Earth and field observations underpin metapopulation dynamics in complex landscapes: Near-term study on carabids

Jonathan Giezendanner, Damiano Pasetto, Javier Perez-Saez, Cristiana Cerrato, Ramona Viterbi, Silvia Terzago, Elisa Palazzi, and Andrea Rinaldo

*Laboratory of Ecohydrology, École Polytechnique Fédérale de Lausanne, CH-1015 Lausanne, Switzerland; ‡Gran Paradiso National Park, 10135 Torino, Italy; §National Research Council (CNR), Institute of Atmospheric Sciences and Climate, 10133 Torino, Italy; and †Dipartimento di Ingegneria Civile Edile e Ambientale (DICEA), Università di Padova, 35131 Padova, Italy

Contributed by Andrea Rinaldo, April 4, 2020 (sent for review November 27, 2019; reviewed by Robert J. Fletcher, Jr and David A. Keith)

Understanding risks to biodiversity requires predictions of the spatial distribution of species adapting to changing ecosystems and, to that end, Earth observations integrating field surveys prove essential as they provide key numbers for assessing landscape-wide biodiversity scenarios. Here, we develop, and apply to a relevant case study, a method suited to merge Earth/field observations with spatially explicit stochastic metapopulation models to study the near-term ecological dynamics of target species in complex terrains. Our framework incorporates the use of species distribution models for a reasoned estimation of the initial presence of the target species and accounts for imperfect and incomplete detection of the species presence in the study area. It also uses a metapopulation fitness function derived from Earth observation data subsuming the ecological niche of the target species. This framework is applied to contrast occupancy of two species of carabids (*Pterostichus flavofemoratus*, *Carabus depressus*) observed in the context of a large ecological monitoring program carried out within the Gran Paradiso National Park (GPNP, Italy). Results suggest that the proposed framework may indeed exploit the hallmarks of spatially explicit ecological approaches and of remote Earth observations. The model reproduces well the observed in situ data. Moreover, it projects in the near term the two species’ presence both in space and in time, highlighting the features of the metapopulation dynamics of colonization and extinction, and their expected trends within verifiable timeframes.

Monitoring, understanding, and predicting changes in biodiversity and in the geographical distribution of species have become increasingly important to mainstream ecology (1–6). As climate change, possibly human induced (7), redefines the geographical distribution of species (8, 9), alterations in habitat structures appear ubiquitously as dominant drivers (10, 11). Improving our capability to reliably describe and monitor the changes in biodiversity, species distribution, and habitat in space and time is thus fundamental toward informed decisions on landscape, biodiversity, and conservation management (12–14). To that end, theory can be instructive (1, 15–25), in particular if meant to link directly Earth observations (EOs) to in situ ecological data on presence, absence, and abundance providing estimates of the geographical distribution of a species within a landscape (12, 26, 27). Wherever the movement of individuals across structured habitats is concerned, by necessity metapopulation dynamics are relevant. Within that context, understanding the links between a complex and heterogeneous landscape acting as the ecological substrate and the hosted population dynamics is central (4, 28–31). In complex topographies such as mountainous regions shaped by fluvial processes (32) the core of landscape dynamics in complex landscapes: Near-term study on carabids

Significance

We propose a general framework for population viability analysis aimed at near-term ecological forecasting of suitable focus species’ presence in heterogeneous landscapes. Our method is based on remotely acquired and objectively manipulated Earth observations, coupled with hard-gained field evidence and landscape-explicit stochastic models of population dynamics. The integration of field data on the geographical niche plays a major role in calibration and in defining the impact of imperfect detection of species’ presence. Our case study deals with the persistence of two carabid species in the Gran Paradiso National Park (Italy), singled out as important biodiversity indicators. Results suggest the effectiveness of our framework in near-term forecasting the dynamics of species presence in space/time, leading to testable predictions of ecological responses.
probability of presence of a species at equilibrium (43, 44), implying that the presence of a species in the landscape reflects its niche while other processes, like dispersal, colonization, or migration, do not have major effects (45). In species distribution models, most processes are implicitly incorporated into the habitat suitability and hence into the species’ niche (46). Physical barriers, sources–sinks, and population and metapopulation dynamics are thus not modeled explicitly (45), although models have been developed to integrate certain features of the relevant dynamics with SDMs (21, 35, 47, 48).

Here, we focus explicitly on the linkage of attributes of the landscape matrix with stochastic metapopulation dynamics (6, 16, 20, 25, 49) where the landscape is subdivided into cells, as is commonplace when EO rasters are used. Each cell constitutes a patch sensu Hanski, from which a focus species can colonize other empty patches or go extinct. This allows us to characterize precisely the abiotic conditions within each cell based on EO data and derive a species’ fitness, i.e., the local habitat suitability. Once the niche of species is estimated, the metapopulation model can, in time, reflect the unfolding colonization of geographical regions and account explicitly for physical barriers that prevent it. Moreover, a method estimating this niche by calibrating the parameters driving processes relevant to metapopulations has to account for the gaps in data in space and time as well as the imperfect detection of species (31, 50, 51). Our method merges the use of SDMs (52) with the demands of stochastic metapopulation models (1, 15) (see SI Appendix, section 4 and Materials and Methods).

