Journal of Biogeography (J. Biogeogr.) (2013)

ORIGINAL ARTICLE

Analysing and mapping species range dynamics using occupancy models

Marc Kéry[1]*, Gurutzeta Guillera-Arroita[2,3] and José J. Lahoz-Monfort[2,3]

[1]Swiss Ornithological Institute, 6204 Sempach, Switzerland, [2]School of Mathematics, Statistics and Actuarial Science, University of Kent, Canterbury, CT2 7NF, UK, [3]School of Botany, University of Melbourne, Parkville, Vic. 3010, Australia

ABSTRACT

Aim Our aims are: (1) to highlight the power of dynamic occupancy models for analysing species range dynamics while accounting for imperfect detection; (2) to emphasize the flexibility to model effects of environmental covariates in the dynamics parameters (extinction and colonization probability); and (3) to illustrate the development of predictive maps of range dynamics by projecting estimated probabilities of occupancy, local extinction and colonization.

Location Switzerland.

Methods We used data from the Swiss breeding bird survey to model the Swiss range dynamics of the European crossbill (*Loxia curvirostra*) from 2000 to 2007. Within-season replicate surveys at each 1 km^2 sample unit allowed us to fit dynamic occupancy models that account for imperfect detection, and thus estimate the following processes underlying the observed range dynamics: local extinction, colonization and detection. For comparison, we also fitted a model variant where detection was assumed to be perfect.

Results All model parameters were affected by elevation, forest cover and elevation-by-forest cover interactions and exhibited substantial annual variation. Detection probability varied seasonally and among years, highlighting the need for its estimation. Projecting parameter estimates in environmental or geographical space is a powerful means of understanding what the model is telling about covariate relationships. Geographical maps were substantially different between the model where detection was estimated and that where it was not, emphasizing the importance of accounting for imperfect detection in studies of range dynamics, even for high-quality data.

Main conclusions The study of species range dynamics is among the most exciting avenues for species distribution modelling. Dynamic occupancy models offer a robust framework for doing so, by accounting for imperfect detection *and* directly modelling the effects of covariates on the parameters that govern distributional change. Mapping parameter estimates modelled by spatially indexed covariates is an under-used way to gain insights into dynamic species distributions.

Keywords
Colonization, European crossbill, extinction, distribution dynamics, *Loxia curvirostra*, occupancy model, range dynamics, site-occupancy model, Switzerland, R-package 'unmarked'.

*Correspondence: Marc Kéry, Swiss Ornithological Institute, 6204 Sempach, Switzerland.
E-mail: marc.kery@vogelwarte.ch

INTRODUCTION

The study of species distributions is an area of great scientific interest in ecology (e.g. Graham *et al.*, 2006; Pearman *et al.*, 2008) and evolutionary biology (e.g. Kearney *et al.*, 2009). Species distribution modelling is also an important tool for biodiversity conservation, as knowledge about species distributions is vital for informing and prioritizing conservation action (e.g. Hirzel *et al.*, 2006) and for effective conservation planning (e.g. Araújo & Williams, 2000; Wintle *et al.*, 2005).

© 2013 Blackwell Publishing Ltd

doi:10.1111/jbi.12087

Journal of Biogeography

In this context, predictive models of species distributions are even more critical in the face of climate change, because they allow an understanding of how the distributions of species might change under different projected climatic scenarios (e.g. Hannah *et al.*, 2005; Kearney *et al.*, 2010).

Over the last few decades many species distribution modelling approaches have been proposed for static distributions (for a comparison of some methods see Guisan & Zimmermann, 2000; Elith *et al.*, 2006); however, there is only one, usually referred to as 'occupancy modelling' (MacKenzie *et al.*, 2002; Tyre *et al.*, 2003), that formally accounts for one of the hallmarks of ecological data: imperfect detection (Yoccoz *et al.*, 2001; Kéry, 2002; Kéry & Schmidt, 2008), an issue also relevant for sessile organisms such as plants (Chen *et al.*, 2013). This framework allows the joint modelling of species occupancy probability as a function of spatially indexed covariates and of the detection process, resulting in an extension of the traditional logistic regression model to account for species detectability. Despite often being disregarded (e.g. Araújo & Guisan, 2006), explicitly accounting for imperfect detection can be crucial when modelling species distributions (Kéry, 2011). Otherwise, rather than capturing the actual distribution of the species, the area where the species is detected is estimated, thus leading to incorrect inferences regarding habitat relationships (Tyre *et al.*, 2003; Gu & Swihart, 2004; MacKenzie, 2006), a problem which is greatest when detectability is itself a function of the habitat covariates (Kéry, 2010; Kéry *et al.*, 2010a; Kéry & Schaub, 2012). Disregarding imperfect detection has also been shown to decrease model predictive performance of where a species would be detected (Rota *et al.*, 2011).

Perhaps even more important and exciting is the study of the change of species distributions, or range dynamics. In contrast to the study of static distributions, considerably fewer modelling methods have been proposed so far for the study of range dynamics (e.g. Morin & Thuiller, 2009; Pagel & Schurr, 2012). In this paper we aim to highlight one particularly powerful framework for the analysis of range dynamics: dynamic occupancy models (MacKenzie *et al.*, 2003), also known as multiple-season occupancy models. These extend the static (single-season) occupancy model of MacKenzie *et al.* (2002) and Tyre *et al.* (2003) to allow estimation of occupancy change among discrete time steps by explicitly modelling the processes underlying occupancy dynamics in a Markovian manner. Thus, the occurrence of a species at each time step is described as a function of occurrence during the previous time step and parameters for the probability that a previously unoccupied site becomes occupied (colonization) and that a previously occupied site becomes unoccupied (local extinction). Provided replicate observations of 'presence/absence' (more accurately 'detection/non-detection') are available for at least some of the sites within survey seasons, when the occurrence state is assumed not to change, the model allows imperfect detection to be accounted for, thus making the estimates of occupancy, colonization and local extinction robust against the presence of false absence records (i.e. false negatives). Model parameters can be expressed as functions of covariates, including those describing the dynamics of species occupancy (colonization, local extinction), which can be described using site- and time-specific characteristics. Estimates of all covariate relationships can be projected onto a region of interest, such as an entire country, obtaining predictive maps of all components of range dynamics, a powerful and yet under-utilized method to visualize the results.

We believe that the potential of the dynamic occupancy modelling framework to model species range dynamics, as well as the value of visually expressing the inferences drawn as a map, is far from widely appreciated. In this paper we aim to draw attention to this class of models and we illustrate its power by means of analyses of the range dynamics of a bird species in Switzerland. One of the main objectives of this study is to emphasize the ability to produce predictive maps. In particular, we generate maps of spatially explicit colonization and local extinction probabilities for a species at the scale of an entire country.

MATERIALS AND METHODS

Study area, study species and data collection

Switzerland is a small (41,285 km^2) country in western Europe with an extensive elevational gradient reaching from 200 to 4600 m a.s.l. Forests cover *c.* 30% of the country mostly in the form of rather small mosaic-like patches at lower elevations, beyond which they are more extensive, especially at medium elevations (800–1800 m a.s.l.).

The European crossbill (*Loxia curvirostra* Linnaeus, 1758) is a spruce seed-eating finch of coniferous forests, which often exhibits substantial interannual dynamics of abundance and local occurrence, depending on the annually and regionally varying production of spruce cones and being most abundant and widespread in a region during mast years. Crossbills are widespread in Switzerland from the lowlands up to the tree line, but are most abundant at medium elevations where conifers are particularly common (Schmid *et al.*, 1998).

We used data collected in the period 2000–2007 in the Swiss breeding bird survey MHB (Monitoring Häufige Brutvögel; Schmid *et al.*, 2004), where 267 quadrats of 1 km^2 laid out as a grid across the country are surveyed 2–3 times every breeding season (mid-April to mid-July). Experienced volunteer observers (a total of 315 serving in 2000–2007) follow a quadrat-specific survey route of typical length 3–8 km (average 5 km) and record all visual and acoustic detections of all potential breeding bird species. Crossbills commonly make flights above the canopy and call frequently, making their presence in a quadrat or short sampling route fairly easy to detect when the birds are active. In this study we reduced the actual counts to simple detection/non-detection observations. A subset of these data from 2001 to 2004 was analysed by Royle & Kéry (2007).

© 2013 Blackwell Publishing Ltd

Dynamic occupancy modelling

Dynamic occupancy models explicitly describe the mechanisms underlying occupancy dynamics and the detection process. The sampling protocol assumed here is one in which detection/non-detection data are collected at a number of sampling sites in various sampling seasons (breeding seasons in our example), with repeat surveys within each season. The design therefore involves two nested levels of replication and can be interpreted as a robust design (Pollock, 1982; Williams *et al.*, 2002). A key model assumption is that the system is closed within seasons but occupancy is allowed to change between seasons. It is also assumed that false detections are absent.

Site occupancy dynamics are modelled as a two-state first-order Markov chain (Fig. 1). State transitions are governed by the probabilities of colonization γ and local extinction ε, and only depend on the occupancy status at the previous time step. This structure gives rise to the following relationship for occupancy probability at time $t + 1$:

$$\psi_{t+1} = \psi_t \cdot (1 - \varepsilon) + (1 - \psi_t) \cdot \gamma.$$

The probability that a site is occupied at $t + 1$ is the sum of two probabilities: (1) the probability that the site was occupied at t and the species did not become locally extinct in the interval between t and $t + 1$ plus (2) the probability that the site was empty at t and the species colonized it between t and $t + 1$. The Markovian dependence accommodates the autocorrelation in occupancy probability, as in an AR(1) autoregressive time-series model. Note that multiple-season occupancy data can also be analysed by separately fitting single-season occupancy models to each season. This implies an assumption of independence in the occupancy status of sites between seasons, equivalent to assuming $\varepsilon = 1 - \gamma$ (MacKenzie *et al.*, 2006, pp. 205–207).

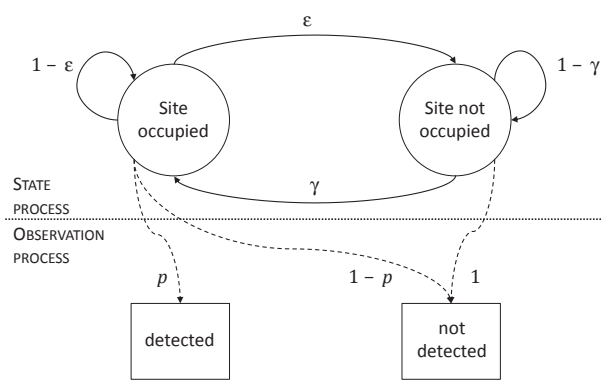


Figure 1 Hidden Markov model for site occupancy dynamics under imperfect detection. Circles are the states representing the occupancy status of a particular site. Squares represent the possible observations. Solid arrows represent the transitions describing occupancy dynamics. Dashed arrows represent the observation process at each survey replicate. The probability that the Markov chain is initialized in the state 'Site occupied' is the initial occupancy psi1.

In the vast majority of applications, species may remain undetected at sites where present. This means that the state 'occupied' in the Markov chain is not perfectly observed. In the dynamic occupancy modelling framework imperfect detection is accounted for by an explicit description of the observation process. Detection/non-detection data from repeat surveys at occupied sites are modelled as a series of independent Bernoulli trials with probability p. The model is therefore a 'hidden Markov model' (MacDonald & Zucchini, 1997); that is, a Markov model in which, rather than observing the states directly, an output dependent on the state is observed.

In its simplest form, the model involves four probability parameters: occupancy in the first season ψ_1, colonization γ, local extinction ε and detectability p. The probabilistic description of the data according to the model, and therefore the construction of its likelihood function, is straightforward. For instance, let us assume a two-season study with three replicate surveys per site and season. The probability of observing in site i a detection history $\boldsymbol{h}_i = 110\ 000$ is

$$\Pr(\boldsymbol{h}_i = 110\,000) = \psi_1 p^2 (1 - p) \times \{\varepsilon + (1 - \varepsilon)(1 - p)^3\};$$

that is, the probability that the site was initially occupied and the species was detected in two replicate surveys and missed in one, and either the site became empty in the second season or it remained occupied and the species was missed in all surveys. MacKenzie *et al.* (2003) provide more examples, along with a matrix form of the likelihood function and alternative parameterizations.

The model allows species range dynamics to be studied by incorporating site-specific covariates to describe the parameters governing occupancy (ψ, ε, γ). Colonization and local extinction probabilities can also be expressed as a function of interseason-specific covariates to reflect their time variation. Very conveniently, the model allows differences in the detection process to be modelled as a function of site- and survey-specific characteristics. Covariates can be incorporated following a generalized linear modelling (GLM) approach. When dealing with binary data, like here, the most commonly used link function is the logit, which leads to a logistic regression. Note that the linear predictor can include both continuous and discrete explanatory variables (factors) as well as polynomial relationships with the covariates (e.g. quadratic and interactions).

Analysis of European crossbill data in Switzerland

There are free ready-made software packages that allow dynamic occupancy models to be fitted in the frequentist method of inference (e.g. MARK, White & Burnham, 1999; PRESENCE, Hines, 2012). Models can also be easily implemented in tools for Bayesian inference such as WinBUGS (Lunn *et al.*, 2000), OpenBUGS (Lunn *et al.*, 2009) or JAGS (Plummer, 2003); code can be found in Royle & Kéry (2007), Royle & Dorazio (2008) and Kéry & Schaub (2012). In our study we used the function *colext* in the new R-package UNMARKED (Fiske & Chandler, 2011) to obtain maximum-likelihood estimates

© 2013 Blackwell Publishing Ltd

(MLEs). The package comes with an extensive tutorial on *colext*.

We fitted models with increasing complexity in terms of covariates and used Akaike's information criterion (AIC) for model selection. Covariates were fitted with a linear and a quadratic effect to allow for possibly non-monotonic relationships. We started by fitting a model with constant parameters and compared it with a model where colonization, extinction and detection were time-dependent. Then we successively added to detection probability (p) the linear and quadratic effects of survey date, elevation, forest cover and finally of the interaction terms between elevation and forest cover, as well as the effect of route length. We kept the detection model structure that was best in terms of AIC and went on to model the parameters describing occupancy dynamics. We fitted models of increasing complexity to the initial occupancy parameter (ψ_1), in terms of elevation, forest cover and the interactions of these covariates. After that, in one variant we followed the same step-up modelling approach first for the probability of colonization (γ) and then for the probability of local extinction (ε) and in another, in reverse order. For simplicity, within each model component (i.e. p, ψ_1, γ, ε) we introduced covariates in a fixed sequence. On identifying a preferred model, we then tried all possible single-term deletions (in a backwards stepwise fashion) and retained the overall most parsimonious model (see Appendix S1).

We used the *predict* function in UNMARKED to obtain predictions of all primary model parameters (initial occupancy, colonization, local extinction and detection) and their associated statistical uncertainty, which we then used to produce maps and plots of the environmental relationships. UNMARKED yields estimates of uncertainty for the annual average values of extinction, colonization and detection, which are a function of the primary parameters. To characterize the uncertainty in the estimates of annual average occupancy probability, we used a nonparametric bootstrap with 1000 replicates coded in R. To assess the goodness of fit of the best model, we used a parametric bootstrapping approach (Kéry *et al.*, 2005) based on a chi-square test statistic for the AIC-best model, which yielded a P-value of 0.88, indicating no evidence of lack of fit of our model as a description of our data set. Finally, to illustrate the effect of disregarding imperfect detection we re-ran the analysis as above but imposing $p = 1$. Such a 'traditional' metapopulation model can be fitted easily by duplicating the aggregated annual data (i.e. whether a crossbill was *ever* detected at a site during a year). We adopted the analogous model selection strategy as above.

