Ecological Modelling 363 (2017) 192–206

Contents lists available at ScienceDirect

Ecological Modelling

journal homepage: www.elsevier.com/locate/ecolmodel

A Bayesian geostatistical approach to modeling global distributions of *Lygodium microphyllum* under projected climate warming

John M. Humphreys*, James B. Elsner, Thomas H. Jagger, Stephanie Pau

Department of Geography, Florida State University, 113 Collegiate Loop, Tallahassee, FL 32306-2190, United States

ARTICLE INFO

Article history:
Received 7 March 2017
Received in revised form 10 August 2017
Accepted 8 September 2017

Keywords:
Lygodium microphyllum
Invasive species
Species distribution model
Global warming
Biogeography
Geostatistics

ABSTRACT

Species distribution modeling aimed at forecasting the spread of invasive species under projected global warming offers land managers an important tool for assessing future ecological risk and for prioritizing management actions. The current study applies Bayesian inference and newly available geostatistical tools to forecast global range expansion for the ecosystem altering invasive climbing fern *Lygodium microphyllum*. The presented modeling framework emphasizes the need to account for spatial processes at both the individual and aggregate levels, the necessity of modeling non-linear responses to environmental gradients, and the explanatory power of biotic covariates. Results indicate that *L. microphyllum* will undergo global range expansion in concert with anthropogenic global warming and that the species is likely temperature and dispersal limited. Predictions are undertaken for current and future climate conditions assuming both limited and unlimited dispersal scenarios.

1. Introduction

Invasive plants pose a major threat to the biological integrity of the world's native natural communities (Pysek et al., 2012; Wilcove et al., 1998; Allen and Bradley, 2016). Non-indigenous plants adversely effect the suitability of habitat for native wildlife, disrupt successional trajectories, alter disturbance patterns, and compete with native plants for access to light and other resources (Gordon, 1998; Ehrenfeld, 2010; Pysek et al., 2012; Olden et al., 2004; McKinney and Lockwood, 1999). Exacerbating the effects of invasive plants, rising temperatures resulting from anthropogenic climate change may increase the risk, extent, and intensity of invasion by non-indigenous species (Rogers and McCarty, 2000; Bradley et al., 2010; Ayllón et al., 2013; Hulme, 2016). Because prevention is the less costly and more effective alternative to post-invasion treatment, intervention at the onset of, or prior to, infestation by invasive plants is the most viable and economically sound approach to safeguarding native function and diversity (Leung et al., 2002; Hulme, 2016; Hobbs et al., 2006). For these reasons, species distribution modeling (SDM) aimed at predicting the range of non-indigenous species under projected climate warming offers land managers an important tool for assessing invasion risk and for prioritizing management actions (Buckley, 2008).

SDM is increasingly used to anticipate the expansion of species in both environmental and geographic space, however, extrapolating the presence of invasive organisms to other locations, conditions, or times poses several challenges (Elith and Leathwick, 2009; Elith et al., 2010; Václavík and Meentemeyer, 2012; Elith, 2015). Among these modeling challenges is the need to consider the extent to which invasive species have obtained equilibrium within their introduced range (Guisan and Thuiller, 2005; Soberón and Nakamura, 2009; Václavík and Meentemeyer, 2012), their potential to exhibit non-linear climatic tolerances (Huntley et al., 1995; Austin et al., 2006; Austin, 2007), and the degree to which they interact with other organisms (Jablonski, 2008; Wiens, 2011; Leach et al., 2016).

Although climatic factors and evolutionary histories largely drive biogeographic patterns at regional or continental scales (Pearson et al., 2003), the realized distribution of an invasive species over its introduced range may not fully reflect that organism's true environmental tolerances, as time since introduction, dispersal limitation, biotic interactions, and other influences may constrain ecological opportunity (Guisan and Thuiller, 2005; Soberón and Nakamura, 2009; Václavík and Meentemeyer, 2012). It is critical that SDM efforts contemplate that given sufficient opportunity invasive species may spread to areas that are suitable for colonization but not yet accessible. Even when a species is at or near equilibrium, an a priori assumption of a mean response to underlying environmental gradients could potentially bias results in instances where the species actually experiences a skewed or non-

* Corresponding author.
E-mail address: jmh09@my.fsu.edu (J.M. Humphreys).

http://dx.doi.org/10.1016/j.ecolmodel.2017.09.005
0304-3800/

linear response (Austin et al., 1990; Rydgren et al., 2003; Austin, 2007). Likewise, accounting for the role of biotic variables in shaping species distributions may be as important to understanding and modeling an organism's geographic distribution as is accounting for climatic or edaphic considerations (Pellissier et al., 2010; Leach et al., 2016). By incorporating information pertaining to the locations of allied or competing cohorts, models may explain more of the variance observed in the focal species distribution than can be accounted for by abiotic factors alone.

Beyond the equilibrium assumption, non-linear responses to environmental gradients, and biotic factors, understanding the reciprocity of pattern and process over environmental and geographic space requires the explicit examination of spatial structure and scale (Wiens, 1989). Because feedbacks between pattern and process influence the functioning of both organisms and communities, the spatial arrangement of a species encodes information about the ecology that shapes it (Hurtt and Pacala, 1995; Brown et al., 2011; Velázquez et al., 2015). A major goal of SDM is to reveal this ecology through analysis of geography. When a given geographic distribution cannot be fully articulated as functions of abiotic and biotic variables, then remaining uncertainty must be quantified. Indeed, performing SDM without consideration of latent spatial structure (spatial variability due to unconsidered covariates and spatial correlation errors) may result in coefficient estimates or predictions with significant error (Hoeting, 2009; Dormann, 2007; Renner et al., 2015). Accounting for latent spatial structure can be complex because numerous factors contribute to the spatial arrangement of documented occurrences; including not only species-specific ecological and evolutionary processes such as dispersal or competition, but also biased data collection (Kadmon et al., 2003; Dormann, 2007). As one example, sampling or observer bias influences the distribution of occurrence data and is often introduced when species records are dis-proportionally sampled from particular habitat types (Dennis and Thomas, 2000) or from areas that are more readily accessible by humans (Kadmon et al., 2004; Renner et al., 2015). It is therefore necessary that models account for the "unknown" influences that arise from spatial dependency, unmeasured variables, and biased data collection as well as the "known" physical and ecological parameters that shape species distributions.