EO data are employed to characterize habitat suitability in space and time by reproducing observed spatial and temporal presences of the in situ data for the focus species. As SDMs prove particularly apt at depicting a species’ static probability of presence, we make use of a reasoned (Materials and Methods) generalized linear model (GLM) to estimate the initial presence in the landscape. A stochastic metapopulation framework (spatially explicit patch occupancy model [SPOM]) (6, 16, 25, 49, 53) (Materials and Methods) is employed to dynamically model the presence of the focus species in time (23, 52). A suitably modified metapopulation fitness (i.e., suitability of the habitat for the focus species) is defined based on the core concept of GLMs, which we calibrate by using an iterated filtering (IF) scheme (54), a technique particularly powerful when applied to partially observed Markov processes, tailored to accommodate partial and imperfect observation data in space and time (50, 51, 53), without the need for observations in each modeled path or timestep. An application to a case study then showcases the capabilities of the framework by replicating the observed spatial and temporal dynamics of two species of carabids viewed as significant biodiversity indicators—an outcome of a large field observations program in the Gran Paradiso National Park in the Italian Alps (Materials and Methods) (Figs. 1 and 2).

Results

The output of the model (Materials and Methods) is shown for both focus species, Pterostichus flavofemoratus and Carabus depressus, simulated via properly calibrated parameters (Table 1 and Materials and Methods) contrasting field sampling and by layers of a filtering algorithm (SI Appendix, section 4 and Materials and Methods). Note that the process-based results on the species are to be read with caution given the relatively short case study (Discussion) and mainly show the possible extent of interpretation of the method.

Fig. 3 shows the two species’ simulated average occupancy in the landscape for the sampled years, computed from 480 replicates of the stochastic algorithm tackling metapopulation dynamics (Materials and Methods), along with the relative changes in occupancy highlighting possible differences arising in time. For both species, the modeled average occupancy matches systematically the observed ratios from pitfall trapping samples (Materials and Methods) in nearly all plots. This suggests that the calibration process identified a parameter combination that results in an appropriate fitness function driving the metapopulation dynamics. Fluctuations occur around multiple stable cores, where the pressure of colonization lessens. The presence of C. depressus appears more patchy than the one of P. flavofemoratus, with clusters of high occupancy and surrounding lower occupancy maintained by the colonization process.

For both species, the modeled average occupancy matches systematically the observed ratios from pitfall trapping samples (Materials and Methods) in nearly all plots. This suggests that the calibration process identified a parameter combination that results in an appropriate fitness function driving the metapopulation dynamics. Fluctuations occur around multiple stable cores, where the pressure of colonization lessens. The presence of C. depressus appears more patchy than the one of P. flavofemoratus, with clusters of high occupancy and surrounding lower occupancy maintained by the colonization process.

Fig. 4 A and B shows a comparison between the modeled (k, red line with simulation envelope) and in situ (black dots) number of times the species was observed in each plot and
year. The modeled $k$ (the number of times per year and plot a species is counted) is derived by drawing 1,000 samples from the binomial distribution $P(k = K|W) = (W = 1)B(p_{TP}, n) + (W = 0)B(p_{FP}, n)$ for each particle, plot, and year (with the values of estimated true positive $[p_{TP}]$ and false positive $[p_{FP}]$ displayed in Table 1), knowing the output of the model ($W$, the binary occupancy state) and the number of times the plot was sampled in the given year ($n$). The displayed median $k$ and the simulation envelope are computed from the joint distribution.

The modeled $k$ of $P. flavofemoratus$ encapsulates the in situ data within errors in almost all of the plots, with a few plots displaying a large envelope, but with the median usually close to the observed data. Because for $C. depressus$ the uncertainty on the false negative data ($1 - p_{TP}$, Table 1) is higher, this translates directly into higher uncertainty above the median in the modeled value of $k$. Decreasing and successively increasing trends in the observations, as seen in Fig. 4A and B, plot $f$ for the Gran Piano site and plot $e$ for Vaudalletaz, are well captured by the modeled ones through changes in the simulation envelope and, in some cases, by significant variations in the median.

Fig. 4C and D shows the distribution of the data within the modeled simulation envelope. For both species, most of the data are around the 50% quantile, with 88.3% ($P. flavofemoratus$) and 85.0% ($C. depressus$) of the observations within 95% of the model envelope. For $P. flavofemoratus$, the highest source of deviation from the median is at quantile 100%, where the model underestimates the certainty of occupancy of the plot. Overestimation of the presence (left-hand side of the plot in Fig. 4C and D) is less frequent. For $C. depressus$ there appears to be a notable underestimation of the occupancy (right-hand side of the plot in Fig. 4D).

Fig. 3C shows the two focus species’ occupancy of the landscape in time. Overall, the two species appear to be stable, as expected over these timescales of observations, with a significant increase in presence over time for $P. flavofemoratus$ and a decrease for $C. depressus$. The observed trends can mostly be related to the changes in fitness (SI Appendix, section 5), although they may be observed only at the landscape level, with little to no trends observed at the pixel level. The average occupancy appears to be lower for $C. depressus$ than for $P. flavofemoratus$, owing to a more clustered fitness function along with a lower base fitness.