RESULTS

There was strong evidence for time-dependence in colonization, extinction and detection probability. Detection probability was affected by survey date, elevation, forest cover (both linear and quadratic) and by all interaction terms between elevation and forest cover, but not by route length. The latter suggests that heterogeneity in route length was not an issue for our analysis. Initial occupancy, colonization and extinction probability were also affected by elevation and forest, including quadratic and interaction terms. Interestingly, interactions between the two environmental covariates were important for all parameters of the dynamic occupancy model (see Appendix S1 in the Supporting Information for model selection results and Table 1 for best model parameter estimates). Using the traditional metapopulation model resulted in a similar most parsimonious model, but substantially different parameter estimates.

There was considerable annual variation in all parameters, including occupancy probability (Fig. 2). The annual means were in the range of 0.31–0.42 (mean 0.40) for occupancy, 0.30–0.62 (mean 0.43) for colonization, 0.13–0.61 (mean 0.35) for local extinction and 0.31–0.61 (mean 0.49) for detection. Detection probability varied not only by year but also over the breeding season, although the magnitude of the interannual variation was greater than was the seasonal variation (Fig. 3). In the traditional metapopulation model, ignoring detection, occupancy was underestimated while extinction and colonization tended to be over- and underestimated, respectively (Fig. 2).

There were complex relationships between the responses and the continuous covariates elevation and forest cover (Fig. 4). Initial occupancy (year 2000) and detection were greatest at mid–high elevations with mid–high forest cover. Colonization was lowest at low elevations irrespective of forest cover and at high elevation with low forest cover. Extinction was greatest at low elevations and with low forest cover. Figure 5 contains a sample of predictive maps of all model components of crossbill range dynamics in Switzerland (the full set of maps, under both the occupancy and the traditional metapopulation model, can be found in Appendix S2). These maps translate the complex relationships among the responses and the covariates (environmental space), shown in Fig. 4, into the spatial domain of Switzerland (geographical space). For instance, as a broad observation, colonization probability is estimated to be lower in the north of the country compared to the south, while extinction probability is higher. This is consistent with the occupancy pattern, which indicates that crossbill occupancy is higher in the south. Figure 6 shows the uncertainty for the prediction of detection probability (Fig. 5, bottom right), illustrating how production of uncertainty maps is straightforward. The results of the analysis assuming perfect detection (i.e. $p = 1$) demonstrate substantial discrepancies in the predictions of range dynamics when imperfect detection is disregarded. This is shown most clearly in the difference maps (Appendix S2, right panels), which are highly patterned for all parameters and can be thought of as maps of bias for the model that does not accommodate imperfect detection.

The total range of a species can be estimated simply by adding up the predicted values of occupancy over the spatial domain studied. For the whole of Switzerland, this yielded values of between 11,869 and 17,542 km^2 for the occupancy

© 2013 Blackwell Publishing Ltd

Table 1 Parameter estimates (maximum-likelihood estimates, MLEs, with asymptotic standard errors, ASEs) of the best dynamic occupancy (Dynocc) model, in terms of Akaike's information criterion (AIC), for the range dynamics of the European crossbill (*Loxia curvirostra*) in Switzerland from 2000 to 2007. For comparison, MLEs and ASEs are also given for the best traditional, 'naïve' metapopulation model that assumes perfect detection (i.e. $p = 1$). Terms not in the model selected by AIC are denoted by –. Models contain parameters for elevation (elev), forest cover (forest), year and survey date (date). Dots denote interactions.

Parameter	Dynocc model (p estimated) MLE	ASE	'Naïve' model ($p = 1$ fixed) MLE	ASE
Initial occupancy (ψ_1)				
Intercept	−0.199	0.243	−0.244	0.281
elev	1.869	0.354	1.994	0.338
elev^2	–	–	−0.140	0.341
forest	0.813	0.255	0.215	0.360
forest^2	−0.142	0.197	0.005	0.210
elev.forest	0.548	0.254	0.501	0.277
elev.forest^2	−0.621	0.243	−0.444	0.276
elev^2.forest	–	–	1.018	0.373
elev^2.forest^2	–	–	–	–
Colonization (γ)				
year2000	−0.432	0.345	−0.917	0.295
year2001	0.487	0.345	0.006	0.239
year2002	0.204	0.363	0.047	0.252
year2003	−0.834	0.510	−0.812	0.314
year2004	−0.708	0.434	−0.236	0.260
year2005	−0.106	0.426	−0.672	0.281
year2006	−0.511	0.520	−0.125	0.240
elev	0.509	0.227	0.939	0.163
elev^2	−0.841	0.313	−0.769	0.230
forest	−0.825	0.324	−0.357	0.209
forest^2	0.092	0.265	0.034	0.168
elev.forest	0.563	0.273	0.033	0.137
elev.forest^2	0.031	0.268	−0.151	0.132
elev^2.forest	1.466	0.406	0.959	0.213
elev^2.forest^2	−0.833	0.399	−0.314	0.206
Extinction (ε)				
year2000	0.442	0.322	0.742	0.278
year2001	−0.453	0.377	−0.181	0.326
year2002	−1.889	0.484	−1.176	0.312
year2003	−1.240	0.400	−0.411	0.250
year2004	−0.983	0.330	−0.572	0.286
year2005	−0.962	0.537	0.188	0.270
year2006	−1.565	0.471	−0.915	0.329
elev	−1.102	0.225	−0.972	0.193
elev^2	–	–	0.074	0.192
forest	−0.616	0.169	−0.907	0.184
forest^2	–	–	0.323	0.118
elev.forest	0.524	0.198	−0.687	0.218
elev.forest^2	–	–	0.356	0.164
elev^2.forest	–	–	–	–
elev^2.forest^2	–	–	–	–
Detection				
year2000	0.386	0.175	–	–
year2001	−0.348	0.204	–	–
year2002	−0.123	0.180	–	–
year2003	0.168	0.160	–	–
year2004	−0.188	0.180	–	–
year2005	0.461	0.176	–	–
year2006	−0.783	0.213	–	–
year2007	0.110	0.165	–	–
elev	0.575	0.122	–	–
elev^2	−0.091	0.135	–	–
forest	0.107	0.162	–	–
forest^2	−0.042	0.090	–	–
date	0.037	0.057	–	–
date^2	0.117	0.048	–	–
elev.forest	0.243	0.192	–	–
elev.forest^2	−0.021	0.116	–	–
elev^2.forest	0.651	0.199	–	–
elev^2.forest^2	−0.243	0.130	–	–

model and between 6770 and 13,107 km^2 for the traditional metapopulation model ignoring imperfect detection, representing proportional underestimates in the latter of around 31% on average.

DISCUSSION

The study of range dynamics is highly topical in ecology and conservation, as it attempts to shed light on the *processes* governing the changes in the observed patterns of species distributions. Such understanding is particularly relevant in the face of current and future climatic and environmental change. We have showcased dynamic occupancy models (MacKenzie *et al.*, 2003) as a flexible and rigorous framework for modelling range dynamics. These models contain a mechanistic description of changes in occurrence as a function of previous occurrence and parameters for colonization and local extinction probability, which are embedded in a description of the observation process that accommodates false-negative detection errors. Unlike species distribution models that do not accommodate imperfect detection, this modelling framework results in parameters that have a more explicit biological meaning, as the bias that would be induced by imperfect detection on the estimators of occupancy and its dynamics is avoided (Kéry *et al.*, 2010a). In addition, covariate effects can be flexibly introduced into all model parameters, and site-specific parameter estimates and their uncertainties (e.g. standard errors) can be projected onto a larger region to produce a map, for instance, of the species distribution, but also of detection (Olea & Mateo-Tomás, 2011; Chen *et al.*, 2013), colonization and extinction probabilities. In this study we have illustrated the production of predictive maps of all these quantities. Of course, as with any modelling technique, extrapolations beyond the sampled range of the covariates or geographical extent have to be interpreted with care (Elith & Leathwick, 2009) and were avoided here by masking mapped areas beyond the normal elevational range of the modelled data.

© 2013 Blackwell Publishing Ltd

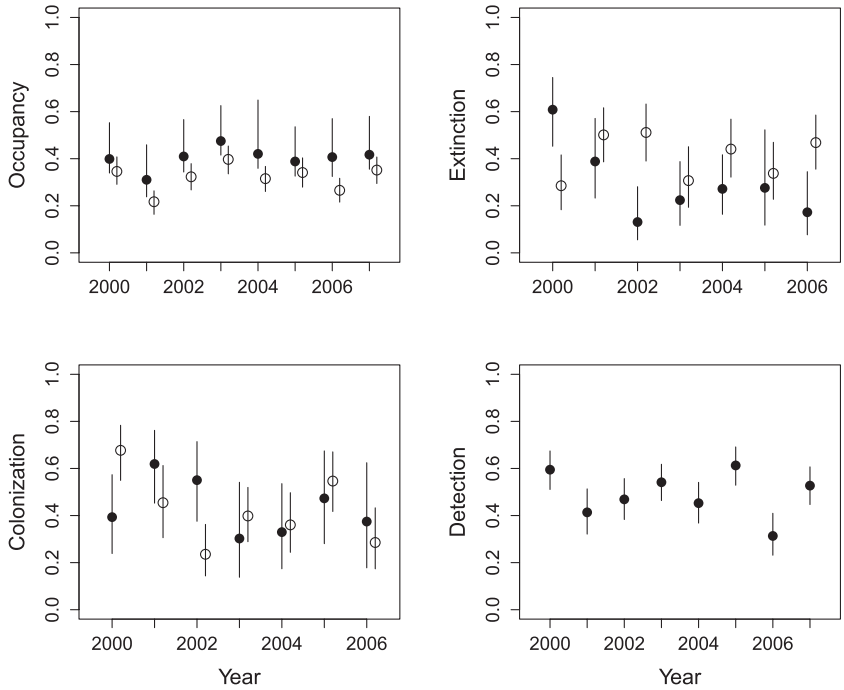


Figure 2 Annual variation in range dynamics and detection parameters for the European crossbill (*Loxia curvirostra*) in Switzerland from 2000 to 2007 under the dynamic occupancy model where detection probability (p) is estimated (filled circles) and under the traditional metapopulation model with the assumption that $p = 1$ (open circles). Lines represent 95% confidence intervals based on asymptotic standard errors except for occupancy probability, where they are based on nonparametric bootstrapping (1000 replicates).

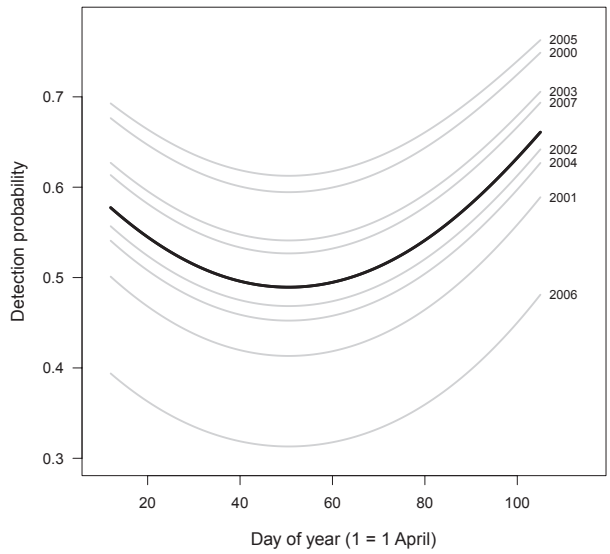


Figure 3 Relationship between detection probability, year and survey date for the European crossbill (*Loxia curvirostra*) in Switzerland from 2000 to 2007. The thick black line shows the average over all years (figure based on parameter estimates in Table 1). Predictions were made at average values of the other covariates (i.e. elevation and forest).

Imperfect detection in species distribution modelling

Almost all current species distribution models make strong assumptions about species detectability (p): that it is either perfect ($p = 1$) or at least that it does not vary over the time interval covered by the study and the environmental gradients surveyed. By explicitly accounting for the detection process, site-occupancy models do not require such, usually untested, assumptions. On the other hand, the site-occupancy modelling framework requires replicate observations to be made at least at some sites and for at least some time periods, within which closure (i.e. no change in the occurrence status of a site) can be assumed. These requirements are not as restrictive as they may first appear. Additionally, most importantly and in spite of assertions to the contrary (e.g. Gómez-Rodríguez *et al.*, 2012), replicate observations need not be made at all sites nor during all primary periods (i.e. breeding seasons in our study). The model may be fitted to data sets that lack replicates for some sites and/or times. The information about detectability is obtained from the sites and times where replication is available and the model assumes that detection probability at sites and times without replication is identical to those where there is replication or, alternatively, variation in detection may be modelled using covariates that are also available at those sites or times. As an example, the analyses reported in Kéry *et al.* (2010a,b) had replicated observations for only 30–60% of the sites. Second, replicates may sometimes be deduced from databases containing records from multiple species (e.g. Kéry *et al.*, 2010b). And third, the closure assumption may sometimes be relaxed: if during the periods within which closure is assumed the species moves in and out of the sample plots, and surveys are conducted such that those movements can be considered random, the occupancy estimator remains unbiased; it simply needs to be interpreted as probability of use, rather than of permanent occurrence (MacKenzie *et al.*, 2006, p. 105). Closure issues can be entirely avoided by modelling occupancy via estimation of the rate of detections within single survey visits (Garrard *et al.*, 2008; Guillera-Arroita *et al.*, 2011, 2012).

© 2013 Blackwell Publishing Ltd

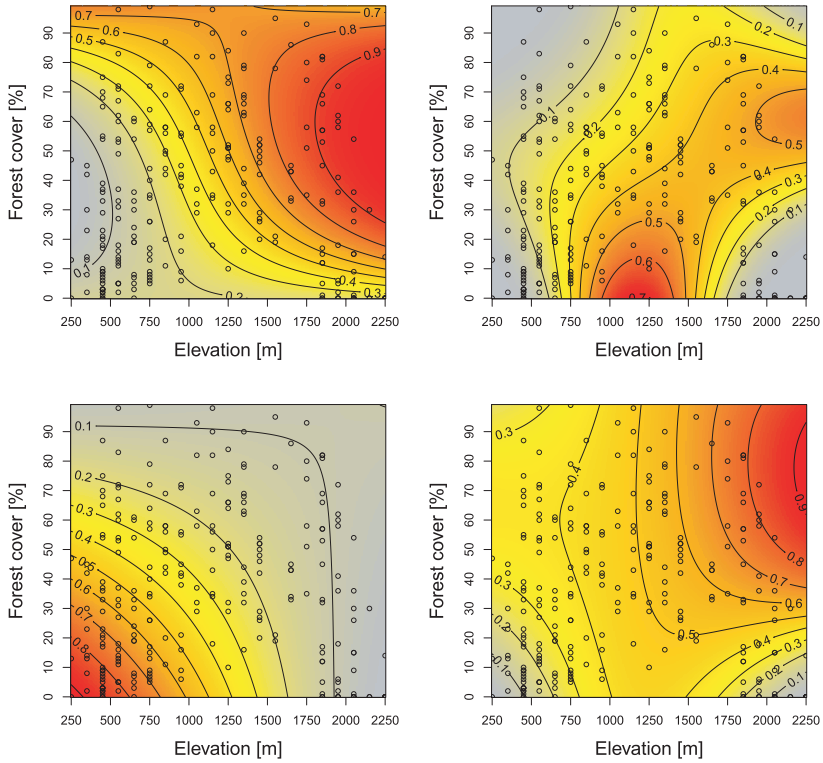


Figure 4 Predicted response in environmental space: joint effects of elevation and forest cover on the range dynamics and detection parameters for the European crossbill (*Loxia curvirostra*) in Switzerland from 2000 to 2007 (based on parameter estimates in Table 1): first-year occupancy probability (top left), colonization probability (top right), extinction probability (bottom left) and detection probability (bottom right). Plots for colonization, extinction and detection are averaged over all study years. Detection probability was computed for the mean survey date. The range of grey to red colour corresponds to a range of low to high probability (same scale as in Fig. 5). Circles represent the observed covariate values at the 267 sample plots.

