Use of stochastic patch-occupancy models in the California red-legged frog for Bayesian inference regarding past events and future persistence

Nicolas Alcala [1*], Alan E Launer [2], Michael F Westphal [3], Richard Seymour [4], Esther M Cole [2,†], and Noah A Rosenberg [1,†]

1 Department of Biology, 371 Serra Mall, Stanford University, Stanford, CA 94305-5020, USA
2 Land Use and Environmental Planning, Stanford University, 3160 Porter Drive, Suite 200, Palo Alto, CA 94304-8442, USA
3 US Bureau of Land Management, Hollister Field Office, 20 Hamilton Court, Hollister, CA 95023, USA
4 Environmental Consultant to the Stanford Conservation Program, 3160 Porter Drive, Suite 200, Stanford University, Palo Alto, CA 94304-8442, USA

* Correspondence to Nicolas Alcala, 371 Serra Mall, Stanford University, Stanford, CA 94305-5020, USA. E-mail: nalcala@stanford.edu
† Equal contribution

Keywords: amphibians, connectivity, metapopulations, statistics, streams

Article impact statement: Patch-occupancy data can be used to detect disturbances and to predict the impact of management strategies.

Abstract. Assessing causes of population decline has critical importance for management of threatened species. Stochastic patch occupancy models (SPOMs) are popular tools for understanding spatial and temporal dynamics of populations when presence/absence data in multiple habitat patches are available. We develop a Bayesian Markov chain method that extends existing SPOMs by focusing on past environmental changes that might have altered occupancy patterns prior to the beginning of data collection. We apply the method to assess causes of population decline in the

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/cobi.13192.

California red-legged frog in three creeks: *in situ* die-off, or residual impact of past source population loss. Despite having no occupancy data for the 20-30 years between the hypothetical event leading to population decline and the first data, we discriminate among hypotheses, finding evidence of increased *in situ* die-off in two creeks. Although the creeks have comparably many occupied segments, owing to different extinction–colonization dynamics, we predict an eightfold difference in persistence probabilities of their populations to 2030. Adding a source population leads to a greater predicted persistence probability than does decreasing the *in situ* die-off, emphasizing that reversing the deleterious impacts of a disturbance need not be the most efficient management strategy. We expect that the method will be useful for studying dynamics and evaluating management strategies of many species.

Introduction

Occupancy models, which consider observations of presence or absence of a species across habitat patches, have been used to advance ecological theory and to conserve threatened species. They have appeared in ecological hypothesis tests concerning metapopulations (Hanski 1994), species invasions (Yackulic *et al.* 2012), disease dynamics (Adams *et al.* 2010), species distributions (Gormley *et al.* 2011), population trends (With & King 1999), abiotic relationships (Cole & North 2014), and community-level interactions (Welsh *et al.* 2006).

Stochastic patch occupancy models (SPOMs; Gyllenberg & Silvestrov 1994; Hanski 1994) are a family of occupancy models that describe transitions between occupancy states in terms of extinction and colonization. SPOMs have advanced to accommodate imperfect detection, demographic dynamics, sparse datasets, or spatially explicit data. Some early SPOMs considered incidence-function methods, in which the stationary probability of occupancy is used to infer extinction and colonization rates (Hanski 1994; ter Braak *et al.* 1998), assuming they are time-

independent. To address this oversimplification, several studies developed Markov chain models that permit time-varying extinction and colonization probabilities dependent on values from other patches (O'Hara *et al.* 2002; ter Braak & Etienne 2003; Moilanen 2004), and that accommodate imperfect detection (MacKenzie *et al.* 2003; Johnson *et al*. 2013). A recent method with demographic dynamics is spatially explicit and allows imperfect detection (Sutherland *et al.* 2014).

Our study was motivated by the population decline of the federally threatened California red-legged frog (*Rana draytonii*) and a desire to recover an *R. draytonii* metapopulation on Stanford lands. The species has experienced a 70% reduction in historical range (Hayes & Jennings 1988; Fisher & Shaffer 1996). On Stanford lands, *R. draytonii* was reported in 23 stream segments in 1997, and only in 12 in 2012. Factors at multiple spatial scales might influence the populations: habitat loss (Davidson et al. 2002), predation by exotic species (Lawler et al. 1999), disease (Fisher et al. 2012), and climate change (Davidson et al. 2002).

We found existing SPOM methods to be imperfectly suited to our system. Our dataset has missing data during the sampling period 1997-2015, but also has "missing data" predating 1997, when events relevant to the decline likely took place. Many current spatially implicit methods use a Bayesian framework and do allow years of missing data (see reviews of Royle & Dorazio 2008 and Bailey *et al.* 2014, and Fiske *et al.* 2011 and Kéry & Schaub 2011 for implementations), but do not estimate parameters for those years. Although Markov methods can infer parameters during missing years, they have not implemented the possibility of inferring changes in parameter values during the missing-data period (O'Hara *et al.* 2002; ter Braak & Etienne 2003; Risk *et al.* 2011).

Here, we develop a Bayesian Markov chain method for inference under a novel SPOM that permits substantial missing data, handles imperfect detection, and explicitly models time periods predating sampling, allowing temporal parameter changes owing to disturbance prior to sampling. The method enables inference of extinction and colonization rates and detection probabilities,

hypothesis testing, and extinction risk prediction. First, we derive the likelihood of model parameters given occupancy data. Next, we incorporate prior parameter information to obtain posterior probabilities and credible intervals. Parameter distributions are used for probabilistic imputation of missing occupancy data, and for estimating future probabilities of extinction. We compare two causes of decline in *R. draytonii* that would have occurred before the sampling period: increased *in situ* die-off (e.g. due to disease or introduced predators), and source population loss (e.g. through habitat loss). We evaluate possible management actions and their impact on extinction risk. This extended example illustrates the potential of the new method for a variety of problems.

Methods

DATA

From 1997 to 2015, up to two visual encounter surveys per year were completed in prespecified segments of three creeks known to contain California red-legged frog (Fig. 1). The creeks are often partially dry during the summer dry season and are continuously wet during the winter wet season. The proportions of a creek that have pools, riffles, and runs vary over time, shifting with flood events; the locations of these features also vary.

Surveys focused on presence and absence of adult frogs, the life stage most closely linked to long-term population persistence in r-selected species such as *R. draytonii* (Biek *et al.* 2002*;* Vonesh & De la Cruz 2002). Summer surveys, when water levels are low, maximize detection probability of adult frogs and staff safety.

We designed our model for the annual life cycle of *R. draytonii*. Breeding and dispersal happen during the wet season; breeding occurs in permanent or seasonal ponds or stream pools, then tadpoles metamorphose in subsequent months. Dispersal of metamorphosed frogs occurs both

upstream and downstream, with mean dispersal distance 50-500m (Lannoo 2005), and across interstitial habitats (Bulger *et al.* 2003). We assumed that the population was closed during summer (Appendix S1); therefore, surveys that did not detect *R. draytonii* were regarded as detection failures, if occupancy was detected in the same segment in other surveys of the same year.

MODEL

General SPOM

In our model, a linear habitat is divided into N patches; the distance between patches i and j is denoted d_{ij}. In a time period t, incremented as a discrete variable, each patch is in one of two states: occupied or unoccupied. Because our system has an annual cycle, we consider "years" rather than a generic time unit. We denote by p_j the per-patch detection probability of the species during survey j, by $J_{i,t}$ the number of surveys in patch i and year t, by $Y_{i,j,t}$ the detected occupancy (1 for presence, 0 for absence) in the jth survey of patch i in year t, and by $\mathbf{Y}_t=(Y_{1,1,t}, \ldots, Y_{N,1,t}\ Y_{1,2,t}, \ldots, Y_{N,J_{N,t},t})$ the vector of observations of all patches in all surveys of year t..

