*, 11, 1301–1310. <https://doi.org/10.1016/j.ecolind.2011.02.003>
- Barbet-Massin, M., Jiguet, F., Albert, C. H., & Thuiller, W. (2012). Selecting pseudo-absences for species distribution models: How, where and how many? *Methods in Ecology and Evolution*, 3, 327–338. <https://doi.org/10.1111/j.2041-210X.2011.00172.x>
- Barrat, A., Barthélemy, M., Pastor-Satorras, R., & Vespignani, A. (2004). The architecture of complex weighted networks. *Proceedings of the National Academy of Sciences of the United States of America*, 101, 3747–3752. <https://doi.org/10.1073/pnas.0400087101>
- Barve, N., Barve, V., Jiménez-Valverde, A., Lira-Noriega, A., Maher, S. P., Peterson, A. T., ... Villalobos, F. (2011). The crucial role of the accessible area in ecological niche modeling and species distribution modeling. *Ecological Modelling*, 222, 1810–1819. <https://doi.org/10.1016/j.ecolmodel.2011.02.011>
- BFS (2013). *Waldmischungsgrad der Schweiz*. Berne, Switzerland: BFS.
- BFS (2015). *Statistik der Bevölkerung und der Haushalte (STATPOP)*. Berne, Switzerland: BFS.
- Boyce, M. S., Vernier, P. R., Nielsen, S. E., & Schmiegelow, F. K. A. (2002). Evaluating resource selection functions. *Ecological Modelling*, 157, 281–300. [https://doi.org/10.1016/S0304-3800\(02\)00200-4](https://doi.org/10.1016/S0304-3800(02)00200-4)
- Broennimann, O., Randin, C., Zimmermann, N. E., & Guisan, A. (2003). *Swiss Eco-Climatic GIS data*. Lausanne, Switzerland: Ecospat Spatial Ecology Group, University of Lausanne.
- Bulluck, L., Fleishman, E., Betrus, C., & Blair, R. (2006). Spatial and temporal variations in species occurrence rate affect the accuracy of occurrence models. *Global Ecology and Biogeography*, 15, 27–38. <https://doi.org/10.1111/j.1466-822X.2006.00170.x>
- Bunn, A. G., Urban, D. L., & Keitt, T. H. (2000). Landscape connectivity: A conservation application of graph theory. *Journal of Environmental Management*, 59, 265–278. <https://doi.org/10.1006/jema.2000.0373>
- Clauzel, C., Girardet, X., & Foltete, J. C. (2013). Impact assessment of a high-speed railway line on species distribution: Application to the European tree frog (*Hyla arborea*) in Franche-Comte. *Journal of Environmental Management*, 127, 125–134. <https://doi.org/10.1016/j.jenvman.2013.04.018>
- Csardi, G., & Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal Complex Systems*, 1695, 1–9.
- Deiner, K., Bik, H. M., Mächler, E., Seymour, M., Lacoursière-Roussel, A., Altermatt, F., ... Bernatchez, L. (2017). Environmental DNA metabarcoding: Transforming how we survey animal and plant communities. *Molecular Ecology*, 26, 5872–5895.
- Dufloy, R., Avon, C., Roche, P., & Bergès, L. (2018). Combining habitat suitability models and spatial graphs for more effective landscape conservation planning: An applied methodological framework and a species case study. *Journal for Nature Conservation*, 46, 38–47. <https://doi.org/10.1016/j.jnc.2018.08.005>