Imperfect detection also has implications regarding the methods traditionally used for model validation in other species distribution modelling approaches. For instance, one could compute the area under the receiver operating characteristic curve (AUC) for the ability to predict *detections* (Rota *et al.*, 2011). This can be taken as a vague indication of goodness of fit but it is not a direct measure of predictive performance for species occurrences. Rather, it expresses how well a model predicts the product of occurrence probability and conditional detection probability (Yackulic *et al.*, 2012). Other goodness-of-fit tests are available, such as the one used in this study, but there is scope for further research in this area.

Species detectability is the product of various mechanisms, including local abundance, animal behaviour and surveyor skills, with the variation in abundance often being a key factor leading to its spatial or temporal variation (Kéry *et al.*, 2010a). By using the dynamic occupancy modelling framework we were able to investigate the range dynamics of the European crossbill accounting for its imperfect detection. Disregarding the issue of imperfect detection when modelling range dynamics from detection/non-detection data can lead to biased inferences, with false absences inducing negative bias in the estimators of occupancy and survival probabilities ($1-\varepsilon$) and the slope of covariate relationships and a positive bias in turnover rates (Royle & Kéry, 2007; Kéry & Schaub, 2012). Analyses of presence-only data are similarly affected by imperfect detection. Consider the extreme case of the species preferring habitat B over A, but being much harder to detect in B. No amount of sophisticated modelling will prevent us from concluding that A is highly preferred over B. The inability to disentangle the occupancy and detection processes would then be propagated to the estimation of the processes underlying occupancy dynamics (Yackulic *et al.*, 2012).

Note that, while the model accounts for false negatives, it assumes that there are no false positives in the data. We believe that this assumption is reasonable for our study, as the European crossbill is the only *Loxia* species present in Switzerland and is easily identifiable owing to its distinctive calls, which are hard to misidentify. However, misidentifications leading to false positives may not be rare in other studies or for other species and it has been shown that these can induce severe bias in the estimators of occupancy, colonization and extinction (McClintock *et al.*, 2010). Accounting for this source of error can therefore be critical in some applications. Miller *et al.* (2011) propose techniques to do this for single-season occupancy models that can be readily incorporated into the dynamic occupancy modelling framework.

Although large-scale studies of species range dynamics have traditionally disregarded imperfect detection, two recent exceptions that account for this issue using the dynamic occupancy model are Eraud *et al.* (2007) and Altwegg *et al.* (2008), both studying bird species with expanding ranges. However, neither of these studies produces predictive maps for the parameters that govern range dynamics, like we do for the crossbill in Switzerland, and which we believe provide a useful tool for understanding and communicating the results of these analyses.

© 2013 Blackwell Publishing Ltd

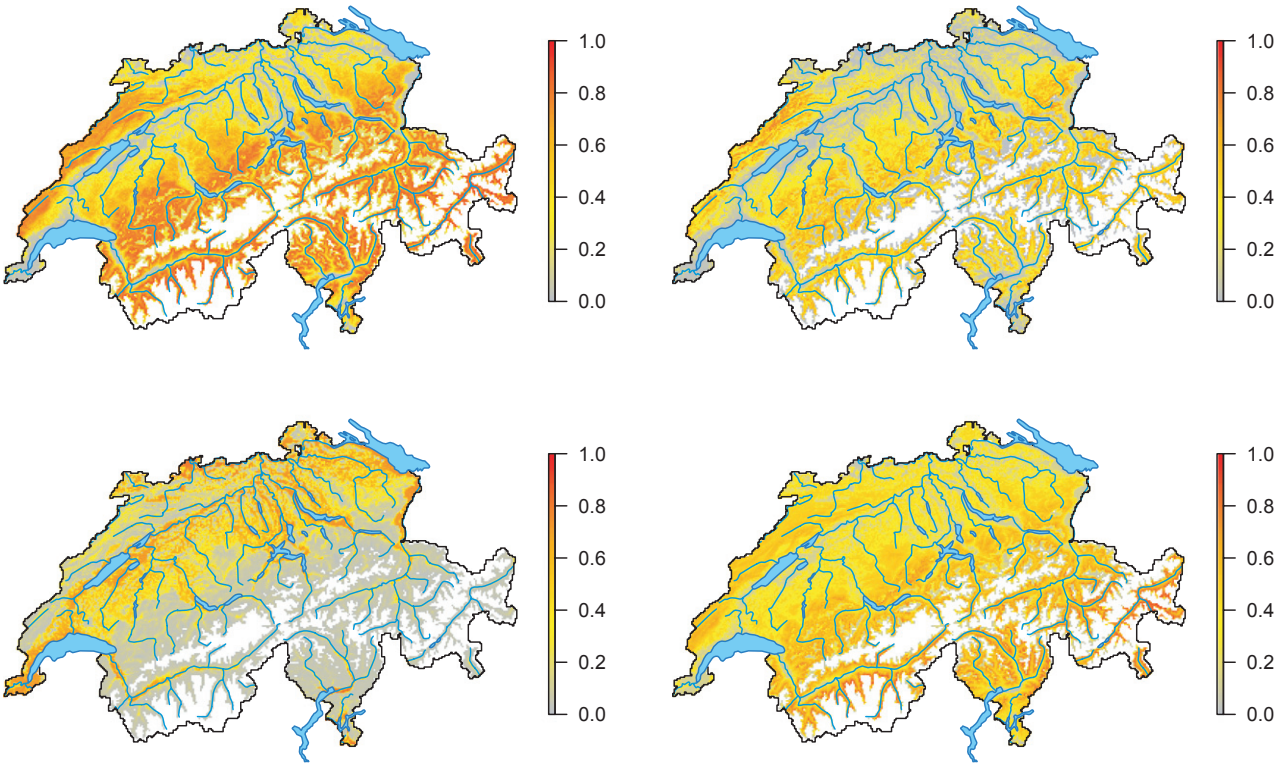


Figure 5 Predicted response in geographical space: visualization of the range dynamics of the European crossbill (*Loxia curvirostra*) in Switzerland in 2002–2003. Occupancy probability in 2003 (top left), colonization probability for the interval 2002–2003 (top right), extinction probability for the interval 2002–2003 (bottom left), and detection probability in 2003 (computed at the mean survey date; bottom right). Areas at elevations greater than 2250 m a.s.l. are masked in white.

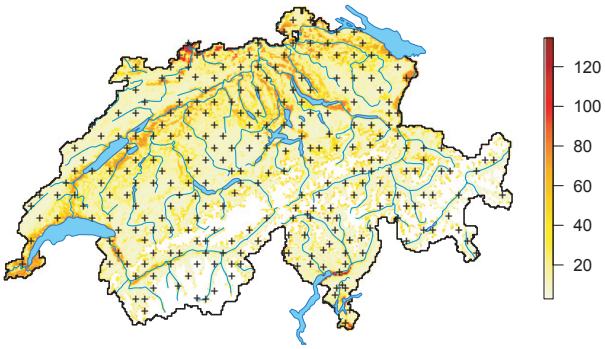


Figure 6 Map of the prediction uncertainty (as the coefficient of variation, in %) in detection probability of the European crossbill (*Loxia curvirostra*) in Switzerland in 2003 (asymptotic standard errors; see Fig. 5, bottom right). Crosses show the locations of the 267 sampled quadrats.

Modelling of range dynamics

The concept of range dynamics involves different aspects. Let us first consider a case in which the overall extent of occurrence of the species is relatively stable. Even if the probabilities of colonization (γ) and local extinction (ε) are constant within the timeframe of the study, this does not necessarily imply that such system is at equilibrium in the sense that species occupancy may still be in a transient stage. In this connection, we would like to draw attention to a common misconception in the interpretation of the parameters governing Markovian dynamics, namely that one can extract conclusions regarding the increase/decline in the number of sites occupied by a species simply by comparing the value of colonization and local extinction probabilities. For instance, Eraud *et al.* (2007, p. 1083) claim that 'local extinction was lower than local colonization, thus explaining the positive trend in site occupancy', while Altwegg *et al.* (2008, p. 582) state that 'colonization exceeded extinction, and [the species] thus occupied more sites over time'. Relating the increase in species occupancy over time simply to the fact that colonization γ is greater than local extinction ε is incorrect. In the absence of time-variation in the range dynamics parameters, site occupancy will tend over time to an equilibrium value $\psi_{eq} = \gamma/(\gamma + \varepsilon)$. This has implications not always realized by all. The fact that γ is greater than ε only implies that occupancy will tend to a value $\psi_{eq} > 0.5$. Likewise, $\varepsilon > \gamma$ does not imply that the species will disappear from the study area, only that $\psi_{eq} < 0.5$. The actual direction of the change in occupancy over time will depend on whether initial occupancy ψ_1 is above or below ψ_{eq}. For example, a decline in occupancy is experienced if $\psi_1 > \psi_{eq}$, regardless of γ being larger or smaller than ε. Note also that, even if a system has reached the equilibrium, this does not mean that the

© 2013 Blackwell Publishing Ltd

occupancy status of individual sites does not change, but that the overall proportion of occupied sites remains constant.

Checking whether the estimated current occupancy and the computed equilibrium occupancy probabilities match is thus one way of telling whether a species is at equilibrium. In the crossbill example, taking the mean values of the colonization and extinction probabilities yields an equilibrium occupancy of 0.57. This is more than the mean observed occupancy of 0.42 and might lead us to conclude that the Swiss range of the crossbill will increase in the future. However, this ignores the annual variation in colonization and extinction, a further driver of range dynamics. In the short term, crossbills exhibit a highly dynamic distribution in Switzerland which changes from year to year in response to local factors (e.g. spruce mast) and to coarser scale factors (e.g. irruptions from other European countries; Schmid *et al.*, 1998). A possible approach to gauge the 'equilibrium' question in this case would be to model annual variability in extinction and colonization as random effects and then conduct stochastic projections of the system ahead in time to explore the characteristics of the 'equilibrium' regime attained under the current mean and variance of the colonization and extinction probabilities. This could be carried out more conveniently within the Bayesian framework, e.g. using WinBUGS or JAGS (Kéry & Schaub, 2012).

Finally, our study has focused on exploring the processes governing range dynamics for a species with considerable dispersal potential, which is well established across the area of the study, so we can assume independence in the occupancy status of our sites. However, the occupancy dynamics modelling framework lends itself to extensions that allow the investigation of spatially correlated range changes, such as in the spread of invasive species. This has previously been done by incorporating in the dynamics parameters a covariate representing the interaction of time and distance from the point of release (MacKenzie *et al.*, 2006, pp. 201–203) and more recently using a more mechanistic model which incorporates a latent auto-covariate that reflects the influence of the occupancy status of neighbouring cells (Bled *et al.*, 2011). One further important avenue for progress is to model range dynamics by combining presence–absence data and time series of population counts in a spatially explicit model of population dynamics and dispersal (e.g. Pagel & Schurr, 2012). Nevertheless, such models require much more and different data (i.e. population counts), while the data required for occupancy modelling are 'cheaper' to obtain.

Dynamic occupancy model of European crossbills in Switzerland

Fitting the dynamic occupancy model to data for the European crossbill at a 1 km^2 scale in Switzerland, it became obvious that detection probability varied in space and time and was therefore an important source of variation in the observed data. There were pronounced spatial patterns in detection probability related to the environment (elevation, forest cover) which, if not accounted for, could have led to biased inferences about the parameters driving range dynamics and their relationship with covariates. The discrepancies with the predictions obtained assuming perfect detection illustrate the relevance of accounting for the detection process. The variation in crossbill detectability is likely to reflect differences in abundance, being higher in areas of mid–high elevation with high forest cover where occupancy is also higher. There has not been any formal assessment in the literature about the effects of abundance-induced detection heterogeneity for the dynamic occupancy model; however, one could expect bias to be induced in the estimators, as happens in static models where occupancy can be underestimated (Royle, 2006; Dorazio, 2007). By incorporating covariates that are likely to reflect abundance, we are at least partially controlling for this in our study.

Our crossbill models do not contain a very elaborate suite of explanatory variables, but simply elevation and forest cover. We emphasize, though, that the main aim of our study is to illustrate the use of dynamic occupancy models and the ability to produce maps including those of colonization and extinction probabilities. Although elevation explains a large amount of the variability in occurrence in any species in such a mountainous country as Switzerland, and forest cover is similarly a powerful explanation for a forest species, we would not want to argue that just two covariates, plus their interactions, are always enough to produce useful maps of species distributions and their dynamic components. In our study we did not have information about the proportion of conifers per quadrat, which we would expect to be a better predictor for crossbill occupancy. Nevertheless, the interactions between elevation and forest cover may partly have accommodated the fact that the proportion of conifers is a function of elevation, being much higher at medium to high elevations. Not surprisingly, interactions between elevation and forest cover proved significant in all model components. Our model reflects a situation in which the species spreads out from a consistent core into more marginal areas in 'good' years, retracting to the core during 'bad' years. Note that, if the relationship with the environment is expected to change over time, interactions between environmental variables and year can be added to the model.

CONCLUSIONS

We have emphasized that any model of abundance or occurrence with spatially indexed covariates is a species distribution model that allows distribution maps to be obtained simply by projecting its estimates onto a wider region or even an entire country. In particular, in dynamic occupancy models, maps of the dynamics parameters underlying change in distributions may be drawn as well; these may give important insights for instance into the dynamics of a population of stable distribution at large spatial scales, or into the expansion and contraction areas of species subject to

© 2013 Blackwell Publishing Ltd

directional range changes (e.g. the invasion of the collared dove *Streptopelia decaocto* in the USA; Bled *et al.*, 2011). Different processes can lead to similar occupancy patterns and mapping the dynamic parameters can be a useful tool for better understanding of the species distribution. We note that maps of detection probability (p) can also be produced; these can be thought of as an indication of the relative bias of a modelling method that would have ignored imperfect detection, with larger bias to be expected wherever p is small. Such maps provide an indication of the amount of survey effort needed to detect the species in different areas (Olea & Mateo-Tomás, 2011). It is also relevant to realize that the very same covariates that affect detectability may also affect the biological states and their dynamics, as illustrated by our crossbill study; for species distributions this is almost to be expected, as detection probability will often reflect the underlying spatial distribution of population abundance, potentially governed by the same effects as occupancy and its range dynamics. Finally, within the realm of parametric statistical models, it is straightforward to produce maps of the uncertainty of the predictions (e.g. standard error or confidence interval ranges) for all parameters considered. Quantifying and honestly communicating the uncertainty in species distribution maps is a greatly under-appreciated but very important issue.