We model sequential extinction-colonization dynamics, assuming that occupancy measurement precedes extinction. First, patches are surveyed, providing $\mathbf{Y}_t$. Following Mackenzie et al. (2003), survey results depend on the detection probability p_j and the occupancy state $\mathbf{z}_t=(z_{1,t},\ldots,z_{N,t})$. Given that the species is present ($z_{i,t}=1$), the probability of observing presence in survey j of patch i in year t is p_j, and the probability of observing absence is $1\text{-}p_j$. The probability of $Y_{i,j,t}=1$ given that the species is absent ($z_{i,t}=0$) is 0, and the probability of $Y_{i,j,t}=0$ given species absence is 1. We consider surveys to be independent, so $\mathbf{Y}_t$ has probability equal to the product of the probabilities of the $Y_{i,j,t}$ across surveys j and patches i. We henceforth assume that the p_j are

constant and equal to p for all surveys; the framework can easily be extended to allow the detection probability to vary spatially and temporally.

Extinction then occurs, representing the dry season in which some patches become empty. We denote by E_i the extinction probability of patch i, the probability it converts from occupied to unoccupied in the extinction phase of a discrete time unit (Fig. 2a). Following Hanski and Ovaskainen (2000), we assume extinction rates are inversely proportional to population size, and colonized patches instantly reach carrying capacity, K_i for patch i. Thus, $E_i=e/K_i$, where e is a global extinction parameter constant across patches; note that e but not E_i might exceed 1.

After extinction, colonization occurs, representing the wet season. We denote by $C_{i,t}$ the colonization probability of patch i in year t (Fig. 2a). Following Hanski and Ovaskainen (2000), we assume that colonization rates are proportional to the total number of migrants entering from occupied patches, and we assume an exponential dispersal kernel. Thus, the colonization rate of patch i at time t is $C_{i,t}=c\sum_{j\neq i}\exp(-\alpha d_{ij})K_j z'_{j,t}$, where c is the global colonization parameter, a constant across patches independent of i and t, α is the inverse of the mean dispersal distance of the species, and $z'_{j,t}$ is the occupancy of patch j after the extinction phase of year t. Although K_j values are unaffected by the extinction phase, $C_{i,t}$ is indirectly affected because extinction leads to $z'_{j,t}=0$ for some segments j, and thus decreases $C_{i,t}$. Also note that c but not $C_{i,t}$ might exceed 1.

We compute the model likelihood in Appendix S2. From E_i and $C_{i,t}$, we compute $\boldsymbol{\Phi}_t$, the $2^N\times 2^N$ transition matrix from all possible states in year t to states in year $t+1$, where entry ϕ_{tkl} represents the transition probability from state k to state l. Denoting by $\boldsymbol{\Phi}_{1997}$ the 1×2^N probability vector of all 2^N possible states in 1997, by $\mathbf{q}_t$ the $2^N\times 1$ column vector whose elements correspond to the probabilities of observing $\mathbf{Y}_t$ given each state, and by $D(\mathbf{q}_t)$ the $2^N\times 2^N$ diagonal matrix whose diagonal entries are the elements of $\mathbf{q}_t$ (Appendix S2), the exact likelihood function of model parameters $\boldsymbol{\Theta}_\mathbf{0}$ given a series of detections measured from data $\mathbf{Y}_{1997},\ldots,\mathbf{Y}_{2016}$ follows (eq. S2.8):

$$\mathcal{L}(\boldsymbol{\Theta_0}|\mathbf{Y_{1997}}, \ldots, \mathbf{Y_{2016}}) = \boldsymbol{\phi_{1997}} \left[\prod_{t=1997}^{2015} \boldsymbol{D}(\mathbf{q}_t)\boldsymbol{\phi}_t(\boldsymbol{\Theta_0}) \right] \mathbf{q}_T. \quad (1)$$

Eq. 1 matches eq. 5 of Mackenzie et al. (2003), but our computation of $\boldsymbol{\phi}_t$ is related to that of O'Hara *et al.* (2002) and ter Braak & Etienne (2003) instead of that from Mackenzie et al. (2003)..

To reduce computation times when the number of segments N is large, we built an approximate likelihood function, denoted $\tilde{\mathcal{L}}(\boldsymbol{\Theta_0}|\mathbf{Y}_1, \ldots, \mathbf{Y}_T)$ (eq. S5.7; Appendix S5, Fig. S8), which only considers the most likely occupancy states.

Two hypothetical causes of population decline

We built models corresponding to two hypothetical causes of population decline for *R. draytonii*: *in situ* die-off or source population loss (Fig. 2b,c). In both models, we assume that an event changed the model parameters in the past, before the first survey that produced occupancy data. We assume that parameters were constant for a long time before the event of interest, so the initial occupancy state has limited impact on occupancy at the time of the event. We also assume that parameters were constant after the event.

Under *hypothesis 1* (*in situ* die-off), population declines result from sudden mortality increases in all patches at time t_D prior to the first sampling time. *In situ* die-off could result from disease, introduced predators, or reductions in habitat quality or availability. We assume that population sizes are equal in all patches. We label the per-patch population size before *in situ* die-offs K_D and denote the subsequent size K. Because extinction and colonization are functions of population size, we assume that *in situ* die offs increase population extinction rates and decrease colonization rates. Appendix S3 derives the likelihood under hypothesis 1 (eq. S3.3) from the general likelihood (eq. 1).

Under *hypothesis 2* (source population loss), habitat destruction or local extinction leads to loss of a neighboring source population, e.g., a pond not subject to seasonal disappearance, at time t_L before the first sampling time. Stanford lands are bordered on three sides by urban developments that have increased in density and spatial scope over the last 50 years. We hypothesize that a source population of size K_L was formerly near the current habitat, at distance d_L, and that it became extinct at time t_L. We assume that this source population was simply present and not subject to extinction/recolonization before the loss: a pond, for example. We assume that all patches other than the source population have the same population size. The difference in hypothesis 2 compared to hypothesis 1 is that the past event is localized rather than occurring as a global event affecting all patches similarly. Under hypothesis 2, global parameters e and c do not change following the source population loss, nor do the within-patch population sizes or extinction rates, each of which depends only on e and the population size. However, the patch colonization rates—which depend on dispersal from the source population—do change. The source population loss hypothesis encodes reduced colonization by preventing population rescue from the source population.

We compute the likelihood of the two hypotheses in Appendix S3. We denote by t_0 the initial year, and by $\boldsymbol{\Theta}_h$ the parameters under hypothesis h. Under hypothesis 1, $\boldsymbol{\Theta}_h=(K_D, t_D)$, and under hypothesis 2, $\boldsymbol{\Theta}_h=(K_L, d_L, t_L)$. The probability vector of all possible states in t_0 is $\boldsymbol{\phi}_{t0}$. Because transition probabilities after the event do not depend on $\boldsymbol{\Theta}_h$, denoting by t_e the event date, t_D or t_L, we have:

$$\mathcal{L}\left(\boldsymbol{\Theta}_h, \mathbf{z}_{t_0} \middle| \mathbf{Y}_{1997}\right) = \boldsymbol{\phi}_{t_0}\left[\prod_{t=t_0}^{t_e-1} \boldsymbol{D}(\mathbf{q}_t)\boldsymbol{\phi}_t(\boldsymbol{\Theta}_0, \boldsymbol{\Theta}_h)\right]\left[\prod_{t=t_e}^{1996} \boldsymbol{D}(\mathbf{q}_t)\boldsymbol{\phi}_t(\boldsymbol{\Theta}_0)\right]\mathbf{q}_{1997}. \quad (2)$$

Eq. 2 is computed from eq. S3.3 under hypothesis 1, or from eq. S3.5 under hypothesis 2. The null hypothesis is a special case of hypothesis 1 with K_D=1 and can be computed from eq. 2.