- Elith, J. (2002). Quantitative methods for modeling species habitat: comparative performance and an application to Australian plants. In S. Ferson, & M. Burgman (Eds.), *Quantitative methods for conservation biology* (pp. 39–58). New York: Springer.
- Elith, J., & Leathwick, J. (2017). *Boosted Regression Trees for ecological modeling*. R documentation. Retrieved from <https://cran.r-project.org/web/packages/dismo/vignettes/brt.pdf>
- Elith, J., Leathwick, J. R., & Hastie, T. (2008). A working guide to boosted regression trees. *Journal of Animal Ecology*, 77, 802–813. <https://doi.org/10.1111/j.1365-2656.2008.01390.x>
- EMPA (2011). *Aufbereitung von flächendeckenden Grundlagen für die Schallausbreitungsmodellierung in den Bereichen Meteorologie und Bodeneigenschaften*. Dübendorf, Switzerland: EMPA.
- ESRI (2016). *ArcGIS Desktop: Release 10.4.1*. Redlands, CA: Environmental Systems Research Institute.
- Etherington, T. R. (2016). Least-cost modelling and landscape ecology: Concepts, applications, and opportunities. *Current Landscape Ecology Reports*, 1, 40–53. <https://doi.org/10.1007/s40823-016-0006-9>
- Foltête, J.-C. (2019). How ecological networks could benefit from landscape graphs: A response to the paper by Spartaco Gippoliti and Corrado Battisti. *Land Use Policy*, 80, 391–394. <https://doi.org/10.1016/j.landusepol.2018.04.020>
- Freeman, L. C. (1978). Centrality in social networks conceptual clarification. *Social Networks*, 1, 215–239. [https://doi.org/10.1016/0378-8733\(78\)90021-7](https://doi.org/10.1016/0378-8733(78)90021-7)
- Gillham, N. W. (2001). *A life of Sir Francis Galton: From African exploration to the birth of eugenics*. Oxford, UK: Oxford University Press.
- Guisan, A., Broennimann, O., Engler, R., Vust, M., Yoccoz, N. G., Lehmann, A., & Zimmermann, N. E. (2006). Using niche-based models to improve the sampling of rare species. *Conservation Biology*, 20, 501–511. <https://doi.org/10.1111/j.1523-1739.2006.00354.x>
- Guisan, A., & Zimmermann, N. E. (2000). Predictive habitat distribution models in ecology. *Ecological Modelling*, 135, 147–186. [https://doi.org/10.1016/S0304-3800\(00\)00354-9](https://doi.org/10.1016/S0304-3800(00)00354-9)
- Hanski, I. (1992). Inferences from ecological incidence functions. *The American Naturalist*, 139, 657–662. <https://doi.org/10.1086/285349>
- Hanski, I. (1998). Metapopulation dynamics. *Nature*, 396, 41. <https://doi.org/10.1038/23876>
- Hijmans, R. J., Phillips, S., Leathwick, J., & Elith, J. (2017) *Package 'dismo'*. *dismo: species distribution modeling*. R package version 1.1-4. Retrieved from <https://CRAN.R-project.org/package=dismo>
- Hodgson, J. A., Moilanen, A., Wintle, B. A., & Thomas, C. D. (2011). Habitat area, quality and connectivity: Striking the balance for efficient conservation. *Journal of Applied Ecology*, 48, 148–152. <https://doi.org/10.1111/j.1365-2664.2010.01919.x>
- Jenkins, D. G., Brescacin, C. R., Duxbury, C. V., Elliott, J. A., Evans, J. A., Grablo, K. R., ... Williams, S. E. (2007). Does size matter for dispersal distance? *Global Ecology and Biogeography*, 16, 415–425. <https://doi.org/10.1111/j.1466-8238.2007.00312.x>