ACKNOWLEDGEMENTS

We thank the volunteers who conduct the fieldwork in the Swiss Monitoring Häufige Brutvögel (MHB), H. Schmid for providing the crossbill data set, J. Guélat for help with mapping, and V. Keller, three anonymous referees and the editors for helpful comments on the manuscript.

REFERENCES

Altwegg, R., Wheeler, M. & Erni, B. (2008) Climate and the range dynamics of species with imperfect detection. *Biology Letters*, **4**, 581–584.

Araújo, M.B. & Guisan, A. (2006) Five (or so) challenges for species distribution modelling. *Journal of Biogeography*, **33**, 1677–1688.

Araújo, M.B. & Williams, P.H. (2000) Selecting areas for species persistence using occurrence data. *Biological Conservation*, **96**, 331–345.

Bled, F., Royle, J.A. & Cam, E. (2011) Hierarchical modeling of an invasive spread: the Eurasian Collared-Dove *Streptopelia decaocto* in the United States. *Ecological Applications*, **21**, 290–302.

Chen, G., Kéry, M., Plattner, M., Ma, K. & Gardner, B. (2013) Imperfect detection is the rule rather than the exception in plant distribution studies. *Journal of Ecology*, **101**, 183–191.

Dorazio, R.M. (2007) On the choice of statistical models for estimating occurrence and extinction from animal surveys. *Ecology*, **88**, 2773–2782.

Elith, J. & Leathwick, J.R. (2009) Species distribution models: ecological explanation and prediction across space and time. *Annual Review of Ecology, Evolution, and Systematics*, **40**, 677–697.

Elith, J., Graham, C.H., Anderson, R.P. *et al.* (2006) Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, **29**, 129–151.

Eraud, C., Boutin, J.M., Roux, D. & Faivre, B. (2007) Spatial dynamics of an invasive bird species assessed using robust design occupancy analysis: the case of the Eurasian collared dove (*Streptopelia decaocto*) in France. *Journal of Biogeography*, **34**, 1077–1086.

Fiske, I. & Chandler, R. (2011) unmarked: an R package for fitting hierarchical models of wildlife occurrence and abundance. *Journal of Statistical Software*, **43**, 1–23.

Garrard, G.E., Bekessy, S.A., McCarthy, M.A. & Wintle, B.A. (2008) When have we looked hard enough? A novel method for setting minimum survey effort protocols for flora surveys. *Austral Ecology*, **33**, 986–998.

Gómez-Rodríguez, C., Bustamante, J., Díaz-Paniagua, C. & Guisan, A. (2012) Integrating detection probabilities in species distribution models of amphibians breeding in Mediterranean temporary ponds. *Diversity and Distributions*, **18**, 260–272.

Graham, C.H., Moritz, C. & Williams, S.E. (2006) Habitat history improves prediction of biodiversity in rainforest fauna. *Proceedings of the National Academy of Sciences USA*, **103**, 632–636.

Gu, W. & Swihart, R.H. (2004) Absent or undetected? Effects of non-detection of species occurrence on wildlife–habitat models. *Biological Conservation*, **116**, 195–203.

Guillera-Arroita, G., Morgan, B.J.T., Ridout, M.S. & Linkie, M. (2011) Species occupancy modeling for detection data collected along a transect. *Journal of Agricultural, Biological, and Environmental Statistics*, **16**, 301–317.

Guillera-Arroita, G., Ridout, M.S., Morgan, B.J.T. & Linkie, M. (2012) Models for species-detection data collected along transects in the presence of abundance-induced heterogeneity and clustering in the detection process. *Methods in Ecology and Evolution*, **3**, 358–367.

Guisan, A. & Zimmermann, N. (2000) Predictive habitat distribution models in ecology. *Ecological Modelling*, **135**, 147–186.

Hannah, L., Midgley, G.F., Hughes, G. & Bomhard, B. (2005) The view from the Cape: extinction risk, protected areas, and climate change. *BioScience*, **55**, 231–242.

Hines, J.E. (2012) *PRESENCE 5.5 – Software to estimate patch occupancy and related parameters*. USGS-PWRC. Available at: http://www.mbr-pwrc.usgs.gov/software/presence.html.

Hirzel, A., Le Lay, G., Helfer, V., Randin, C. & Guisan, A. (2006) Evaluating the ability of habitat suitability models to predict species presences. *Ecological Modelling*, **199**, 142–152.

Kearney, M., Porter, W.P., Williams, C., Ritchie, S. & Hoffmann, A.A. (2009) Integrating biophysical models and evolutionary theory to predict climatic impacts on species'

© 2013 Blackwell Publishing Ltd

ranges: the dengue mosquito *Aedes aegypti* in Australia. *Functional Ecology*, **23**, 528–538.

Kearney, M.R., Wintle, B.A. & Porter, W.P. (2010) Correlative and mechanistic models of species distribution provide congruent forecasts under climate change. *Conservation Letters*, **3**, 203–213.

Kéry, M. (2002) Inferring the absence of a species: a case study of snakes. *Journal of Wildlife Management*, **66**, 330–338.

Kéry, M. (2010) *Introduction to WinBUGS for ecologists – a Bayesian approach to regression, ANOVA, mixed models and related analyses*. Academic Press, Burlington, MA.

Kéry, M. (2011) Towards the modeling of true species distributions. *Journal of Biogeography*, **38**, 617–618.

Kéry, M. & Schaub, M. (2012) *Bayesian population analysis using WinBUGS: a hierarchical perspective*. Academic Press, Waltham, MA.

Kéry, M. & Schmidt, B.R. (2008) Imperfect detection and its consequences for monitoring for conservation. *Community Ecology*, **9**, 207–216.

Kéry, M., Royle, J.A. & Schmid, H. (2005) Modeling avian abundance from replicated counts using binomial mixture models. *Ecological Applications*, **15**, 1450–1461.

Kéry, M., Gardner, B. & Monnerat, C. (2010a) Predicting species distributions from checklist data using site-occupancy models. *Journal of Biogeography*, **37**, 1851–1862.

Kéry, M., Royle, J.A., Schmid, H., Schaub, M., Volet, B., Häfliger, G. & Zbinden, N. (2010b) Site-occupancy distribution modeling to correct population-trend estimates derived from opportunistic observations. *Conservation Biology*, **24**, 1388–1397.

Lunn, D., Spiegelhalter, D., Thomas, A. & Best, N. (2009) The BUGS project: evolution, critique and future directions. *Statistics in Medicine*, **28**, 3049–3067.

Lunn, D.J., Thomas, A., Best, N. & Spiegelhalter, D. (2000) WinBUGS – A Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing*, **10**, 325–337.

MacDonald, I. & Zucchini, W. (1997) *Hidden Markov and other models for discrete-valued time series*. Chapman & Hall, New York.

MacKenzie, D.I. (2006) Modelling the probability of resource use: the effect of, and dealing with, detecting a species imperfectly. *Journal of Wildlife Management*, **70**, 367–374.

MacKenzie, D.I., Nichols, J.D., Lachman, G.B., Droege, S., Royle, J.A. & Langtimm, C.A. (2002) Estimating site occupancy rates when detection probability rates are less than one. *Ecology*, **83**, 2248–2255.

MacKenzie, D.I., Nichols, J.D., Hines, J.E., Knutson, M.G. & Franklin, A.D. (2003) Estimating site occupancy, colonization and local extinction when a species is detected imperfectly. *Ecology*, **84**, 2200–2207.

MacKenzie, D.I., Nichols, J.D., Royle, J.A., Pollock, K.H., Bailey, L.L. & Hines, J.E. (2006) *Occupancy estimation and modelling: inferring patterns and dynamics of species occurrence*. Elsevier Publishing, Amsterdam.

McClintock, B.T., Bailey, L.L., Pollock, K.H. & Simons, T.R. (2010) Unmodeled observation error induces bias when inferring patterns and dynamics of species occurrence via aural detections. *Ecology*, **91**, 2446–2454.

Miller, D.A., Nichols, J.D., McClintock, B.T., Campbell Grant, E.H., Bailey, L.L. & Weir, L.A. (2011) Improving occupancy estimation when two types of observation error occur: non-detection and species misidentification. *Ecology*, **92**, 1422–1428.

Morin, X. & Thuiller, W. (2009) Comparing niche- and process-based models to reduce prediction uncertainty in species range shifts under climate change. *Ecology*, **90**, 1301–1313.

Olea, P.P. & Mateo-Tomás, P. (2011) Spatially explicit estimation of occupancy, detection probability and survey effort needed to inform conservation planning. *Diversity and Distributions*, **17**, 714–724.

Pagel, J. & Schurr, F.M. (2012) Forecasting species ranges by statistical estimation of ecological niches and spatial population dynamics. *Global Ecology and Biogeography*, **21**, 293–304.

Pearman, P.B., Guisan, A., Broennimann, O. & Randin, C.F. (2008) Niche dynamics in space and time. *Trends in Ecology and Evolution*, **23**, 149–158.

Plummer, M. (2003) JAGS: a program for analysis of Bayesian graphical models using Gibbs sampling. *Proceedings of the 3rd International Workshop on Distributed Statistical Computing (DSC 2003), March 20–22* (ed. by K. Hornik, F. Leisch and A. Zeileis). Vienna, Austria. Available at: http://www.ci.tuwien.ac.at/Conferences/DSC-2003/Proceedings/Plummer.pdf.

Pollock, K.H. (1982) A capture-recapture design robust to unequal probability capture. *Journal of Wildlife Management*, **46**, 752–757.

Rota, C.T., Fletcher, R.J., Evans, J.M. & Hutto, R.L. (2011) Does accounting for imperfect detection improve species distribution models? *Ecography*, **34**, 659–670.

Royle, J.A. (2006) Site occupancy models with heterogeneous detection probabilities. *Biometrics*, **62**, 97–102.

Royle, J.A. & Dorazio, R.M. (2008) *Hierarchical modelling and inference in ecology: the analysis of data from populations, metapopulations and communities*. Academic Press, San Diego, CA.

Royle, J.A. & Kéry, M. (2007) A Bayesian state-space formulation of dynamic occupancy models. *Ecology*, **88**, 1813–1823.

Schmid, H., Luder, R., Naef-Daenzer, B., Graf, R. & Zbinden, N. (1998) *Schweizer Brutvogelatlas. Verbreitung der Brutvögel in der Schweiz und im Fürstentum Liechtenstein 1993–1996*. Swiss Ornithological Institute, Sempach, Switzerland.

Schmid, H., Zbinden, N. & Keller, V. (2004) *Überwachung der Bestandsentwicklung häufiger Brutvögel in der Schweiz*. Swiss Ornithological Institute, Sempach, Switzerland.

Tyre, A.J., Tenhumberg, B., Field, S.A., Niejalke, D., Parris, K. & Possingham, H.P. (2003) Improving precision and

© 2013 Blackwell Publishing Ltd

reducing bias in biological surveys: estimating false-negative error rates. *Ecological Applications*, **13**, 1790–1801.

White, G.C. & Burnham, K.P. (1999) Program MARK: survival estimation from populations of marked animals. *Bird Study*, **46**, 120–139.

Williams, B.K., Nichols, J.D. & Conroy, M.J. (2002) *Analysis and management of animal populations*. Academic Press, San Diego, CA.

Wintle, B., Elith, J. & Potts, J. (2005) Fauna habitat modelling and mapping: a review and case study in the Lower Hunter Central Coast region of NSW. *Austral Ecology*, **30**, 719–738.

Yackulic, C.B., Chandler, R., Zipkin, E.F., Royle, J.A., Nichols, J.D., Grant, E.H.C. & Veran, S. (2013) Presence-only modelling using MAXENT: when can we trust the inferences? *Methods in Ecology and Evolution*, doi:10.1111/2041-210x.12004.

Yoccoz, N.G., Nichols, J.D. & Boulinier, T. (2001) Monitoring of biological diversity in space and time. *Trends in Ecology and Evolution*, **16**, 446–453.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

Appendix S1 Model selection results.
Appendix S2 Predictions of all main parameters under the dynamic occupancy model and under a traditional metapopulation model where detection is implicitly assumed to be equal to 1.

BIOSKETCH

Marc Kéry is a statistical ecologist currently with the Swiss Ornithological Institute. He is particularly interested in capture–recapture models for inference about abundance, distribution and species richness in ecology and their applications to ecology, monitoring and conservation biology.

Gurutzeta Guillera-Arroita is currently a Research Fellow at the University of Melbourne and a member of the UK's National Centre for Statistical Ecology. She is interested in the development and application of statistical models to study wildlife populations, with a focus on optimal study design, as well as on optimal decision-making for conservation problems.

José J. Lahoz-Monfort is currently a Research Fellow at the University of Melbourne and a member of the UK's National Centre for Statistical Ecology. His interests include the modelling of animal demography, population dynamics and species distributions in the context of applied ecology and biodiversity conservation.

Editor: Michael Patten

© 2013 Blackwell Publishing Ltd

SUPPORTING INFORMATION

Analysing and mapping species range dynamics using occupancy models

Marc Kéry, Gurutzeta Guillera-Arroita and José J. Lahoz-Monfort

Journal of Biogeography

Appendix S1 Model selection results.

Model selection results for European crossbill (*Loxia curvirostra*) range dynamics in Switzerland from 2000 to 2007 with two variants: colonization first or extinction first. Model selection was approached sequentially. After an initial comparison of a fully time-dependent and a fully constant model, we first optimized the submodel for detection (p) and then that for initial (first-year) occupancy (ψ_1), colonization (γ) and local extinction probability (ε) (or, in a variant, extinction before colonization). At each step, we retained the best model structure,based on Akaike's information criterion, from the previous step. At the end of each section, all possible single-term deletions were tried (subject to the rules of marginality, McCullagh & Nelder, 1989), and if they resulted in a more parsimonious model, it was retained. Information for the most parsimonious model in each section is printed in bold-face. Column headings denote the number of estimated parameters (Npar), the maximum value of the negative log-likelihood (NLL), AIC score (AIC) and the difference between the AIC and the lowest-AIC model in each subsection of the table (ΔAIC). Effects are denoted Y (year), e (elevation), f (forest cover), d (survey date) and r (route length). Quadratic terms are denoted by a superscript 2 and interactions by dots. Ellipses (...) denote identical structure of a submodel as in the model above.

