Article Type: Research paper

# Constructing a hybrid species distribution model from standard large-scale distribution data

Alexander Singer[a,1,*], Oliver Schweiger[b], Ingolf Kühn[b, c, d], Karin Johst[a]

[a]Department of Ecological Modelling, Helmholtz Centre for Environmental Research – UFZ, Permoserstr. 15, 04318 Leipzig, Germany

[b]Department of Community Ecology, Helmholtz Centre for Environmental Research – UFZ, Theodor-Lieser-Str.4, 06120 Halle, Germany

[c]Geobotany and Botanical Garden, Martin-Luther-University Halle-Wittenberg, Am Kirchtor 1, 06108 Halle, Germany

[d]German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Deutscher Platz 5e, Leipzig, Germany

[*]Correspondence address:
Dr. Alexander Singer
Swedish Species Information Centre
Swedish University of Agricultural Sciences
Box 7007, 75007 UppsalaSweden
Email: alexander.singer@slu.se

[1]Present address: Swedish Species Information Centre, Swedish University of Agricultural Sciences, Box 7007, 75007 Uppsala, Sweden

**Abstract**

Species range shifts under climate change have predominantly been projected by models correlating species observations with climatic conditions. However, geographic range shifting may depend on biotic factors such as demography, dispersal and species interactions. Recently suggested hybrid models include these factors. However, parameterization of hybrid models suffers from lack of detailed ecological data across many taxa. Further, it is methodologically unclear how to upscale ecological information from scales relevant to ecological processes to the coarser resolution of species distribution data (often 100km$^2$ or even 2500 km$^2$). We tackle these problems by developing a novel modelling and calibration framework, which allows hybrid model calibration from (static) presence-absence data that is available for many species. The framework improves understanding of the influence of biotic processes on range projections and reveals critical sources of uncertainty that limit projection reliability. We demonstrate its performance for the case of the butterfly Titania's Fritillary (*Boloria titania)*.

**Keywords**

Biotic interaction, colonization, extinction, range projection, process-based, dispersal

**Abbreviations**

C-SDM: correlative species distribution model

G-ECM: grid cell extinction-colonization model

H-SDM: hybrid species distribution model

## 1 Introduction

Projections of species distributions under changing environmental conditions are needed to support the conservation of biodiversity (Dawson et al., 2011; Pereira et al., 2010). Most models (such as bio-climatic envelope models, synonyms: ecological niche models, habitat models or species distribution models) statistically correlate species observations and environmental conditions (Elith and Leathwick, 2009; Guisan and Zimmermann, 2000; Yalcin and Leroux, 2017). Because of their correlative calibration approach, we refer to them as correlative species distribution models (C-SDMs). The correlative approach takes advantage of the type of widely available data, such as raster maps of species occurrence and environmental factors, usually derived from atlas data. The models, however, often for technical reasons (Singer et al., 2016), ignore key biotic mechanisms that affect species distributions (Urban et al., 2016).

Recent methodological advances suggest ways to enhance structural realism of species distribution models by incorporating biotic factors (Bocedi et al., 2014; Cabral et al., 2017; Evans et al., 2016; Kissling et al., 2012; Schurr et al., 2012; Talluto et al., 2016; Zurell, 2017) and have been shown to improve range projections (Zurell et al., 2016). One approach are hybrid species distribution models (H-SDMs; Dormann et al., 2012; also called niche population models Fordham et al., 2013). H-SDMs are based on C-SDMs that constitute filtering by the abiotic environment but add further relevant biotic processes or factors, such as interspecific interactions (Kissling et al., 2010; Meier et al., 2012; Schweiger et al., 2012), individual variability and local adaptation (Swab et al., 2015), dispersal or transport (Buse and Griebeler, 2011; Chapman et al., 2016; De Cáceres and Brotons, 2012; Kramer-Schadt et al., 2004), or demography (Keith et al., 2008). These studies used additional ecological

knowledge to parameterize the biotic processes, and showed differences in range projections compared to such based on abiotic environmental information only.

However, for many species, ecological process knowledge is lacking. Therefore, it would be desirable to parameterize H-SDMs from widely available data similarly to C-SDMs. Clearly such an endeavor has its limitations, because model projections cannot be more reliable than the information from which they are constructed. Missing knowledge on model and process structure can compromise projection accuracy, while missing knowledge to calibrate the model parameters results in projection uncertainty (Singer et al., 2016). Nevertheless, we will show that, hybrid models can improve species range projections based on generally available biogeographic information such as presence-absence data.

In this study, we develop a hybrid modelling framework consisting of six steps that allows including and calibrating demographic processes and dispersal from currently available maps on species occupancy and environmental variables. To demonstrate its merits the framework was exemplarily applied to project the range of the holarctic butterfly Titania's Fritillary (*Boloria titania*), almost monophagously feeding as larvae on Adderwort (*Bistorta officinalis*). We analyzed the calibration framework with respect to its ability to shrink parameter and structural uncertainty, and studied the consequences of remaining uncertainty for projection reliability.

## 2 Material and methods

We developed a generic framework (Figure 1) to construct H-SDMs from currently available data (left boxes in Figure 1). The type of data that is generally available for species distribution modelling comprises variables on abiotic environmental conditions (third left box in Figure 1) and maps of observed species occupancy patterns (fourth left box in Figure 1). However, information on biotic processes is often limited to general theory (first left box in

Figure 1), information on interspecific dependencies (second left box in Figure 1), or expert knowledge (fifth left box in Figure 1).

In sec. 2.1 we describe major challenges arising from data limitations, and suggest solutions. In sec. 2.2, we operationalize these solutions in a protocol for the H-SDM model construction and calibration.

## 2.1 Major challenges and our solutions

**Challenge 1: Lack of information**

We identify two types of knowledge gaps that can limit projection reliability of species distributions: structural gaps and data gaps.

Structural gaps, e.g. concerning the relevant environmental drivers or the formulation of ecological response processes to these drivers, compromise the structure of the model resulting in structural uncertainty. Data gaps or uncertainty of data measurements compromise model parameterization resulting in parameter uncertainty. Both structural and parameter uncertainty reduce projection reliability (Singer et al., 2016).

**Solution 1: Knowledge gap analysis**

Identifying knowledge gaps prior to implementing the model helps addressing them adequately by (1) filling the gaps with complementary information (e.g. from targeted empirical studies; Bierman et al., 2010; Manceur and Kühn, 2014), by (2) assessing their impact on projection reliability or by (3) projecting metrics that are robust in spite of missing knowledge (Gould et al., 2014). Focusing here on models with a defined endpoint, i.e. projection of spatiotemporal species occurrence, we are mainly concerned with the second aspect.

Consequences of structural gaps can be assessed by implementing alternative model formulations (as suggested in pattern oriented modelling; Grimm and Railsback, 2011). To shrink the number of alternative models, model formulations can be weighted or even rejected, based on their plausibility. Plausibility criteria might comprise statistical parsimony, ecological theory, expert knowledge (Gallien et al., 2010) or post hoc sensitivity analyses (Saltelli et al., 2000). Consequences of data gaps can be assessed by parameter uncertainty analyses (Saltelli et al., 2000). Finally, the joint impact of structural and data gaps on projection reliability can be estimated in ensemble projections of the alternative models and parameterizations.

**Challenge 2: Coarse spatial resolution impedes detailed description of ecological processes**

The coarse spatial resolution of available species distribution data exceeds the fine spatial resolution of ecological processes (Soberon and Nakamura, 2009). Therefore, distribution data contains only up-scaled information about spatial ecological processes and biotic responses to the environment.

**Solution 2a: Grid cell – extinction colonization model (G-ECM) aggregates local population dynamics and dispersal**

We follow suggestions of applying a metapopulation approach to implement ecological dynamics in species distribution models (Talluto et al., 2017; Thuiller et al., 2013). For metapopulations at landscape scale, local dynamics within patches are aggregated and patch-occupancy patterns reflect small-scale ecological processes (Etienne et al., 2004; Grimm et al., 2004; Hanski, 1994). This metapopulation approach has been transferred to larger spatial scales of geographic grid cells (e.g. Buse and Griebeler, 2011; De Cáceres and Brotons, 2012). We implement the approach as a grid cell – extinction colonization model (G-ECM;

see section 2.2.4) that tracks species occupancy dynamics at geographic grid cells. However, we point out that, in contrast to metapopulations (Hanski, 1992; ter Braak et al., 1998), scaling rules of ecological processes are unclear at the biogeographic scale (Barwell et al., 2014; Henle et al., 2014; Pearson et al., 2004). Therefore, a strict ecological interpretation of the extinction and colonization parameters is not possible at the large geographic grid scale.

**Solution 2b: Interspecific dependency as a geographic filter for species distributions**

Small-scale interspecific interactions can affect species distributions at large scales, and prior knowledge on interspecific interactions can refine projections of species distributions (Kissling et al., 2012; Singer et al., 2016; Wisz et al., 2013). Particularly, if one species depends on other host species to complete its lifecycle, the distribution of the dependent species is constrained by presence of the host species. Consequently, we suggest limiting the distribution of a species by the distribution of its host species, i.e. applying the spatial distribution of host species as a geographic filter (Schweiger et al., 2012, 2008). This filtering approach is suitable for predominantly unidirectional interactions (called here interspecific dependency; e.g., commensalism, mutualism or parasitism with weak feedbacks). A recently suggested advancement (Staniczenko et al., 2017) based on Bayesian networks promises to account for more complex interspecific interactions (e.g., competition, facilitation or predator-prey interactions). However, caution should be taken concerning interspecific interactions with strong feedbacks because feedbacks can lead to complex effects on range-dynamics (Singer et al., 2016, 2013).

**Challenge 3: Lack of temporal information**

Temporal information is essential to calibrate dynamics of species distributions (De Cáceres and Brotons, 2012). In extinction-colonization models, parameters should be estimated from turnover data or a set of occupancy maps taken at different points in time (Etienne et al.,

2004; Moilanen, 2004, 1999; O'Hara et al., 2002; ter Braak and Etienne, 2003). Recent methods to calibrate dynamic species distribution models apply hierarchical likelihood-based techniques (Cabral and Schurr, 2010; Evans et al., 2016; Marion et al., 2012; Pagel and Schurr, 2012; Talluto et al., 2017). These methods require detail in spatiotemporal occupancy (Talluto et al., 2017) or abundance time series (Pagel and Schurr, 2012). However, in atlas data, temporally resolved information is usually pooled to gain spatial coverage. The current lack of time series data across large extents impedes grid cell turnover estimates for most species (but see Schurr et al., 2012 for suggestions to improve the situation).

**Solution 3a: Sequential H-SDM calibration**

We use an H-SDM (consisting of a C-SDM coupled with a G-ECM) and parameterize C-SDM and G-ECM in subsequent steps, which is a common approach (Dormann et al., 2012). However, sequential parameter estimation can introduce bias in model calibration. Particularly, in the initial calibration step, the C-SDM might confound patterns from biological processes for effects from abiotic filters (Schurr et al., 2012). As a consequence, the G-ECM calibration in the second step might not be based on the full biotic information, or both calibration steps might pick up the biotic information, resulting in double-accounting for the biotic impact (Gallien et al., 2010). Nevertheless, this assumption is widely used in H-SDM approaches (e.g. Holloway et al., 2016; Keith et al., 2008; Meier et al., 2012; Regan et al., 2012). To assess and cope with potential bias in parameter estimates, detailed analysis of the dynamic model projections have to be performed.

**Solution 3b: Equilibrium calibration followed by simulations of the stochastic spatiotemporal dynamics**

We resort to calibrating the G-ECM assuming the species is in equilibrium and its equilibrium distribution is represented by its distribution map. The equilibrium assumption has been

successfully applied to calibrate dynamic metapopulation models from single patch occupancy patterns (pioneered by Hanski, 1994) and to calibrate a demographic plant species distribution model from a single abundance pattern (Cabral and Schurr, 2010). Yet, we point out that calibration of a dynamic model from a single snapshot can cause biased parameter estimates (Moilanen, 2000). This may happen if the species distribution is not in equilibrium, such as for invasive or currently range changing species. Therefore, after equilibrium calibration we proceed with simulations of the explicit stochastic occupancy dynamics of the H-SDM.