BAYESIAN PARAMETER ESTIMATION

Estimation proceeds in two steps.

1. Using the whole occupancy dataset, we infer the detection probability p and the unknown parameters shared in both hypotheses: the mean dispersal distance $1/\alpha$ and the global extinction and colonization parameters, e and c. The Bayesian estimation uses Bayes' theorem to obtain posterior distributions of the parameters given the dataset by multiplying the model likelihood (eq. 1) with the prior distributions of model parameters (Appendix S2). The mode of the joint posterior distribution is then used to obtain maximum *a posteriori* estimates of $\widetilde{\alpha^{-1}}$, $\tilde{e}$, and $\tilde{c}$, in each creek; 95% credible intervals (CI) use the 2.5% and 97.5% quantiles of marginal posterior distributions. In addition, we show in Appendix S4 how to perform missing data imputation using the posterior.

2. We infer the unknown parameters distinctive to the two hypotheses: under hypothesis 1, the population size before infection K_D and the infection timing t_D, and under hypothesis 2, the size of the lost habitat K_L, its distance to the creek d_L, and the loss timing t_L. We multiply the prior distribution of the parameters by the likelihood under the hypothesis (eq. 2) to obtain the posterior distribution of $\boldsymbol{\Theta}_h$ (Appendix S3). Maximum *a posteriori* parameter estimates and CI are computed as in step 1. Our method provides a posterior distribution for the initial occupancy $\mathbf{z}_{1882}$; however, this value is not of interest, and we integrate the joint posterior for the other parameters over all possible values of $\mathbf{z}_{1882}$.

Appendix S5 provides the numerical implementation of the inference method and an approximate method for the case of a large number of patches, which we used for San Francisquito Creek. We implemented the method in software MIDASPOM (Data S3).

Parametrization

Model parameters and prior distributions appear in Table 1 (details in Appendices S2 and S3). The mean dispersal distance $1/\alpha$ is unknown and assumed equal for all creeks. The detection probability p and the global extinction and colonization parameters e and c are also unknown and differ by creek. The size K_D of the populations under hypothesis 1, the source population size K_L under hypothesis 2, and the timing t_D of the hypothetical increased *in situ* die-off and t_L of the hypothetical source population loss, are also unknown, and we assume that events occurred between 1902 and 1982 (Padgett-Flohr & Hopkins 2009; Fofonoff et al. 2017; Google Historical Imagery). Occupancies before t_D and t_L are unknown, and we assume that the species was present in 1882 (Fellers 2005; Hayes & Jennings 1988) so that likelihood computation incorporates at least 20 years before t_D and t_L. The position of the hypothetical source population under hypothesis 2 is unknown; we assume its distance d_L to the first creek segment was between 200m and 4km. Elements of vector $\boldsymbol{\phi}_{1997}$ correspond to possible states $\mathbf{z}_{1997}$, and their values reflect prior probabilities of each state in 1997; we consider a Bernoulli prior.

All other parameters are assumed known. Patches correspond to creek segments, and distances d_{ij} between segment pairs are assumed to be known from geographic data. Because the habitat is linear, and because distances between midpoints of consecutive segments are similar (Data S1), we assume that consecutive segments have fixed distance d. The relative carrying capacities K_i are treated as known from geographic data, assuming proportionality between patch area and population size (Hanski 2000). As a first approximation, we assumed that $K_1=K_2=\ldots=K_N=1$ (Data S1).

Hypothesis testing

To quantify fit of hypotheses to data, we performed Bayesian model selection. For hypotheses i and j, we computed Bayes factors by numerically integrating under each hypothesis the product of

the model likelihood and the prior probabilities of parameters, considering all possible parameter values:

$$B_{i,j} = \frac{\Pr(\mathbf{Y}_{1997}|H_i)}{\Pr(\mathbf{Y}_{1997}|H_j)}$$

$$= \frac{\int_{\mathbf{\Theta}_i} \sum_{\mathbf{z}_{t_0}} \mathcal{L}(\mathbf{\Theta}_i, \mathbf{z}_{t_0}|\mathbf{Y}_{1997}) \Pr(\mathbf{\Theta}_i) \Pr(\mathbf{z}_{t_0}) d\mathbf{\Theta}_i}{\int_{\mathbf{\Theta}_j} \sum_{\mathbf{z}_{t_0}} \mathcal{L}(\mathbf{\Theta}_j, \mathbf{z}_{t_0}|\mathbf{Y}_{1997}) \Pr(\mathbf{\Theta}_j) \Pr(\mathbf{z}_{t_0}) d\mathbf{\Theta}_j} . \quad (3)$$

where i and j equal 0 under H_0, 1 under H_1, and 2 under H_2, the likelihood is computed from eq. 2, and the prior probabilities from eqs. S3.7-S3.9 (Appendix S3). We interpret B as follows (Jeffreys 1998): $|\log_{10}(B)|<0.5$ indicates little support for either hypothesis, $0.5<|\log_{10}(B)|<1$ indicates substantial evidence for one hypothesis, and $1<|\log_{10}(B)|$ indicates strong evidence.

EXTINCTION PROBABILITY UNDER MANAGEMENT SCENARIOS

We predict future extinction probabilities in each creek under each of four management scenarios.

1. No management.

2. Control of local causes of mortality within every stream segment (e.g. non-native predator removal) to increase carrying capacity K. This scenario is equivalent to reversing the deleterious effects of the population decline that occurs under hypothesis 1.

3. Habitat creation: a population of size K_S is restored at distance d_S from the creek. This scenario is equivalent to reversing the deleterious effects of the decline that occurs with source population loss in hypothesis 2.

4. A combined strategy reducing local causes of mortality and creating new habitat.

To implement these scenarios, we performed Monte Carlo simulations from the current state $\mathbf{p}_{2015}$ using patch extinction and colonization probabilities E_i and C_i computed using e and c values sampled from their joint posterior distribution (eq. S2.13). For each creek, we recorded the proportion of simulations with all segments unoccupied (total extinction of the creek) from 2016 to 2065.

To account for the difficulty of implementing management plans, for each management scenario we chose parameter values that correspond to reversing some but not all effects of disturbance events (hypotheses 1 and 2; parameter estimates in Table S2). For scenario 2 above, this meant considering an increased local population size smaller than the estimated size before the event ($\widetilde{K}_D \geq 1.45$ for hypothesis 1, from Fig. 4). We thus tested the impact of a small increase of population size (K=1 to 1.05), and a larger increase (K=1 to 1.25). For scenario 3, this meant considering an increased source population size smaller than the estimated size before the event ($\widetilde{K}_L = 66.1$ for hypothesis 2, from Fig. S4). We thus tested the impact of the addition of a small population (K=1) either near (200m) or farther from (600m) the creek.