- Jordán, F., Magura, T., Tóthmérész, B., Vasas, V., & Kódöböcz, V. (2007). Carabids (Coleoptera: Carabidae) in a forest patchwork: A connectivity analysis of the Bereg Plain landscape graph. *Landscape Ecology*, 22, 1527–1539. <https://doi.org/10.1007/s10980-007-9149-8>
- Jordán, F., & Scheuring, I. (2004). Network ecology: Topological constraints on ecosystem dynamics. *Physics of Life Reviews*, 1, 139–172. <https://doi.org/10.1016/j.plrev.2004.08.001>
- Kauffman, S. A. (1993). *The origins of order: self-organization and selection in evolution*. Oxford, UK: Oxford University Press.
- Kéry, M., & Schaub, M. (2012). Estimation of occupancy and species distributions from detection/nondetection data in metapopulation designs using site-occupancy models. In M. Kéry, & M. Schaub (Eds.), *Bayesian population analysis using WinBUGS* (pp. 413–461). Boston, MA: Academic Press.
- Kool, J. T., Moilanen, A., & Treml, E. A. (2013). Population connectivity: Recent advances and new perspectives. *Landscape Ecology*, 28, 165–185. <https://doi.org/10.1007/s10980-012-9819-z>
- MacKenzie, D. I., Nichols, J. D., Lachman, G. B., Droege, S., Royle, J. A., & Langtimm, C. A. (2002). Estimating site occupancy rates when detection probabilities are less than one. *Ecology*, 83, 2248–2255. [https://doi.org/10.1890/0012-9658\(2002\)083\[2248:Esorw d\]2.0.Co;2](https://doi.org/10.1890/0012-9658(2002)083[2248:Esorw d]2.0.Co;2)
- MacKenzie, D. I., Nichols, J. D., Royle, J. A., Pollock, K. H., Bailey, L. L., & Hines, J. E. (2006). *Occupancy estimation and modeling: Inferring patterns and dynamics of species occurrence*. San Diego, CA: Elsevier.
- McRae, B. H. (2006). Isolation by resistance. *Evolution*, 60, 1551–1561. <https://doi.org/10.1111/j.0014-3820.2006.tb00500.x>
- Müller-Jentsch, D. (2012). *Wie dicht ist die Schweiz besiedelt?* Avenir Suisse. Retrieved from <https://www.avenir-suisse.ch/wie-dicht-ist-die-schweiz-besiedelt/>
- Naimi, B., Hamm, N. A., Groen, T. A., Skidmore, A. K., & Toxopeus, A. G. (2014). Where is positional uncertainty a problem for species distribution modelling? *Ecography*, 37, 191–203. <https://doi.org/10.1111/j.1600-0587.2013.00205.x>
- OFEV (2010). *Inventaire fédéral des bas-marais d'importance nationale*. Berne, Switzerland: OFEV.
- OFEV (2011). *Régions biogéographiques de Suisse*. Neuchâtel, Switzerland: OFEV.
- OFS (2010). *Statistique de la superficie selon nomenclature 2004 – Occupation du sol (Land Cover)*. Berne, Switzerland: OFS.
- Oliphant, T. E. (2006). *A guide to NumPy*. USA: Trelgol Publishing.
- Pellet, J., Hoehn, S., & Perrin, N. (2004). Multiscale determinants of tree frog (*Hyla arborea* L.) calling ponds in western Switzerland. *Biodiversity and Conservation*, 13, 2227–2235. <https://doi.org/10.1023/B:BIOC.0000047904.75245.1f>
- Prevedello, J. A., & Vieira, M. V. (2009). Does the type of matrix matter? A quantitative review of the evidence. *Biodiversity and Conservation*, 19, 1205–1223. <https://doi.org/10.1007/s10531-009-9750-z>
- R Development Core Team (2016). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Roth, U., Schwick, C., & Spichtig, F. (2010). *Zustand der Landschaft in der Schweiz. Zwischenbericht Landschaftsbeobachtung Schweiz (LABES)*. Berne, Switzerland: ARE.
- Saura, S., & Pascual-Hortal, L. (2007). A new habitat availability index to integrate connectivity in landscape conservation planning: Comparison with existing indices and application to a case study. *Landscape and Urban Planning*, 83, 91–103. <https://doi.org/10.1016/j.landurbplan.2007.03.005>