### 2.2 Protocol of the generic modelling and calibration framework

These challenges and our solutions are the basis for the six steps of our generic hybrid modelling framework (Figure 1, right): (1) identifying knowledge gaps and accounting for them in a set of plausible alternative model scenarios, (2) accounting for interspecific dependency, (3) estimating the abiotic species niche through environmental filtering in a C-SDM, (4) modelling the ecological dynamics based on a G-ECM (taking into account both interspecific dependency - step 2 and abiotic conditions - step 3) and calibrating the parameters assuming the system is in equilibrium, (5) simulating the stochastic spatiotemporal dynamics of occupancy of the H-SDM for different calibrated parameter combinations to further confine parameter ranges, and (6) repeating the previous steps for the alternative model scenarios identified in step 1. With each step, we aim at extracting additional information on the species' environmental response from the species spatial occupancy pattern. These steps are now explained in detail and can be followed in Figure 1.

[INSERT FIGURE 1 AROUND HERE]

### 2.2.1 Step 1: Establishing alternative calibration scenarios

From an analysis of knowledge gaps, we suggest constructing "model scenarios" that describe alternative model formulations in accordance with the available knowledge about the system. For example, if lack of knowledge on a species dispersal pattern impedes resolving the dispersal mechanism, different likely dispersal kernels should be considered in alternative model scenarios. If several knowledge gaps exist, the study design should account for interactions among sources of uncertainties (e.g. in a full-factorial design).

### 2.2.2 Step 2: Accounting for interspecific dependency

Known interspecific dependencies of the species are taken into account to limit its spatial extent. In all modeling steps only grid cells are accessible on which host species are present.

### 2.2.3 Step 3: Modelling abiotic influence

C-SDMs are used to estimate the suitability of grid cells. A large variety of modelling algorithms exists ranging from more traditional regression-based methods, such as Generalized Linear Models or Generalized Additive Models, to more advanced machine-learning methods, like Boosted Regression Trees or MAXENT, with different advantages and shortcomings (e.g. Elith et al., 2006). The choice of the algorithm usually depends on the requirements like interpolation ability or model transferability (e.g. Heikkinen et al., 2012) and the reliability of absence data, e.g. presence-only, presence-background or presence-absence methods. Within our framework we need reliable absence data (see below) and thus respective methods should be used. In all modeling approaches we suggest to consider interactions among environmental variables as well as non-linear responses of species to environmental variables. We strongly recommend to avoid overfitting by proper variable selection or regularization procedures. The output of presence-absence methods usually are occurrence probabilities which depend on the prevalence of the species. For the subsequent

G-ECM calibration (step 4), resulting occurrence probabilities should be centered at a prevalence of 0.5. Therefore, we suggest weighting absences to ensure a prevalence of 0.5 (Maggini et al., 2006).

We interpret the projected probability of species occurrence as the suitability of abiotic environmental conditions for the species, which we call grid cell suitability $H_i(t)$ of grid cell $i$ at time $t$. $H_i(t)$ covers the spatial environmental heterogeneity among grid cells and has the potential to consider environmental change over time.

With grid cell suitability $H_i(t)$ we link abiotic information with ecological processes (step 4), following previous approaches (Buse and Griebeler, 2011; De Cáceres and Brotons, 2012; Meier et al., 2010; Swab et al., 2015):

### 2.2.4 Step 4: Modelling ecological dynamics and equilibrium calibration

To track ecological dynamics, we implement a grid cell-extinction colonization model (G-ECM) at the spatial scale of geographic grid cells. In the G-ECM, the occupancy of a grid cell can change from two stochastic population dynamical processes. These are species extinction from a previously occupied cell or colonization of a previously empty cell. Both population dynamical processes are applied in each modelled time step and act simultaneously. They are defined as follows:

**Extinction:** A species in grid cell $i$ goes extinct at time $t$ with extinction probability

$$E_i(t) = \min\left(1, e \cdot \left(1 - H_i(t)\right)\right) \quad (1)$$

where $H_i(t)$ denotes the cell suitability estimated in the C-SDM (step 3) for all accessible cells (step 2). Eq. 1 assumes a linear decline of $E_i(t)$ with increasing grid cell suitability $H_i(t)$. Model parameter $e$ describes the ability of the species to cope with local habitat conditions and scales with time step length and grid cell size.

**Colonization:** An empty accessible (step 2) cell *i* is colonized by a species in time step *t* with probability

$$C_i(t) = 1 - (1 - c)^{I_i(t)} \quad (2)$$

where *c* is a parameter that indicates an individual´s ability to establish. $I_i(t)$ denotes the number of immigrants arriving at the empty cell *i*. We assume that immigrants can originate only from grid cells that are occupied in the previous time step *t-1*. Therefore,

$$I_i(t) = M \cdot \sum_{j \neq i} g(d_{ij}) \cdot J_j(t-1) \quad (3)$$

where *M* is the number of emigrants per grid cell, $J_j(t-1)$ is occupancy of cell *j* at time step *t-1*. $g(d_{ij})$ denotes a dispersal kernel (depending on distance $d_{ij}$ between grid cells *i* and *j*). Explicit modelling of dispersal is rare in species distribution modelling (Holloway and Miller, 2017), although the inclusion of probabilistic dispersal kernels $g(d_{ij})$ can improve projections of species distributions (Holloway et al., 2016).

**Equilibrium calibration**: We follow the incidence function approach (Hanski, 1994) that is technically suitable to calibrate the model parameters from the equilibrium state of grid cell occupancy $\bar{J}_i$ (Etienne et al., 2004):

$$\bar{J}_i = C_i/(C_i + E_i) \quad (4)$$

where we assume constancy of grid cell occupancy $J_i$ during time steps *t* and *t-1*, which is a reasonable approximation for the quasi-stationary equilibrium state (Hanski 1999). Inserting equations (1 - 3) into (4) we derive

$$\bar{J}_i = \frac{1-(1-c)^{M \cdot \sum_{j \neq i} g(d_{ij}) \cdot J_j}}{1-(1-c)^{M \cdot \sum_{j \neq i} g(d_{ij}) \cdot J_j + \min(1, e \cdot (1 - H_i(t)))}} \quad (5)$$

Eq. 5 links grid cell suitability as well as demographic and dispersal parameters to equilibrium grid cell occupancy $\bar{J}_i$. Parameter values can be estimated from eq. 5 by minimizing the negative log likelihood

279 $\bar{L}(P, O) = -\sum_{i=all\ occupied\ cells\ in\ O} \ln\left(\overline{J_i}(P)\right) + \sum_{i=all\ empty\ cells\ in\ O} \ln\left(1 - \overline{J_i}(P)\right)$ (6)

280 for the observed occupancy *O* and parameter set *P*.

281 As mentioned above, our modelling framework assumes reliable presence-absence data. False
282 absences could affect G-ECM calibration in several ways (see Moilanen, 2002 in the context
283 of metapopulation model calibration). False absences of high quality grid cells might increase
284 estimates of grid cell extinction rate. Additionally, distances among occupied cells would
285 appear larger due to missing occupied cells, which might result in over-estimation of dispersal
286 distances and colonization success.

287 The calibration can be supported by potentially available ecological information (e.g. expert
288 or anecdotal knowledge on species dispersal or local extinction risk, Figure 1 bottom left).
289 This often coarse information can indicate reasonable parameter ranges and filter out
290 unrealistic parameter estimates.

291 **Selection of parameter sets:** We suggest a hierarchical latin hypercube design to estimate
292 likely parameter values from eq. 6 applying rejection sampling. Latin hypercube rejection
293 sampling (LHS) is commonly applied for the calibration of complex ecological models
294 (Hartig et al., 2011; Jakoby et al., 2014). The structure of eq. 5 allows splitting the parameter
295 space in two lower dimensional parameter spaces: one for the colonization and extinction
296 related parameters *c* and *e*, the other containing dispersal kernel parameters. This splitting
297 reduces computation time, because estimation of complex dispersal kernels for long distances
298 can be computationally costly. However, the splitting destroys homogeneity of the LHS. To
299 avoid under-sampled areas of the parameter space, the sample size in each of the split samples
300 has to be high.

301 In tests, we found that G-ECM calibration can suffer from equifinality (i.e. different
302 parameter combinations might equally likely fit the data). Equifinality leads to uncertainty

about the best suitable parameterization. To account for this uncertainty, we suggest selecting several 'best' parameter sets that result in high negative log likelihood values (eq. 6). The amount of best parameter sets should be adjusted according to flatness of the likelihood function and available computational power for the subsequent model simulations.

#### 2.2.5 Step 5: Simulating stochastic spatial occupancy dynamics with the H-SDM

To simulate the stochastic spatial dynamics of species, extinction from and colonization of grid cells are implemented as Bernoulli random processes with probabilities according to eq. 1 and 2 and parameterized with the 'best' parameter sets resulting from the equilibrium calibration procedure (step 4). Model simulations are performed for each parameter set separately. Each simulation starts from the observed distribution data and is run until a dynamic equilibrium is reached. The stochastic simulation is replicated to project the stochastic distribution of model outcomes. To evaluate projection reliability, model outcomes are compared to the observed distribution.

#### 2.2.6 Step 6: Analyzing alternative model scenarios

For each model scenario (step 1), interspecific dependences are applied as geographic filters (step2), the respective model formulation is calibrated (step 3 and 4) and simulated (step 5) independently. Subsequently, projections from the alternative model scenarios can be aggregated to reflect projection uncertainty. They can also be analysed comparatively to identify how each of the alternative model scenarios contributes to projection uncertainty.

## 3 Case study

We aimed to analyse the distribution of Titania's fritillary.

## 3.1 Available data

### 3.1.1 Species geographic distribution

Presence and absence data of the host plant *Bistorta officinalis* DELARBRE (Adderwort) was taken from the database on Atlas Florae Europaeae (AFE - Jalas and Suominen, 1979), compiled by the AFE secretariat at the Finish Museum of Natural History (dark gray dots in Figure 2A). Distribution data for *Boloria titania* ESPER (Titania's Fritilary) was taken from a database which constituted also the basis for the 'Distribution Atlas of Butterflies in Europe' (Kudrna et al., 2011) (red crosses in Figure 2A). In order to run the model for both species at the same spatial resolution, butterfly distribution data from about 7000 georeferenced localities were aggregated to the 50 km x 50 km CGRS grid used by AFE. For the butterfly and its host plant, the distribution data can be assumed to provide a good representation of true presences and absences, given the large spatial resolution, the aggregation of data from several decades, and a reasonably large sampling effort.

### 3.1.2 Environmental variables

We used monthly interpolated climate data (Fronzek et al., 2012), originally provided via the ALARM project (Settele et al., 2005) at a 10 arcmin grid resolution and aggregated it to the CGRS grid used by AFE. In accordance with Settele et al. (2008) we used aggregated climate variables: mean annual accumulated growing degree days with a base temperature of 5°C until August, range of annual temperature (°C), range of annual precipitation (mm) and soil water content for the upper horizon (0.5 m). Soil water content was taken from the dynamic vegetation model LPJ-GUESS (Hickler et al., 2009, 2004) and represented a process-based water balance in terrestrial systems. We used averaged values for the period 1971-2000 for the climate data to match the time span used for butterfly occurrence data.

### 3.2 Application of the generic modelling and calibration framework

The modelling and calibration framework was implemented in statistical language R.

#### 3.2.1 Step 1: Establishing alternative model scenarios

We accounted for uncertainty in model structure and data, considering four sources of uncertainty. For each of these sources of uncertainty, we assumed two alternative hypotheses leading to a full-factorial design of 16 alternative models.

1) **Distributional data on Titania's Fritillary**

It can be expected that agricultural land-use and corresponding habitat loss constrains the range of the Alpine Titania's Fritillary population and excludes butterflies from lower (particularly Northern) Alpine regions. Therefore, the observed occurrence might not reflect the butterfly's abiotic and dispersal limited niche. To account for related uncertainties, we considered two occurrence scenarios:

**Occupancy observed (OO):** The available distributional data

**Occupancy land-use corrected (OL):** Available distributional data plus grid cells at lower altitudes in the Alps that would have been climatically suitable according to the C-SDM (yellow area in Figure 2A).