As all cofactors have been normalized (SI Appendix, section 2), a unit change in a parameter directly influences the contribution of the cofactors to the fitness relative to the other explanatory variables. For $P. flavofemoratus$, most of the variables

### Table 1. Calibrated parameter values for the two species of interest

|                  | $P. flavofemoratus$ | $C. depressus$ |
|------------------|---------------------|---------------|
| $p_{TP}$         | 79%                 | 63%           |
| $p_{FP}$         | 1%                  | 1%            |
| $c$              | 187.6               | 733.1         |
| $E$              | $1.5 \times 10^{-3}$| $1.2 \times 10^{-2}$ |
| $D$              | 189.9               | 173.0         |
| $\beta_{Temperature}$ | 2.1               | 3.4            |
| $\beta_{Wetness}$ | 3.6                | −0.5          |
| $\beta_{Greenness}$ | 3.2               | −0.3          |
| $\beta_{Forest}$  | 0.4                 | 0.8           |
| $\beta_{Northness}$ | 1.54              | −3.61         |
| $\beta_{Eastness}$ | −2.4             | −2.7          |
| $\beta_{Slope}$   | −2.5                | −4.0          |
| $\beta_{Forest}$  | −2.7                | 1.4           |
| $\beta_{Forest}$  | −3.8                | 1.1           |

The fitness function multiplies each covariate by the parameter in the following way (cf. Eq. 4, Materials and Methods):

$$f_i(t) = \frac{1}{1 + \exp \left( -\alpha_0 - \sum_j \beta_j x_j(t) \right)} \in (0, 1).$$

The first row consists of the parameters meant for the relation between the measures and the model (probability of true and false positives) and the main parameters of the metapopulation model (colonization and extinction rates and dispersal distance). The second and third rows show the parameters generating the logistic fitness function. The second row shows the base fitness value ($\alpha_0$) and the parameters multiplying a time-varying covariate (temperature and tasseled cap indexes; Fig. 2A–D), and the third row shows the parameters multiplying the static explanatory variables (eastness, northness, slope, and presence of forest Fig. 2E–G).
A. C. depressus appears to be most sensitive to the presence of vegetation (greenness), with a negative parameter value explaining the observed trend.

Discussion

The proposed framework shows how EO data can be selected and used in conjunction with partially observed in situ data to model the presence of a metapopulation in space and time. The resulting calibration process converges to parameter combinations that generate initial spatial occupancy comparable to the one proposed by the SDM (SI Appendix, Fig. S15) and subsume the spatial dynamics of the metapopulation processes.

Static variables such as slope and aspect (northness and eastness) appear to be essential for characterizing the species’ niche. However, temporally varying state variables emerge as key players in predicting temporal trends in species occupancy. Because the static variables do not change in time, the observed changes are stemming only from two sources: 1) colonization and extinction processes related to the metapopulation parameters [c, e, and D (c.f. Materials and Methods, Metapopulation model - SPOM)], and 2) time-varying climatic drivers. The metapopulation parameters explain most of the stochastic variations, given that the observed trends are not due to the model rebalancing the occupancy to match parameter fluctuations (see SI Appendix, section 4 and Materials and Methods on model spin-up). The observed trends are therefore related to the variation of the time-dependent drivers. At the plot level, the simulated results show that the model is capable of reproducing, within a 95%
simulation envelope, most of the species’ dynamics (Fig. 4). Similarly, the modeled envelope is produced by the SPOM stochastic structure, and the variations are induced by the variation in time of the covariates. An example is given in plot f of the valleys of Gran Piano and Lauson (Fig. 4 A and B), where the model envelope increases as the observed presence increases.

The results at plot level corroborate the assumption that simulations reflect rather closely the observations, yielding a global interpretation of the species presence. The stochastic variations in occupancy occur around stable cores, which calls to mind the classic metapopulation approach centered around patches, but highlights the importance of modeling the landscape explicitly, as most of the changes occur at the edge of the stable cores or between them and not within a patch of homogeneous quality. These fluctuations, occurring in the marginal areas of habitat patches, could lead to the identification of leading and trailing edges if propagated in time (25, 55, 56). At a landscape level, little changes are observed considering the time frame and the domain’s scale. In a period of 8 y, perturbations of the environment in a protected area are hardly extreme and mostly observed in terms of variations in the number of individuals captured (to be interpreted with caution because of the imperfect detection) and not in terms of major shifts in geographical space of the species’ distribution. This is precisely what is expected from near-term ecological predictions (14).

Because sampling reflects the species’ activity–density, a species observed and accounts for this in the calibration. This problem is that in 37% of the cases of its level of activity. Here, the model follows the assumption that in 37% of the cases C. depressus was indeed present but not observed and accounts for this in the calibration. This problem is less evident the case of P. flavofemoratus, where over all plots the largest ratio of observed to sampled is higher.