Results

ESTIMATION OF SHARED MODEL PARAMETERS

We estimated mean dispersal distance $\widetilde{\alpha^{-1}} = 175\text{m}$ (Fig. S1), with CI $[125\text{m}, 425\text{m}]$ (Table S2). In addition, we estimated that probabilities of detection are similar in Matadero and Deer Creeks (Fig. 3a, 3c), $\tilde{p} = 0.77$ in Matadero (CI $[0.69, 0.82]$) and $\tilde{p} = 0.75$ in Deer (CI $[0.64,0.81]$), and lower in San Francisquito (Fig. 3e), with $\tilde{p} = 0.69$ (CI $[0.57,0.77]$).

The extinction and colonization parameter estimates differ among creeks (Fig. 3b, 3d, 3f). In Matadero Creek, point estimates of both parameters are small, $\tilde{e} = 0.12$, $\tilde{c} = 0.46$. The marginal

posterior distributions are narrow, indicating that the dataset is informative about extinction and colonization dynamics; e has CI $[0.06, 0.24]$, c has CI $[0.22, 0.96]$.

By contrast, point estimates in Deer Creek are large: $\tilde{e} = 0.39$ and $\tilde{c} = 1.34$. The marginal posterior distributions of e (CI $[0.21, 0.52]$) and c are also wider (CI $[0.75, 1.87]$). Using an informative or uninformative prior for the missing data in the initial year (1998) leads to similar point estimates and credible intervals (Fig. S2).

In San Francisquito Creek, the extinction point estimate is the largest of all creeks, $\tilde{e} = 0.47$, and the colonization estimate lies between those of Matadero and Deer Creeks, $\tilde{c} = 0.81$. The marginal posterior distributions of e and c are both moderately large (CI $[0.33, 0.62]$ for e, (CI $[0.43, 1.30]$ for c). Fig. S3 gives an assessment of the accuracy of the estimation in Fig. 3.

HYPOTHESIS TESTING

The Bayes factor strongly supports the *in situ* die-off hypothesis H_1 in Deer Creek [Table S3; $\log_{10}(B_{0,1})$=-0.944; $\log_{10}(B_{0,2})$=-0.387; $\log_{10}(B_{1,2})$=0.558] and San Francisquito Creek [$\log_{10}(B_{0,1})$=-44.616; $\log_{10}(B_{0,2})$=0.087; $\log_{10}(B_{1,2})$=44.680]. In Matadero Creek, it does not reject the null hypothesis H_0 [$\log_{10}(B_{0,1})$=0.058; $\log_{10}(B_{0,2})$=-0.011; $\log_{10}(B_{1,2})$=-0.069].

Accuracy of the hypothesis testing presented in Table S3 is assessed in Fig. S9.

PARAMETER ESTIMATION

For Deer Creek, the posterior probability of K_D plateaus when it exceeds a threshold $\widetilde{K}_D$ of 1.45; for San Francisquito Creek, it is highest for $K \geq \widetilde{K}_D = 95.5$ (Fig. 4). These values indicate that current declines in Deer and San Francisquito Creeks can respectively be explained by decreases of at least 31% and 98% of the population size in each segment due to increased *in situ* die-off. The posterior

distribution of the event timing is relatively flat for Deer Creek, indicating that the data contain little information about this quantity (Fig. S6 (c)), and largest in 1982 for San Francisquito Creek (Fig. S6(e)).

For Matadero Creek, because H_0 was not rejected, parameter estimation under H_1 is not relevant (Figs. S5 and S6). Parameter estimations under H_2 in the creeks appear in Figs. S4 and S6.

EXTINCTION PROBABILITY UNDER MANAGEMENT SCENARIOS

The Deer Creek population is more likely to go extinct than that of Matadero Creek, having an eightfold higher extinction probability by 2030 (black line in Fig. 5a and 5b). The San Francisquito Creek population is probably already extinct since 2008 (Fig. S7).

All management interventions decrease the extinction probability in Deer and Matadero Creeks. In San Francisquito Creek, only source population creation restores the population and lowers the future extinction probability below 1. In the three creeks, source population creation 200m from the creek generates the lowest extinction probability.

Discussion

We have presented a method that uses temporal patch occupancy data to infer extinction and colonization rates and to test hypotheses using a Bayesian framework. The method tests for changes in parameter values that occurred before the first occupancy survey. It is able to test hypotheses of past disturbance and to estimate future population trajectories.

A key innovation is that our method detects disturbances that predate data collection. Many spatially implicit SPOM methods (Royle & Dorazio 2008; Bailey *et al.* 2014) use a Bayesian framework, allow arbitrary missing data, and detect temporal parameter changes; however, they

neither detect changes before the first survey nor test explicit hypotheses concerning such changes. Although the spatially explicit SPOMSIM (Moilanen 2004) and the method from O'Hara *et al.* (2002) permit years of missing data, they do not allow temporal parameter changes or missing data years before the first survey. The spatially explicit Bayesian method from ter Braak and Etienne (2003) allows years of missing data prior to data collection by setting a prior for the initial occupancy state, but it does not implement parameter changes prior to data collection. We extend their framework by using the whole dataset to estimate current parameter values and the first year of data to estimate past values.

PAST INFLUENCES ON CALIFORNIA RED-LEGGED FROGS

Our method identified local variation in *R. draytonii* populations. Despite the proximity of three creeks, we found differences in extinction and colonization (Fig. 3). Local variation in parameter estimates might reflect differences in community structure, habitat quality, or disease dynamics. For example, nonnative predators are most abundant in San Francisquito Creek (AEL and EMC, unpublished data).

We estimated the mean dispersal distance of the species at 175m (Fig. S1). This value accords with movement studies in Point Reyes, California, which reported a median of 185m (Fellers & Kleeman 2007); note, however, that Point Reyes frogs breed in freshwater ponds rather than creeks.

We compared two past population decline scenarios: *in situ* die-off vs. source population loss. We found that an *in situ* die-off starting between the 1960s and the early 1980s is the most likely explanation for decline in Deer and San Francisquito Creeks (Table S3). Several factors could contribute to *in situ* die-off; unfortunately, our model cannot disentangle disease and introduced

predators because these phenomena could have had similar impacts. Infection by Bd (*Batrachochytrium dendrobatidis*), a fungal pathogen, is widespread in amphibians in California (Padgett-Flohr & Hopkins 2009) and worldwide (Fisher et al. 2012). Bd infections were observed in 1961 on the Stanford campus (Padgett-Flohr & Hopkins 2009), but might have begun in California as early as the late 1800s (Huss et al. 2013). Bullfrogs (*Lithobates catesbeianus*), one non-native predator of *R. draytonii*, were introduced into California around 1900 (Stebbins & McGinnis 2012), and signal crayfish (*Pacifastacus leniusculus*) were introduced in San Francisco watersheds as early as 1898 (Fofonoff et al. 2017). Non-native predators present in Stanford creeks, including mosquitofish (*Gambusia affinis*) (Lawler et al 1999), bullfrogs (Moyle 1973), and signal crayfish (Allan and Tennent 2000), have contributed to similar *R. draytonii* declines elsewhere in California. The population decline trajectory suggests extinction debt, where local extinction occurs after substantial delay following habitat degradation (Kuussaari et al 2009).