2) **Butterfly dispersal kernel**

Species specific information on the butterfly's dispersal behavior was not available. To evaluate potential impact of rare long distance dispersal (Chesson and Lee, 2005; Hastings et al., 2004), which can significantly affect projected species distributions (Holloway et al., 2016), we compared two alternative dispersal kernels:

**Negative exponential kernel (exp):**

$$g^{exp}(d_{ij}) = \frac{A_{cell}\exp(-\alpha d_{ij})}{\frac{2\pi}{\alpha^2}\cdot((1+\alpha d_{min})\cdot\exp(-\alpha d_{min}) - (1+\alpha d_{max})\cdot\exp(-\alpha d_{max}))} \quad (7)$$

**Powerlaw kernel (pow):**

$$g^{pow}(d_{ij}) = \frac{A_{cell} {d_{ij}}^{-x}}{\frac{2\pi}{(2-x)} \cdot ({d_{max}}^{2-x} - {d_{min}}^{2-x})} \quad (8)$$

where $d_{ij}$ denoted the centre to centre distance, $A_{cell}$ meant the cell area of 50km x 50km = 2500km$^2$, minimum and maximum distances of cell centers $d_{min}$ = 33km (was smaller than 50km due to few smaller cells that corrected for the planar CGRS grid projection), $d_{max}$ = 4509km. Calibration parameters $x$ and $\alpha$ were related to dispersal distance. The continuous dispersal kernels were adapted to the grid structure by an approximate normalization for total area (Chipperfield et al., 2011).

**3) Geographic extent for the equilibrium calibration procedure**

The European Alps, the Carpathian mountains, and to lower extent the Baltic states make up the main distributional range of Titania's Fritillary (red crosses in Figure 2A). Consequently, at European scale, most grid cells are empty. Grid cells might be unoccupied because they are (i) climatically unsuitable or (ii) unreachable due to dispersal limitations. Both reasons should be distinguished during calibrations (Soberon and Nakamura, 2009). Finally, the butterfly also occurs in Russian and Belorussian areas for which occupancy maps are not available. To test the influence of the spatial extent on model calibration and projection we considered the two scenarios:

**Extent all (EA):** the entire area for which data was available (including an observed population in the Baltics).

**Extent Central Europe (EC):** a smaller area around the currently observed range in the Alps (area enclosed by blue line in Figure 2A). EC excludes the Baltic population. In this scenario the number of observed presences and absences is roughly equal.

**4) Alternative host plant scenarios**

Distribution of the butterfly's obligate host plant *B. officinalis* strongly determines, which grid cells are accessible for the butterfly (step 2). However, the impact of ecological processes on the distribution of *B. officinalis* is largely unknown.

Therefore, prior to modelling the butterfly we modelled and projected distributions of its obligate host plant, following the same methodology (Supplementary material Appendix A). We considered two alternative host plant projections, which differed in dispersal kernels (similar to the butterfly kernels)

**P1**: negative exponential host plant dispersal kernel

**P2**: power law host plant dispersal kernel

to account for uncertainty in long-distance plant dispersal (see also Supplementary material Appendix A for further details)

### 3.2.2 Step 2: Accounting for interspecific dependency

To account for the butterfly's obligate host plant dependence, the butterfly C-SDM (sec.3.2.3) and equilibrium calibration of the G-ECM (sec. 3.2.4) were restricted to grid cells where its obligate host plant was present in the observed data (Schweiger et al., 2008). Similarly, in the H-SDM projections (sec. 3.2.5), butterflies could only colonize and survive in grid cells where the host plant had been projected to be present by the host plant H-SDM (Supplementary Material Appendix A).

### 3.2.3 Step 3: Modelling abiotic influence (C-SDM)

As one of the modelling approaches combining both high prediction accuracy and transferability (Heikkinen et al., 2012), we used boosted regression trees for calibrating the C-SDM. We assumed a binomial error structure and used a logit link function. Boosted regression trees were constructed with a relatively slow learning rate of 0.005, to obtain

417 optimal model fits (Elith et al., 2008). We allowed up to three-way interactions among
418 climate variables. To avoid overfitting, we identified the appropriate number of trees
419 contributing to the final model by analyzing 10-fold cross-validated predictive deviance (Elith
420 et al., 2008). We also weighted absences to ensure a prevalence of 0.5 (see Maggini et al.,
421 2006).

422 **3.2.4 Step 4: Modelling ecological dynamics and equilibrium calibration**

423 We calibrated parameters *e* (eq. 1), *c* and *M* (eq. 2-3) in the G-ECM as well as either *x* or $\alpha$
424 depending on the dispersal kernel scenario (eq.7 or 8) using eq. (5 and 6) as described in sec.
425 2.2.4 (using function "improvedLHS" from R- package "lhs"). For the hierarchical LHS, we
426 first generated 400 parameter combinations for the dispersal parameters *M* and *x* or $\alpha$
427 respectively. We secondly generated 400 parameter combinations of *e* and *c*. The total sample
428 contained 160000 parameter sets from the full factorial combination of the two samples.
429 Initial parameter ranges are shown in Figure 3. We selected the ten 'best' parameter sets that
430 resulted in the ten lowest negative log likelihood values to exemplify the variability of the
431 outcome due to parameter correlation.

432 **3.2.5 Step 5: Simulating stochastic spatial occupancy dynamics**

433 We performed stochastic dynamic projections with the H-SDM according to sec. 2.2.5 for the
434 ten best parameter sets. Each simulation ran for 1000 time steps to ensure that the model
435 reached equilibrium conditions. For each parameter set, we repeated the stochastic simulation
436 100 times. Note, because *this* study focused on testing performance of the calibration method
437 only, we did not project the model under climate change conditions (i.e. $H_i(t)$ is constant over
438 time in our case study).

#### 3.2.6 Step 6: Analyzing alternative model scenarios

The first three sources of uncertainties (see step 1) resulted in eight model scenarios that affected butterfly calibration, resulting in 8 x 10 suitable parameter sets. Projection models were then parameterized with each of these parameter sets and additionally the two alternative dynamic host plant projections (fourth source of uncertainty, see step 1), which resulted in 160 differently parameterized models in total. Each of these stochastic models were repeated 100 times, which summed to in total 16000 simulation runs.

### 3.3 Statistical analysis

#### 3.3.1 Model performance

We evaluated discriminative model performance (Lawson et al., 2014) by calculating the area under the receiver operating characteristic curve (AUC - R-package ROCR) for the C-SDM calibration (step 3), the equilibrium calibration procedure using the G-ECM (step 4) and the dynamic occupancy projections of the H-SDM (step 5). Since we do not translate the resulting occurrence probabilities of a particular step into presence-absence data but rather use them as direct input for the subsequent steps, we rely on AUC as a threshold-independent measure of model performance. In addition to AUC, we also considered probabilistic versions of accuracy and sensitivity (Bennett et al., 2013; Lawson et al., 2014) if appropriate and visually compared spatial model projections to observed geographic occupancy data. For the C-SDM, AUC calculations were based on 10-fold cross validation for the data restricted to host plant presence. For projection results from the G-ECM equilibrium calibration, AUC values were calculated separately for each selected parameter set. For the H-SDM, we estimated the probability of grid cell occupancy for each of the modeled calibration scenarios at the end of the simulation, after it had reached equilibrium from 100 replicates, before calculating the respective AUC values. For the G-ECM equilibrium calibration and the dynamic occupancy

projections with the H-SDM, standard procedures for cross-validation are not applicable since random exclusion of grid cells would disturb the structure of spatial grid cell connectivity, and potentially impact model calibration (Moilanen, 2002).

### 3.3.2 Parameter sets from the equilibrium calibration procedure

We analysed correlations among model scenarios and selected parameters. Parameter correlations were linearly decomposed by PCA and associated to model scenarios (R-package vegan; see also Borcard et al., 2008). From inspection of scatter plots among parameter values (see Supplementary material Appendix B, Fig. B.1), we derived a non-linear combination that accounted for relations among model parameters:

$$z = \frac{M \cdot c}{e} \qquad (9)$$

The aggregated parameter $z$ can be interpreted as turn-over.

### 3.3.3 Parameter sets from the stochastic dynamic occupancy projections

From the stochastic dynamic H-SDM projections (Figure 1, step 5), we calculated the quasi-stationary means and standard deviations of the number of occupied grid cells and compared them with the observed number of occupied cells. These statistics aggregated the 100 repetitions, the spatial extent of projections and the last 50 time steps of the simulation. By assuming quasi-stationarity, we reflected the current situation, where indeed the species exists and is observed (Kudrna et al., 2011). Quasi-stationarity was achieved by excluding parameter combinations from the analysis where the butterfly population went extinct during simulation runs. This procedure eliminated some extreme and ecologically unrealistic parameter combinations. Spatially resolved probabilities of grid cell occupancy were calculated as means over the projected presences and absences at the end of the simulation from all replicates of a modelling scenario.

## 4 Results

[INSERT FIGURE 2 AROUND HERE]

### 4.1 Calibration of the C-SDM

AUC of the C-SDM was 0.90. The C-SDM projection (step 3) covered the current range of the butterfly, however, it further indicated climatically suitable conditions in other mountainous areas and in Scandinavia (Figure 2B), which are unoccupied. The potential Scandinavian range is currently not suitable due to the lack of the host plant (see gray dots in Figure 2A and Schweiger et al., 2008).

### 4.2 Equilibrium calibration procedure

When parameterized with the selected suitable parameter sets of the equilibrium calibration procedure (step 4), the G-ECM equilibrium solution (eq. 5) closely reproduced the observed occupancy pattern (for an example see Figure 2C). For each of the selected parameter sets, probabilistic accuracy (i.e. the average probability to correctly project butterfly occupancy on a grid cell where host plants occur) ranged between 0.93 and 0.94. In contrast, sensitivity (the average probability to correctly project only butterfly presence) ranged between 0.68 and 0.81. This indicated that the model better projected absences than presences. The AUC values for all 80 selected parameter sets (10 best sets for 8 butterfly-related modelling scenarios) were 0.96. High probabilities of butterfly occupancy could be expected only around the butterfly's current range (Figure 2C).

[INSERT FIGURE 3 AROUND HERE]

In all model scenarios, initial ranges for the parameters were reduced (Figure 3). Calibrated ranges for $M$, $e$ and $c$ were rather similar whereas calibrated ranges of dispersal distance

related parameters $x$ and $\alpha$ varied among the scenarios. This variation could not be attributed to one specific calibration scenario (Figure 3).

The equilibrium calibration procedure delivered different parameter sets that performed equally well (Table 1 and Figure 3). This is not surprising as eq. (4) and (5), respectively, can generate equal mean cell occupancy with different but correlated extinction and colonization parameters.

[INSERT TABLE 1 AROUND HERE]

Variation of suitable parameter values could be explained as follows. Firstly, we identified a dynamic and a static solution in the equally well performing parameter sets (Fig. 3, Supplementary material Appendix B, Fig. B1). The dynamic solution was characterized by turnover in grid cell occupancy (i.e. non-zero values of $e$ and $c$). In the static solution, turn-over was suppressed (i.e. nearly zero values of $e$ and $c$). As the static solution only reproduces the C-SDM, these parameter sets were excluded from further analysis.

[INSERT FIGURE 4 AROUND HERE]

Secondly, a principal component analysis (Figure 4) that reduced dimensionality of the parameter space revealed the correlation among the parameters (Table 1). The first two axes explained more than 80% of variance. The first axis was related to colonization parameters $M$, $c$ and $\alpha$ or $x$, the second axis to the extinction parameter $e$. The aggregated variable $z$ (eq. 8) subsumed the correlations. Using the aggregated parameter $z$ in the PCA, parameter sets

533 arranged along the *z*-direction, but clustered (parallel shift) according to specific α values
534 (Figure 4C, D).
535

536 **4.3 Stochastic dynamic occupancy projections**

537 Stochastic dynamic occupancy projections of the H-SDM (step 5) using the best parameter
538 sets of the equilibrium calibration procedure (i.e., the 10 best sets of step 4 for each of the 16
539 model scenarios) performed worse than projections of the equilibrium calibration procedure.
540 Mean of all AUC values calculated for each of the 160 projected parameter sets was 0.81
541 (standard deviation 0.15) and increased to 0.84 (±0.12), if we used only the single best
542 parameter set of each projection scenario.
543

544 Interestingly, in the dynamic stochastic occupancy projections, the butterfly exceeded the
545 observed range and occupied regions further north to the Alps (Figure 2D). Moreover, the
546 dynamic projections suggested only a very low occupancy in the Baltics in contrast to the
547 observations (compare crosses in the Baltics (North-East) in Figure 2A and that are not
548 reflected in Figure 2D)
549

550 [INSERT FIGURE 5 AROUND HERE]
551

552 The projected numbers of occupied grid cells varied among sets of selected parameters, and
553 scattered around the observed occupancy (Figure 5). The aggregated parameter *z* was a good
554 predictor for projected occupancy (Figure 5I, J). Butterfly occupancy to a minor degree also
555 depended on the underlying host plant projection, where we found that projection P1 caused
556 butterfly occupancy to be slightly lower than projection P2.