The two species’ fitness (SI Appendix, Figs. S19 and S24) are closely correlated and the lower deviations in occupancy observed in Fig. 3C. The model evaluates the probability of false negative C. depressus to be quite high (37%) compared to the one of P. flavofemoratus (21%). The model here reflects the sampling (SI Appendix, Fig. S8), where C. depressus has been observed fewer times than P. flavofemoratus. Nevertheless, a question remains on whether the occurrences of species not observed reflect true absence or undersampling. Because sampling reflects the species’ activity–density, a species less active at certain times of the year will be less present in the observed. However, the model is meant to reflect the species average presence in the landscape during the year, regardless of its level of activity. Here, the model follows the assumption that in 37% of the cases C. depressus was indeed present but not observed and accounts for this in the calibration. This problem is less evident the case of P. flavofemoratus, where over all plots the largest ratio of observed to sampled is higher.

The two species’ fitness (SI Appendix, Figs. S19 and S24) are closely correlated and the lower deviations in occupancy observed in Fig. 3C. The model evaluates the probability of false negative C. depressus to be quite high (37%) compared to the one of P. flavofemoratus (21%). The model here reflects the sampling (SI Appendix, Fig. S8), where C. depressus has been observed fewer times than P. flavofemoratus. Nevertheless, a question remains on whether the occurrences of species not observed reflect true absence or undersampling. Because sampling reflects the species’ activity–density, a species less active at certain times of the year will be less present in the observations. However, the model is meant to reflect the species average presence in the landscape during the year, regardless of its level of activity. Here, the model follows the assumption that in 37% of the cases C. depressus was indeed present but not observed and accounts for this in the calibration. This problem is less evident the case of P. flavofemoratus, where over all plots the largest ratio of observed to sampled is higher.

The two species’ fitness (SI Appendix, Figs. S19 and S24) are closely correlated and the lower deviations in occupancy observed in Fig. 3C. The model evaluates the probability of false negative C. depressus to be quite high (37%) compared to the one of P. flavofemoratus (21%). The model here reflects the sampling (SI Appendix, Fig. S8), where C. depressus has been observed fewer times than P. flavofemoratus. Nevertheless, a question remains on whether the occurrences of species not observed reflect true absence or undersampling. Because sampling reflects the species’ activity–density, a species less active at certain times of the year will be less present in the observations. However, the model is meant to reflect the species average presence in the landscape during the year, regardless of its level of activity. Here, the model follows the assumption that in 37% of the cases C. depressus was indeed present but not observed and accounts for this in the calibration. This problem is less evident the case of P. flavofemoratus, where over all plots the largest ratio of observed to sampled is higher.

The two species’ fitness (SI Appendix, Figs. S19 and S24) are closely correlated and the lower deviations in occupancy observed in Fig. 3C. The model evaluates the probability of false negative C. depressus to be quite high (37%) compared to the one of P. flavofemoratus (21%). The model here reflects the sampling (SI Appendix, Fig. S8), where C. depressus has been observed fewer times than P. flavofemoratus. Nevertheless, a question remains on whether the occurrences of species not observed reflect true absence or undersampling. Because sampling reflects the species’ activity–density, a species less active at certain times of the year will be less present in the observations. However, the model is meant to reflect the species average presence in the landscape during the year, regardless of its level of activity. Here, the model follows the assumption that in 37% of the cases C. depressus was indeed present but not observed and accounts for this in the calibration. This problem is less evident the case of P. flavofemoratus, where over all plots the largest ratio of observed to sampled is higher.

The two species’ fitness (SI Appendix, Figs. S19 and S24) are closely correlated and the lower deviations in occupancy observed in Fig. 3C. The model evaluates the probability of false negative C. depressus to be quite high (37%) compared to the one of P. flavofemoratus (21%). The model here reflects the sampling (SI Appendix, Fig. S8), where C. depressus has been observed fewer times than P. flavofemoratus. Nevertheless, a question remains on whether the occurrences of species not observed reflect true absence or undersampling. Because sampling reflects the species’ activity–density, a species less active at certain times of the year will be less present in the observations. However, the model is meant to reflect the species average presence in the landscape during the year, regardless of its level of activity. Here, the model follows the assumption that in 37% of the cases C. depressus was indeed present but not observed and accounts for this in the calibration. This problem is less evident the case of P. flavofemoratus, where over all plots the largest ratio of observed to sampled is higher.

The two species’ fitness (SI Appendix, Figs. S19 and S24) are closely correlated and the lower deviations in occupancy observed in Fig. 3C. The model evaluates the probability of false negative C. depressus to be quite high (37%) compared to the one of P. flavofemoratus (21%). The model here reflects the sampling (SI Appendix, Fig. S8), where C. depressus has been observed fewer times than P. flavofemoratus. Nevertheless, a question remains on whether the occurrences of species not observed reflect true absence or undersampling. Because sampling reflects the species’ activity–density, a species less active at certain times of the year will be less present in the observations. However, the model is meant to reflect the species average presence in the landscape during the year, regardless of its level of activity. Here, the model follows the assumption that in 37% of the cases C. depressus was indeed present but not observed and accounts for this in the calibration. This problem is less evident the case of P. flavofemoratus, where over all plots the largest ratio of observed to sampled is higher.