Capable of accommodating both fixed and random effects, Bayesian hierarchical models offer a flexible approach to SDM. Their tiered configuration allows for incorporation of "known" environmental and ecological variables as well as the "unknown" effects associated with latent processes like spatial correlation. Within the hierarchical model framework, latent structural processes can be assimilated into models via random effect terms that serve to quantify the uncertainty remaining after accounting for the effects of fixed covariates (Elith and Leathwick, 2009; Elsner et al., 2016). Until recently, fitting of Bayesian hierarchical models has been restricted to computationally demanding (read: potentially slow) Markov chain Monte Carlo (MCMC) simulation; however, integrated nested Laplace approximation (INLA) uses accurate approximations to the marginal posterior densities for the hyper parameters and latent variables providing a fast alternative to MCMC (Rue et al., 2009). Moreover, INLA when coupled with random effect terms as Gaussian random fields are shown to produce greater predictive accuracy than use of generalized additive models, logistic regression, and maximum entropy methods (Golding and Purse, 2016).

Although INLA provides a newly accessible alternative to MCMC, the specification of Gaussian random fields can still become quite computationally expensive, particularly over large study domains with dense matrices like those of interest to biogeographers and frequently encountered during SDM. To help overcome this issue, Lindgren et al. (2011) prescribe the use of approximate weak solutions to stochastic partial differential equations (SPDE) as a means of linking Gaussian random fields in the Matérn class to discreetly indexed Gaussian Markov random fields. Because properties of the Gaussian Markov random field enable estimation over sparse matrices, the SPDE approach permits the modeling of spatial random effects over a triangulated mesh; thereby, negating the need for dense grids and easing computational demand. The SPDE approach builds on the efficiency afforded through INLA by facilitating construction of complex point process models in a flexible and economical manner. From the perspective of spatial ecology, point process models allow for the extraction of ecologically relevant information pertaining to both pattern and process (Illian et al., 2013; Simpson et al., 2015; Fithian et al., 2015; Renner et al., 2015; Velázquez et al., 2016), which is critical for modeling species such as *Lygodium microphyllum*.

Indigenous to the pantropics of Africa, Asia, and Oceania (Pemberton and Ferriter, 1998), *L. microphyllum* (Old world climbing fern) is classified as a Noxious Weed by the United States Department of Agriculture (USDA et al., 2012) and a "Category One" invasive by the Florida Exotic Pest Plant Council, meaning that the plant impacts natural communities through the displacement of native species, the changing of community structure, and alteration of ecological function (Florida Exotic Pest Plant Council, 2015). Since first collected as isolated specimens in the mid-1960s (Beckner, 1968), *L. microphyllum* has spread throughout Florida's southern peninsula where it has enveloped canopy trees, enshrouded herbaceous marshes, entangled wildlife, and altered disturbance regimes by carrying fire into non-pyrogenic communities (Roberts, 1996; Wu et al., 2006; Volin et al., 2004). Because of its injurious effects to native biota, *L. microphyllum* has been reasonably well-studied in southern Florida. In addition to landscape-level models aimed at predicting growth rates and the probability of occupancy over the greater Florida Everglades (Volin et al., 2004; Wu et al., 2006; Fujisaki et al., 2010), greenhouse studies have investigated the fern's physiological response to changing hydrology (Gandiaga et al., 2009) and freezing temperature (Hutchinson and Langeland, 2014). The findings from these investigations highlight several important biophysical and ecological characteristics; *L. microphyllum* tolerates wetland habitats (Volin et al., 2010; Gandiaga et al., 2009), exhibits a tendency to cluster or aggregate within specific distance thresholds (Wu et al., 2006), is often located in proximity to other invasive species (Rodgers et al., 2014), and is likely limited by temperature (Hutchinson and Langeland, 2014).

The goal of the current study is to forecast the distribution of *L. microphyllum* under current and projected climate conditions through adoption of a Bayesian geostatistical perspective. Though our central ecological concern is the potential for the invasive to spread throughout the southeastern United States, we choose to model the species globally in order to include observations from both its native and introduced range. Key goals of the study include demonstrating the potential for random effects to estimate non-linear climatic tolerances, clustering distance thresholds, and spatial errors. In support of these goals, spatial models and non-linear terms are compared to their non-spatial and linear counterparts.

2. Materials and methods

2.1. Overview

The flowchart shown in Fig. 1 depicts the model development process adopted for this study beginning with species occurrence data and moving downward through geographic and environmental considerations, iterative model evaluation, and prediction. Incorporated data types and data sources are described in the

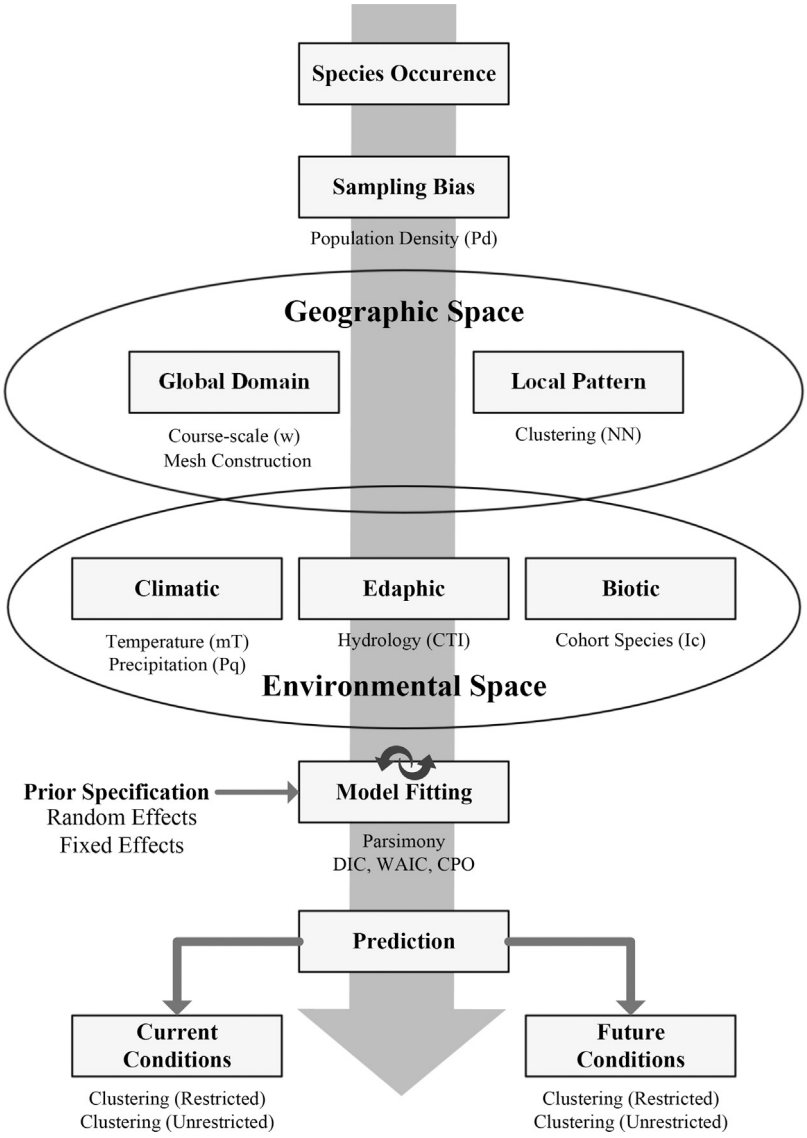

Fig. 1. Model construction work flow. Flowchart depicts model development process beginning with species occurrence data [top] and moving downward through geographic and environmental considerations, iterative model evaluation, and prediction.