Model	Npar	NLL	AIC	ΔAIC
(a) Modelling time dependence in parameters				
$\psi_1(.)\gamma(.)\varepsilon(.)p(.)$	4	2630	4729	57.0
$\psi_1(.)\gamma(\mathrm{Y})\varepsilon(\mathrm{Y})p(\mathrm{Y})$	**23**	**2313**	**4672**	**0.0**
(b) Modelling patterns in detection probability (p)				

$\psi_1(.)\gamma(\mathrm{Y})\varepsilon(\mathrm{Y})p(\mathrm{Y})$	23	2313	4672	361.8
$\psi_1(.)\gamma(\mathrm{Y})\varepsilon(\mathrm{Y})p(\mathrm{Y+d})$	24	2303	4653	343.1
$\psi_1(.)\gamma(\mathrm{Y})\varepsilon(\mathrm{Y})p(\mathrm{Y+d+d^2})$	25	2300	4650	340.5
$\psi_1(.)\gamma(\mathrm{Y})\varepsilon(\mathrm{Y})p(\mathrm{Y+d+d^2+r})$	26	2300	4651	341.5
$\psi_1(.)\gamma(\mathrm{Y})\varepsilon(\mathrm{Y})p(\mathrm{Y+e+d+d^2+r})$	27	2232	4518	207.8
$\psi_1(.)\gamma(\mathrm{Y})\varepsilon(\mathrm{Y})p(\mathrm{Y+e+e^2+d+d^2+r})$	28	2224	4503	193.6
$\psi_1(.)\gamma(\mathrm{Y})\varepsilon(\mathrm{Y})p(\mathrm{Y+e+e^2+f+d+d^2+r})$	29	2185	4429	118.9
$\psi_1(.)\gamma(\mathrm{Y})\varepsilon(\mathrm{Y})p(\mathrm{Y+e+e^2+f+f^2+d+d^2+r})$	30	2163	4387	76.9
$\psi_1(.)\gamma(\mathrm{Y})\varepsilon(\mathrm{Y})p(\mathrm{Y+e+e^2+f+f^2+d+d^2+r+e.f})$	31	2159	4380	70.6
$\psi_1(.)\gamma(\mathrm{Y})\varepsilon(\mathrm{Y})p(\mathrm{Y+e+e^2+f+f^2+d+d^2+r+e.f+e.f^2})$	32	2150	4364	54.0
$\psi_1(.)\gamma(\mathrm{Y})\varepsilon(\mathrm{Y})p(\mathrm{Y+e+e^2+f+f^2+d+d^2+r+e.f+e.f^2+e^2.f})$	33	2129	4323	13.2
$\psi_1(.)\gamma(\mathrm{Y})\varepsilon(\mathrm{Y})p(\mathrm{Y+e+e^2+f+f^2+d+d^2+r+e.f+e.f^2+e^2.f+e^2.f^2})$	34	2121	4310	0.0
$\psi_1(.)\gamma(\mathrm{Y})\varepsilon(\mathrm{Y})p(\mathrm{Y+e+e^2+f+f^2+d+d^2+e.f+e.f^2+e^2.f+e^2.f^2})$	**33**	**2121**	**4308**	**-1.9**
(c) Modelling patterns in first-year occupancy probability (ψ_1)				
$\psi_1(.)\gamma(\mathrm{Y})\varepsilon(\mathrm{Y})p(\mathrm{Y+e+e^2+f+f^2+d+d^2+e.f+e.f^2+e^2.f+e^2.f^2})$	33	2121	4308	30.5
$\psi_1(\mathrm{e})\gamma(\mathrm{Y})\varepsilon(\mathrm{Y})p(...)$	34	2106	4280	2.3
$\psi_1(\mathrm{e+e^2})\gamma(\mathrm{Y})\varepsilon(\mathrm{Y})p(...)$	35	2106	4282	4.1
$\psi_1(\mathrm{e+e^2+f})\gamma(\mathrm{Y})\varepsilon(\mathrm{Y})p(...)$	36	2103	4279	1.3
$\psi_1(\mathrm{e+e^2+f+f^2})\gamma(\mathrm{Y})\varepsilon(\mathrm{Y})p(...)$	37	2103	4280	2.9
$\psi_1(\mathrm{e+e^2+f+f^2+e.f})\gamma(\mathrm{Y})\varepsilon(\mathrm{Y})p(...)$	38	2103	4281	4.1
$\psi_1(\mathrm{e+e^2+f+f^2+e.f+e.f^2})\gamma(\mathrm{Y})\varepsilon(\mathrm{Y})p(...)$	39	2100	4278	0.3
$\psi_1(\mathrm{e+e^2+f+f^2+e.f+e.f^2+e^2.f})\gamma(\mathrm{Y})\varepsilon(\mathrm{Y})p(...)$	40	2099	4277	0.0
$\psi_1(\mathrm{e+e^2+f+f^2+e.f+e.f^2+e^2.f+e^2.f^2})\gamma(\mathrm{Y})\varepsilon(\mathrm{Y})p(...)$	41	2099	4279	2.0
$\psi_1(\mathrm{e+f+f^2+e.f+e.f^2})\gamma(\mathrm{Y})\varepsilon(\mathrm{Y})p(...)$	**38**	**2100**	**4276**	**-1.2**
Model selection variant 1: colonisation before extinction				
(d 1) Modelling patterns in colonization probability (γ) before extinction				
$\psi_1(\mathrm{e+f+f^2+e.f+e.f^2})\gamma(\mathrm{Y})\varepsilon(\mathrm{Y})p(...)$	38	2100	4276	44.4
$\psi_1(...)\gamma(\mathrm{Y+e})\varepsilon(\mathrm{Y})p(...)$	39	2091	4260	27.8

$\psi_1(...)\gamma(\mathrm{Y+e+e^2})\varepsilon(\mathrm{Y})p(...)$	40	2085	4249	17.6
$\psi_1(...)\gamma(\mathrm{Y+e+e^2+f})\varepsilon(\mathrm{Y})p(...)$	41	2085	4251	19.6
$\psi_1(...)\gamma(\mathrm{Y+e+e^2+f+f^2})\varepsilon(\mathrm{Y})p(...)$	42	2082	4247	15.4
$\psi_1(...)\gamma(\mathrm{Y+e+e^2+f+f^2+e.f})\varepsilon(\mathrm{Y})p(...)$	43	2076	4238	5.8
$\psi_1(...)\gamma(\mathrm{Y+e+e^2+f+f^2+e.f+e.f^2})\varepsilon(\mathrm{Y})p(...)$	44	2075	4239	6.9
$\psi_1(...)\gamma(\mathrm{Y+e+e^2+f+f^2+e.f+e.f^2+e^2.f})\varepsilon(\mathrm{Y})p(...)$	45	2072	4234	2.2
$\psi_1(...)\gamma(\mathrm{Y+e+e^2+f+f^2+e.f+e.f^2+e^2.f+e^2.f^2})\varepsilon(\mathrm{Y})p(...)$	**46**	**2070**	**4232**	**0.0**
(e 1) Modelling patterns in extinction probability (ε) after colonization				
$\psi_1(...)\gamma(\mathrm{Y+e+e^2+f+f^2+e.f+e.f^2+e^2.f+e^2.f^2})\varepsilon(\mathrm{Y})p(...)$	46	2070	4232	13.5
$\psi_1(...)\gamma(...)\varepsilon(\mathrm{Y+e})p(...)$	47	2065	4224	5.6
$\psi_1(...)\gamma(...)\varepsilon(\mathrm{Y+e+e^2})p(...)$	48	2065	4226	7.6
$\psi_1(...)\gamma(...)\varepsilon(\mathrm{Y+e+e^2+f})p(...)$	49	2061	4220	1.7
$\psi_1(...)\gamma(...)\varepsilon(\mathrm{Y+e+e^2+f+f^2})p(...)$	50	2061	4222	3.6
$\psi_1(...)\gamma(...)\varepsilon(\mathrm{Y+e+e^2+f+f^2+e.f})p(...)$	51	2058	4218	0.0
$\psi_1(...)\gamma(...)\varepsilon(\mathrm{Y+e+e^2+f+f^2+e.f+e.f^2})p(...)$	52	2058	4220	2.0
$\psi_1(...)\gamma(...)\varepsilon(\mathrm{Y+e+e^2+f+f^2+e.f+e.f^2+e^2.f})p(...)$	53	2058	4222	4.0
$\psi_1(...)\gamma(...)\varepsilon(\mathrm{Y+e+e^2+f+f^2+e.f+e.f^2+e^2.f+e^2.f^2})p(...)$	54	2057	4223	4.6
$\psi_1(...)\gamma(...)\varepsilon(\mathrm{Y+e+f+f^2+e.f})p(...)$	50	2058	4216	-1.9
$\psi_1(...)\gamma(...)\varepsilon(\mathrm{Y+e+f+e.f})p(...)$	**49**	**2058**	**4214**	**-3.9**
Model selection variant 2: extinction before colonisation				
(d 2) Modelling patterns in extinction probability (ε) before colonization				
$\psi_1(\mathrm{e+f+f^2+e.f+e.f^2})\gamma(\mathrm{Y})\varepsilon(\mathrm{Y})p(...)$	38	2100	4276	18.7
$\psi_1(...)\gamma(\mathrm{Y})\varepsilon(\mathrm{Y+e})p(...)$	39	2092	4262	4.8
$\psi_1(...)\gamma(\mathrm{Y})\varepsilon(\mathrm{Y+e+e^2})p(...)$	40	2090	4260	2.9
$\psi_1(...)\gamma(\mathrm{Y})\varepsilon(\mathrm{Y+e+e^2+f})p(...)$	**41**	**2088**	**4257**	**0.0**
$\psi_1(...)\gamma(\mathrm{Y})\varepsilon(\mathrm{Y+e+e^2+f+f^2})p(...)$	42	2088	4259	1.5
$\psi_1(...)\gamma(\mathrm{Y})\varepsilon(\mathrm{Y+e+e^2+f+f^2+e.f})p(...)$	43	2087	4261	3.2
$\psi_1(...)\gamma(\mathrm{Y})\varepsilon(\mathrm{Y+e+e^2+f+f^2+e.f+e.f^2})p(...)$	44	2087	4261	3.6

$\psi_1(...)\gamma(\mathrm{Y})\varepsilon(\mathrm{Y+e+e^2+f+f^2+e.f+e.f^2+e^2.f})\,p(...)$	45	2086	4261	3.9
$\psi_1(...)\gamma(\mathrm{Y})\varepsilon(\mathrm{Y+e+e^2+f+f^2+e.f+e.f^2+e^2.f+e^2.f^2})\,p(...)$	46	2085	4262	4.5
(e 2) Modelling patterns in colonization probability (γ) after extinction				
$\psi_1(...)\gamma(...)\varepsilon(\mathrm{Y+e+e^2+f})\,p(...)$	41	2088	4257	37.6
$\psi_1(...)\gamma(\mathrm{Y+e})\varepsilon(...)\,p(...)$	42	2084	4252	32.1
$\psi_1(...)\gamma(\mathrm{Y+e+e^2})\varepsilon(...)\,p(...)$	43	2076	4238	17.9
$\psi_1(...)\gamma(\mathrm{Y+e+e^2+f})\varepsilon(...)\,p(...)$	44	2075	4239	18.8
$\psi_1(...)\gamma(\mathrm{Y+e+e^2+f+f^2})\varepsilon(...)\,p(...)$	45	2073	4237	16.6
$\psi_1(...)\gamma(\mathrm{Y+e+e^2+f+f^2+e.f})\varepsilon(...)\,p(...)$	46	2068	4228	7.8
$\psi_1(...)\gamma(\mathrm{Y+e+e^2+f+f^2+e.f+e.f^2})\varepsilon(...)\,p(...)$	47	2067	4228	8.0
$\psi_1(...)\gamma(\mathrm{Y+e+e^2+f+f^2+e.f+e.f^2+e^2.f})\varepsilon(...)\,p(...)$	48	2062	4221	0.9
$\psi_1(...)\gamma(\mathrm{Y+e+e^2+f+f^2+e.f+e.f^2+e^2.f+e^2.f^2})\varepsilon(...)\,p(...)$	**49**	**2061**	**4220**	**0.0**

REFERENCE

McCullagh, P. & Nelder, J.A. (1989) *Generalized linear models*. Chapman and Hall, London.

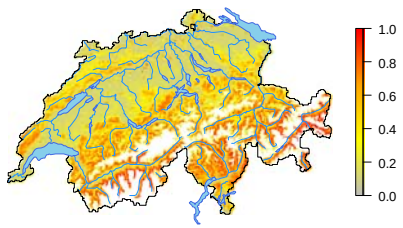
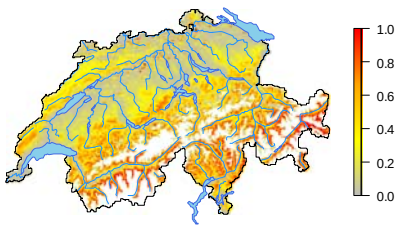
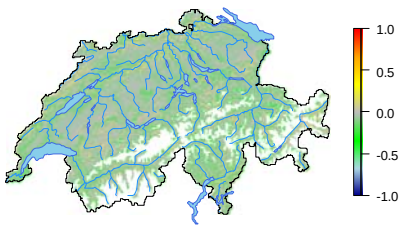
SUPPORTING INFORMATION

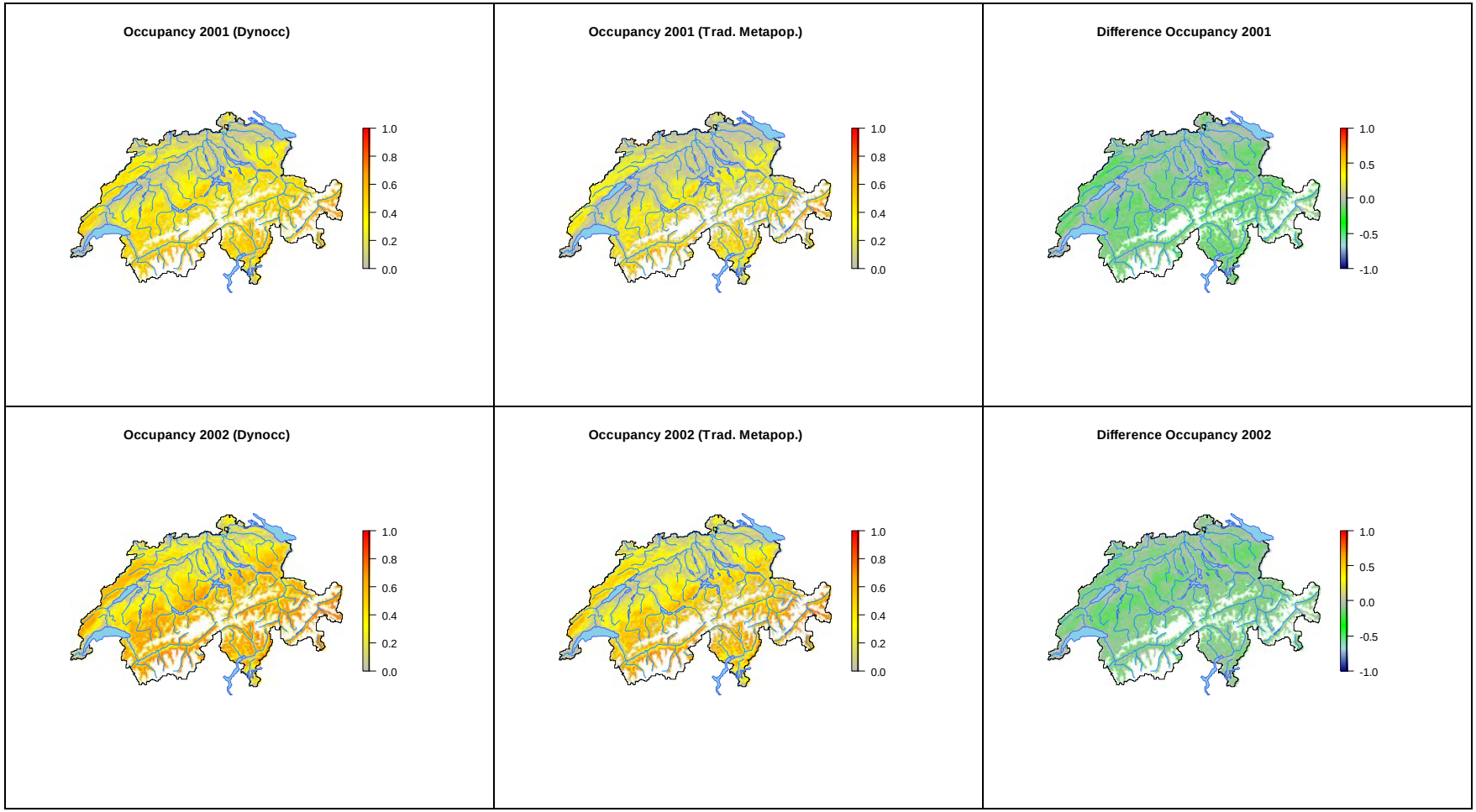
Analysing and mapping species range dynamics using occupancy models