557 Most importantly, for some parameter combinations ($z$, $\alpha$ or $x$ and host plant projection), the
558 model closely projected the observed butterfly occupancy (see Figure 5).

559 **5 Discussion**

560 We suggest a generic framework to calibrate hybrid species distribution models (H-SDMs)
561 from maps of species occupancy and variables describing environmental conditions (sec. 2.2).
562 In contrast to other H-SDMs (reviewed in Fordham et al., 2013), our framework requires only
563 weak (or incomplete) independent information on biotic processes and factors.
564 Tackling the problem of lacking mechanistic biotic information about the species spatial
565 population dynamics, model construction was confronted with three major challenges
566 concerning (i) spatial and (ii) temporal scales as well as (iii) process detail (which in fact can
567 encompass several independent processes). Our suggested solutions enable the construction of
568 model-based projections of species distribution dynamics, even if critical dynamic
569 information is lacking. The lack of knowledge compromises the resolution of biotic processes
570 in the model, and therefore the reliability of the model projections. However, due to the
571 included model scenario analysis (steps 1 and 6) and the detailed dynamic analysis of
572 different parameter sets (step 5) the framework is able to identify parameter sets that are in
573 accordance with the available data and to estimate their impact on model projections. Thus, it
574 provides crucial insight in the impact of biotic factors on the spatial dynamics of species and
575 can reveal critical sources of uncertainty. In the following we discuss the merits and
576 limitations of the framework on the example of the case study.

577 **5.1 Performance of the modelling framework**

578 The suitability model (C-SDM, step 3) reproduced the current distribution of the butterfly in
579 the European Alps, the Carpathian Mountains and in the Baltics (Figure 2B) and corroborated

580 results from Schweiger et al. (2008) that environmentally suitable habitat also exists far
581 beyond the presently observed distribution, which however is inaccessible due to host plant
582 absence (Figure 2A). Projections of the equilibrium solution (step 4) performed better because
583 they revealed dispersal limitation, which restricted butterfly projections to areas close to the
584 observed butterfly range (Figure 2C). The stochastic dynamic projections of the H-SDM (step
585 5) indicated a potential for colonization of mountainous regions adjacent to the Alps, where
586 the butterfly currently is not present (Figure 2D). However, they also showed absence of the
587 butterfly in the Baltic area where it is presently observed. We hypothesize that these
588 discrepancies to the observed occupancy pattern could be caused by parameter uncertainty
589 (e.g. from unresolved correlations), by the impact of land-use excluding the butterfly from
590 potentially suitable areas, by a secondary host plant or that the Baltic *Boloria titania*
591 population might be a sink in the species colonization-extinction dynamics. These points are
592 discussed in the following.

593 **5.2 Parameter uncertainty and correlation**

594 There are two sources of parameter uncertainty in the calibration: (i) alternative states and (ii)
595 equifinality, as explained below. They might also be the reason, why we could not identify a
596 strong impact of calibration scenarios on the parameter values (Figure 4).
597 We identified two alternative solutions (see sec. 4.2): a static solution without colonization-
598 extinction dynamics (corresponding rates are nearly zero) and a dynamic solution with
599 colonization-extinction dynamics (and corresponding nonzero rates).
600 The static solution represented the C-SDM by suppressing dynamics. The existence of the
601 dynamic solution indicates that the abiotic niche described by the C-SDM could not explain
602 the entire distribution of the butterfly and thus points to the importance of biotic processes for
603 this distribution.

The dynamic solution could not be uniquely resolved by model calibration. Instead we found alternative parameter sets (Figure 3) that equally well explained the observed butterfly occupancy pattern (equifinality). Equifinality (Beven and Freer, 2001) indicates over-parameterization (Dormann et al., 2012) and is the result of insufficient information content in the data to parameterize the biotic processes.

In this study, one possible reason for a reduction in information content could have been the sequential model calibration. If the calibrated C-SDM had incorrectly attributed biotic information to environmental factors (e.g. due to covariation), we had to assume reduced explanatory power of the latter calibrated dynamic H-SDM. Therefore, we particularly aimed at avoiding over-fitting while constructing the C-SDM (sec. 3.2.3). We assume that the C-SDM calibration had not strongly reduced information content for the latter H-SDM equilibrium calibration.

Instead, we consider correlation among the dynamic parameters as the main calibration problem. The derived parameter $z$ (eq. 9) reflects this correlation as it can be interpreted as a descriptor of grid cell turnover (i.e. the ratio between rates of local colonization of empty grid cells and extinction from occupied grid cells). In the equilibrium calibration procedure (step 4), this ratio cannot be resolved further, because equal numbers of grid cells had to be colonized and vacated in order to keep occupancy constant. Single snapshot occupancy data provide too limited information to resolve correlated biotic processes (Gu and Swihart, 2003). Therefore, the subsequent stochastic dynamic occupancy projections (step 5) are essential to gauge consequences of alternative suitable parameterization for model projections.

### 5.3 Dispersal ability

The stochastic dynamic occupancy projections produced variations in the average grid cell occupancy (Figure 5) even if the turn-over parameter $z$ was constant. This variance arose from

628 insufficient knowledge about dispersal in combination with two simplifications of the
629 equilibrium calibration procedure compared to the full dynamics of the H-SDM. The
630 equilibrium calibration, firstly calibrated colonization probability of empty grid cells from the
631 species dispersal ability during a single time step (eq. 5). Dispersal in several steps was
632 ignored. Secondly, being based on the first moment approximation (eq. 4), the equilibrium
633 calibration procedure could not account for variability from immigration-extinction
634 stochasticity (Hanski, 1994). In contrast, the dynamic projections allowed for subsequent
635 random colonization events that were not immediately compensated by extinctions. Thus,
636 variance in projected occurrence that could not be resolved by parameter $z$ can be attributed to
637 uncertainty about dispersal abilities of the butterfly including its stochasticity. Our analysis
638 therefore identifies dispersal as a critical source of uncertainty in our case study.

639 **5.4 Projected range expansion around the European Alps: Does land use limit the**
640 **butterfly range?**

641 The stochastic dynamic occupancy projections of the H-SDM led to an extended range
642 beyond the currently observed butterfly range in the European Alps (compare Figure 2A to
643 D). This challenges our assumption that currently the butterfly population is in equilibrium.
644 However, given the long-term and intensive monitoring effort, we trust the observed absences
645 of *Boloria titania* and believe that the population indeed is in equilibrium.
646 More likely, land use might have limited the observed species range but was not explicitly
647 considered in the model. We found that taking into account land use impact in the equilibrium
648 calibration procedure affected the dispersal related parameters ($\alpha$ and $x$; Figure 3). The
649 slightly expanded geographic range in mountainous regions, predicted by the dynamic
650 projections (Figure 2D) compared to the equilibrium calibration procedure (Figure 2C), likely
651 indicates the butterfly's potential geographic range without impact of anthropogenic land use.

652 Thus, our analysis identifies land use as a potentially important factor for the distribution of
653 *Boloria titania*, which agrees with findings for high latitude butterflies (Eskildsen et al.,
654 2013).

655 **5.5 The Baltic *Boloria titania* is potentially a sink population**

656 The stochastic dynamic occupancy projections provided new important insight in the
657 existence of the Baltic butterfly population. We found that the butterfly population in the
658 Baltic States was projected very likely to go extinct. This result conflicts with the observed
659 presence of Titania's Fritillary in the Baltic States. There are two likely explanations for this
660 discrepancy. Firstly, the Baltic butterfly population is potentially oligophagous and might
661 utilize *Viola* species as a secondary host plant (see supplementary information in Pöyry et al.,
662 2008 for the observed feeding behaviour in Finland and Northern Europe). Since we do not
663 have reliable distribution data for *Viola* species nor detailed information about the possible
664 density-dependent benefits of a secondary host plant, we could not consider such effects in
665 the model. A secondary host plant might enhance the suitable butterfly area.
666 Secondly, the Baltic population might be connected to larger populations in Russia (Kudrna et
667 al., 2011). For these populations, reliable occupancy data is not available, which restricts the
668 spatial extent of our study. However, the impact of the spatial extent on parameter estimates
669 (including or not the Baltic population and the Eastern border - see Figure 3) was low because
670 only grid cells at the Eastern border of the modeled area should be affected (Moilanen, 2002)**.**
671 Given the fact that ignoring the Eastern butterfly range in our dynamic occupancy projections,
672 the Baltic population went extinct, we hypothesize that the Baltic population could be a sink
673 population, connected to a source east of the modeled area (Moilanen, 2002 on biased
674 occupancy projections due to disconnection of patches). Source-sink dynamics have been

675 considered one potential source of bias to static species distribution models but can be
676 revealed with demographic approaches (Pagel and Schurr, 2012).

677 **6 Conclusions**

678 We present a hybrid modelling and calibration framework to project species distributions,
679 taking into account demographic processes and dispersal. The framework is distinct in its
680 attempt to be calibrated from widely available data on geographic distributions of species and
681 environmental factors. Relying on rather low data requirements, the framework can
682 potentially be applied to a wide range of species. However, the small information base also
683 reveals knowledge gaps that impact model projection reliability.
684 A problematic knowledge gap is the lack of data to inform spatial population dynamics (e.g.
685 time series of species distributions or species specific functional trait information).
686 Particularly, the available information is not sufficient to resolve correlation among
687 colonization and extinction processes. This causes uncertainty in the species dispersal ability
688 and subsequently in the range projections, even assuming temporally constant environmental
689 conditions. Under environmental change, we expect uncertainties to even increase.
690 To cope with the lack of knowledge, we suggest within the framework extensive and detailed
691 analyses of the data and structural gaps (step 1) and of the results of the modelling and
692 calibration steps 2-6 (Figure 1). In particular, step 5 (the stochastic dynamic projections of the
693 H-SDM) is essential as it selects the dynamic biotic processes (species extinction and
694 colonization rates including dispersal distances) that most likely reflect the data. The model
695 scenario analysis (step 6) is important as it allows evaluating projection reliability. Model
696 scenarios that provide biased projections of current occupancy should not be ignored, but
697 thoroughly investigated as they provide insights in confounding factors and achievable
698 projection reliability.

699 Executing the steps of the framework narrows the parameter range and reveals critical
700 knowledge gaps that can compromise projection reliability (e.g. the lack of data to inform the
701 population dynamics or dispersal processes). Further, it can explain mechanisms that drive the
702 propagation of uncertainties. With the presently available data, the framework cannot be
703 expected to provide highly reliable quantitative projections of species distributions. Instead, it
704 can enhance mechanistic understanding of the species range dynamics, estimate the reliability
705 of species range projections, and reveal, which additional data would improve projections.
706 This is often the best possible achievement to support management facing data limitations
707 (Singer et al., 2011). Very importantly, this framework provides potential agendas for field
708 related research to improve and tailor the collection of biotic parameters.

709 **Declaration of interest**

710 none.

711 **Acknowledgements**

712 IK acknowledges support of the project EU BON (Building the European Biodiversity
713 Observation Network), funded by the European Union under the 7th Framework programme
714 [Contract No. 308454].