- Sinatra, R., Wang, D., Deville, P., Song, C., & Barabási, A.-L. (2016). Quantifying the evolution of individual scientific impact. *Science*, 354, aaf5239. <https://doi.org/10.1126/science.aaf5239>
- Smith, M. A., & Green, D. M. (2005). Dispersal and the metapopulation paradigm in amphibian ecology and conservation: Are all amphibian populations metapopulations? *Ecography*, 28, 110–128. <https://doi.org/10.1111/j.0906-7590.2005.04042.x>
- Straka, M., Paule, L., Ionescu, O., Štofik, J., & Adamec, M. (2012). Microsatellite diversity and structure of Carpathian brown bears (*Ursus arctos*): Consequences of human caused fragmentation. *Conservation Genetics*, 13, 153–164. <https://doi.org/10.1007/s10592-011-0271-4>
- Swisstopo (2016). *Catalogue des objets swissTLM3D 1.4*. Berne, Switzerland: Swisstopo.
- Swisstopo (2018). *swissALTI3D: The high precision digital elevation model of Switzerland*. Berne, Switzerland: Swisstopo.
- Taylor, P. D., Fahrig, L., Henein, K., & Merriam, G. (1993). Connectivity is a vital element of landscape structure. *Oikos*, 68, 571–573. <https://doi.org/10.2307/3544927>
- Thuiller, W., Georges, D., Engler, R., & Breiner, F. (2016). *biomod2: ensemble platform for species distribution modeling*. R package version 3.3-7. Retrieved from <https://CRAN.R-project.org/package=biomod2>
- Thuiller, W., & Münkemüller, T. (2010). Habitat suitability modeling. In A. P. Møller, W. Fiedler, & P. Berthold (Eds.), *Effects of climate change on birds* (pp. 77–85). New York, NY: Oxford University Press.
- Urban, D. L., Minor, E. S., Treml, E. A., & Schick, R. S. (2009). Graph models of habitat mosaics. *Ecology Letters*, 12, 260–273. <https://doi.org/10.1111/j.1461-0248.2008.01271.x>
- Van Buskirk, J. (2005). Local and landscape influence on amphibian occurrence and abundance. *Ecology*, 86, 1936–1947. <https://doi.org/10.1890/04-1237>
- Van Buskirk, J. (2011). Amphibian phenotypic variation along a gradient in canopy cover: Species differences and plasticity. *Oikos*, 120, 906–914. <https://doi.org/10.1111/j.1600-0706.2010.18845.x>
- van Langevelde, F., & Jaarsma, C. F. (2009). Modeling the effect of traffic calming on local animal population persistence. *Ecology and Society*, 14, 39. <https://doi.org/10.5751/ES-03061-140239>
- van Rossum, G. (1995). *Python tutorial, Technical Report CS-R9526*. Amsterdam, The Netherlands: Centrum voor Wiskunde en Informatica (CWI).
- Van Strien, M. J., & Grêt-Regamey, A. (2016). How is habitat connectivity affected by settlement and road network configurations? Results from simulating coupled habitat and human networks. *Ecological Modelling*, 342, 186–198. <https://doi.org/10.1016/j.ecolmodel.2016.09.025>
- van Strien, M. J., Keller, D., Holderegger, R., Ghazoul, J., Kienast, F., & Bolliger, J. (2014). Landscape genetics as a tool for conservation planning: Predicting the effects of landscape change on gene flow. *Ecological Applications*, 24, 327–339. <https://doi.org/10.1890/13-0442.1>
- Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of ‘small-world’ networks. *Nature*, 393, 440. <https://doi.org/10.1038/30918>
- Zanini, F., Pellet, J., & Schmidt, B. R. (2009). The transferability of distribution models across regions: An amphibian case study. *Diversity and Distributions*, 15, 469–480. <https://doi.org/10.1111/j.1472-4642.2008.00556.x>
- Ziółkowska, E., Ostapowicz, K., Radeloff, V. C., & Kuemmerle, T. (2014). Effects of different matrix representations and connectivity measures on habitat network assessments. *Landscape Ecology*, 29, 1551–1570. <https://doi.org/10.1007/s10980-014-0075-2>