The two species’ fitness (SI Appendix, Figs. S19 and S24) are closely correlated and the lower deviations in occupancy observed in Fig. 3C. The model evaluates the probability of false negative C. depressus to be quite high (37%) compared to the one of P. flavofemoratus (21%). The model here reflects the sampling (SI Appendix, Fig. S8), where C. depressus has been observed fewer times than P. flavofemoratus. Nevertheless, a question remains on whether the occurrences of species not observed reflect true absence or undersampling. Because sampling reflects the species’ activity–density, a species less active at certain times of the year will be less present in the observations. However, the model is meant to reflect the species average presence in the landscape during the year, regardless of its level of activity. Here, the model follows the assumption that in 37% of the cases C. depressus was indeed present but not observed and accounts for this in the calibration. This problem is less evident the case of P. flavofemoratus, where over all plots the largest ratio of observed to sampled is higher.

The two species’ fitness (SI Appendix, Figs. S19 and S24) are closely correlated and the lower deviations in occupancy observed in Fig. 3C. The model evaluates the probability of false negative C. depressus to be quite high (37%) compared to the one of P. flavofemoratus (21%). The model here reflects the sampling (SI Appendix, Fig. S8), where C. depressus has been observed fewer times than P. flavofemoratus. Nevertheless, a question remains on whether the occurrences of species not observed reflect true absence or undersampling. Because sampling reflects the species’ activity–density, a species less active at certain times of the year will be less present in the observations. However, the model is meant to reflect the species average presence in the landscape during the year, regardless of its level of activity. Here, the model follows the assumption that in 37% of the cases C. depressus was indeed present but not observed and accounts for this in the calibration. This problem is less evident the case of P. flavofemoratus, where over all plots the largest ratio of observed to sampled is higher.

The two species’ fitness (SI Appendix, Figs. S19 and S24) are closely correlated and the lower deviations in occupancy observed in Fig. 3C. The model evaluates the probability of false negative C. depressus to be quite high (37%) compared to the one of P. flavofemoratus (21%). The model here reflects the sampling (SI Appendix, Fig. S8), where C. depressus has been observed fewer times than P. flavofemoratus. Nevertheless, a question remains on whether the occurrences of species not observed reflect true absence or undersampling. Because sampling reflects the species’ activity–density, a species less active at certain times of the year will be less present in the observations. However, the model is meant to reflect the species average presence in the landscape during the year, regardless of its level of activity. Here, the model follows the assumption that in 37% of the cases C. depressus was indeed present but not observed and accounts for this in the calibration. This problem is less evident the case of P. flavofemoratus, where over all plots the largest ratio of observed to sampled is higher.

The two species’ fitness (SI Appendix, Figs. S19 and S24) are closely correlated and the lower deviations in occupancy observed in Fig. 3C. The model evaluates the probability of false negative C. depressus to be quite high (37%) compared to the one of P. flavofemoratus (21%). The model here reflects the sampling (SI Appendix, Fig. S8), where C. depressus has been observed fewer times than P. flavofemoratus. Nevertheless, a question remains on whether the occurrences of species not observed reflect true absence or undersampling. Because sampling reflects the species’ activity–density, a species less active at certain times of the year will be less present in the observations. However, the model is meant to reflect the species average presence in the landscape during the year, regardless of its level of activity. Here, the model follows the assumption that in 37% of the cases C. depressus was indeed present but not observed and accounts for this in the calibration. This problem is less evident the case of P. flavofemoratus, where over all plots the largest ratio of observed to sampled is higher.
Appendix, section 1). This means that observations must not provide the basis for describing their demography and interactions, including the role/functions of the various species competing in the ecosystems in which they are embedded. Inferring species functionality and recovery sensu ref. 60 is thus complementary to this approach. This general framework, however, is based on much easier to collect Earth observations and presence/absence data and resolves a number of practical and theoretical issues for the generalized application of the metapopulation approach. Unlike recovery targets for a focus species (60), which involve the need to gather notable ecological information, like size, density, and demographic structure of its population, this framework does not require the kind of intensive field work necessary should ecological species interactions be defining. In this framework, the role of the species of species recovery is surrogated by the probability of occupancy, the end result of our framework, which also provides opportunities for letting ecological theory inform conservation practice.

Conclusions

The framework shown here must be seen as a proof of concept of a dynamic landscape-explicit metapopulation model driven by Earth observations intended for mountain species epitomizing, however, complex landscapes (i.e., substrates for ecological interactions). The proposed framework allows the estimation of the initial distribution of the species in the landscape. In a first step the initial occupancy is obtained by an SDM and, then, by calibrating the metapopulation model through the iterated filtering process, ensuring the occupancy in the landscape to be at equilibrium for the chosen fitness and metapopulation parameters. The calibration process identifies the parameters most suited for the species, which, in turn, permits the identification of trends in the spatial variability of the species. The framework permits the estimation of the spatiotemporal gaps between in situ observations. The proposed method permits us to extrapolate the information from sampled sites to unsampled sites and to capture the process in time. The method does not require all modeled sites to be observed as was previously the case in other studies (50, 51) and offers an integrated method to estimate the uncertainty of the observations (similar to ref. 50). Here, for instance, the metapopulation results are the only source of information for the occupancy from years 2008 to 2011, with the years 2006, 2007, 2012, and 2013 serving as reference for the calibration. Additionally, the framework manages to capture the metapopulation process of the species, thus potentially, with only EO data available, or with projected simulated EO data, permitting one to project the presence of these species into the future. Given the dynamical nature of the model, the presented spatial distribution serves as starting point for the simulation, which would then be driven by the EO data and the model dynamics.