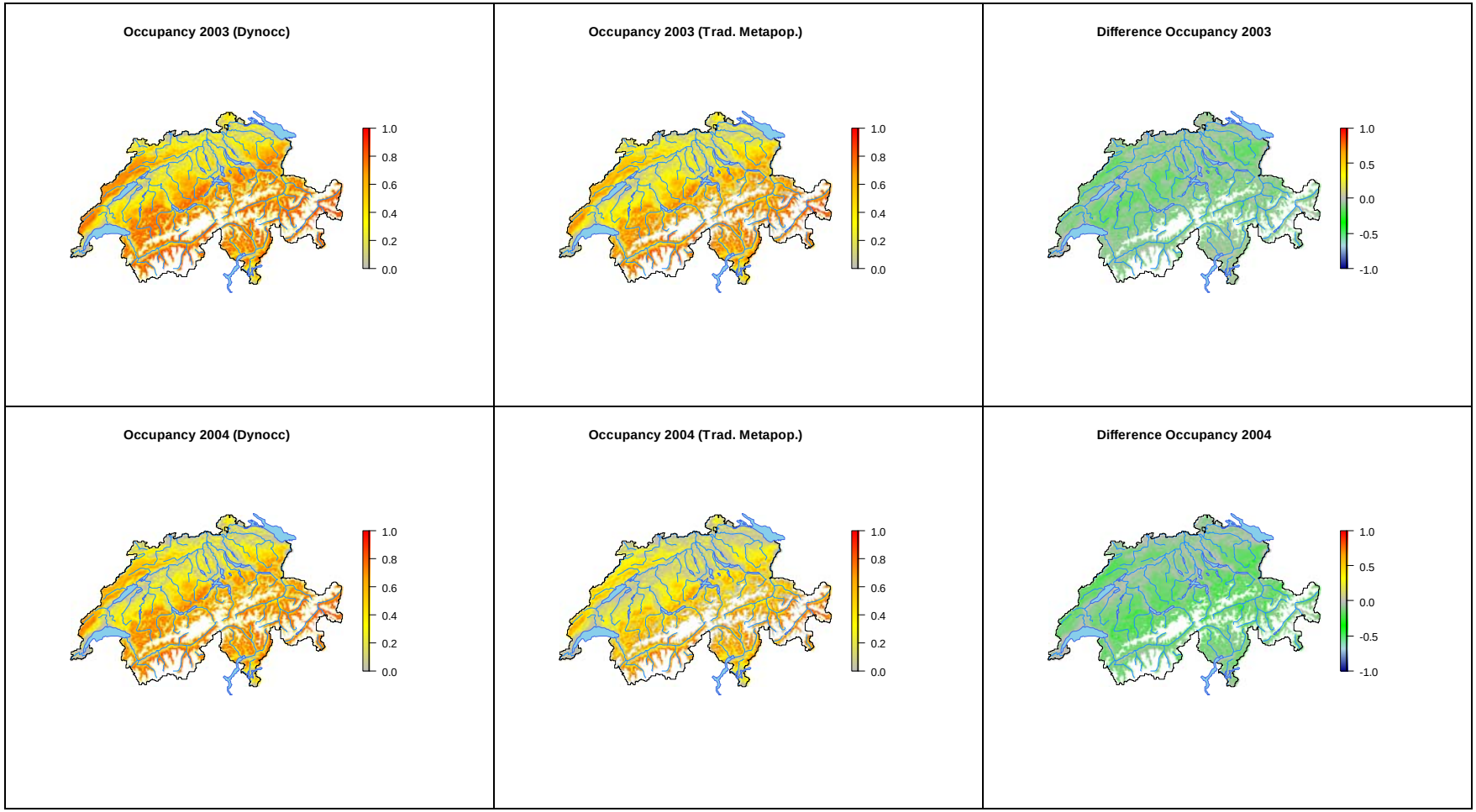
Marc Kéry, Gurutzeta Guillera-Arroita and José J. Lahoz-Monfort

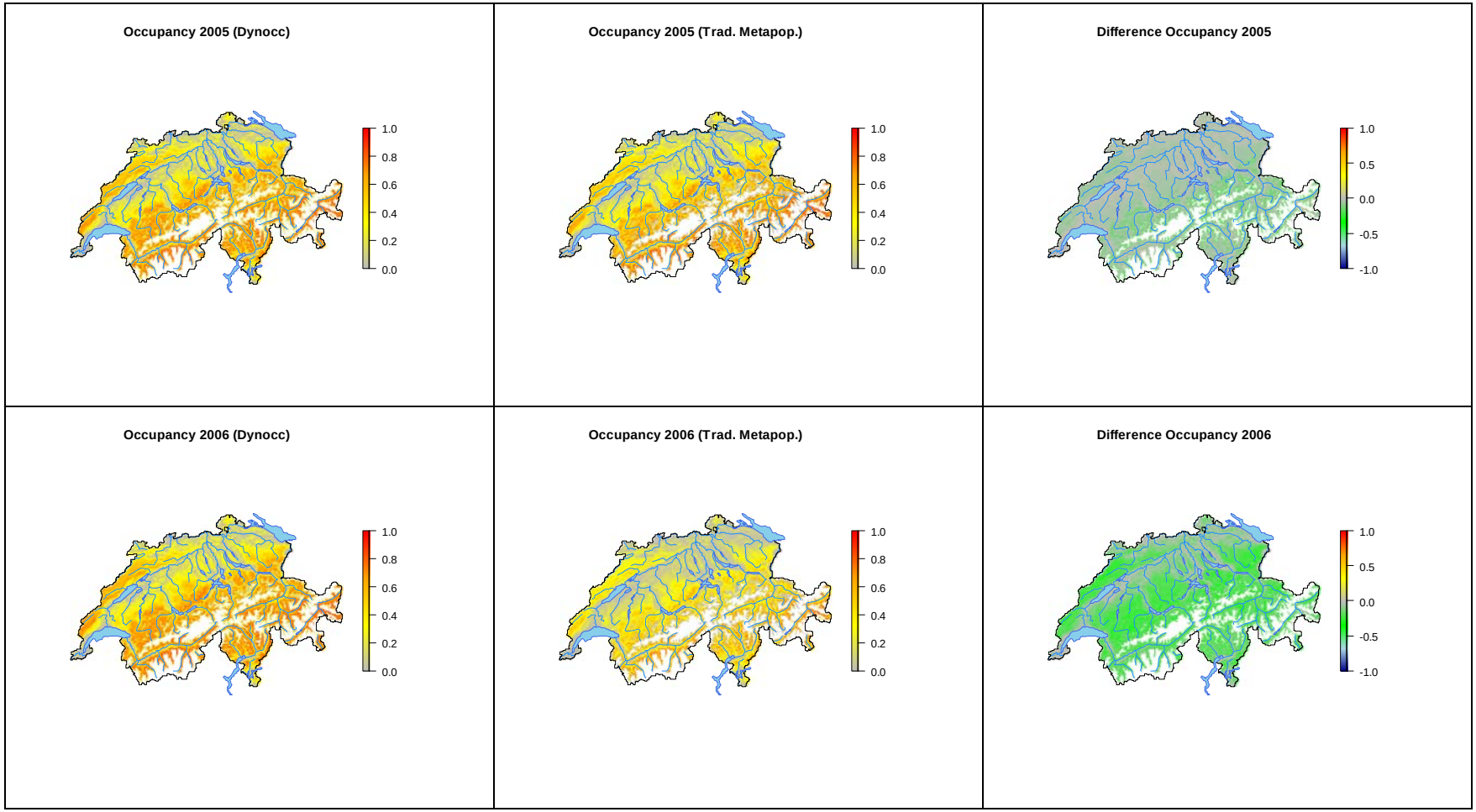
Journal of Biogeography

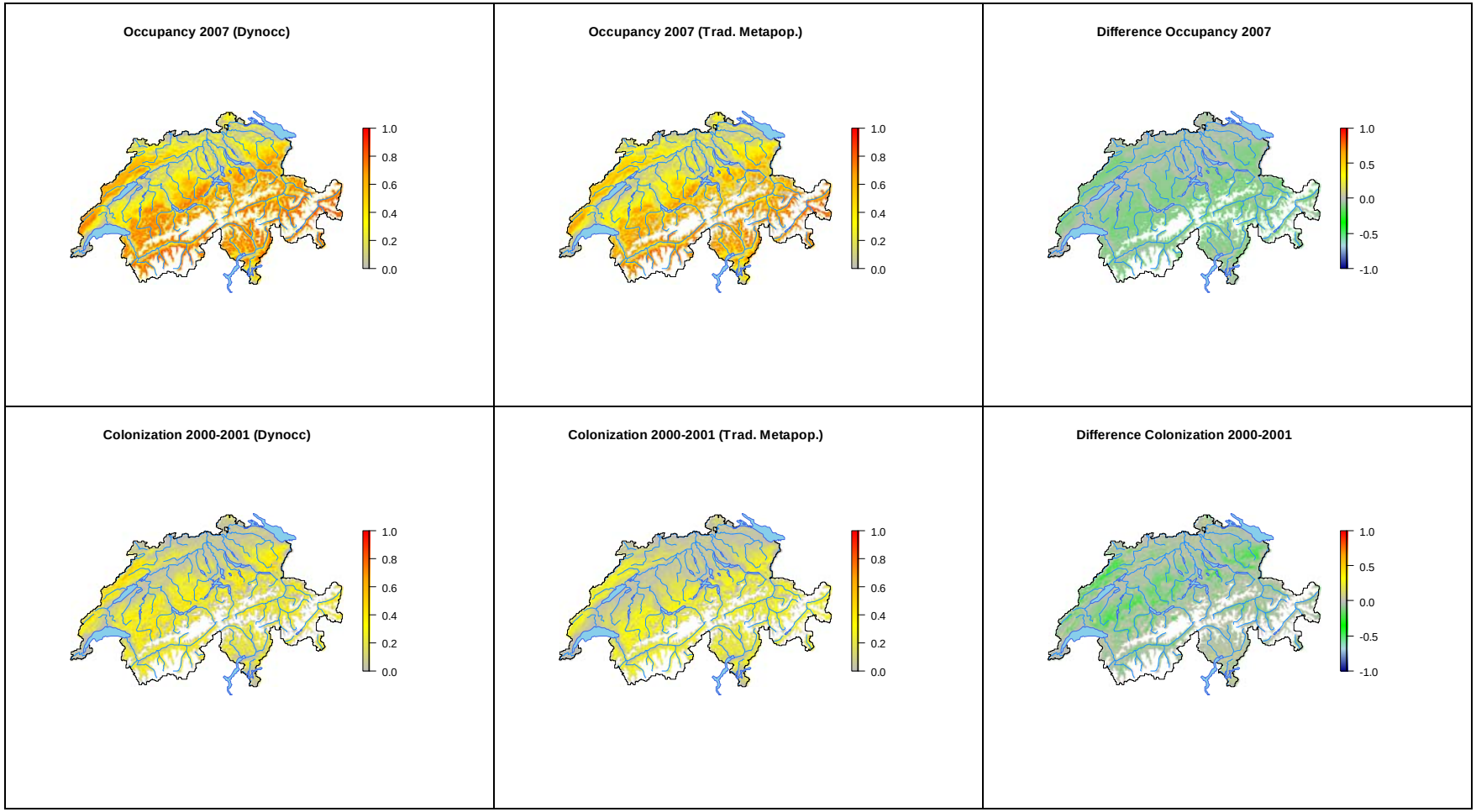
Appendix S2 Range dynamics of the European crossbill (*Loxia curvirostra*) in Switzerland from 2000 to 2007. Predictions of all main parameters under the dynamic occupancy model (left panel) and under a traditional metapopulation model where detection is implicitly assumed to be equal to 1 (central panel). The right panels show the absolute difference between the maps of the two models. Detection maps are shown only for the dynamic occupancy model because there is no detection parameter in the traditional metapopulation model. Areas above 2250 m a.s.l. are masked in white.

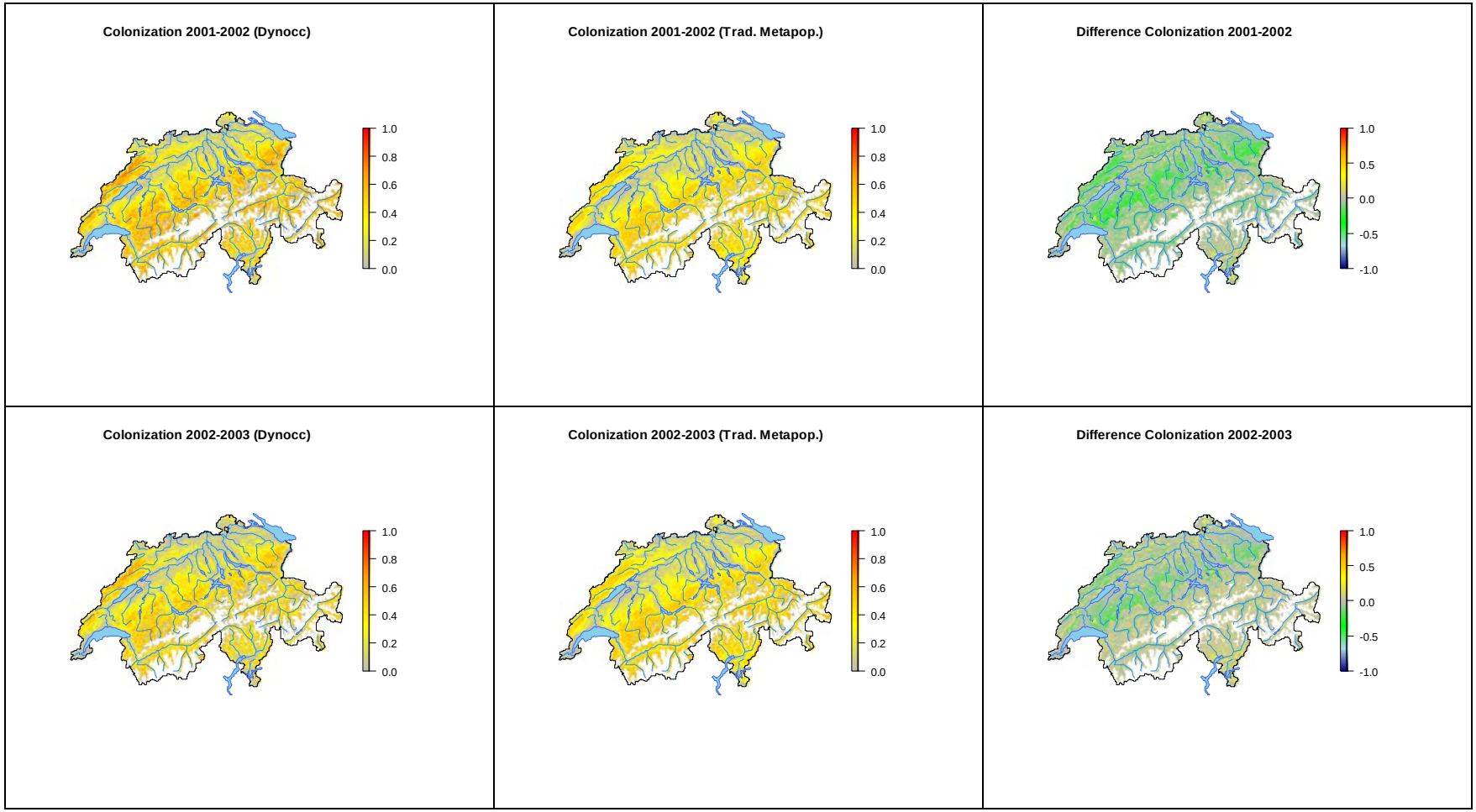
M1: Dynamic occupancy model	M2: Traditional metapopulation model	Map of absolute differences (M2 − M1)
		