715 **References**

716 Barwell, L.J., Azaele, S., Kunin, W.E., Isaac, N.J.B., 2014. Can coarse-grain patterns in insect
717 atlas data predict local occupancy? Divers. Distrib. 20, 895–907. doi:10.1111/ddi.12203
718 Bennett, N.D., Croke, B.F.W., Guariso, G., Guillaume, J.H.A., Hamilton, S.H., Jakeman, A.J.,
719 Marsili-Libelli, S., Newham, L.T.H., Norton, J.P., Perrin, C., Pierce, S.A., Robson, B.,
720 Seppelt, R., Voinov, A.A., Fath, B.D., Andreassian, V., 2013. Characterising

721 performance of environmental models. Environ. Model. Softw. 40, 1–20.
722 doi:10.1016/j.envsoft.2012.09.011
723 Beven, K., Freer, J., 2001. Equifinality, data assimilation, and uncertainty estimation in
724 mechanistic modelling of complex environmental systems using the GLUE
725 methodology. J. Hydrol. 249, 11–29. doi:10.1016/S0022-1694(01)00421-8
726 Bierman, S.M., Butler, A., Marion, G., Kühn, I., 2010. Bayesian image restoration models for
727 combining expert knowledge on recording activity with species distribution data.
728 Ecography 33, 451–460. doi:10.1111/j.1600-0587.2009.05798.x
729 Bocedi, G., Palmer, S.C.F., Pe'er, G., Heikkinen, R.K., Matsinos, Y.G., Watts, K., Travis,
730 J.M.J., 2014. RangeShifter: A platform for modelling spatial eco-evolutionary dynamics
731 and species' responses to environmental changes. Methods Ecol. Evol. 5, 388–396.
732 doi:10.1111/2041-210X.12162
733 Borcard, D., Gillet, F., Legendre, P., 2008. Numerical Ecology with R, Applied Spatial Data
734 Analysis with R. Springer, New York. doi:10.1007/978-0-387-78171-6
735 Buse, J., Griebeler, E.M., 2011. Incorporating classified dispersal assumptions in predictive
736 distribution models - A case study with grasshoppers and bush-crickets. Ecol. Modell.
737 222, 2130–2141. doi:10.1016/j.ecolmodel.2011.04.010
738 Cabral, J.S., Schurr, F.M., 2010. Estimating demographic models for the range dynamics of
739 plant species. Glob. Ecol. Biogeogr. 19, 85–97.
740 Cabral, J.S., Valente, L., Hartig, F., 2017. Mechanistic simulation models in macroecology
741 and biogeography: state-of-art and prospects. Ecography 40, 267–280.
742 doi:10.1111/ecog.02480
743 Chapman, D.S., Makra, L., Albertini, R., Bonini, M., Páldy, A., Rodinkova, V., Šikoparija,
744 B., Weryszko-Chmielewska, E., Bullock, J.M., 2016. Modelling the introduction and

745 spread of non-native species: international trade and climate change drive ragweed
746 invasion. Glob. Chang. Biol. 22, 3067–3079. doi:10.1111/gcb.13220
747 Chesson, P., Lee, C.T., 2005. Families of discrete kernels for modeling dispersal. Theor.
748 Popul. Biol. 67, 241–256.
749 Chipperfield, J.D., Holland, E.P., Dytham, C., Thomas, C.D., Hovestadt, T., 2011. On the
750 approximation of continuous dispersal kernels in discrete-space models. Methods Ecol.
751 Evol. 2, 668–681. doi:10.1111/j.2041-210X.2011.00117.x
752 Dawson, T.P., Jackson, S.T., House, J.I., Prentice, I.C., Mace, G.M., 2011. Beyond
753 predictions: biodiversity conservation in a changing climate. Science 332, 53–58.
754 doi:10.1126/science.1200303
755 De Cáceres, M., Brotons, L., 2012. Calibration of hybrid species distribution models: The
756 value of general-purpose vs. targeted monitoring data. Divers. Distrib. 18, 977–989.
757 doi:10.1111/j.1472-4642.2012.00899.x
758 Dormann, C.F., Schymanski, S.J., Cabral, J., Chuine, I., Graham, C., Hartig, F., Kearney, M.,
759 Morin, X., Römermann, C., Schröder, B., Singer, A., 2012. Correlation and process in
760 species distribution models: Bridging a dichotomy. J. Biogeogr. 39, 2119–2131.
761 doi:10.1111/j.1365-2699.2011.02659.x
762 Elith, J., Graham, C.H., Anderson, R.P., Dudik, M., Ferrier, S., Guisan, A., Hijmans, R.J.,
763 Huettmann, F., Leathwick, J.R., Lehmann, A., Li, J., Lohmann, L.G., Loiselle, B.A.,
764 Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., Overton, J.M., Peterson, A.T.,
765 Phillips, S.J., Richardson, K., Scachetti-Pereira, R., Schapire, R.E., Soberon, J.,
766 Williams, S., Wisz, M.S., Zimmermann, N.E., 2006. Novel methods improve prediction
767 of species' distributions from occurrence data. Ecography 29, 129–151.
768 doi:10.1111/j.2006.0906-7590.04596.x

769 Elith, J., Leathwick, J.R., 2009. Species Distribution Models: Ecological Explanation and
770 Prediction Across Space and Time. Annu. Rev. Ecol. Evol. Syst. 40, 677–697.
771 doi:10.1146/annurev.ecolsys.110308.120159
772 Elith, J., Leathwick, J.R., Hastie, T., 2008. A working guide to boosted regression trees. J.
773 Anim. Ecol. 77, 802–813.
774 Eskildsen, A., Roux, P.C., Heikkinen, R.K., Høye, T.T., Kissling, W.D., Pöyry, J., Wisz,
775 M.S., Luoto, M., 2013. Testing species distribution models across space and time : high
776 latitude butterflies and recent warming. Glob. Ecol. Biogeogr. 22, 1293–1303.
777 doi:10.1111/geb.12078
778 Etienne, R.S., ter Braak, C.J.F., Vos, C.C., 2004. Application of Stochastic Patch Occupancy
779 Models to Real Metapopulations, in: Hanski, I., Gaggiotti, O.E. (Eds.), Ecology,
780 Genetics and Evolution of Metapopulations. Academic Press, Amsterdam, pp. 105–132.
781 doi:10.1016/B978-012323448-3/50007-6
782 Evans, M.E.K., Merow, C., Record, S., McMahon, S.M., Enquist, B.J., 2016. Towards
783 Process-based Range Modeling of Many Species. Trends Ecol. Evol. 31, 860–871.
784 doi:10.1016/j.tree.2016.08.005
785 Fordham, D.A., Akçakaya, H.R., Araújo, M.B., Keith, D.A., Brook, B.W., 2013. Tools for
786 integrating range change, extinction risk and climate change information into
787 conservation management. Ecography 36, 956–964. doi:10.1111/j.1600-
788 0587.2013.00147.x
789 Fronzek, S., Carter, T.R., Jylhä, K., 2012. Representing two centuries of past and future
790 climate for assessing risks to biodiversity in Europe. Glob. Ecol. Biogeogr. 21, 19–35.
791 doi:10.1111/j.1466-8238.2011.00695.x
792 Gallien, L., Münkemüller, T., Albert, C.H., Boulangeat, I., Thuiller, W., 2010. Predicting

793 potential distributions of invasive species: where to go from here? Divers. Distrib. 16,
794 331–342. doi:10.1111/j.1472-4642.2010.00652.x
795 Gould, S.F., Beeton, N.J., Harris, R.M.B., Hutchinson, M.F., Lechner, A.M., Porfirio, L.L.,
796 Mackey, B.G., 2014. A tool for simulating and communicating uncertainty when
797 modelling species distributions under future climates. Ecol. Evol. 4, 4798–4811.
798 doi:10.1002/ece3.1319
799 Grimm, V., Lorek, H., Finke, J., Koester, F., Malachinski, M., Sonnenschein, M., Moilanen,
800 A., Storch, I., Singer, A., Wissel, C., Frank, K., 2004. META-X: Generic Software for
801 Metapopulation Viability Analysis. Biodivers. Conserv. 13, 165–188.
802 doi:10.1023/B:BIOC.0000004317.42949.f7
803 Grimm, V., Railsback, S.F., 2011. Pattern-oriented modelling: a "multi-scope" for predictive
804 systems ecology. Philos. Trans. R. Soc. B Biol. Sci. 367, 298–310.
805 doi:10.1098/rstb.2011.0180
806 Gu, W., Swihart, R.K., 2003. Are patch occupancy data sufficient for inferring
807 metapopulation dynamics using spatially explicit patch occupancy models? Acta Zool.
808 Sin. 49, 787–794.
809 Guisan, A., Zimmermann, N.E., 2000. Predictive habitat distribution models in ecology. Ecol.
810 Modell. 135, 147–186. doi:10.1016/S0304-3800(00)00354-9
811 Hanski, I., 1994. A practical model of metapopulation dynamics. J. Anim. Ecol. 63, 151–162.
812 doi:10.2307/5591
813 Hanski, I., 1992. Inferences from Ecological Incidence Functions. Am. Nat. 139, 657–662.
814 Hartig, F., Calabrese, J.M., Reineking, B., Wiegand, T., Huth, A., 2011. Statistical inference
815 for stochastic simulation models - theory and application. Ecol. Lett. 14, 816–827.
816 doi:10.1111/j.1461-0248.2011.01640.x

817 Hastings, A., Cuddington, K., Davies, K.F., Dugaw, C.J., Elmendorf, S., Freestone, A.,
818 Harrison, S., Holland, M., Lambrinos, J., Malvadkar, U., Melbourne, B.A., Moore, K.,
819 Taylor, C., Thomson, D., 2004. The spatial spread of invasions: new developments in
820 theory and evidence. Ecol. Lett. 8, 91–101. doi:10.1111/j.1461-0248.2004.00687.x
821 Heikkinen, R.K., Marmion, M., Luoto, M., 2012. Does the interpolation accuracy of species
822 distribution models come at the expense of transferability? Ecography 35, 276–288.
823 doi:10.1111/j.1600-0587.2011.06999.x
824 Henle, K., Potts, S., Kunin, W., Matsinos, Y., Simila, J., Pantis, J., Grobelnik, V., Penev, L.,
825 Settele, J. (Eds.), 2014. Scaling in Ecology and Biodiversity Conservation. Pensoft
826 Publishers, Sofia. doi:10.3897/ab.e1169
827 Hickler, T., Fronzek, S., Araújo, M.B., Schweiger, O., Thuiller, W., Sykes, M., 2009. An
828 ecosystem-model-based estimate of changes in water availability differs from water
829 proxies that are commonly used in species distribution models. Glob. Ecol. Biogeogr. 18,
830 304–313.
831 Hickler, T., Smith, B., Sykes, M.T., Davis, M.B., Sugita, S., Walker, K., 2004. Using a
832 generalized vegetation model to simulate vegetation dynamics in northeastern USA.
833 Ecology 85, 519–530.
834 Holloway, P., Miller, J.A., 2017. A quantitative synthesis of the movement concepts used
835 within species distribution modelling. Ecol. Modell. 356, 91–103.
836 doi:10.1016/j.ecolmodel.2017.04.005
837 Holloway, P., Miller, J.A., Gillings, S., 2016. Incorporating movement in species distribution
838 models: how do simulations of dispersal affect the accuracy and uncertainty of
839 projections? Int. J. Geogr. Inf. Sci. 30, 2050–2074. doi:10.1080/13658816.2016.1158823
840 Jakoby, O., Grimm, V., Frank, K., 2014. Pattern-oriented parameterization of general models

for ecological application: Towards realistic evaluations of management approaches. Ecol. Modell. 275, 78–88. doi:10.1016/j.ecolmodel.2013.12.009

Jalas, J., Suominen, J. (Eds.), 1979. Polygonaceae, in: Atlas Florae Europaeae. Distribution of Vascular Plants in Europe. The Committee for Mapping the Flora of Europe & Societas Biologica Fennica Vanamo, Helsinki, p. 71 pp. [maps 384–pp. 478].

Keith, D.A., Akçakaya, H.R., Thuiller, W., Midgley, G.F., Pearson, R.G., Phillips, S.J., Regan, H.M., Araújo, M.B., Rebelo, T.G., 2008. Predicting extinction risks under climate change: coupling stochastic population models with dynamic bioclimatic habitat models. Biol. Lett. 4, 560–563. doi:10.1098/rsbl.2008.0049

Kissling, W.D., Dormann, C.F., Groeneveld, J., Hickler, T., Kühn, I., Mcinerny, G.J., Montoya, J.M., Römermann, C., Schiffers, K., Schurr, F.M., Singer, A., Svenning, J.C., Zimmermann, N.E., O'Hara, R.B., 2012. Towards novel approaches to modelling biotic interactions in multispecies assemblages at large spatial extents. J. Biogeogr. 39, 2163–2178. doi:10.1111/j.1365-2699.2011.02663.x

Kissling, W.D., Field, R., Korntheuer, H., Heyder, U., Böhning-Gaese, K., 2010. Woody plants and the prediction of climate-change impacts on bird diversity. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 365, 2035–2045. doi:10.1098/rstb.2010.0008

Kramer-Schadt, S., Revilla, E., Wiegand, T., Breitenmoser, U., 2004. Fragmented landscapes, road mortality and patch connectivity: modelling influences on the dispersal of Eurasian lynx. J. Appl. Ecol. 41, 711–723.