Materials and Methods

The developed framework is demonstrated on data originating from a large ecological monitoring program called the “Biodiversity Monitoring Project” performed in the Gran Paradiso National Park (GPMP), a protected area located in the Italian Alps (Fig. 1A). In this study, data on invertebrates (Coleoptera Carabidae, Coleoptera Staphylinidae, Araneae, Hymenoptera Formicidae, Orthoptera, Lepidoptera Rhopalocera) and birds (Aves) have been routinely gathered since 2006 to gain an extended understanding on the park’s five valleys (Gran Piano, San Besso, Lauson, Orvieilles, Vaudalet-. . . taz; Fig. 18), in a 2-y sampling, 4-y break scheme; i.e., the processed data are available in 2006, 2007, 2012, and 2013. The project monitors species presence and abundance by considering a series of circular plots (5 to 7 in each valley, for a total of 30; Fig. 18) of a radius of 100 m. These plots are separated by 200 m in elevation along a transect of the valley (Fig. 1C).

We here focus on carabids, one of the collected taxon. Carabids, or ground beetles, are a family of terrestrial arthropods (Carabidae taxa) composed of more than 40,000 species worldwide (61). Carabids are among the most studied taxa in mountain ecosystems (61), in part because of the common use as bioindicators and in agriculture (SI Appendix, section 1A). Their study has significantly advanced our understanding of conservation biology and landscape and population ecology (SI Appendix, section 1B). Carabids have been confirmed as a taxon following metapopulation dynamics (61), which frame them as good candidates for this study (SI Appendix, section 1E). It has been shown that environmental drivers such as temperature, humidity, or presence of vegetation drive these species (61). EO data capturing these landscape features are therefore good candidates to model their niche.

During the 4 y of sampling, 90 different species of carabids have been identified, in various frequencies of occurrence and plots. We here focus on 2 species, which display different occurrence dynamics: P. flavofemoratus, relatively stable in terms of occurrences, and C. depressus, with more fluctuations (62). The analyses of the 2 species’ in situ data are presented in SI Appendix, section 3.

We aim at modeling near-term, interannual dynamics of the carabids, rather than seasonal ones. This choice for the temporal resolution of the model output is justified by three reasons: First, carabids have a life cycle which spans over 1 y (SI Appendix, section 1C), and most changes in density are observed between years and not within the same year (SI Appendix, section 1D). We have thus aggregated the field data by plots and year, focusing on presence rather than density, and as such reduce the uncertainty associated with categorizing each individual. For each year and plot, the number of times a certain species is found is counted (k) along with the number of times the plot was sampled (n), which corresponds to the number of times each trap in a plot was visited in a given year. This provides a good estimate of the reliability with which a species has been identified and captured in a given plot, assuming that the species have the same activity during the sampling period. Issues on imperfect detection have been considered (50, 51).

The raster EO data have been selected based on the carabid ecology described in the literature (SI Appendix, section 1E) to best portray the carabids’ niche, i.e., a multidimensional function describing the suitability of the habitat (34, 40, 42, 45). The data collection raster scale imposes a natural constraint on the choice of pixel size of 180 m, taken as the fundamental surrogate for unit range (patch) size, thus neutralizing geometric uncertainty (63).

Because carabids are quite sensitive to temperature (61), this climatic driver is the first environmental descriptor to be considered (64), with complementary topographic data (slope and aspect) to characterize the niche allowed by the environment (65). Additionally, carabids are affected by humidity, light, and the presence of vegetation (12, 61). To quantify these factors, we consider the tasseled cap indexes (65), which are a series of transformations (linear combinations) of the Landsat bands (66) designed to best quantify the degree of greenness, wetness, and brightness present in each pixel. Greenness is linked to the amount of vegetation present, wetness to humidity, and brightness to solar radiation. The tasseled cap indexes have recently been proposed as an interesting candidate to represent the variables of interest in the context of modeling beetle presence (67). Finally, forest presence has been mentioned multiple times as an important descriptor of carabid presence (61, 68) and is therefore included among explanatory variables (69). The preprocessing and generation of EO data are detailed in SI Appendix, section 2, along with a data analysis in SI Appendix, section 3 which did not yield the need to remove any covariate beforehand, letting the calibration process decide their individual importance (cf. calibration below and results).

Modeling Framework.

Metapopulation model—SPOM. Metapopulation theory is concerned with the persistence of a focus species in a given landscape, assumed to either survive due to the balance between colonization and extinction processes or go extinct. In SPOM (6, 16, 31, 53), each species is characterized by a set of parameters that defines its traits and the fitness function (performance of the species). SPOM is a Markov chain, where, for each cell i and simulation time t, the probabilities of colonization and extinction events are computed from the current presence W(t) as
where $\Delta t$ is the simulation timestep and $E_i(t)$ and $C_i(t)$ are the extinction and colonization rates (with dimension $T^{-1}$) for cell $i$ at time $t$. The colonization and extinction mechanisms are directly related to a fitness function, $f_i(t)$ (see below), which measures the suitability of patch $i$ for the species. The local extinction rate on a cell $i$ is inversely proportional to the fitness; i.e., $E_i(t) = e / f_i(t)$, where $e$ is the extinction constant. The colonization rate of an unoccupied cell is driven by the sum of the contributions from surrounding occupied cells,

$$C_i(t) = c \sum_{j \neq i} W_{ij}(t) \exp\left(\sum_{j} \beta_j k'_j \right) / D,$$

where $d_{ij}$ is the distance between cells $i$ and $j$, $D$ the dispersal distance, and $c$ the colonization constant.