**How to cite this article:** Ortiz-Rodríguez DO, Guisan A, Holderegger R, van Strien MJ. Predicting species occurrences with habitat network models. *Ecol Evol*. 2019;9:10457–10471. <https://doi.org/10.1002/ece3.5567>

**APPENDIX A1**

The 13 amphibian species of which presence records in the Swiss Plateau were used in this study

Species	Common name
<i>Hyla arborea</i>	European Tree Frog
<i>Alytes obstetricans</i>	Midwife Toad
<i>Bombina variegata</i>	Yellow-bellied Toad
<i>Bufo bufo</i>	Common Toad
<i>Epidalea calamita</i>	Natterjack Toad
<i>Ichthyosaura alpestris</i>	Alpine Newt
<i>Lissotriton helveticus</i>	Palmate Newt
<i>Pelophylax</i> sp. ( <i>P. lessonae</i> + <i>P. esculentus</i> )	Green Frog complex (Pool Frog and Edible Frog)
<i>Pelophylax ridibundus</i>	Lake Frog
<i>Rana dalmatina</i>	Agile Frog
<i>Rana temporaria</i>	Grass Frog
<i>Triturus carnifex</i>	Italian crested newt
<i>Triturus cristatus</i>	Northern crested Newt

**APPENDIX A2**

Final set of 20 predictor variables used in habitat suitability modeling (at a resolution of 1 ha), classified in three main categories

Category	Predictor	Variable type
Human influence	Density of traffic	Continuous
	Density of railways	Continuous
	Total noise at nighttime	Continuous
	Population density	Continuous
	Agriculture density	Continuous
	Arable land	Binary
	Green settlements	Binary
	Grey settlements	Binary
	Meadows and farm pastures	Binary
	Orchards, vineyards, horticulture	Binary
Natural landscape features	Deciduous forest coverage	Binary
	Mixed forest coverage	Binary
	Coniferous forest coverage	Binary
	Density of forest (general)	Continuous
	Distance to forest edge	Continuous
	Presence of rivers	Binary
	Slope	Continuous
Climatic variables	Mean summer precipitation	Continuous
	Mean annual direct solar radiation	Continuous
	Mean annual temperature	Continuous

**APPENDIX A3**

Predictor variables in habitat suitability modeling, their content and source

Predictor	Content	Source
Density of traffic	Individual vehicle traffic for 2010	NPVM with tunnels removed (ARE, 2010)
Density of railways	Density of rail network	SwissTLM3D (Swisstopo, 2016)
Total noise at nighttime	Nighttime rail noise combined with nighttime street noise	EMPA (2011)
Population density	Statistics on Swiss population, geolocated	STATPOP (BFS, 2015)
Agriculture density	Density of agricultural areas, derived from an aggregate of the four main categories of agricultural land use	Arealstatistik (OFS, 2010)
Arable land	Agricultural area taken from point estimates on 72 land use categories	Arealstatistik (OFS, 2010)
Green settlements	Area of green spaces in settlements taken from point estimates of 72 land use categories	Arealstatistik (OFS, 2010)
Grey settlements	Area of grey (sealed areas and buildings) areas taken from 72 land use categories	Arealstatistik (OFS, 2010)
Meadows and pastures	Area of meadows and pastures taken from point estimates of 72 land use categories	Arealstatistik (OFS, 2010)
Orchards, vineyards, horticulture	Area of orchards, vineyards and horticulture taken from point estimates of 72 land use	Arealstatistik (OFS, 2010)
Deciduous forest coverage	Occurrence of deciduous forests	Waldmischungsgrad (BFS, 2013)

Continues

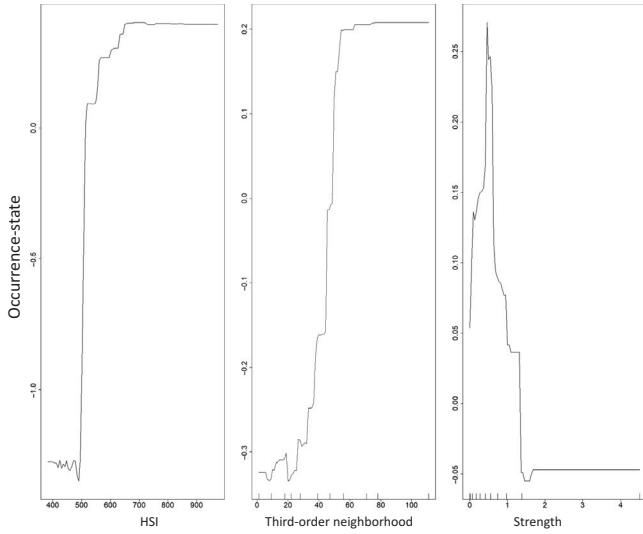
## APPENDIX A3 (Continued)