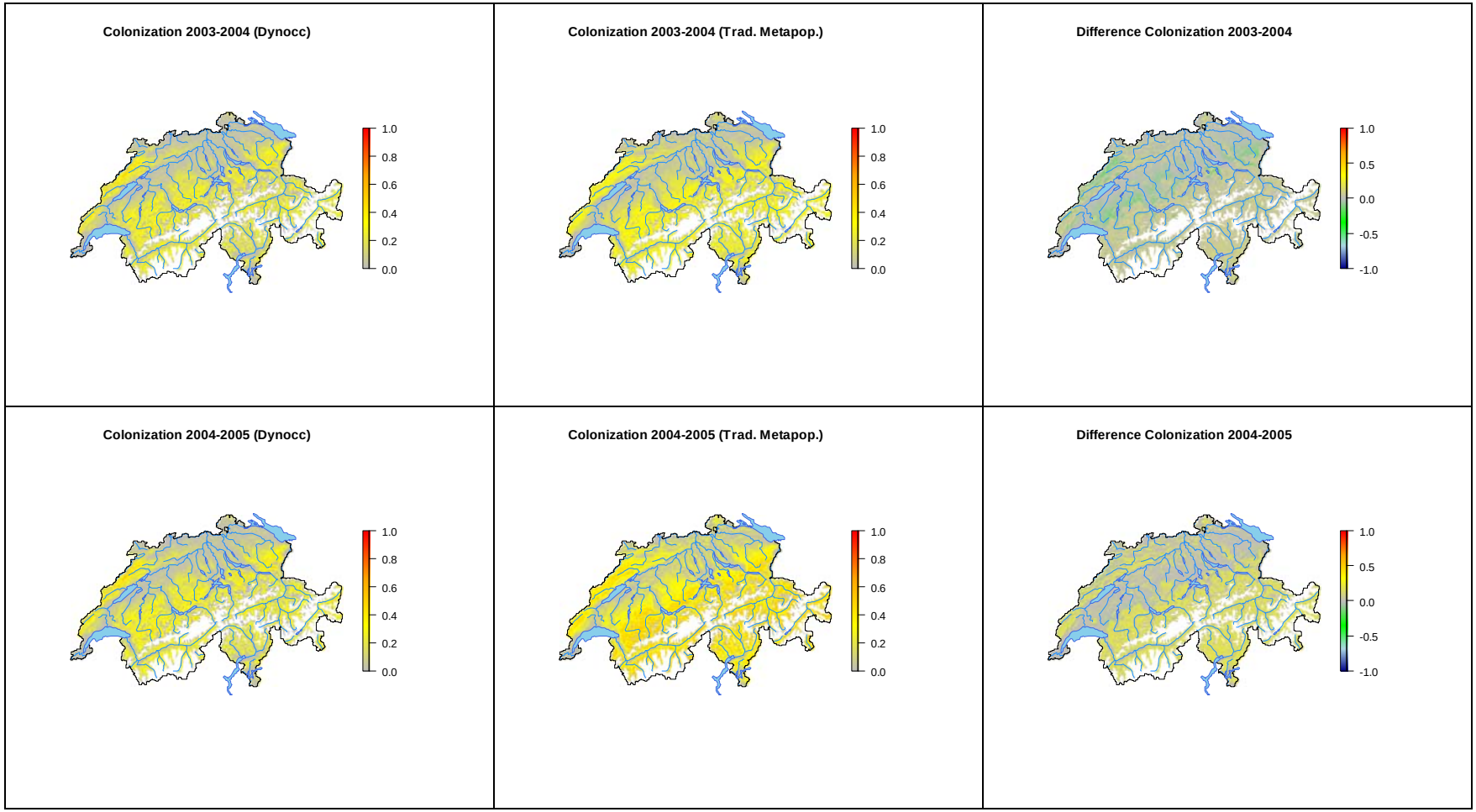

Occupancy 2001 (Dynocc)
1.0
0.8
0.6
0.4
0.2
0.0
Occupancy 2001 (Trad. Metapop.)
1.0
0.8
0.6
0.4
0.2
0.0
Difference Occupancy 2001
1.0
0.5
0.0
-0.5
-1.0
Occupancy 2002 (Dynocc)
1.0
0.8
0.6
0.4
0.2
0.0
Occupancy 2002 (Trad. Metapop.)
1.0
0.8
0.6
0.4
0.2
0.0
Difference Occupancy 2002
1.0
0.5
0.0
-0.5
-1.0

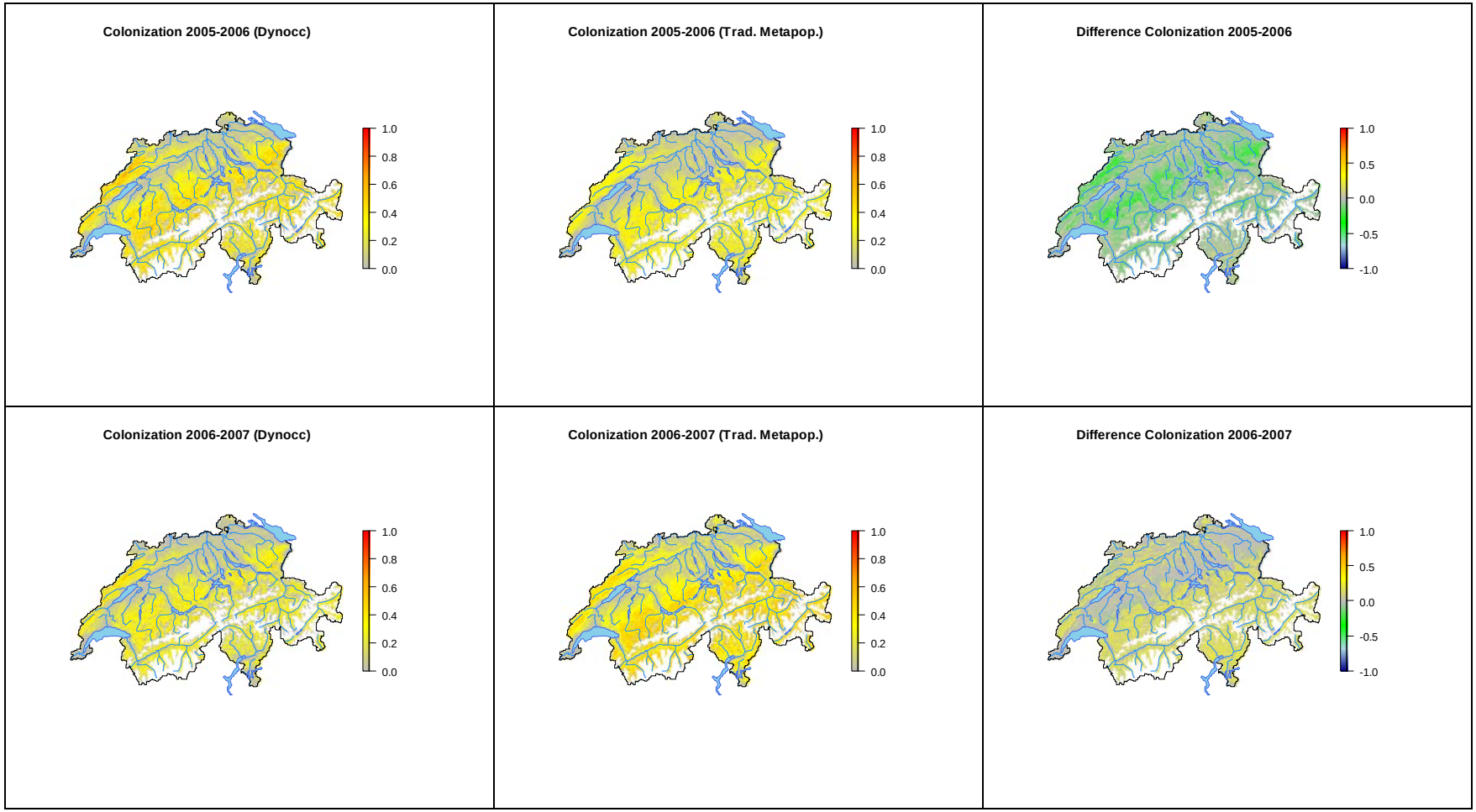

Occupancy 2003 (Dynocc)
1.0
0.8
0.6
0.4
0.2
0.0
Occupancy 2003 (Trad. Metapop.)
1.0
0.8
0.6
0.4
0.2
0.0
Difference Occupancy 2003
1.0
0.5
0.0
-0.5
-1.0
Occupancy 2004 (Dynocc)
1.0
0.8
0.6
0.4
0.2
0.0
Occupancy 2004 (Trad. Metapop.)
1.0
0.8
0.6
0.4
0.2
0.0
Difference Occupancy 2004
1.0
0.5
0.0
-0.5
-1.0

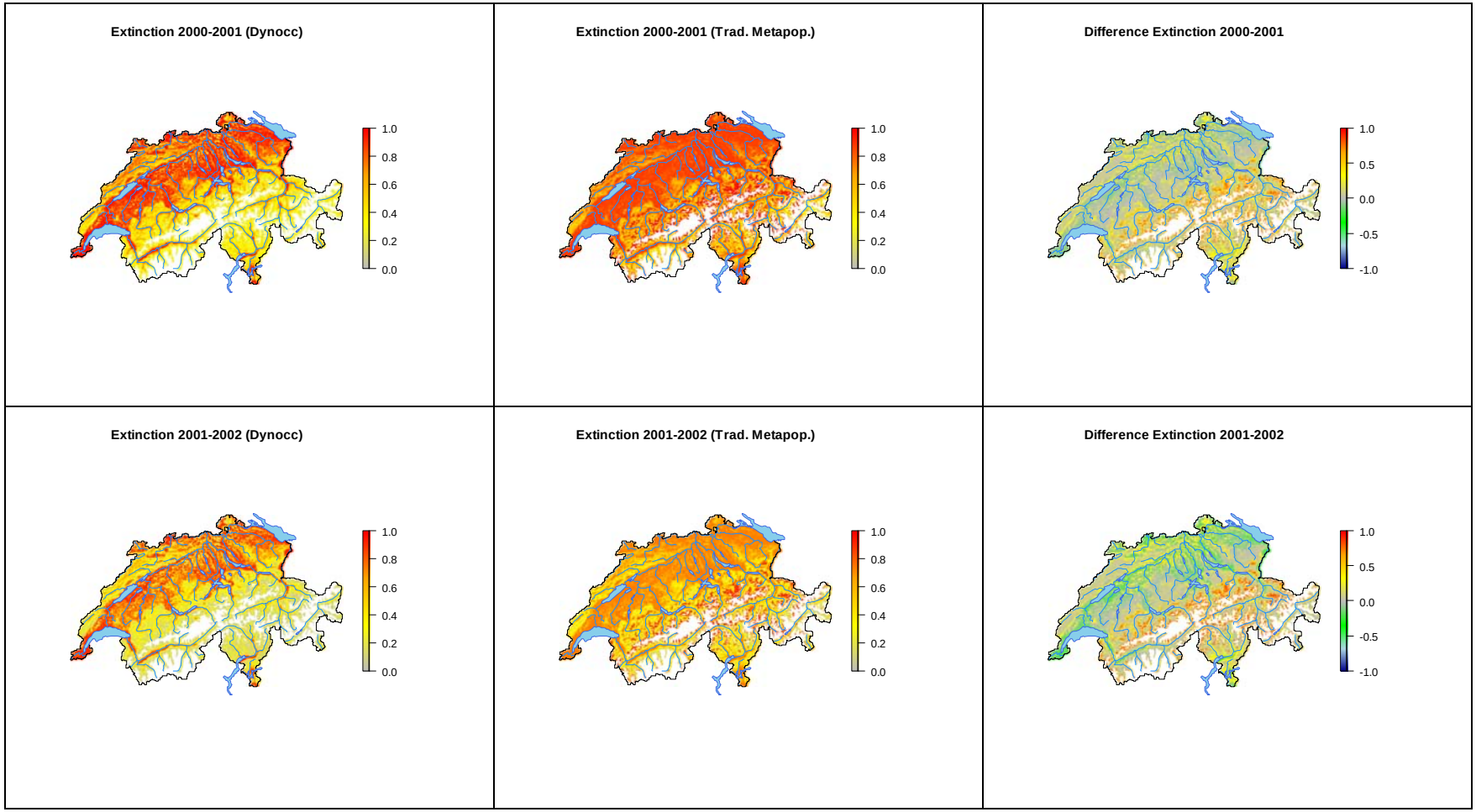

Occupancy 2005 (Dynocc)
1.0
0.8
0.6
0.4
0.2
0.0
Occupancy 2005 (Trad. Metapop.)
1.0
0.8
0.6
0.4
0.2
0.0
Difference Occupancy 2005
1.0
0.5
0.0
-0.5
-1.0
Occupancy 2006 (Dynocc)
1.0
0.8
0.6
0.4
0.2
0.0
Occupancy 2006 (Trad. Metapop.)
1.0
0.8
0.6
0.4
0.2
0.0
Difference Occupancy 2006
1.0
0.5
0.0
-0.5
-1.0