**Fitness.** We represent the species suitability at cell $i$ and time $t$ by the fitness function (performance of the species) $f_i(t)$ inspired by the nonlinear logistic regression functions adopted in GLMs (16, 49),

$$f_i(t) = \frac{1}{1 + \exp\left(-\alpha_0 - \sum_j \beta_j k'_j \right)} \in (0, 1),$$

where $k'_j(t)$ is the value of the EO cofactor $j$ at position $i$ and time $t$, $\alpha_0$ the intercept, and $\beta_j$ the parameter associated with the EO cofactor $j$.

**Initial distribution estimation.** Given the discrete nature of spatial sampling, the carabid's initial (2006) distribution in the GPNP needs to be estimated, for which an SDM is used. SDMs are a range of statistical models used to assess the probability of presence of a focus species in a landscape given in situ data and, often, EO data (41, 46).

Here, a subtype of the SDM, a GLM, is used (52, 70). A GLM attempts to model the presence and absence data by linearly combining explanatory variables (71) as

$$g(\pi_i) = \eta_i = \alpha_0 + \sum_j \beta_j k'_j,$$

where $\alpha_0$ and $\beta_j$ are the GLM's parameters. We then use a logit link function to link the linear predictor ($\eta_i$) to the probability of presence ($\pi_i$), assuming a binomial distribution (71):

$$g(\pi_i) = \eta_i, \quad \mu_i = g^{-1}(\eta_i) = \frac{1}{1 + \exp(-\eta_i)}.$$  

The species presence is estimated based on the 2006 sampled vs. observed carabids data and the EO data. The sampled and observed data describe the number of times a species was sampled in a plot, vs. the number of times it was actually found in the given plot, and permit us to use a binomial distribution for the calibration of the probability of presence,

$$P(k_i | n_i, \mu_i) = \binom{n_i}{k_i} \mu_i^{k_i} (1 - \mu_i)^{n_i - k_i},$$

where $\mu_i$ is the probability of the species predicted by the GLM, $k_i$ the number of times the species was observed in the plot in $n_i$ times the plot was sampled, and $n_i - k_i$ the number of times the plot was found empty. The values of $\alpha_0$ and $\beta_j$ are calibrated by maximum-likelihood estimation (70, 71).

Once the GLM is calibrated, the output probability of presence is used (SI Appendix, Fig. S15). Random initial occupation landscapes are generated by sampling the produced probability of presence. These initial occupancy maps are then further refined through the iterated filtering algorithm (see below) and spin-up period, as it is important to ensure that the initial presence is relaxed around the quasi-equilibrium imposed by the metapopulation parameters (ref. 72 and SI Appendix, section 4C).

**Iterated filtering algorithm.** The IF algorithm has been proposed as a way to calibrate partially observed Markov processes (POMP) (54) using a frequentist approach based on maximum-likelihood estimation. Here, we make use of algorithm IF2 described in ref. 54. IF2 proposes to search for the maximum likelihood by spawning a given number of particles which all perform the Markov process multiple times (called iterations; SI Appendix, Fig. S16). Each particle is associated with certain values for each parameter. At each step of the Markov process, the parameters are first perturbed, and then the step is performed. After the step, if in situ data are available, a new set of particles is drawn by randomly sampling from the pool of particles based on a weighted distribution proportional to the likelihood, meaning one particle can be selected multiple times if it outperforms the other particles, or by randomness (73). A measurement model then links the model states to the observations (see below). The process then goes on by perturbing the parameters and running the Markov step.

The perturbation of the parameters is done by sampling from a Gaussian distribution around the values of the parameters (transformation of the sampling space can be performed for only positive parameters (log), bounded parameters (logistic, etc.) and with an SD which decreases in time to ensure the convergence of the parameters. The convergence of IF2 to the maximum likelihood estimation has been proved (54).

**IF2 Applied to SPOM.** We here sought to apply IF2 to SPOM. We propose a likelihood function which takes advantage of the in situ data structure, capitalizing on the yearly sampled vs. observed data. We make use of a binomial function with a probability depending on the model prediction and integrating imperfect detection:

$$L(t) = \prod_i \left\{ \begin{array}{ll} B(n_i(t), k_i(t), \rho_{TP}) & \text{if } W_i(t) = 0 \\ B(n_i(t), k_i(t), \rho_{FP}) & \text{if } W_i(t) = 1 \end{array} \right.$$  

with $\rho_{TP}$ the parameters of the model, $B(n_i(t), k_i(t), \rho_{TP})$ the binomial distribution as described above at time $t$, $W_i(t)$ the state of the model at time $t$ and in patch $i$, and $\rho_{TP}$, $\rho_{FP}$ the species misidentification (probability of false positive) and detection probabilities (probability of true positive), respectively. Note that $1 - \rho_{TP}$ corresponds to $\rho_{TN}$, the probability of true negative and hence the probability of correctly measuring the absence of the species, and $1 - \rho_{FP}$ corresponds to $\rho_{FN}$, the probability of missing the species in the sampling. Note that the likelihood is computed in each pixel and year where and when in situ data are available (one pixel per plot per year). The model itself does not require observation data in each modeled pixel or year to generate a presence value. The two probabilities described here are considered as two parameters calibrated along the other parameters of the model.