PREDICTED DYNAMICS AND CONSERVATION

The method predicts future persistence probabilities. Interestingly, despite similar ratios of extinction and colonization rates in Matadero and Deer Creeks, stochastic dynamics produce a greater 50-year persistence probability in Matadero Creek (Fig. 5). In Deer Creek, large estimated extinction and colonization rates magnify variability in occupancy. Thus, successive “bad years” could lead to chance extinction. In one year, only two segments were occupied in Deer Creek. With extinction parameter e=0.39, because extinction precedes colonization, the probability of complete extinction in the next year depends only on the extinction rate, equaling 0.39^2=0.1521. In Matadero Creek, however, smaller extinction and colonization parameters, and consequent lower occupancy variation, make successive “bad years” unlikely. For example, in 3 recent instances, 3 segments were occupied; with e=0.12, the extinction probability was a relatively safe

0.12^3=0.001728. Note that sensitivity to "bad years" raises concerns for persistence under unpredictable climate change, as California has experienced a warming trend with precipitation deficits and increased incidence of extreme drought (Diffenbaugh et al. 2015), with consequent potential threats to aquatic species (Meyer et al 1999).

Our results suggest conservation actions: a source population <200m from Matadero and Deer Creeks would reduce extinction risk (Fig. 5). The greatest threat is stochasticity of extinction and colonization; adding a source would facilitate survival by population rescue. The Stanford Conservation Program has pursued this management approach in a similar project, having created ponds that support California tiger salamander reproduction, and other similar projects have also observed positive results (Petranka et al 2007; Rannap et al 2009). Interestingly, the preferred conservation action (adding a source) is not the one that most directly reverses the most likely past disturbance (increased *in situ* die-off). Our results can prioritize management actions, as actions targeting Deer Creek likely give the most risk reduction for a given level of effort.

EXTENSIONS

Incorporating life-stage or demographic data, by modeling local population sizes as temporally varying rather than constant, could enable modeling of disturbances affecting specific life stages. It would also enable modeling of phenological shifts in life-history dynamics, such as differing dispersal by life stage. However, when population size data are unavailable, as in the frog example, models integrating local population dynamics are inapplicable (Sutherland *et al.* 2014). Another extension would allow patches to have different carrying capacities K_i, if estimates of K_i are known. This change would simply rescale expressions for the extinction and colonization rates E and C.

Our model permits nonlinear habitats and unequal patch sizes. Computations only require pairwise distances between populations, and could therefore be applied to many geometries. Scenarios beyond increased *in situ* die-off and source population loss can be examined: for example, reductions affecting only certain patches would reduce some population sizes K_i, cyclic droughts would periodically push certain K_i to 0, and so on. A covariate x that influences extinction or colonization can be accommodated by making the associated parameters in patch i at time t, $e_{i,t}$ and $c_{i,t}$, functions of $x_{i,t}$, for example using a logit link $e_{i,t}(x_{i,t})=[\exp(a_0+a_1x_{i,t})]/[1+\exp(a_0+a_1x_{i,t})]$. Posterior distributions of parameters a_0 and a_1 can then be inferred.

CONCLUSIONS

Our method enables assessments of influences on population persistence of complex scenarios, including local (e.g., adding new patches) or global actions (e.g., reducing predation or disease). It can both assess precise scenarios (e.g., specifying new habitat locations) and perform exploratory analyses (e.g., predictions across a range of possible locations). In our system, it finds that the best management action need not be the one that mitigates the original cause of decline. This observation is particularly important for management of species for which threats are hard to address, and suggests that alternative strategies can improve persistence. We encourage use of our framework for advancement of ecological theory and conservation management.

Acknowledgments: We thank the Stanford Creek Monkeys for their tireless field work and data collection over many years. This work was supported by Swiss National Science Foundation Early Postdoc.Mobility fellowship P2LAP3_161869.

Supporting Information

Supplementary methods (Appendices S1 to S5), supplementary tables (Tables S1 to S4), supplementary figures (Fig. S1-S9), tables of segment lengths (Data S1) and segment occupancies (Data S2), and the software MIDASPOM (Data S3) are available online.

Literature Cited

Allan MF, Tennent T. 2000. Evaluation of critical habitat for the California red-legged frog (*Rana aurora draytonii*). *UC Riverside: Center for Conservation Biology*. Retrieved from https://escholarship.org/uc/item/7027275m

Adams MJ, *et al.* 2010. Using occupancy models to understand the distribution of an amphibian pathogen, *Batrachochytrium dendrobatidis*. Ecological Applications **20**:289–302.

Bailey LL, MacKenzie DI, Nichols JD. 2014. Advances and applications of occupancy models. Methods in Ecology and Evolution **5**:1269–1279.

Biek R, Funk WC, Maxell BA, Mills LS. 2002. What is missing in amphibian decline research: insights from ecological sensitivity analysis. Conservation Biology **16**:728–734.

Bulger JB, Scott NJ, Seymour RB. 2003. Terrestrial activity and conservation of adult California red-legged frogs *Rana aurora draytonii* in coastal forests and grasslands. Biological Conservation **110**:85–95.

Cole EM, North MP. 2014. Environmental influences on amphibian assemblages across subalpine wet meadows in the Klamath mountains, California. Herpetologica **70**:135–148.

Davidson C, Shaffer HB, Jennings MR. 2002. Spatial tests of the pesticide drift, habitat destruction, UV-B, and climate-change hypotheses for California amphibian declines. Conservation Biology **16**:1588–1601.

Diffenbaugh NS, Swain DL, Touma D. 2015. Anthropogenic warming has increased drought risk in California. Proceedings of the National Academy of Sciences, **112**:3931–3936.

Fellers GM. 2005. *Rana draytonii* baird and girard 1852, California red-legged frog. In M. J. Lannoo, editor. *Amphibian Declines: the Conservation Status of United States Species*, volume 2. University of California Press, Berkeley, 552–554.

Fellers GM, Kleeman PM. 2007. California red-legged frog (*Rana draytonii*) movement and habitat use: implications for conservation. Journal of Herpetology, **41**:276–286.

Fisher MC, *et al.* 2012. Emerging fungal threats to animal, plant and ecosystem health. Nature **484**:186–194.

Fisher RN, Shaffer HB. 1996. The decline of amphibians in California's Great Central Valley. Conservation Biology **10**:1387–1397.

Fiske I, *et al.* 2011. unmarked: An R package for fitting hierarchical models of wildlife occurrence and abundance. Journal of Statistical Software **43**:1–23.

Fofonoff PW, Ruiz GM, Steves B, Simkanin C, Carlton JT. 2017. National Exotic Marine and Estuarine Species Information System. https://invasions.si.edu/nemesis/calnemo/SpeciesSummary.jsp?TSN=97326 (accessed January 2018).

Gormley AM, *et al.*, 2011. Using presence-only and presence–absence data to estimate the current and potential distributions of established invasive species. Journal of Applied Ecology **48**:25–34.

Gyllenberg M, Silvestrov DS. 1994. Quasi-stationary distributions of a stochastic metapopulation model. Journal of Mathematical Biology **33**:35–70.

Hanski I. 1994. A practical model of metapopulation dynamics. Journal of Animal Ecology **63**:151–162.

Hanski I, Ovaskainen O. 2000. The metapopulation capacity of a fragmented landscape. Nature **404**:755–758.

Hayes MP, Jennings MR. 1988. Habitat correlates of distribution of the California red-legged frog (*Rana aurora draytonii*) and the foothill yellow-legged frog (*Rana boylii*): Implications for management. In RC Szaro, KE Severson, DR Patton, editors, *Proceedings of the Symposium on Management of Amphibians, Reptiles and Small Mammals in North America, At Flagstaff, Arizona*. USDA Forest Service, 144–158.