"Species and environmental data" subsection, data cleaning and preparation are described under "Preprocessing," the model is formally presented under "Model description," with "Model selection and prediction" concluding the materials and methods section.

All analysis and modeling are performed using the open-source R language for statistical computing (R Core Team, 2016) with freely-available data. Copies of the R code and data for this study are available at https://github.com/JMHumphreys/SDM_Lygodium.

2.2. Species and environmental data

Incorporated data includes species occurrence locations (presence-only) from the Global Biodiversity Information Facility (http://gbif.org) and both the EDDMAPS (2016) and IMapInvasives (2015) cooperative databases. In addition to the focal species *L. microphyllum* (25,016 total records before data cleaning), occurrence data are also compiled from these databases for 31 additional species (Table 1) invasive to the United States (88,007 total records before data cleaning) that are used in construction of a biotic covariate. Incorporating information relating to the locations of invasive cohorts may improve model performance over using abiotic factors alone.

Minimum temperature estimates for the coldest month and total precipitation for the wettest quarter from the Intergovernmental Panel on Climate Change Coupled (IPCC) Model Intercomparison Project (CMIP5) are secured from the WorldClim database (Hijmans et al., 2005). In addition to "current conditions" (i.e., average conditions years 1950–2000), global climate model projections for the 8.5 representative concentration pathway (RCP) for the year 2070 is also utilized. RCPs are the greenhouse gas concentration trajectories adopted by the Intergovernmental Panel on Climate Change for the Fifth Assessment Report (Zurek et al., 2008; Pachauri and Meyer, 2014). WorldClim rasters are at a resolution of 2.5 arc-minutes.

Rasters with 30 m resolution are also obtained for elevation (Wickham et al., 2015) and human population density (CIESIN, 2005).

Table 1
List of invasive cohort species and number of records used in construction of biotic model covariate.

	Species	Count of records
1	*Abrus precatorius*	4549
2	*Alternanthera philoxeroides*	1464
3	*Ardisia elliptica*	588
4	*Bischofia javanica*	1514
5	*Callistemon viminalis*	1333
6	*Epipremnum pinnatum*	618
7	*Eragrostis atrovirens*	879
8	*Ficus microcarpa*	787
9	*Hymenachne amplexicaulis*	2920
10	*Ixora coccinea*	448
11	*Limnophila sessiliflora*	251
12	*Mangifera indica*	2729
13	*Melaleuca quinquenervia*	19,490
14	*Neyraudia reynaudiana*	1876
15	*Oryza rufipogon*	5838
16	*Panicum maximum*	7926
17	*Panicum repens*	8168
18	*Pouzolzia zeylanica*	518
19	*Psidium guajava*	5465
20	*Pteris vittata*	2606
21	*Sacciolepis indica*	2193
22	*Schefflera actinophylla*	1790
23	*Schinus terebinthifolia*	2222
24	*Senna pendula*	2289
25	*Solanum diphyllum*	843
26	*Spathodea campanulata*	741
27	*Spirodela punctata*	780
28	*Syngonium podophyllum*	2005
29	*Syzygium cumini*	1287
30	*Terminalia catappa*	1546
31	*Youngia japonica*	2344

2.3. Preprocessing

Presence-only records for *L. microphyllum* are pooled, and then cleaned to remove duplicates and all attached attributes except those for the observed occurrence locations (coordinates). To avoid model over-fitting, point occurrences remaining following duplicate removal are geographically thinned (Verbruggen et al., 2013; Boria et al., 2014; Aiello-Lammens et al., 2015) to the resolution of the coarsest environmental variable (2.5 arc-minutes). Removal of duplicate records and thinning are carried-out using the **dismo** package (Hijmans and Elith, 2016). Modeled *L. microphyllum* occurrences are depicted in Fig. 2. As part of sensitivity analyses, the response of model parameters to geographic thinning over more coarse and finer grid resolutions is evaluated and shown not to appreciably influence coefficient estimates.

Recognizing the propensity of *L. microphyllum* to inhabit wetland environments (Gandiaga et al., 2009; Volin et al., 2004) and that elevation data can be used to identify wetland hydrology (Moore et al., 1993; Humphreys et al., 2016), a Compound Topographic Index (CTI) is constructed from the elevation raster by taking the natural log of the upstream contributing area divided by the tangent of the slope in each cell. The resulting index ranges from zero to twenty-eight with the highest values signifying areas of permanent inundation and the lowest index values characterizing non-wetland regions.

To define the Gaussian random field, a spherical, triangulated mesh is constructed for the global study domain using guidelines provided by Lindgren et al. (2011). Tools available in the **r-INLA** package (Lindgren and Rue, 2015) facilitate mesh construction by providing commands to control the distance between mesh vertices, the maximum length of triangle edges, the minimum permissible angle at edge intersections, and the length of the outer extension. As per Lindgren et al. (2011), the mesh is fashioned to have triangulations with few sharp inner angles. Because both the precision of the Gaussian field and the expense of computation increase proportionally with the number of mesh vertices, the mesh is constructed to have a greater density of vertices over terrestrial regions than over oceans (non-suitable habitat). To improve parameter estimates, quadrature points (background/pseudo-absence points) generated at mesh vertices are supplemented with a regularly-spaced point grid (terrestrial regions only). In total 25,304 quadrature points are used. The final mesh is shown in Fig. 3.

Following construction of the global mesh, a biotic covariate is constructed using the 31 species listed in Table 1. The listed species were selected by the authors as potential invasive cohorts of *L. microphyllum*. That is, the listed species satisfy three criteria; they are invasive to the United States (Florida Exotic Pest Plant Council, 2015), they exhibit a predominantly tropical or sub-tropical distribution, and they have been observed near *L. microphyllum* within Florida wetland habitats (EDDMAPS, 2016; IMapInvasives, 2015). Initially, an uncorrected estimate for species richness across the 31 invasive cohorts is estimated by generating a presence/absence raster for each individual species (1° latitude by 1° longitude resolution) and then summing the individual rasters. Uncorrected estimates of invasive cohort richness are found to range from zero to twenty-seven species; however, due to the invasive cohort occurrences being subject to many of the same latent spatial processes and sampling biases as *L. microphyllum*, the uncorrected estimates are corrected by modeling the counts as a Poisson distributed process using the coarse spatial effect and population density covariate as described in the Model Description section for prediction of *L. microphyllum*. The corrected invasive cohort richness raster includes values from near zero to a maximum of 25.6.