We here add the initial occupancy of the landscape to the particle definition. At the beginning of each iteration, for each particle the spin-up is run with the initial presence contained in each particle, and the final state is conserved for the next iteration. The spin-up period ensures that the observed trends in species presence stem from stochastic fluctuations or environmental forcing, i.e., temporal changes in the cofactors, and not from the model transitioning to a quasi-equilibrium imposed by the parameters (SI Appendix, section 4C). Since the parameters change at each step and iteration, it is important to rerun the spin-up at the beginning of the iteration to ensure the pseudo-equilibrium, but with the parameter change being slow over the calibration process, the final state of the previous iteration is already close to the new pseudo-equilibrium imposed by the sampled parameters, and thus this state is used as initial occupancy for the new parameters and spin-up, whose final state then becomes the new initial presence for the particle. To summarize, each particle contains parameter values $\theta(D, c, e, \alpha_0, \beta)$ and an initial presence $W_0$. SI Appendix, section 4 further details the application.

**Data and Model Accessibility.** The EO data accessibility is detailed in SI Table 2. The in situ data are available on Dryad: https://doi.org/10.5061/dryad.9zw3r22b9. The model implementation can be found on GitHub: https://github.com/Giezendanner/SPOM-P.

**Acknowledgments.** This work was supported by the European Union’s Horizon 2020 Research and Innovation Program under Grant 641762 (Project ECOPOTENTIAL: Improving Future Ecosystem Benefits through Earth Observations). A.B. acknowledges the spinoffs of his European Research Council Advanced Grant RINEC-227612. We thank Ruth Sonnenschein for interesting discussions on Earth observation data, with regard to tasseled cap indexes in particular. We further thank Yoni Gavish and Guy Ziv for interesting discussions and insights on the species distribution modeling tools, as well as Enrico Bertuzzo and Lorenzo Mari for exchanges on population dynamics.

Giezendanner et al.
36. R. K. Colwell, G. C. Hurtt, Nonbiological gradients in species richness and a spurious
Res. Giecendanner et al. 201919580117 www.pnas.org/cgi/doi/10.1073/pnas.1919580117
37. E. Economo, T. Keitt, Network isolation and local diversity in neutral metacommuni-
33. J. Wu, R. Hobbs, Key issues and research priorities in landscape ecology. Ecol-
27. A. Jacquin et al. 2004. I. Hanski, Habitat connectivity, habitat continuity, and metapopulations in dynamic
29. K. A. With, “Metapopulation dynamics: Perspectives from landscape ecology” in
8. J. Lenoir, J. C. Svenning, Climate-related range shifts–A global multidimensional
115. N. Acad. Sci. U.S.A. 2015. I. Hanski, Metapopulation dynamics. J.
14. D. Purves, M. A. Zavala, M. B. Araújo, D. W. Purves, Chasing a moving target: Predicting global climate change-induced shifts in non-equilibrium tree species distributions. J. Ecol. 101, 441–453 (2013).
15. D. I. Mackenzie, J. D. Nichols, J. E. Hines, M. G. Knutson, A. B. Franklin, Estimating site occupancy, colonization, and local extinction when a species is detected imperfectly. Ecology 82, 2200–2207 (2003).
158. A. Moilanen, J. Kéry, Habitat suitability and distribution models. Ecol. Modell. 177, 1813–1823 (2007).
15. I. Hanski, Habitat connectivity, habitat continuity, and metapopulations in dynamic landscapes. Oikos, 118, 299–308 (2004).
114. P. Kempeneers, F. Sedano, L. Seebach, P. Strobl, J. San-Miguel-Ayanz, Data fusion of remote sensing data. Remote Sens. Environ. 128, 106–119 (2013).
24. A. D. Keith, K. C. Weathers, Eds. (Springer, 2005), pp. 95–118.
28. J. Soberón, A. T. Peterson, Interpretation of models of fundamental ecological niches and species’ distributional areas. Biodivers. Inf. 2, 1–10 (2005).
21. J. M. Lawler et al. 2009. A. Guisan, W. Thuiller, Predicting global change impacts on plant species’ distributions: Future challenges. Perspect. Plant Ecol. Evol. Syst. 9, 137–152 (2008).
33. J. Wu, R. Hobbs, Key issues and research priorities in landscape ecology. Ecol. Evol. 15, 15–28 (2015).
57. B. A. Woodcock, “Pitfall trapping in ecological studies” in Landscape Ecology, J. G. C. Grant, K. A. Smith, Eds. (Springer-Verlag, New York, 2013).