Huss M, Huntley L, Vredenburg V, Johns J, Green S. 2013. Prevalence of *Batrachochytrium dendrobatidis* in 120 archived specimens of *Lithobates catesbeianus* (American bullfrog) collected in California, 1924–2007. EcoHealth, **10**:339-343.

Jeffreys H. 1998. *The Theory of Probability*. Oxford University Press.

Johnson DS, Conn PB, Hooten MB, Ray JC, Pond BA. 2013. Spatial occupancy models for large data sets. Ecology **94**:801–808.

Kéry M, Schaub M. 2011. *Bayesian population analysis using WinBUGS: a hierarchical perspective*. Academic Press.

Kuussaari M, *et al*. 2009. Extinction debt: a challenge for biodiversity conservation. Trends in Ecology and Evolution, **24**:564–571.

Lannoo MJ. 2005. *Amphibian Declines: the Conservation Status of United States Species*. University of California Press.

Lawler SP, Dritz D, Strange T, Holyoak M. 1999. Effects of introduced mosquitofish and bullfrogs on the threatened California red-legged frog. Conservation Biology **13**:613–622.

MacKenzie DI, Nichols JD, Hines JE, Knutson MG, Franklin AB. 2003. Estimating site occupancy, colonization, and local extinction when a species is detected imperfectly. Ecology **84**:2200–2207.

Meyer JL, Sale MJ, Mulholland PJ, Poff NL. 1999. Impacts of climate change on aquatic ecosystem functioning and health. JAWRA Journal of the American Water Resources Association ***35***:1373–1386.

Moilanen A. 2004. SPOMSIM: software for stochastic patch occupancy models of metapopulation dynamics. Ecological Modelling **179**:533–550.

Moyle PB. 1973. Effects of introduced bullfrogs, Rana catesbeiana, on the native frogs of the San Joaquin Valley, California. Copeia **1973**:18–22.

O'Hara RB, Arjas E, Toivonen H, Hanski I. 2002. Bayesian analysis of metapopulation data. Ecology **83**:2408–2415.

Padgett-Flohr GE, Hopkins RL, 2009. *Batrachochytrium dendrobatidis*, a novel pathogen approaching endemism in central California. Diseases of Aquatic Organisms **83**:1–9.

Petranka JW, Harp EM, Holbrook CT, Hamel JA. 2007. Long-term persistence of amphibian populations in a restored wetland complex. Biological conservation, **138**:371–380.

Rannap R, Lohmus A, Briggs L. 2009. Restoring ponds for amphibians: a success story. Hydrobiologia, **634**:87–95.

Risk BB, De Valpine P, Beissinger SR. 2011. A robust-design formulation of the incidence function model of metapopulation dynamics applied to two species of rails. Ecology **92**:462–474.

Royle JA, Dorazio RM. 2008. *Hierarchical modeling and inference in ecology: the analysis of data from populations, metapopulations and communities*. Academic Press.

Stebbins RC, McGinnis SM. 2012. *Field guide to amphibians and reptiles of California: revised edition,Vol* 103. University of California Press.

Sutherland CS, Elston DA, Lambin X. 2014. A demographic, spatially explicit patch occupancy model of metapopulation dynamics and persistence. Ecology **95**:3149–3160.

ter Braak CJ, Etienne RS. 2003. Improved bayesian analysis of metapopulation data with an application to a tree frog metapopulation. Ecology **84**:231–241.

ter Braak CJ, Hanski I, Verboom J. 1998. The incidence function approach to modeling of metapopulation dynamics. Modeling spatiotemporal dynamics in ecology. Springer-Verlag, Berlin, Germany, 167–188.

Vonesh JR, De la Cruz O. 2002. Complex life cycles and density dependence: assessing the contribution of egg mortality to amphibian declines. Oecologia **133**:325–333.

Welsh HH, Pope KL, Boiano D. 2006. Sub-alpine amphibian distributions related to species palatability to non-native salmonids in the Klamath mountains of northern California. Diversity and Distributions **12**:298–309.

With KA, King AW. 1999. Extinction thresholds for species in fractal landscapes. Conservation Biology **13**:314–326.

Yackulic CB, *et al.* 2012. Neighborhood and habitat effects on vital rates: expansion of the barred owl in the Oregon Coast Ranges. Ecology **93**:1953–1966.

Table 1. Summary of model parameters.

Variable	Interpretation	Range[a]	Value used[b]	Reference
Parameters shared across hypotheses[b]				
P	Probability of detection	$[0,1]$	Uniform prior	*None*
E	global extinction parameter	$[0,1]$	Uniform prior	*None*
C	global colonization parameter	$[0,1.5]$	Uniform prior	*None*
D	distance between consecutive segments (meters)	$[164,256]$ in MC[c]; $[147,249]$ in DC; $[152,446]$ in SFC	200m	Supporting Information
$1/\alpha$	mean annual dispersal distance	$[50,500]$	Uniform prior	Bulger et al. 2003; Lannoo 2005; Fellers and Kleeman 2007
$z_{1997,i}$	Occupancy state in patch i in 1997	$[0,1]$	Bernoulli prior	None
Hypothesis 1: *in situ* die-off				
t_D	timing of infection or increased predation	$[1902,1982]$	Uniform prior	Padgett-Flohr and Hopkins 2009; Fofonoff et al. 2017
K_D	per-patch population size prior to infection	$[0.1,100]$	Log-uniform prior	*None*
Hypothesis 2: source population loss				
t_L	timing of source population loss	$[1902,1982]$	Uniform prior	Google Historical imagery[d]
K_L	population size of extinct population	$[0.1,100]$	Log-uniform prior	*None*
d_L	distance to source population (m)	$[200,4000]$	Uniform prior	*None*

[a] The ranges reported correspond to the maximal and minimal values of the prior distributions detailed in Appendices S2 and S3.

[b] The values of parameters with priors are inferred using our Bayesian method, and distributions are detailed in Appendices S2 and S3.

[c] MC, Matadero Creek; DC, Deer Creek; SFC, San Francisquito Creek.

[d] For the timing of the source population loss, we used Google Historical Imagery to identify the years when the rapid expansion of human habitation in San Mateo and Santa Clara counties around MC, DC, and SFC occurred.