Kudrna, O., Harpke, A., Lux, K., Pennerstorfer, J., Schweiger, O., Settele, J., Wiemers, M., 2011. Distribution atlas of butterflies in Europe. Gesellschaft für Schmetterlingsschutz, Halle.

Lawson, C.R., Hodgson, J.A., Wilson, R.J., Richards, S.A., 2014. Prevalence, thresholds and

865 the performance of presence-absence models. Methods Ecol. Evol. 5, 54–64.
866 doi:10.1111/2041-210X.12123
867 Maggini, R., Lehmann, A., Zimmermann, N.E., Guisan, A., 2006. Improving generalized
868 regression analysis for the spatial prediction of forest communities. J. Biogeogr. 33,
869 1729–1749.
870 Manceur, A.M., Kühn, I., 2014. Inferring model-based probability of occurrence from
871 preferentially sampled data with uncertain absences using expert knowledge. Methods
872 Ecol. Evol. 5, 739–750. doi:10.1111/2041-210X.12224
873 Marion, G., McInerny, G.J., Pagel, J., Catterall, S., Cook, A.R., Hartig, F., O'Hara, R.B.,
874 2012. Parameter and uncertainty estimation for process-oriented population and
875 distribution models: data, statistics and the niche. J. Biogeogr. 39, 2225–2239.
876 doi:10.1111/j.1365-2699.2012.02772.x
877 Meier, E.S., Kienast, F., Pearman, P.B., Svenning, J.C., Thuiller, W., Araújo, M.B., Guisan,
878 A., Zimmermann, N.E., 2010. Biotic and abiotic variables show little redundancy in
879 explaining tree species distributions. Ecography 33, 1038–1048. doi:10.1111/j.1600-
880 0587.2010.06229.x
881 Meier, E.S., Lischke, H., Schmatz, D.R., Zimmermann, N.E., 2012. Climate, competition and
882 connectivity affect future migration and ranges of European trees. Glob. Ecol. Biogeogr.
883 21, 164–178. doi:10.1111/j.1466-8238.2011.00669.x
884 Moilanen, A., 2004. SPOMSIM: Software for stochastic patch occupancy models of
885 metapopulation dynamics. Ecol. Modell. 179, 533–550.
886 doi:10.1016/j.ecolmodel.2004.04.019
887 Moilanen, A., 2002. Implications of empirical data quality to metapopulation model
888 parameter estimation and application. Oikos 96, 516–530. doi:10.1034/j.1600-

889 0706.2002.960313.x
890 Moilanen, A., 2000. The equilibrium assumption in estimating the parameters of
891 metapopulation models. J. Anim. Ecol. 69, 143–153. doi:10.1046/j.1365-
892 2656.2000.00381.x
893 Moilanen, A., 1999. Patch occupancy models of metapopulation dynamics: Efficient
894 parameter estimation using implicit statistical inference. Ecology 80, 1031–1043.
895 doi:10.1890/0012-9658(1999)080[1031:POMOMD]2.0.CO;2
896 O'Hara, R.B., Arjas, E., Toivonen, H., Hanski, I., 2002. Bayesian Analysis of Metapopulation
897 data. Ecology 83, 2408–2415. doi:10.1890/0012-
898 9658(2002)083[2408:BAOMD]2.0.CO;2
899 Pagel, J., Schurr, F.M., 2012. Forecasting species ranges by statistical estimation of ecological
900 niches and spatial population dynamics. Glob. Ecol. Biogeogr. 21, 293–304.
901 doi:10.1111/j.1466-8238.2011.00663.x
902 Pearson, R.G., Dawson, T.P., Liu, C., 2004. Modelling species distributions in Britain: A
903 hierarchical integration of climate and land-cover data. Ecography 27, 285–298.
904 doi:10.1111/j.0906-7590.2004.03740.x
905 Pereira, H.M., Leadley, P.W., Proença, V., Alkemade, R., Scharlemann, J.P.W., Fernandez-
906 Manjarrés, J.F., Araújo, M.B., Balvanera, P., Biggs, R., Cheung, W.W.L., Chini, L.,
907 Cooper, H.D., Gilman, E.L., Guénette, S., Hurtt, G.C., Huntington, H.P., Mace, G.M.,
908 Oberdorff, T., Revenga, C., Rodrigues, P., Scholes, R.J., Sumaila, U.R., Walpole, M.,
909 2010. Scenarios for global biodiversity in the 21st century. Science 330, 1496–1501.
910 doi:10.1126/science.1196624
911 Pöyry, J., Luoto, M., Heikkinen, R.K., Saarinen, K., 2008. Species traits are associated with
912 the quality of bioclimatic models. Glob. Ecol. Biogeogr. 17, 403–414.

913 doi:10.1111/j.1466-8238.2007.00373.x
914 Regan, H.M., Syphard, A.D., Franklin, J., Swab, R.M., Markovchick, L., Flint, A.L., Flint,
915 L.E., Zedler, P.H., 2012. Evaluation of assisted colonization strategies under global
916 change for a rare, fire-dependent plant. Glob. Chang. Biol. 18, 936–947.
917 doi:10.1111/j.1365-2486.2011.02586.x
918 Saltelli, A., Chan, K., Scott, E.M. (Eds.), 2000. Sensitivity Analysis. Wiley, New York.
919 Schurr, F.M., Pagel, J., Cabral, J.S., Groeneveld, J., Bykova, O., O'Hara, R.B., Hartig, F.,
920 Kissling, W.D., Linder, H.P., Midgley, G.F., Schröder, B., Singer, A., Zimmermann,
921 N.E., 2012. How to understand species' niches and range dynamics: a demographic
922 research agenda for biogeography. J. Biogeogr. 39, 2146–2162. doi:10.1111/j.1365-
923 2699.2012.02737.x
924 Schweiger, O., Heikkinen, R.K., Harpke, A., Hickler, T., Klotz, S., Kudrna, O., Kühn, I.,
925 Pöyry, J., Settele, J., 2012. Increasing range mismatching of interacting species under
926 global change is related to their ecological characteristics. Glob. Ecol. Biogeogr. 21, 88–
927 99. doi:10.1111/j.1466-8238.2010.00607.x
928 Schweiger, O., Settele, J., Kudrna, O., Klotz, S., Kühn, I., 2008. Climate change can cause
929 spatial mismatch of trophically interacting species. Ecology 89, 3472–3479.
930 doi:10.1890/07-1748.1
931 Settele, J., Hammen, V., Hulme, P., Al., E., 2005. ALARM: Assessing LArge-scale
932 environmental Risks for biodiversity with tested Methods. Gaia-Ecological Perspect. Sci.
933 Soc. 14, 69–72.
934 Settele, J., Kudrna, O., Harpke, A., Kühn, I., van Swaay, C., Verovnik, R., Warren, M.,
935 Wiemers, M., Hanspach, J., Hickler, T., Kühn, E., van Halder, I., Veling, K.,
936 Vliegenthart, A., Wynhoff, I., Schweiger, O., 2008. Climatic Risk Atlas of European

937 Butterflies. BioRisk 1, 1–710.
938 Singer, A., Johst, K., Banitz, T., Fowler, M.S., Groeneveld, J., Gutiérrez, A.G., Hartig, F.,
939 Krug, R.M., Liess, M., Matlack, G., Meyer, K.M., Pe'er, G., Radchuk, V., Voinopol-
940 Sassu, A.-J., Travis, J.M.J., 2016. Community dynamics under environmental change:
941 How can next generation mechanistic models improve projections of species
942 distributions? Ecol. Modell. 326, 63–74. doi:10.1016/j.ecolmodel.2015.11.007
943 Singer, A., Salman, M., Thulke, H.H., 2011. Reviewing model application to support animal
944 health decision making. Prev. Vet. Med. 99, 60–67.
945 doi:10.1016/j.prevetmed.2011.01.004
946 Singer, A., Travis, J.M.J., Johst, K., 2013. Interspecific interactions affect species and
947 community responses to climate shifts. Oikos 122, 358–366. doi:10.1111/j.1600-
948 0706.2012.20465.x
949 Soberon, J., Nakamura, M., 2009. Niches and distributional areas: Concepts, methods, and
950 assumptions. Proc. Natl. Acad. Sci. 106, 19644–19650. doi:10.1073/pnas.0901637106
951 Staniczenko, P.P.A., Sivasubramaniam, P., Suttle, K.B., Pearson, R.G., 2017. Linking
952 macroecology and community ecology: refining predictions of species distributions
953 using biotic interaction networks. Ecol. Lett. 20, 693–707. doi:10.1111/ele.12770
954 Swab, R.M., Regan, H.M., Matthies, D., Becker, U., Bruun, H.H., 2015. The role of
955 demography, intra-species variation, and species distribution models in species'
956 projections under climate change. Ecography 38, 221–230. doi:10.1111/ecog.00585
957 Talluto, M. V., Boulangeat, I., Ameztegui, A., Aubin, I., Berteaux, D., Butler, A., Doyon, F.,
958 Drever, C.R., Fortin, M.J., Franceschini, T., Liénard, J., Mckenney, D., Solarik, K.A.,
959 Strigul, N., Thuiller, W., Gravel, D., 2016. Cross-scale integration of knowledge for
960 predicting species ranges: A metamodelling framework. Glob. Ecol. Biogeogr. 25, 238–

961 249. doi:10.1111/geb.12395
962 Talluto, M. V., Boulangeat, I., Vissault, S., Thuiller, W., Gravel, D., 2017. Extinction debt
963 and colonization credit delay range shifts of eastern North American trees. Nat. Ecol.
964 Evol. 1, 182. doi:10.1038/s41559-017-0182
965 ter Braak, C., Hanski, I., Verboom, J., 1998. The incidence function approach to modeling of
966 metapopulation dynamics, in: Modeling Spatiotemporal Dynamics in Ecology. Springer-
967 Verlag, Berlin, Germany, pp. 167–188.
968 ter Braak, C.J.F., Etienne, R.S., 2003. Improved Bayesian Analysis of Metapopulation Data
969 with an Application to a Tree Frog Metapopulation. Ecology 84, 231–241.
970 doi:10.1890/0012-9658(2003)084[0231:IBAOMD]2.0.CO;2
971 Thuiller, W., Münkemüller, T., Lavergne, S., Mouillot, D., Mouquet, N., Schiffers, K.,
972 Gravel, D., 2013. A road map for integrating eco-evolutionary processes into
973 biodiversity models. Ecol. Lett. 16 Suppl 1, 94–105. doi:10.1111/ele.12104
974 Urban, M.C., Bocedi, G., Hendry, A.P., Mihoub, J.-B., Peer, G., Singer, A., Bridle, J.R.,
975 Crozier, L.G., De Meester, L., Godsoe, W., Gonzalez, A., Hellmann, J.J., Holt, R.D.,
976 Huth, A., Johst, K., Krug, C.B., Leadley, P.W., Palmer, S.C.F., Pantel, J.H., Schmitz, A.,
977 Zollner, P.A., Travis, J.M.J., 2016. Improving the forecast for biodiversity under climate
978 change. Science 353, aad8466. doi:10.1126/science.aad8466
979 Wisz, M.S., Pottier, J., Kissling, W.D., Pellissier, L., Lenoir, J., Damgaard, C.F., Dormann,
980 C.F., Forchhammer, M.C., Grytnes, J.-A., Guisan, A., Heikkinen, R.K., Høye, T.T.,
981 Kühn, I., Luoto, M., Maiorano, L., Nilsson, M.-C., Normand, S., Öckinger, E., Schmidt,
982 N.M., Termansen, M., Timmermann, A., Wardle, D.A., Aastrup, P., Svenning, J.-C.,
983 2013. The role of biotic interactions in shaping distributions and realised assemblages of
984 species: implications for species distribution modelling. Biol. Rev. 88, 15–30.