As a final step before model fitting, the climatic, edaphic, and biotic variables are extracted from raster datasets to a combined quadrature and *L. microphyllum* spatial points data frame and evaluated for possible collinearity. Included with the variables extracted from rasters is the calculation of a Euclidean distance variable measuring the distance of each point to the nearest *L. microphyllum* occurrence location (i.e., nearest neighbor). Distances are calculated using the **spatstat** package (Baddeley and Turner, 2005). Collinearity diagnostics for independent variables as described by Belsley et al. (1980) are implemented using the **perturb** package (Hendricks and Pelzer, 2004). Initial conditioning of the variable matrix produces a condition index score of 25.4, which falls below the threshold of 30 proposed by Belsley et al. (1980), suggesting that collinearity will not bias model results.

2.4. Model description

Global observations of *L. microphyllum* occurrence $\mathbb{S}^2$ are modeled as a Bernoulli point process such that $Z(s)$ signifies the presence (1) or absence (0) of the plant at a location s ($s = 1, 2, 3, \ldots, n$) with a probability of presence given as π_s. Generally, this relationship can be expressed as,

$$Z(s) \sim \text{Bernoulli}(\pi_s) \tag{1}$$

$$\text{logit}(\pi_s) = \beta_0 + X_s\beta + \mathcal{W}_s + \epsilon_{NN(s)} + \epsilon_{mT(s)}$$

where β_0 is the intercept, X_s represents the vector of fixed covariates with $\beta = (\beta_1, \ldots, \beta_k)$ being the set of coefficients for each observation (s), and $\mathcal{W}_s$ stands for the spatially structured random effect. In addition to the spatial random effect, which is intended to account for spatial variability related to unconsidered covariates and spatial correlation errors, two non-linear covariates $\epsilon_{NN(s)}$ and $\epsilon_{mT(s)}$ are respectively included to capture spatially structured effects of aggregated individuals (clustering) and minimum temperature. All fixed and non-linear covariates are related to π_s

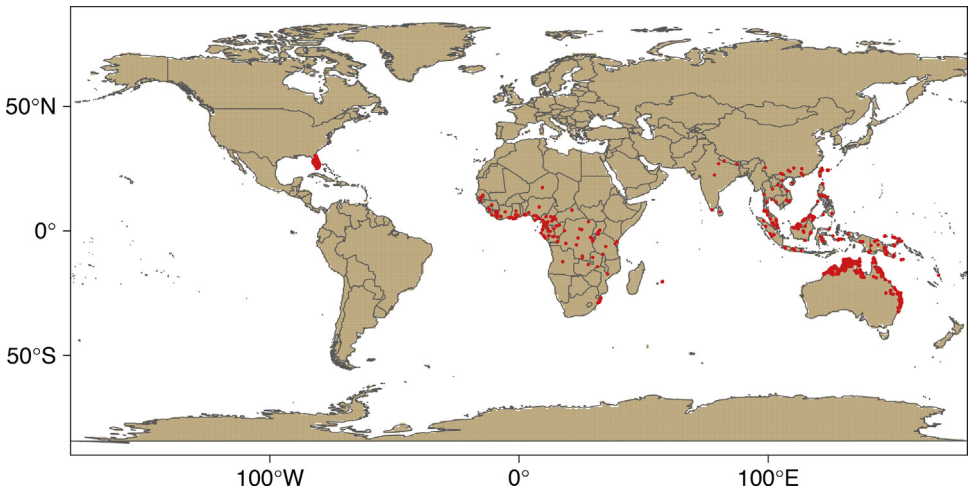

Fig. 2. Global study domain and observation locations. Red points indicate locations of observed *L. microphyllum*. Vertical axis displays degree latitude, horizontal axis degree longitude. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

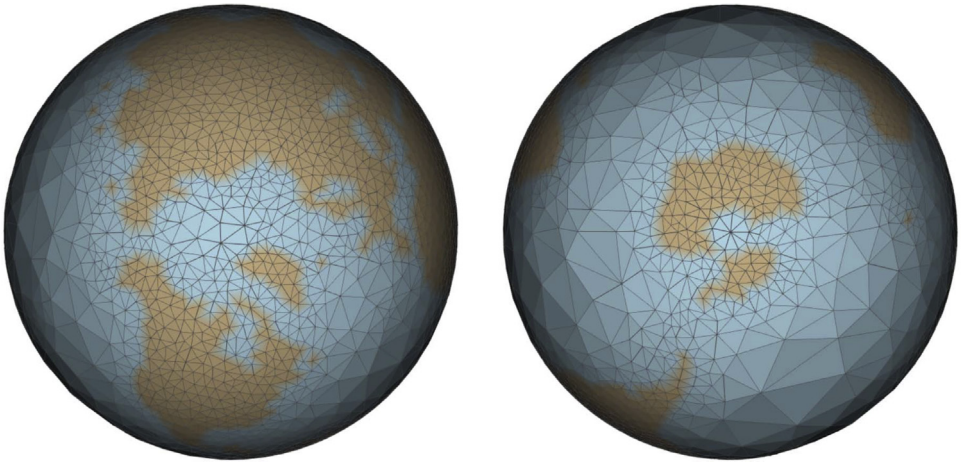

Fig. 3. Constrained refined Delaunay triangulated mesh over which the Gaussian Markov random field is constructed. Terrestrial regions are colored brown and oceanic areas are shown in blue for illustrative purposes. Image at left has been rotated to display the northern hemisphere and the image at right shows the southern hemisphere. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

following the customary logit (logarithm of the odds ratio) link and are discussed in greater detail below.