List of Figures

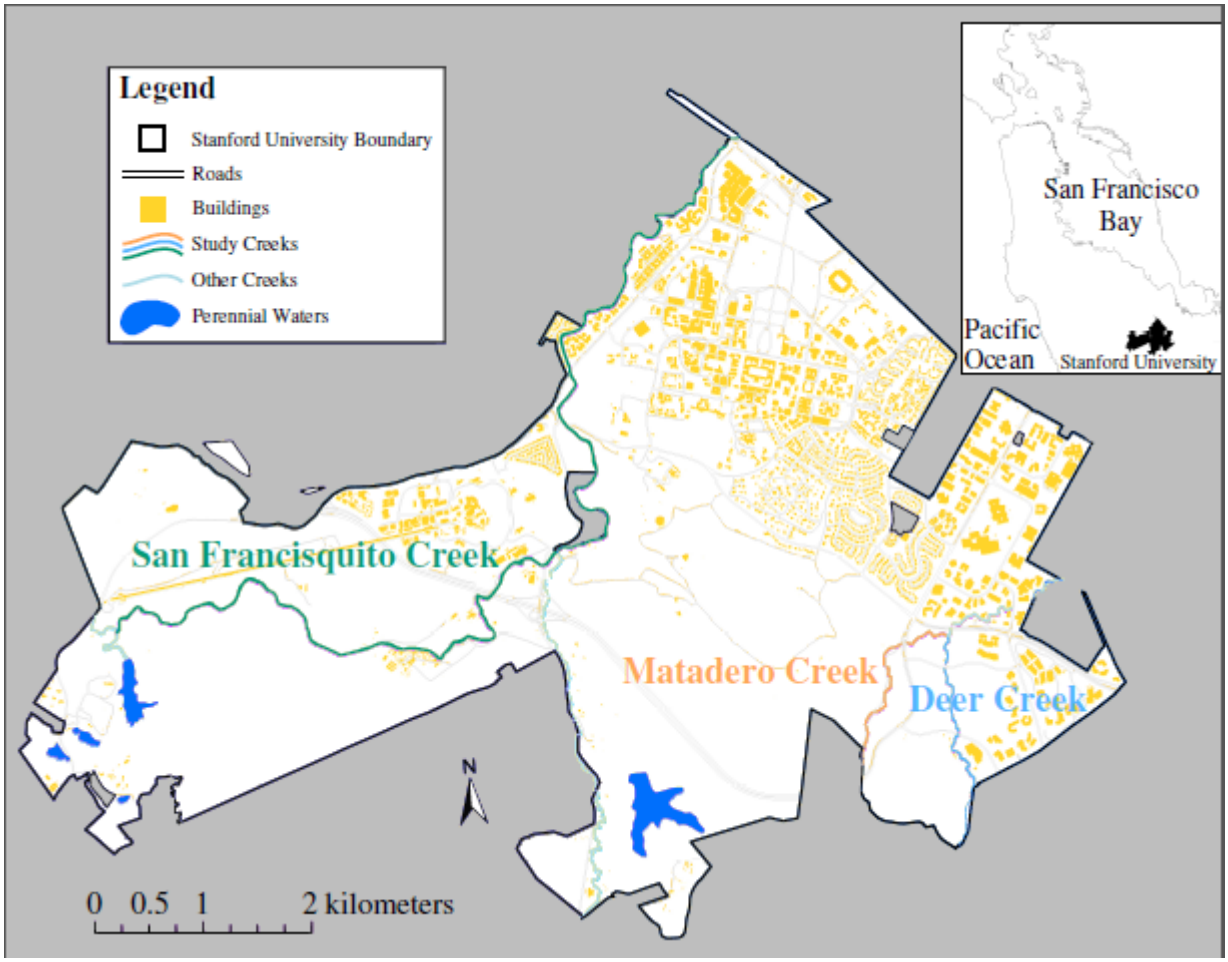

Fig. 1. Survey locations of the California red-legged frog on Stanford University lands. Each creek is treated as a separate linear habitat (Appendix S1): Matadero Creek, orange; Deer Creek, blue; San Francisquito Creek, green. Patches are numbered from upstream to downstream. Stream segments had mean length 202m (standard deviation 41m): 11 segments in Matadero Creek, 10

in Deer Creek, and 26 in San Francisquito Creek (Data S1), with missing data rates 17% in Matadero Creek, 22% in Deer Creek, and 13% in San Francisquito Creek (Data S2).

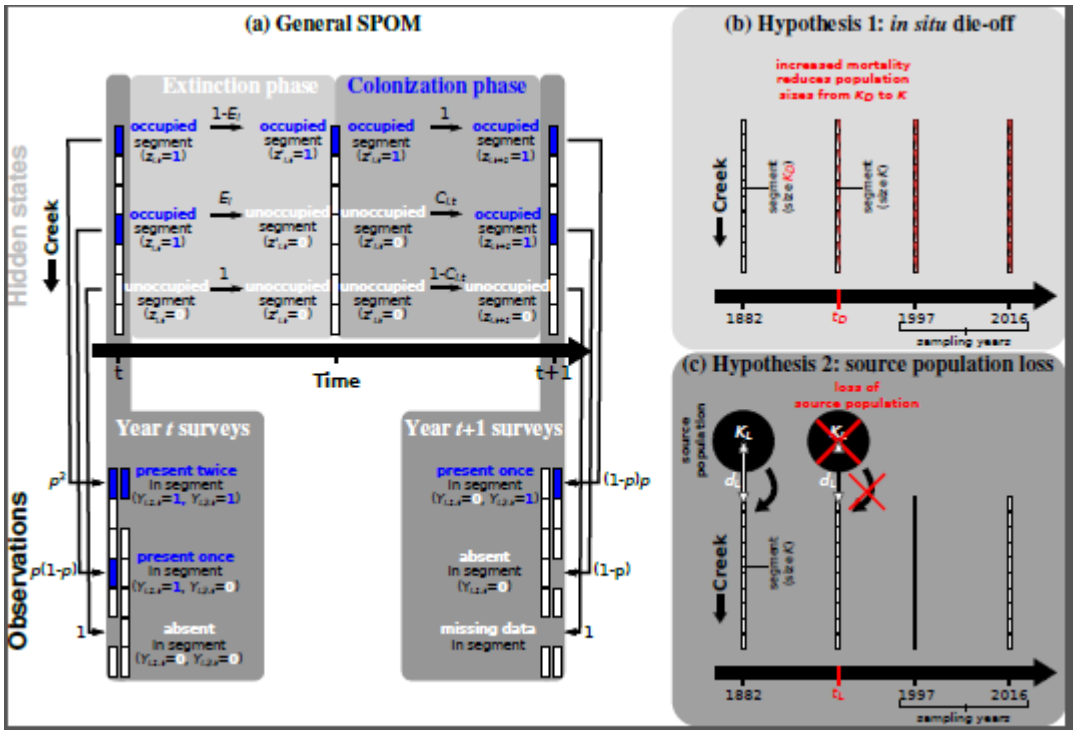

Fig. 2. Schematic representation of the models. (a) General stochastic patch occupancy model (SPOM). Each year t is divided into extinction and colonization phases. Occupancy states at the end of these phases, $z'_{i,t}$ and $z_{i,t+1}$, are hidden random variables. The *observed* occupancy in survey j of patch i in year t, $Y_{i,j,t}$, depends on the detection probability p. Quantities on the arrows indicate probabilities. (b) Hypothesis 1: in year t_D, *in situ* die-off (e.g., due to infection) reduced the population size of each patch from K_D to K. (c) Hypothesis 2: in year t_L, a source population of size K_L at distance d_L from the creek was lost (e.g., due to anthropogenic disturbance). The year considered as the initial condition in our inference is 1882. The downward arrow indicates the direction of flow.