985 doi:10.1111/j.1469-185X.2012.00235.x
986 Yalcin, S., Leroux, S.J., 2017. Diversity and suitability of existing methods and metrics for
987 quantifying species range shifts. Glob. Ecol. Biogeogr. 26, 609–624.
988 doi:10.1111/geb.12579
989 Zurell, D., 2017. Integrating demography, dispersal and interspecific interactions into bird
990 distribution models. J. Avian Biol. 1–12. doi:10.1111/jav.01225
991 Zurell, D., Thuiller, W., Pagel, J., Cabral, J.S., Münkemüller, T., Gravel, D., Dullinger, S.,
992 Normand, S., Schiffers, K.H., Moore, K.A., Zimmermann, N.E., 2016. Benchmarking
993 novel approaches for modelling species range dynamics. Glob. Chang. Biol. 22, 2651–
994 2664. doi:10.1111/gcb.13251
995
996 Supplementary material: Appendices
997

**Figure captions**

**Figure 1** [WIDTH: 2 COLUMS] H-SDM modelling framework. Embedded in a framework of model critique, the H-SDM hierarchically combines two submodels, a correlative species distribution model (C-SDM) and a grid-cell extinction colonization model (G-ECM). In a knowledge survey, available and lacking information is identified and condensed in model scenarios (step 1). Interspecific dependency on host species limits spatial extent (step2). From environmental filtering, the C-SDM projects abiotic grid cell suitability for each geographic grid cell (step 3). This grid cell suitability affects population dynamical processes in the G-ECM. Parametrization via the G-ECM equilibrium solution (step 4) is further improved by stochastic H-SDM projections (step 5). Steps 2 – 5 are repeated in uncertainty and sensitivity analyses (step 6). Models (right column) and data (left column) are linked by modelling steps (central column). For further information refer to sec.2.2.

**Figure 2** [WIDTH: 2 COLUMS] (A) Presence-absence data and model scenarios as well as (B) mean projected probability of butterfly (Boloria titania) occurrence $H_i$ from C-SDM (note that grid cell suitability resulting from the C-SDM does not consider host plant dependency and thus includes grid cells where its host plant Bistorta officinalis is lacking), (C) occurrence $\bar{J}_i$ from all 80 selected suitable parameter sets of the equilibrium calibration procedure, (D) occurrence $\bar{J}_i$ from all 160 selected H-SDM projections which corresponds to 16000 repetitions. In (A): host plant presence (dark gray dots), butterfly presence (red crosses), artificially inflated butterfly occupancy to test land use constraints on butterfly occupancy in model scenario OL (yellow area), confined extend in model scenario EC (surrounded by blue line). In (B and C): red scale indicates the projected probability of butterfly occurrence (dark means higher probability – see color scale).

**Figure 3** [WIDTH: 2 COLUMS] Distribution of parameter values selected by the equilibrium calibration procedure assuming different model scenarios. Box-plots represent the best 50 parameter sets, while red dots indicate the 10 best parameter sets that are used for further analysis. The size of dots indicates the frequency of the selected value. (Repeated sampling of similar values is promoted by the applied hierarchical latin hyper cube). The blue bar at the left of the graph indicates the range of parameter values.

Model scenarios are OO: observed occupancy, OL: land-use corrected occupancy; EA: extent all, EC: extent central Europe. Parameters M, α/x and c relate to grid cell colonization, while parameter e relates to extinction from grid cells.

**Figure 4** [WIDTH: 2 COLUMS] Correlation biplot for butterfly (Boloria titania) exponential (left column) and powerlaw dispersal (right column) kernels. The upper row displays ordination of the model parameters, while in the bottom row the aggregated parameter z is introduced. When using model parameters, the number of migrants spans the first axis, while extinction risk spans the second axis. The other two dispersal parameters seem to contribute to both axes. However, adding directional contributions of colonization related parameters indicates their strongly correlated impact on variance explained along the first axis. Introducing the aggregated parameter z sorts the parameter sets clearer along the two vectors. Explained variance (A- exponential dispersal) axis1: 49%, axis2: 32%; (B – powerlaw dispersal) axis1: 47%, axis2: 35%, (C- exponential dispersal) axis1: 62%, axis2: 38%; (D – powerlaw dispersal) axis1: 63%, axis2: 37%

Model scenarios (blue colour; OO: observed occupancy, OL: land-use corrected occupancy; EA: extent all, EC: extent central Europe) only weakly correlate with selected parameter sets.

**Figure 5** [WIDTH: 2 COLUMS] Projected butterfly (Boloria titania) occupancy vs model parameter values (graphs A-H), trends can hardly be identified. Instead graphs I and J show that butterfly occupancy correlates better with the non-linearly aggregating parameter *z* (eq. 8). Colors indicate the underlying plant projection (red: P1, blue, P2 – see legend). The black dashed line indicates observed occupancy.

1053 **Table captions**

1054 **Table 1** Standard deviation (in bold on the diagonal) and correlation among calibration

1055 parameters (see eq. 5, 7 and 8) resulting from the equilibrium calibration procedure.

1056

**Tables**

Table 1

| Exponential butterfly dispersal kernel | | | | |
|---|---|---|---|---|
| | M | A | e | C |
| M | **80** | | | |
| α | -0.41 | **0.023** | | |
| e | -0.027 | -0.24 | **0.28** | |
| c | -0.53 | 0.21 | 0.26 | **0.040** |
| Powerlaw butterfly dispersal kernel | | | | |
| | M | x | e | c |
| M | **51** | | | |
| x | -0.50 | **0.98** | | |
| e | 0.11 | -0.40 | **0.26** | |
| c | -0.47 | 0.32 | 0.38 | **0.12** |

Ecological theory
Process knowledge
Field data and their uncertainty

**Step 1: Establishing alternative model scenarios**

Collect all relevant knowledge that might inform model structure and parametrization

Identify knowledge gaps
Determine calibration scenarios

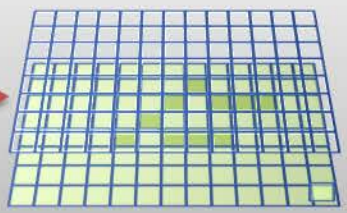
Model scenario 1
Model scenario 2
Model scenario 3
etc.

Interspecific dependency

**Step 2: Accounting for interspecific dependency**

Presence of host species

Determine accessible cells



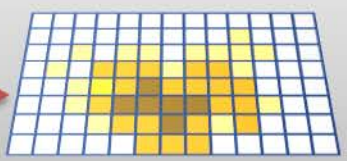
Abiotic environmental conditions

**Step 3: Modelling abiotic influences**

Correlation of species observations and abiotic conditions

Determine abiotic niche

C-SDM projection



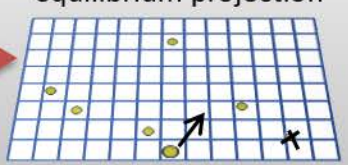
Observed occupancy pattern

**Step 4: Modelling ecological dynamics + equilibrium calibration**

Insert population dynamical processes (extinction/ colonisation )

Determine ecological process parameters from equilibrium

G-ECM equilibrium projection



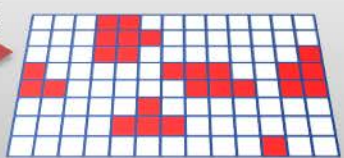
Expert knowledge

**Step 5: Simulating stochastic spatial occupancy dynamics**

Insert ecological processes parameters

Select best ecological process parameters from dynamic projections

H-SDM projection



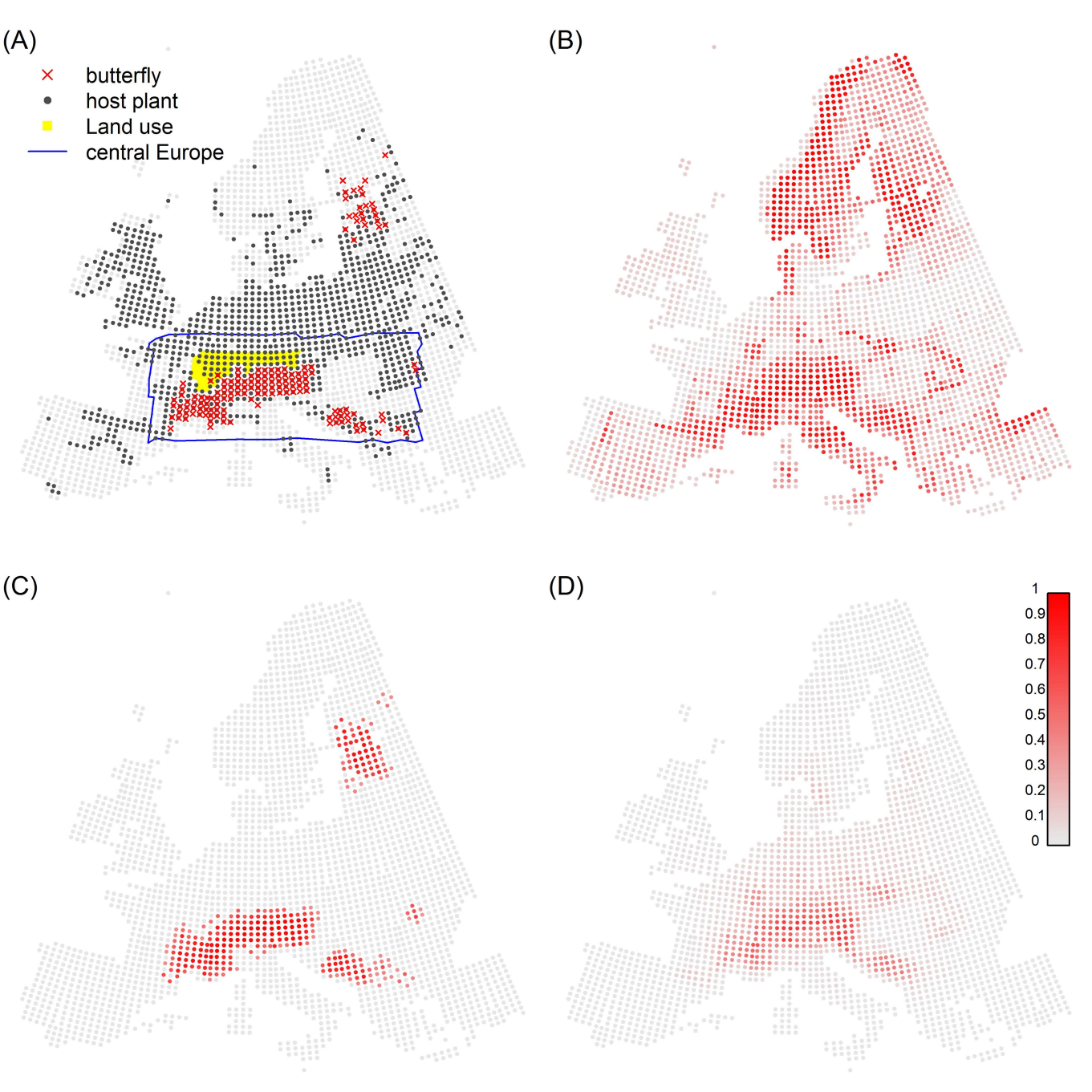
**Step 6: Analyzing alternative model scenarios**