Rather than constructing the latent Gaussian random field across a two-dimensional lattice or dense grid, we opted to specify the field over a three-dimensional triangulated mesh using a finite-dimensional basis function expansion of the form:

$$x(s) = \sum_{i=1}^{n} x_i \phi_i(s), \tag{2}$$

here $x = (x_1, \ldots, x_n)$ is a multivariate Gaussian random vector and the $\{\phi_i(s)\}_{i=1}^{n}$ are independent deterministic basis functions. Estimation of the Gaussian random field is made following the stochastic partial differential equation (SPDE) approach of Lindgren et al. (2011) that allows $\phi_i(s)$ to be defined over our spherical, triangulated mesh representing the Earth's surface (Fig. 3). More narrowly, we incorporate Matérn random fields, which are zero-mean Gaussian stationary, isotropic random fields with covariance function,

$$c(d) = \frac{\sigma^2}{2^{\nu-1}\Gamma(\nu)}(\kappa d)^{\nu} K_{\nu}(\kappa d),\ d \geq 0, \tag{3}$$

where $K_{\nu}(\cdot)$ is the modified Bessel function of the second kind, $\nu > 0$ is the smoothing parameter, $\kappa > 0$ is the range parameter, σ^2 is the marginal variance, and d is the spatial dimension captured as the distance between any two points $\|s_i - s_j\|$. Specifying a smoothing parameter of 2 and a range of 1 ensures that $\nu + d/2$ (α) is an integer value and provides an efficient piecewise linear representation of the Matérn field $x(s)$ (Simpson et al., 2015); specifically, the stationary solution to the stochastic partial differential equation

$$(\kappa^2 - \Delta)^{\alpha/2} x(s) = \mathcal{W}(s), \tag{4}$$

where $(\kappa^2 - \Delta)^{\alpha/2}$ is a pseudo−differential operator and $x(s)$ is the Gaussian field with Matérn covariance.

For each vertex ($s = 1, 2, 3, \ldots, n$) the full hierarchical model can therefore be specified as,

$$Z_s|\pi_s \sim \text{Bernoulli}(\pi_s) \tag{5a}$$

$$\text{logit}(\pi_s) = \beta_0 + X_s\beta + \mathcal{W}_s + \epsilon_{NN(s)} + \epsilon_{mT(s)} \tag{5b}$$

$$\pi(\beta_0) \propto 1 \tag{5c}$$

$$\beta_j \sim \mathcal{N}(0, 1) \tag{5d}$$

$$\mathcal{W} \sim \mathcal{N}(0, \sigma^2_{\mathcal{W}}) \tag{5e}$$

$$\epsilon_{NN} \sim \mathcal{N}(0, \sigma^2_{\epsilon}) \tag{5f}$$

$$\epsilon_{mT} \sim \mathcal{N}(0, \sigma^2_{\epsilon}) \tag{5g}$$

The model's first level (5a) provides the likelihood for *L. microphyllum* presence using a Bernoulli distribution with probability π_s. The model's second level (5b) provides the linear predictor, which we iteratively modify during model fitting to examine various covariates. Ultimately, the selected specification is of the form,

$$\begin{aligned}\text{logit}(\pi_s) = \quad & \beta_0 + \beta_1 Pd(s) \\ & + \beta_2 \text{CTI}(s) + \beta_3 \text{Pq}(s) + \beta_4 \text{Ic}(s) \\ & + \beta_5 \cdot f_{NN} NN(s) + \beta_6 \cdot f_{mT} mT(s) + \mathcal{W}(s),\end{aligned} \tag{6}$$

where β_0 is the intercept, β_1 is the coefficient of human population density measured as the count of individuals per square kilometer. Population density (Pd) is included to correct for sampling biases related to increased frequency of observations in more populous regions. β_2 is the coefficient of the Compound Topographic Index (CTI) or environmental "wetness index," β_3 serves as coefficient for the amount (millimeters) of precipitation (Pq) received at each location during the wettest quarter of the year, and β_4 is the coefficient of the biotic variable referring to invasive cohort richness (Ic). The function $\beta_5 \cdot f_{NN} NN(s)$ describes spatially structured effects connected to the aggregation of individuals (clustering) that occurs below the resolution of $\mathcal{W}(s)$. The function $\beta_6 \cdot f_{mT} mT(s)$ models the minimum temperature of the coldest month at each location. Rather than assuming a distribution or mean response a priori, both *mT* and *NN* are modeled as non-linear, order-one random walks with a corresponding Gaussian vector $x = (x_1, \ldots, x_n)$ constructed assuming independent increments,

$$\Delta x_i = x_i - x_{i+1} \sim \mathcal{N}(0, \tau^{-1}).$$

and the density for x derived from $n-1$ intervals as

$$\begin{aligned}\pi(x|\tau) \quad & \propto \tau^{(n-1)/2} \exp\left\{-\frac{\tau}{2}\sum(\Delta x_i)^2\right\} \\ & = \tau^{(n-1)/2} \exp\left\{-\frac{1}{2} x^T \mathcal{Q} x\right\}.\end{aligned}$$

Here $\mathcal{Q} = \tau\mathcal{R}$ and $\mathcal{R}$ is the matrix reflecting the neighborhood structure of the model (Rue et al., 2009).

As we incorporate the Bayesian paradigm, model levels (5c)–(5g) provide the prior distributions. A flat improper prior is specified for the intercept (β_0). Fixed effects (β_j) are assumed a priori independent with Gaussian prior distributions with density

$$\pi(\theta) = \left(\frac{\tau}{2\pi}\right)^{1/2} \exp\left(-\frac{\tau}{2}(\theta - \mu)^2\right)$$

where the mean and precision are set as $\mu = 0.1$, $\tau = 0.01$.

Following Simpson et al. (2015) and Fuglstad et al. (2015), a Penalized Complexity (PC) framework is adopted for the structured spatial effect ($\mathcal{W}$) with priors drawn from the geometry of the mesh used in construction of the Gaussian random field. In doing so, the PC prior provides a means of ensuring that the magnitude and scale of the spatial effect are tuned to the dimensions of the study domain (Bakka et al., 2016). The PC prior density for the spatial range ρ and the marginal standard deviation σ is

$$P(\rho, \sigma) = dR/2\rho^{-(1-d/2)} \cdot \exp(-R\rho^{-d/2})S \cdot \exp(-S\sigma),$$

where R and S are hyperparameters specified through $P(\rho < \rho_0) = p_\rho$ and $P(\sigma > \sigma_0) = p_\sigma$ with estimates for the lower tail quantile and probability of the range ($\rho_0 = 0.9$ and $p_\rho = 0.9$) as well as the upper tail quantile and probability for the standard deviation ($\sigma_0 = 1$ and $\alpha_\sigma = 0.01$).