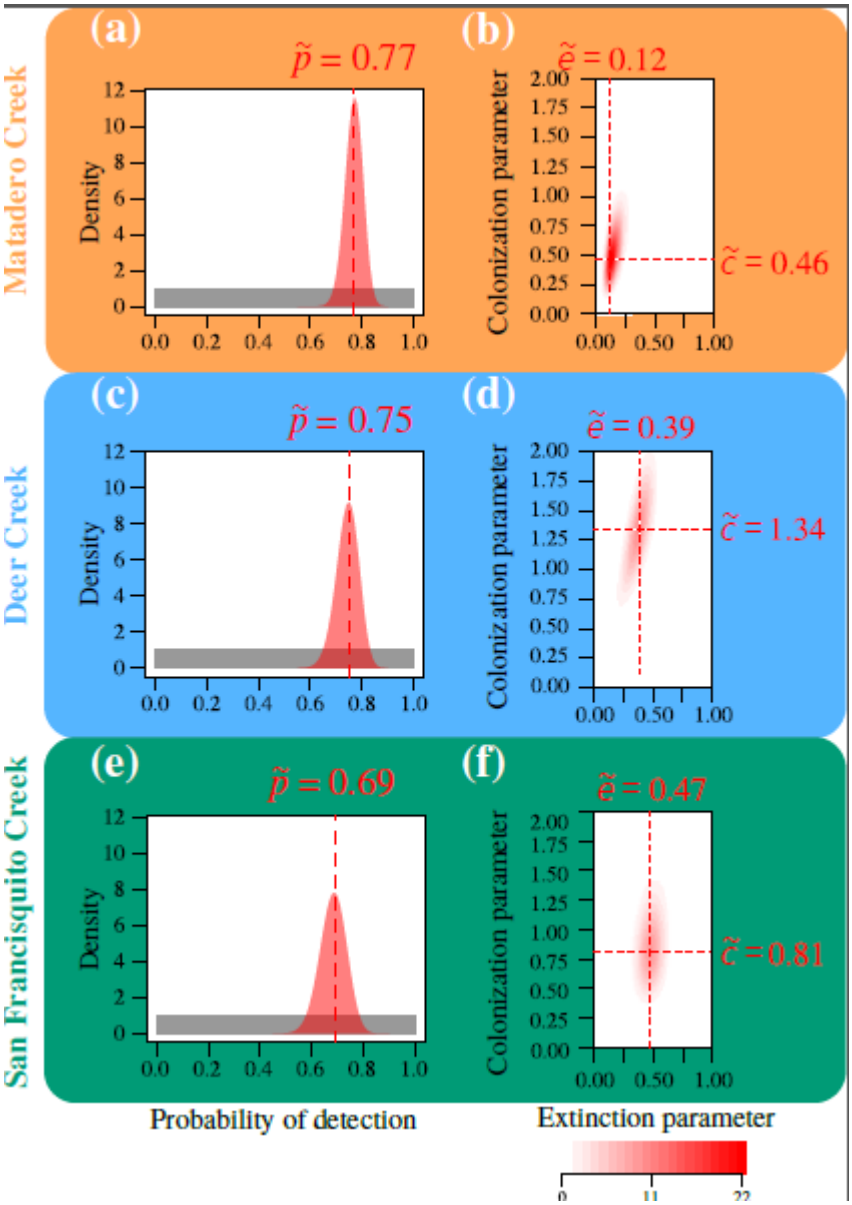

Fig. 3. Bayesian parameter estimation of the probability of detection (*p*), the extinction parameter (*e*), and the colonization parameter (*c*). (A) Probability of detection, Matadero Creek. (B) Joint distribution of extinction and colonization parameters, Matadero Creek. (C) Probability of detection, Deer Creek. (D) Joint distribution of extinction and colonization parameters, Deer Creek. (E) Probability of detection, San Francisquito Creek. (F) Joint distribution of extinction and colonization parameters, San Francisquito Creek. In (A), (C), and (E), the gray area represents the prior distribution (eq. S2.12), the red area represents the posterior distribution given the

observed data (computed from eq. 1 for Matadero and Deer Creek, and from eq. S5.7 for San Francisquito Creek), and the red dotted lines represent the 95% CI. In (B), (D), and (F), shades of red represent the joint posterior distribution, and the black point represents the maximum *a posteriori* estimate ($\tilde{e}$, $\tilde{c}$). An informative prior (Bernoulli; eq. S2.10) was used for the missing data in the initial state of Deer Creek; see Fig. S2 for the posterior distribution under an uninformative prior (eq. S2.9 from Appendix S2).

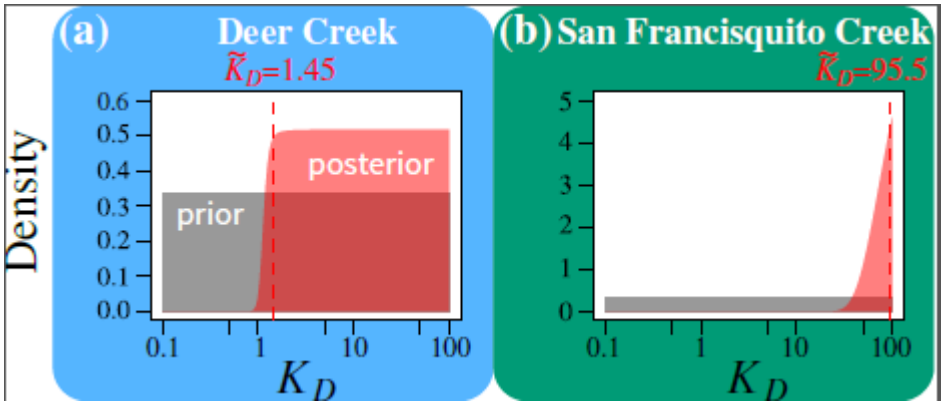

Fig. 4. Bayesian estimation of model parameters under the *in situ* die-off hypothesis (H_1). (A) Deer Creek. (B) San Francisquito Creek. Parameter K_D corresponds to the population size before the event increasing *in situ* die-off. The gray area represents the prior distribution of K_D (eq. S3.7), and the red area represents the posterior given the observed data (computed from eq. S3.11). The red dashed lines represent the mode $\tilde{K}_D$ of the posterior distribution which is used as a point estimate under hypothesis H_1. Model parameters appear in Table 1. This figure only displays parameter estimates under the hypotheses with the greatest support (based on the Bayes factor from Table S3); see Fig. S4 and S6 for parameter estimates under the source population loss hypothesis (H_2), which has less support in the three creeks, and Fig. S5 for parameter estimates under hypothesis H_1 for Matadero Creek, which has less support than the null hypothesis.

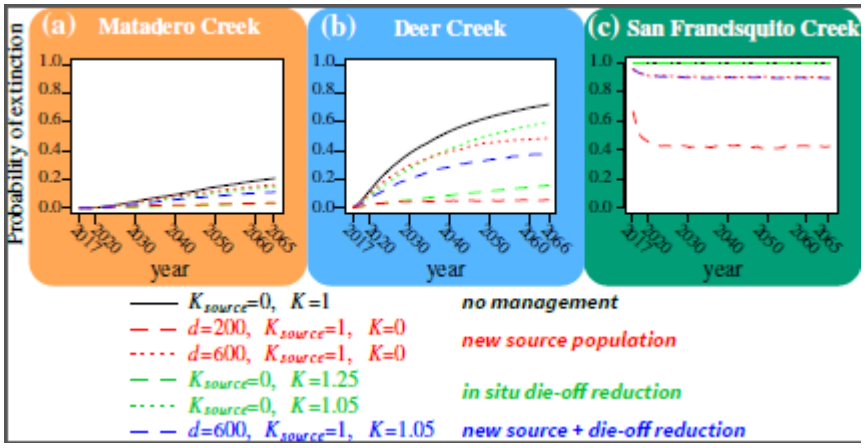

Fig. 5. Cumulative probability of complete within-creek extinction in 50 years, computed using the maximum *a posteriori* estimates of extinction and colonization parameters. (A) Matadero Creek. (B) Deer Creek. (C) San Francisquito Creek. Maximum *a posteriori* extinction and colonization parameters appear in Fig. 3. The extinction probability is obtained by computing the transition probability from the observed state in 2016 to the state where all populations are extinct, in 1, 2, ..., 50 years, assuming that occupancy data are missing after 2016 (from eq. S2.13 in Appendix S2).