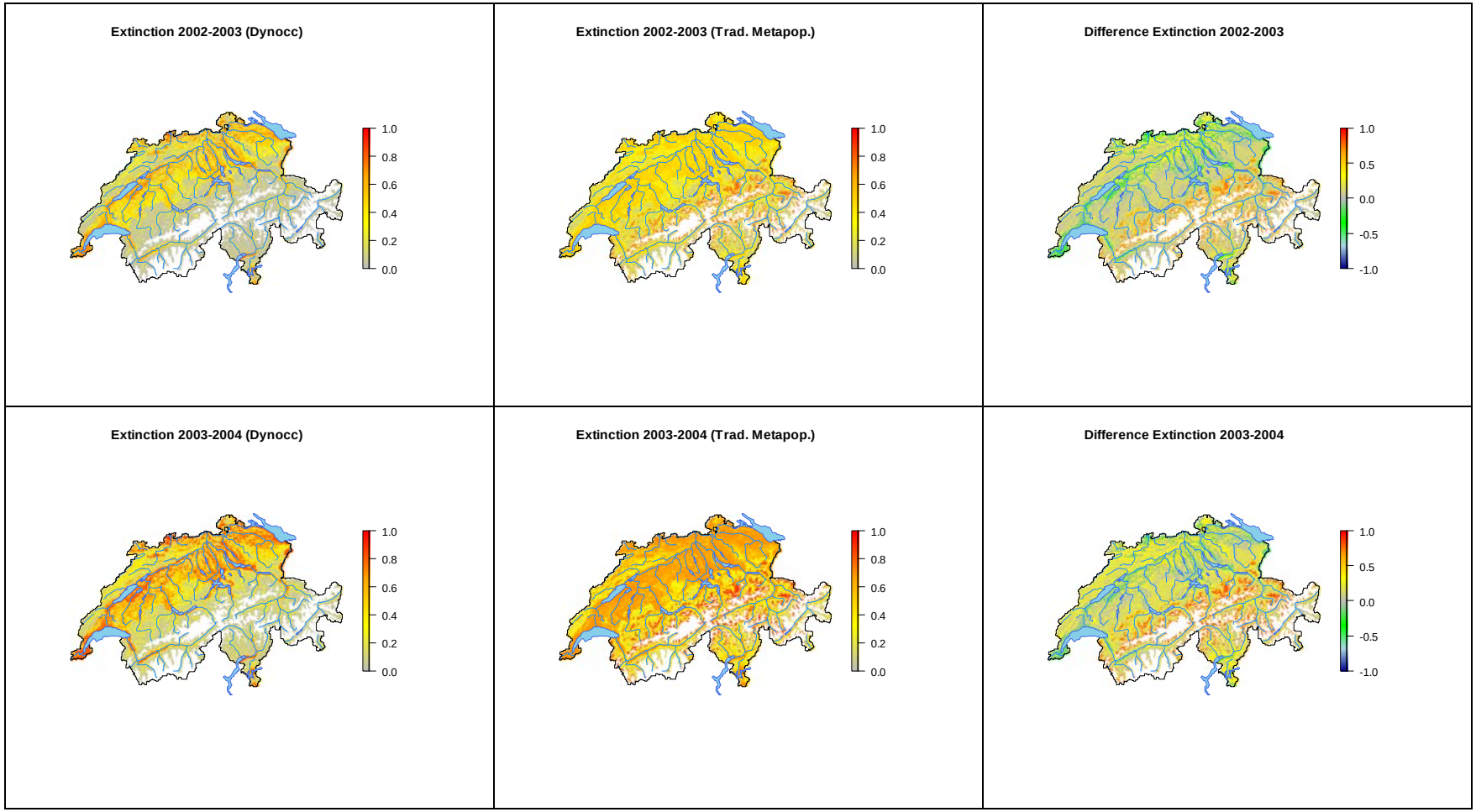

Occupancy 2007 (Dynocc)
1.0
0.8
0.6
0.4
0.2
0.0
Occupancy 2007 (Trad. Metapop.)
1.0
0.8
0.6
0.4
0.2
0.0
Difference Occupancy 2007
1.0
0.5
0.0
-0.5
-1.0
Colonization 2000-2001 (Dynocc)
1.0
0.8
0.6
0.4
0.2
0.0
Colonization 2000-2001 (Trad. Metapop.)
1.0
0.8
0.6
0.4
0.2
0.0
Difference Colonization 2000-2001
1.0
0.5
0.0
-0.5
-1.0

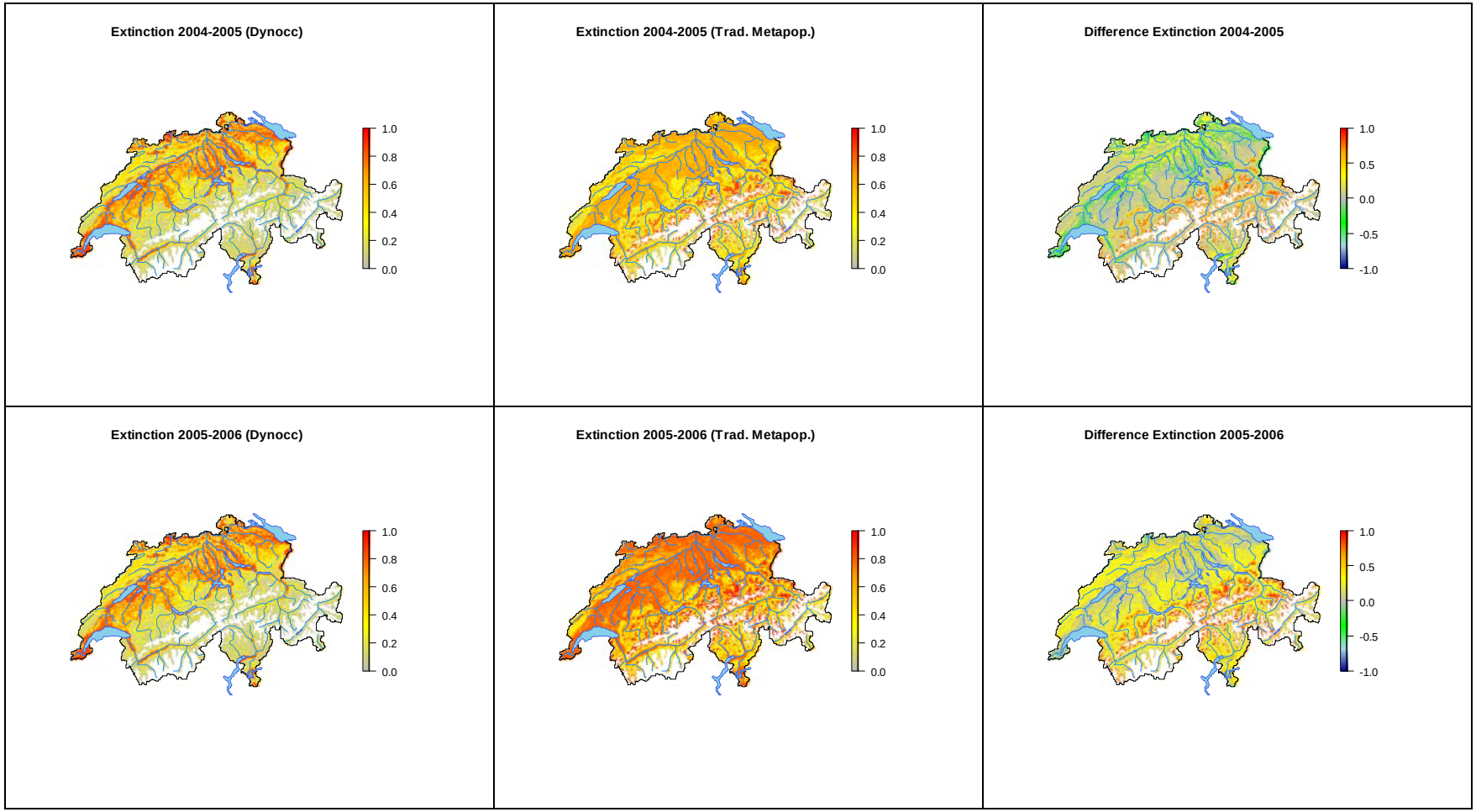

Colonization 2001-2002 (Dynocc)
1.0
0.8
0.6
0.4
0.2
0.0
Colonization 2001-2002 (Trad. Metapop.)
1.0
0.8
0.6
0.4
0.2
0.0
Difference Colonization 2001-2002
1.0
0.5
0.0
-0.5
-1.0
Colonization 2002-2003 (Dynocc)
1.0
0.8
0.6
0.4
0.2
0.0
Colonization 2002-2003 (Trad. Metapop.)
1.0
0.8
0.6
0.4
0.2
0.0
Difference Colonization 2002-2003
1.0
0.5
0.0
-0.5
-1.0

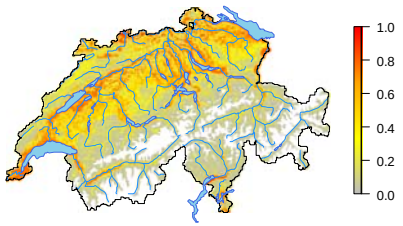

Colonization 2003-2004 (Dynocc)
1.0
0.8
0.6
0.4
0.2
0.0
Colonization 2003-2004 (Trad. Metapop.)
1.0
0.8
0.6
0.4
0.2
0.0
Difference Colonization 2003-2004
1.0
0.5
0.0
-0.5
-1.0
Colonization 2004-2005 (Dynocc)
1.0
0.8
0.6
0.4
0.2
0.0
Colonization 2004-2005 (Trad. Metapop.)
1.0
0.8
0.6
0.4
0.2
0.0
Difference Colonization 2004-2005
1.0
0.5
0.0
-0.5
-1.0

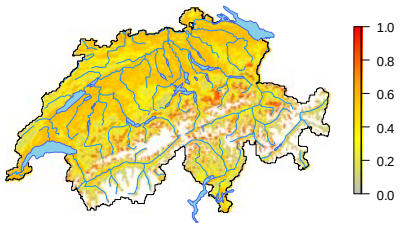

Colonization 2005-2006 (Dynocc)
1.0
0.8
0.6
0.4
0.2
0.0
Colonization 2005-2006 (Trad. Metapop.)
1.0
0.8
0.6
0.4
0.2
0.0
Difference Colonization 2005-2006
1.0
0.5
0.0
-0.5
-1.0
Colonization 2006-2007 (Dynocc)
1.0
0.8
0.6
0.4
0.2
0.0
Colonization 2006-2007 (Trad. Metapop.)
1.0
0.8
0.6
0.4
0.2
0.0
Difference Colonization 2006-2007
1.0
0.5
0.0
-0.5
-1.0

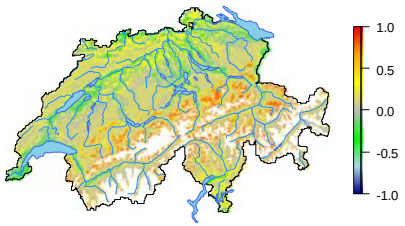

Extinction 2000-2001 (Dynocc)
1.0
0.8
0.6
0.4
0.2
0.0
Extinction 2000-2001 (Trad. Metapop.)
1.0
0.8
0.6
0.4
0.2
0.0
Difference Extinction 2000-2001
1.0
0.5
0.0
-0.5
-1.0
Extinction 2001-2002 (Dynocc)
1.0
0.8
0.6
0.4
0.2
0.0
Extinction 2001-2002 (Trad. Metapop.)
1.0
0.8
0.6
0.4
0.2
0.0
Difference Extinction 2001-2002
1.0
0.5
0.0
-0.5
-1.0

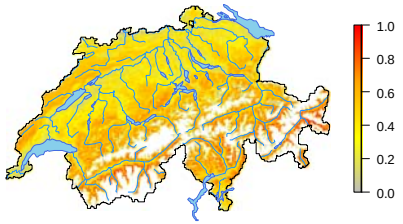

Extinction 2002-2003 (Dynocc)
1.0
0.8
0.6
0.4
0.2
0.0
Extinction 2002-2003 (Trad. Metapop.)
1.0
0.8
0.6
0.4
0.2
0.0
Difference Extinction 2002-2003
1.0
0.5
0.0
-0.5
-1.0
Extinction 2003-2004 (Dynocc)
1.0
0.8
0.6
0.4
0.2
0.0
Extinction 2003-2004 (Trad. Metapop.)
1.0
0.8
0.6
0.4
0.2
0.0
Difference Extinction 2003-2004
1.0
0.5
0.0
-0.5
-1.0


Extinction 2004-2005 (Dynocc)
1.0
0.8
0.6
0.4
0.2
0.0
Extinction 2004-2005 (Trad. Metapop.)
1.0
0.8
0.6
0.4
0.2
0.0
Difference Extinction 2004-2005
1.0
0.5
0.0
-0.5
-1.0
Extinction 2005-2006 (Dynocc)
1.0
0.8
0.6
0.4
0.2
0.0
Extinction 2005-2006 (Trad. Metapop.)
1.0
0.8
0.6
0.4
0.2
0.0
Difference Extinction 2005-2006
1.0
0.5
0.0
-0.5
-1.0


Extinction 2006-2007 (Dynocc)
1.0
0.8
0.6
0.4
0.2
0.0

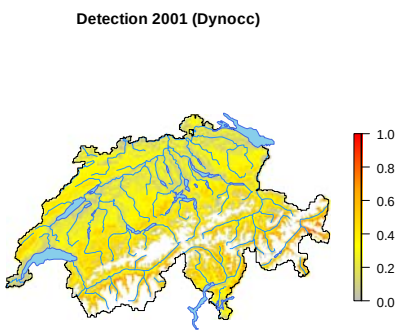
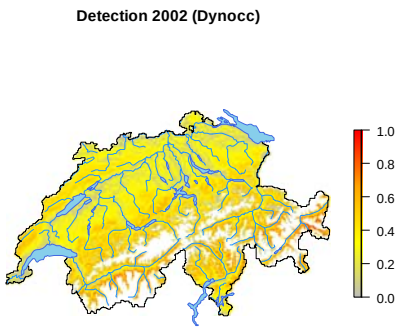

Extinction 2006-2007 (Trad. Metapop.)
1.0
0.8
0.6
0.4
0.2
0.0

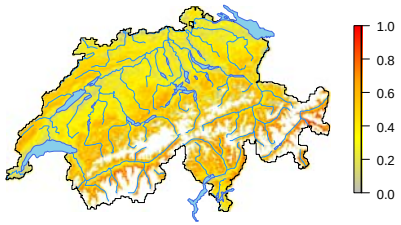
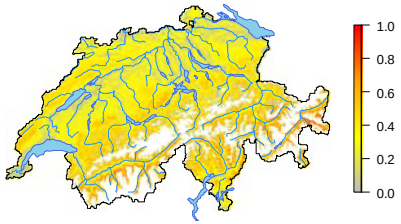

Difference Extinction 2006-2007
1.0
0.5
0.0
-0.5
-1.0

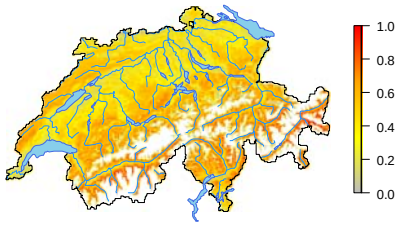
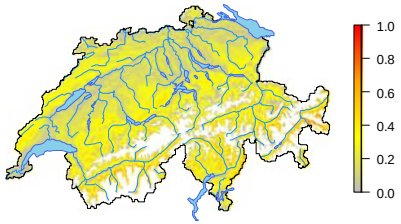

Detection 2000 (Dynocc)
1.0
0.8
0.6
0.4
0.2
0.0

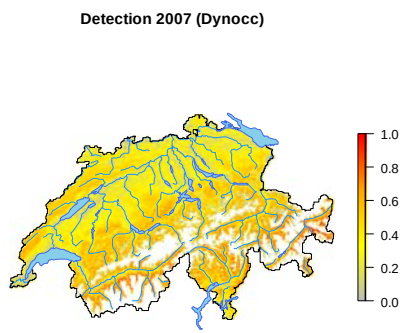
NA

NA

 Detection 2001 (Dynocc) 1.0 0.8 0.6 0.4 0.2 0.0	NA	NA
 Detection 2002 (Dynocc) 1.0 0.8 0.6 0.4 0.2 0.0	NA	NA

Detection 2003 (Dynocc) 	NA	NA
Detection 2004 (Dynocc) 	NA	NA

Detection 2005 (Dynocc) 	NA	NA
Detection 2006 (Dynocc) 	NA	NA

	NA	NA