PC priors are also utilized for the non-linear effects modeled as functions of spatial clustering $\beta_5 \cdot f_{NN} NN(s)$ and minimum temperature $\beta_6 \cdot f_{mT} mT(s)$. These are specified such that the PC prior for precision τ is given as the density

$$\pi(\tau) = \frac{\lambda}{2}\tau^{-3/2}\exp(-\lambda\tau^{-1/2}), \ \tau > 0$$

for $\lambda > 0$ where $\lambda = -\frac{\ln(\alpha)}{u}$ and (u, α) are the parameters for the prior. For $\epsilon_{NN(s)}$ we set $u = 0.1$ and $\alpha = 0.01$. For $\epsilon_{mT(s)}$ we set $u = 3$ and $\alpha = 0.01$ in hopes of reducing interaction with $\mathcal{W}(s)$ (i.e. a "shrinking" prior). In selecting prior parameters (u, α), our assumption is that $P(\sigma > u) = \alpha$, where $u > 0$, $0 < \alpha < 1$, and the standard deviation is $\sigma = 1/\sqrt{\tau}$.

The priors and the likelihood are then combined to obtain the posterior distributions for the model parameters. The integrals can not be solved analytically; therefore, the method of INLA is used as a fast alternative to MCMC simulation for models with a latent Gaussian structure (Rue et al., 2009).

2.5. Model selection and prediction

Model selection is undertaken as an iterative process in which a series of seven candidate models are constructed and then compared. An initial model (Model0) is fit with only an intercept and spatial-effect term (Gaussian random field) to quantify coarse-scale latent structural processes over the study domain. The second model (Model1) includes the intercept and spatial-effect from Model0, as well as three fixed effects (Pd, CTI, Pq) and two random effects (random walks). The random walk terms account for clustering distance among observed *L. microphyllum* locations (NN) and the plant's non-linear response to temperature (mT). The random walk terms assume point estimates between observations to be incrementally independent and provide a means of estimating clustering and temperature as non-linear processes below the resolution of the Gaussian random field.

To explore the significance of non-linear covariates included with Model1, Model2 is fit with the response to temperature (mT) as a fixed effect (linear) and Model3 is executed with both temperature and nearest neighbor (NN) as fixed effects. To gage the influence of the clustering covariate term, Model4 is then fit while excluding NN.

Having explored the significance of climate and edaphic variables and the influence of treating the clustering and temperature covariates as both linear and non-linear effects, the biotic variable for invasive cohort richness (Ic) is added in Model5. Finally, a non-spatial model (Model6) incorporating all covariates is constructed for gaging the explanatory power and significance of the modeled spatial term ($\mathcal{W}$).

Model comparison is undertaken through re-approximation of posterior distributions to ascertain the Watanabe-Akaike information criterion (WAIC), deviance information criterion (DIC), and log conditional predictive ordinance (LCPO). These statistics measure the relative fit of each model given the available data. The WAIC, DIC, and LCPO are scaled such that the lower the value, the better the model. Although the DIC and WAIC are largely comparable under most scenarios, the DIC sometimes underestimates model complexity arising from random effect terms (Gelman et al., 2014); thus, both are used to compare models. Although the data partitioning necessary to calculate the WAIC can sometimes be problematic when evaluating structured data (Watanabe, 2013; Gelman et al., 2014), we recognize that there is no single best solution for comparison of Bayesian hierarchical models with random effect terms. The WAIC does offer a fully Bayesian criterion and has been judged to be a valid approach when evaluating hierarchical and mixture models (Hooten and Hobbs, 2015). Importantly, the LCPO is also ascertained and all three measures agree in selecting model (Model5). The LCPO is a means of leave-one-out cross-validation such that the model with the lowest relative LCPO is considered to exhibit the

Table 2
Comparison of candidate models. Deviance information criterion (DIC), Watanabe-Akaike information criterion (WAIC), log of the conditional predictive ordinance (LCPO), and included covariates. The coarse random effect is represented as $\mathcal{W}$ with covariates included as non-linear functions notated as $f(\cdot)$. Covariates include nearest neighbor (NN), minimum temperature of the coldest month (mT), human population density (Pd), compound topographic index (CTI), precipitation of the wettest quarter (Pq), and invasive cohort richness (Ic).

Model	DIC	WAIC	LCPO	Covariates
Model0	2288.54	2272.87	0.044	$\beta_0 + \mathcal{W}$
Model1	1952.85	1946.8	0.037	$\beta_0 + f(\text{NN}) + f(\text{mT}) + \text{Pd} + \text{CTI} + \text{Pq} + \mathcal{W}$
Model2	1968.62	1963.33	0.038	$\beta_0 + f(\text{NN}) + \text{mT} + \text{Pd} + \text{CTI} + \text{Pq} + \mathcal{W}$
Model3	2061.03	2064.13	0.040	$\beta_0 + \text{NN} + \text{mT} + \text{Pd} + \text{CTI} + \text{Pq} + \mathcal{W}$
Model4	2064.75	2052.97	0.039	$\beta_0 + f(\text{mT}) + \text{Pd} + \text{CTI} + \text{Pq} + \mathcal{W}$
Model5	1931.42	1927.74	0.037	$\beta_0 + f(\text{NN}) + f(\text{mT}) + \text{Pd} + \text{CTI} + \text{Pq} + \text{Ic} + \mathcal{W}$
Model6	2201.34	2205.32	0.042	$\beta_0 + \text{NN} + \text{mT} + \text{Pd} + \text{CTI} + \text{Pq} + \text{Ic}$

best cross-validated skill. Seven candidate models are fit in total, these are summarized with WAIC, DIC, and LCPO in Table 2.

Carrying out formal analytic prediction of *L. microphyllum* under current climate conditions requires several steps. Firstly, all data used in fitting the best performing model (Model5) are duplicated. Next, the response variable indicating presence (1) or absence (0) in the duplicate dataset is set to "NA" and then re-combined with the original data as required by the **r-INLA** package (Lindgren and Rue, 2015). The combined dataset, now twice the length used for initial fitting, is re-fit using the model specification from Model5.

Due to the increased computational demand of performing analytic prediction with a "doubled" dataset, results from the analytic prediction are compared to that forecast using a computationally cheaper alternative method of prediction; specifically, the summation of linear predictor components. This is achieved by projecting the posterior mean of the Gaussian random field to prediction locations and then summing the coincident estimate with the posterior means from the intercept, the clustering and temperature terms, and all fixed effects included with Model5. Finding analytic and linear summation prediction methods to be significantly correlated ($r = 0.99$), the summation of linear components method is utilized to predict *L. microphyllum* occurrence under future climate scenarios. Prediction is therefore accomplished by updating temperature and precipitation projections for the year 2070 (8.5 RCP) while controlling for latent spatial processes and holding all other variables constant.

In addition to prediction of *L. microphyllum* occurrence under current and future climate scenarios, forecasts are also made assuming that the fern has unfettered ecological opportunity and is not restricted by the clustering distances estimated from its realized distribution ($\beta_5 \cdot f_{NN}NN(s)$ in Eq. (6)). To accomplish this, the final model is executed with the nearest *L. microphyllum* distance for all prediction locations set as zero. This ensures that the probability of occurrence contributed by the non-linear clustering term (see, Results) is uniform across the global domain regardless of actual proximity to a realized *L. microphyllum* presence.