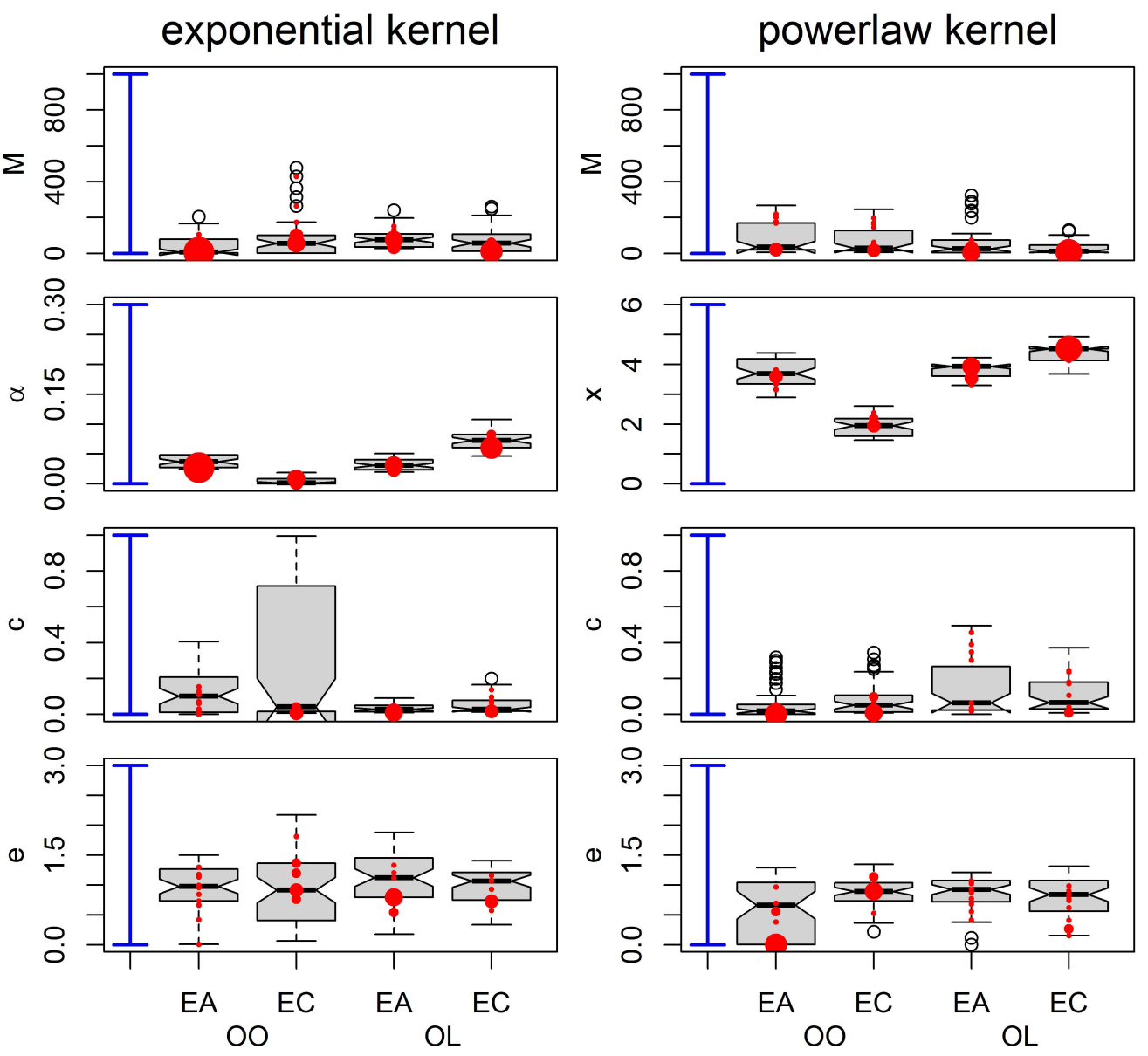
Repeat steps 2 to 5 for the alternative model scenarios

Determine the potential range of projections

Analysis of projection reliability

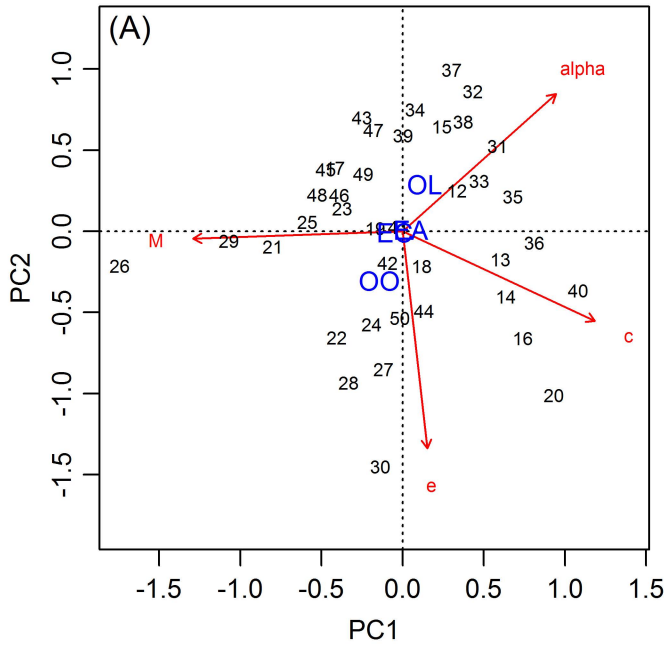
Identification of key processes and parameters

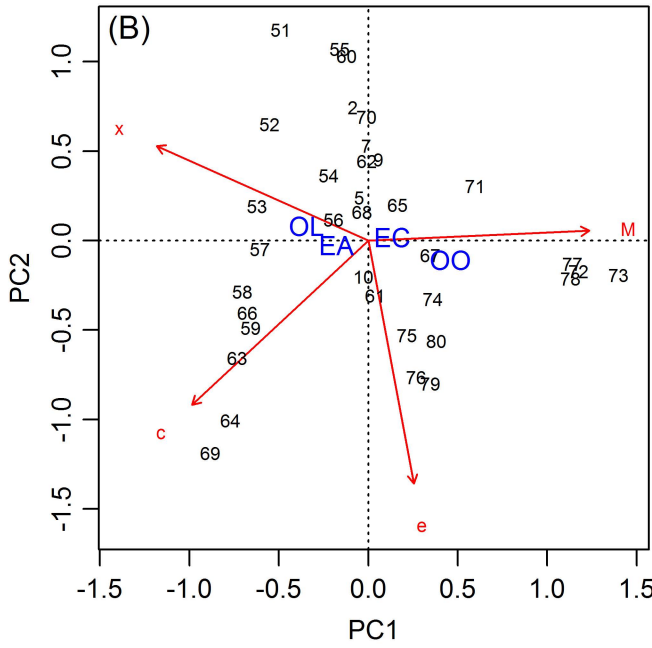

(A)
butterfly
host plant
Land use
central Europe
(B)
(C)
(D)
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0

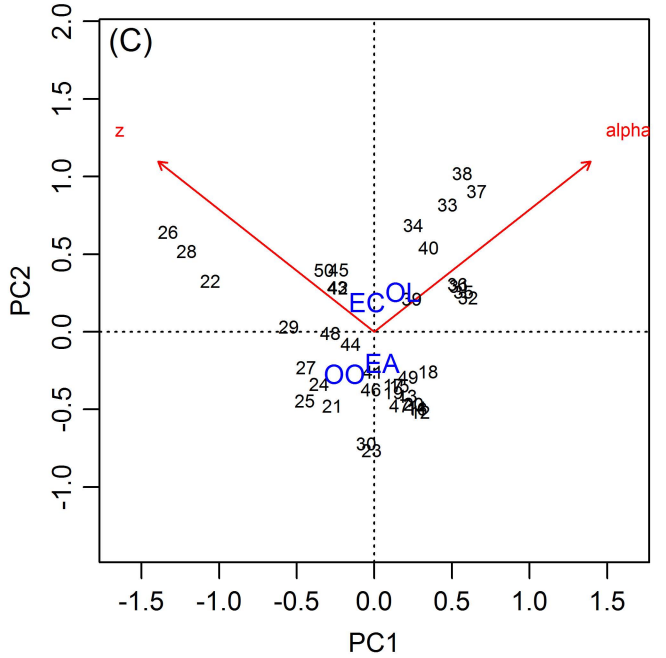

exponential kernel
powerlaw kernel
M
α
c
e
x
0
400
800
0.00
0.15
0.30
2
4
6
0.0
0.4
0.8
1.5
3.0
EA
EC
OO
OL

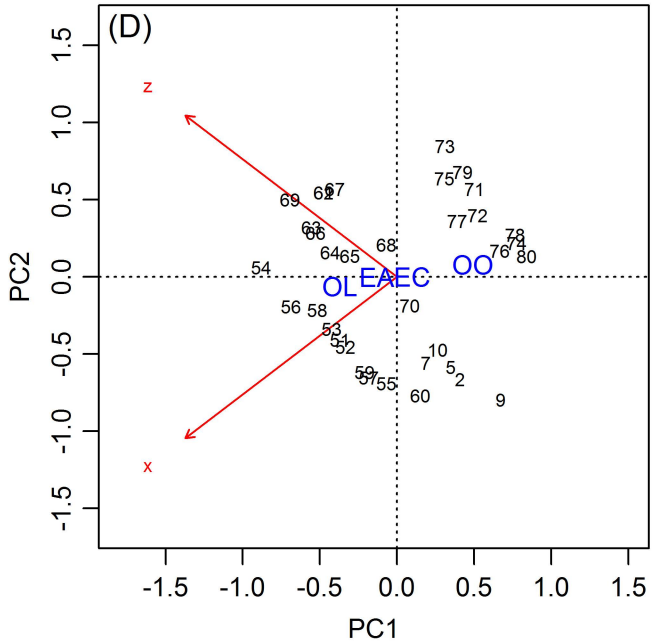
exponential kernel

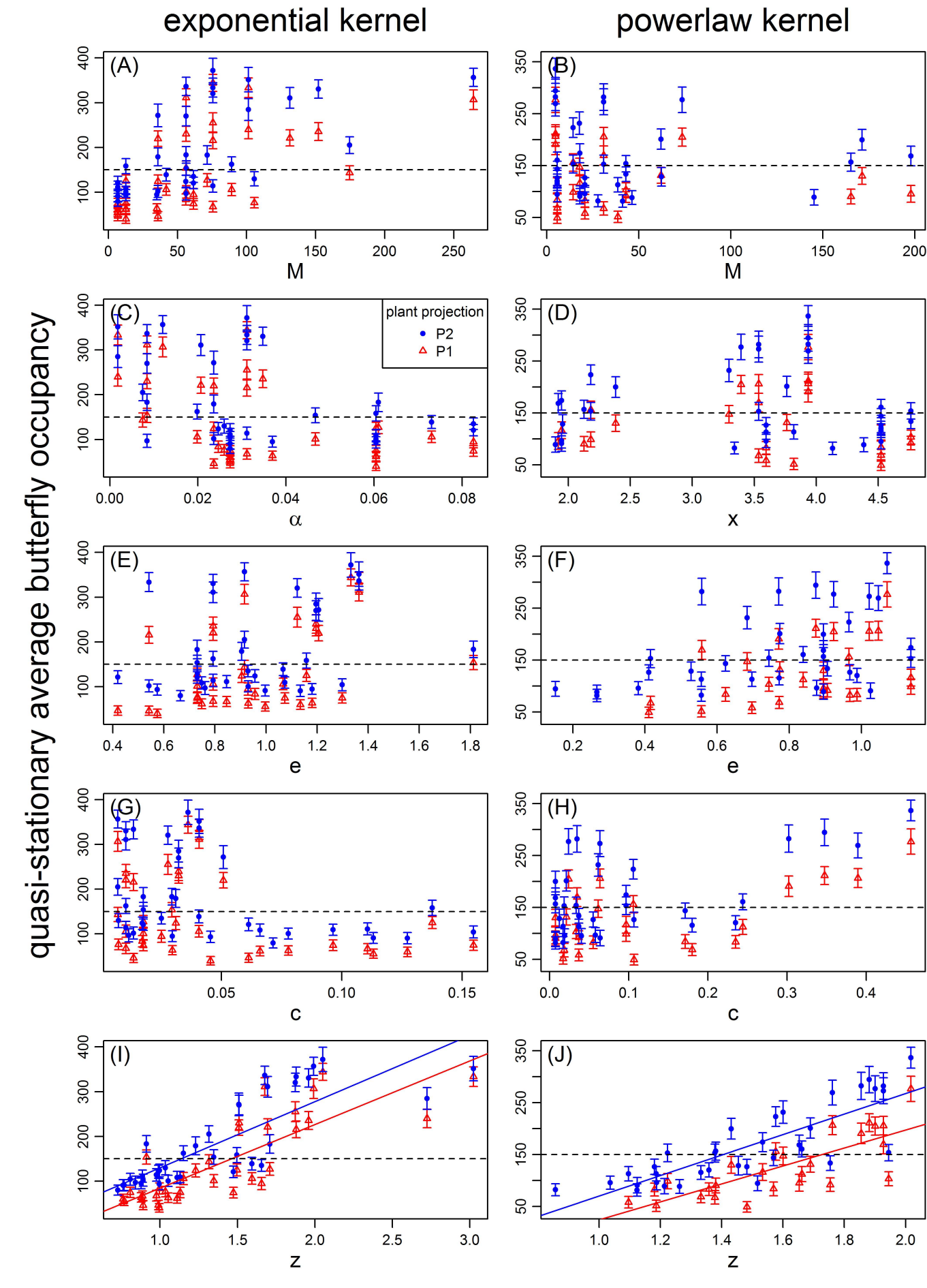
powerlaw kernel




exponential kernel
powerlaw kernel
quasi-stationary average butterfly occupancy
(A)
(B)
(C)
(D)
(E)
(F)
(G)
(H)
(I)
(J)
M
M
α
x
e
e
c
c
z
z
plant projection
P2
P1