Predictor	Content	Source
Mixed forest coverage	Occurrence of mixed forests	Waldmischungsgrad (BFS, 2013)
Coniferous forest coverage	Occurrence of coniferous forests	Waldmischungsgrad (BFS, 2013)
Density of forest	Density of all forest types of Switzerland	Waldmischungsgrad (BFS, 2013)
Distance to forest edge	Distance to forest edges	Waldmischungsgrad (BFS, 2013)
Presence of rivers	Presence of rivers	SwissTLM3D (Swisstopo, 2016)
Slope	Calculated from a digital elevation model	swissALTI3D (Swisstopo, 2018)
Mean summer precipitation	Mean summer precipitation (1961–1990)	Broennimann, Randin, Zimmermann, and Guisan (2003)
Mean annual direct solar radiation	Mean annual direct solar radiation (1961–1990)	Broennimann et al. (2003)
Mean annual temperature	Mean annual temperature (1961–1990)	Broennimann et al. (2003)

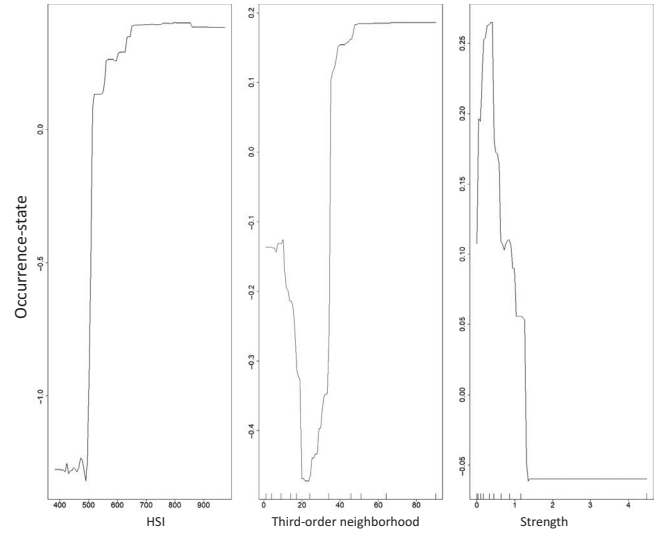
**APPENDIX A4**

Partial dependence plots of the three most important explanatory variables in sample iterations of the four models (Uniform, Traffic, HabSuit, noTopo). (a) Uniform; (b) Traffic; (c) HabSuit; (d) noTopo

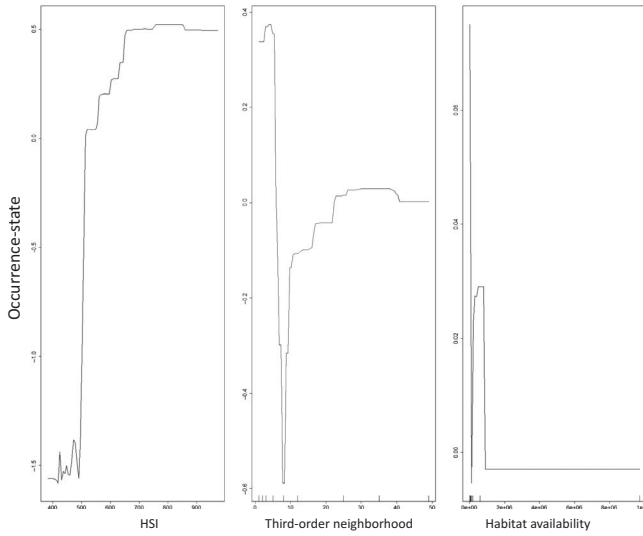
**(a) Uniform**



**(b) Traffic**



**(c) HabSuit**



**(d) noTopo**