3. Results

3.1. Latent spatial processes and fixed effects

Initial fitting of the model with the spatial random effect term (Gaussian random field) in the absence of other environmental and spatial covariates (Model0) allows for the mapping of smoothed random field densities relative to the domain mean (zero) [Fig. 4(A)]. Interpretation of random field densities is undertaken in manner comparable to evaluating mapped model residuals. The map's warm colors highlight regions where densities exceed the mean density for the domain (actual/observed values greater than those predicted by the model) and cool colors display locations where densities fall below the mean (actual/observed values lesser than those predicted by model). The map accentuates regions across Southeast Asia, Australia, Africa, and peninsular Florida where observations of *L. microphyllum* are more concentrated than would be expected assuming the plant to be randomly or uniformly distributed. Of additional note are the areas of relatively low density over tropical and subtropical portions of Central and South America. As shown in Fig. 2, observations of *L. microphyllum* occurrence have yet to be documented in the neotropics, despite the plant's otherwise pantropical prevalence.

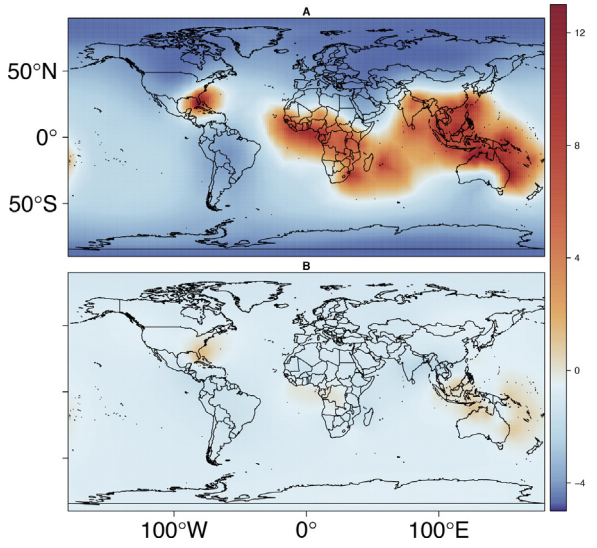

Fig. 4. Mean of the random field. Warm colors (reds) highlight regions where densities exceed the mean density over the domain and cool colors (blues) display locations where densities fall below the mean. Results for model Model0 (A) prior to adding covariates and Model5 (B). Vertical axis displays degree latitude, horizontal axis degree longitude. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

The updated map [Fig. 4 (B)] produced by adding all covariates (Model5) exhibits reduced densities relative to those generated by Model0. Regions of high density previously represented by warm orange and red colors have diminished to cooler blue hues. Stated differently, model spatial residuals have decreased with the addition of explanatory variables. Maps of the spatial random field's standard deviation [Fig. 5] display a comparable succession to that shown by the mean random field plots. Fig. 5 reveals that the standard deviations resulting from Model0 appear substantially reduced by the addition of spatial terms and covariates. That is, dark gray tones signifying areas of elevated standard deviation [Fig. 5(A)] are minimized or eliminated through fitting of the full model [Fig. 5(B)].

Fig. 6 displays the point estimates (black smooth line) and 95% credible intervals (gray lines) for the spatial effects measured from mesh vertices for Model0 [Fig. 6(A)] and Model5 [Fig. 6(B)]. As used here, spatial effects refer to the estimated mean and 95% credible interval for the coarse spatial term ($\mathcal{W}$) used to account for spatial

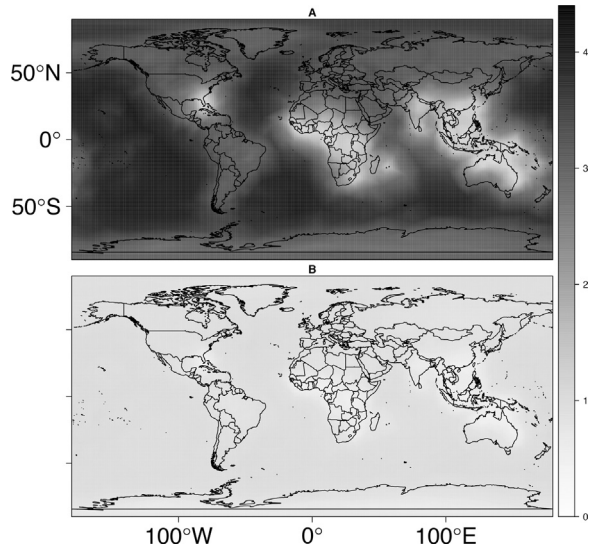

Fig. 5. Standard deviation of the random field. Dark gray tones highlight regions with higher standard deviation and light gray tones display locations with lower standard deviation. Results for model Model0 (A) prior to adding covariates, and Model5 (B) the selected model. Vertical axis displays degree latitude, horizontal axis degree longitude.

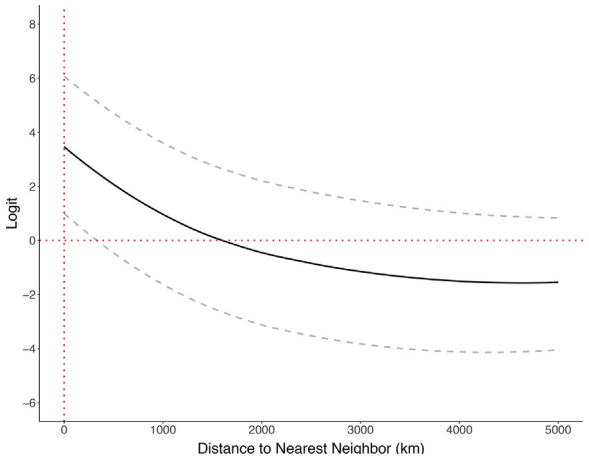

Fig. 7. Non-linear functional relationship of nearest neighbor distance to the Model5 logit (logarithm of odds ratio). Distance units are kilometers. 95% credible interval shown as gray dashed lines. The 2.5% credible interval (lower dashed gray line) becomes negative at 205 km and the logit (bold smooth line) becomes negative at 1585 km.

variability arising from unconsidered covariates and spatial correlation errors. To better interpret the spatial trend, the points are ordinally arranged and arbitrarily indexed along the horizontal axis by the point estimate with negative values at the left and positive values to the right. Note that the vertical axes are scaled differently for ease of interpretation in Fig. 6(A) and (B). The spatial effects, as indicated by the breadth of each plot's credible interval, show that residuals for Model5 are considerably reduced relative to those shown by Model0.

The functional relationship of the nearest neighbor term (NN) to the logit is graphed in Fig. 7. The graph displays a curvilinear response with increased logit values at more proximal locations (e.g.,"clustering") transitioning to a neutral (level) influence with greater distance. The relationship indicates that the probability of *L. microphyllum* occurrence increases non-linearly at distances less than 1585 km, but, only distances less than 205 km remain within the credible interval. At a distance of 0 km, the mean logit is 4.53[(2.02, 4.50) 95% credible interval]. This translates to a 98.93% $[(1/(1+\exp(-4.53))) \times 100\%]$ increase in the probability of *L. microphyllum* occurrence while holding all other covariates constant. Note that the 95% credible interval (gray dashed lines) indicates a low posterior probability of equal occurrence odds at short distances [Fig. 7].

The functional relationship of minimum temperature (mT) to the logit is displayed in Fig. 8. The graph displays a curvilinear response with logit values becoming positive at temperatures equal to or exceeding 3.85 °C. The graphed relationship indicates that the probability of *L. microphyllum* occurrence increases non-linearly above 3.85 °C (0% probability of occurrence), but, then levels at

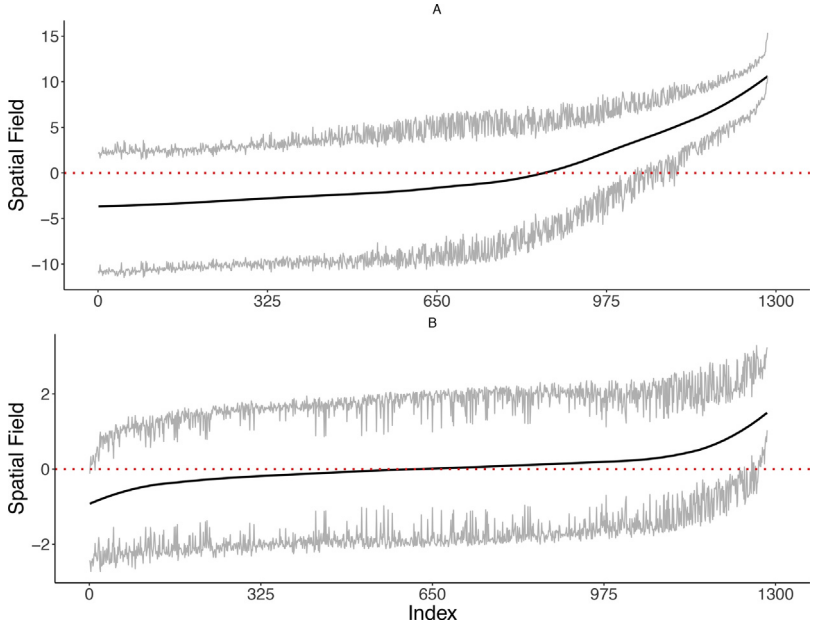

Fig. 6. Point estimates and 95% credible intervals for the spatial effect ($\mathcal{W}$) using the SPDE for Model0 (A) and Model5 (B). The (mesh vertex) points are arranged along the horizontal axis (Index) in increasing order by the point estimate with negative values at the left and positive values to the right. Note that the vertical axes in (A) and (B) is scaled differently. The spatial residuals for Model5 are considerably reduced relative to those produced by Model0.

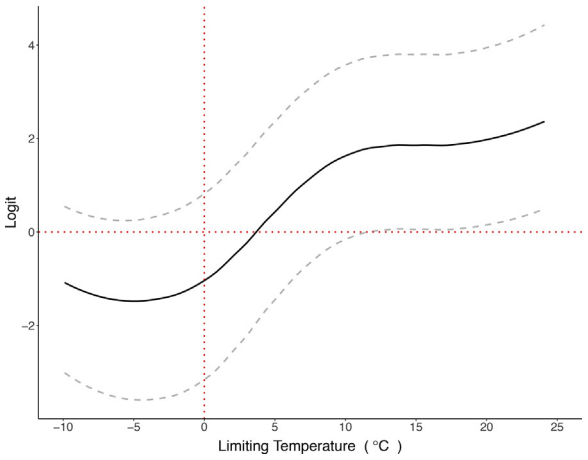

Fig. 8. Functional relationship of minimum temperature of the coldest month (mT) to the Model5 logit. Temperature units are °C. 95% credible interval shown as gray dashed lines. The logit becomes positive at 3.85 °C.

approximately 10 °C. Between 10 °C and the maximum modeled temperature of 24.10 °C, the probability of presence increases from approximately 73.1% to 91.8% holding all other covariates constant. The 95% credible interval (gray dashed lines) indicates that this effect is important.

Table 3 displays the mean, standard deviation, and 95% credible interval for the modeled intercept and linear fixed effects. The posterior probability that the effect is zero is low for all covariates. Population density (Pd), precipitation of the wettest quarter (Pq), and invasive cohort richness (Ic) indicate an elevated probability of occurrence; whereas, increases to the wetness index (CTI) are associated with a decreased probability of occurrence. Precipitation of the wettest quarter is revealed to exhibit the strongest relationship to occurrence with a mean logit of 0.96[(0.50, 1.42) 95% credible interval]. This translates to a 72.3% $[(1/(1+\exp(-0.96))) \times 100\%]$ increase in the probability of *L. microphyllum* occurrence for each millimeter increase in Pq above the mean of 322 mm while holding all other fixed effects constant. Comparatively, each increase in

Table 3
Mean logit, standard deviation, and 95% credible interval (0.025 and 0.975 quantiles) for fixed effects. Population density (Pd), Compound Topographic Index (CTI), precipitation of the wettest quarter (Pq), and invasive cohort richness (Ic). Results from selected model (Model5).

Effect	Mean	sd	0.025 Quantile	0.975 Quantile
Intercept	−9.23	1.54	−12.30	−6.26
Pd	0.33	0.16	0.04	0.66
CTI	−0.82	0.33	−1.47	−0.18
Pq	0.96	0.23	0.50	1.42
Ic	0.74	0.10	0.56	0.93

invasive cohort species richness above the domain-wide average of 0.49 species results in a 67.7% increase in the probability of *L. microphyllum* occurrence while holding other fixed effects constant. The probability of observing *L. microphyllum* jumps 58.3% for each doubling of human population density with other effects held constant. Similarly, application of the Pd variable to correct invasive cohort richness reveals that species richness increases 38.6% for each doubling of population density. In contrast to increased probabilities of presence, chances of occurrence decease by 30.6% for each whole value increase to the CTI beyond the mean index of 14.7. Posterior densities for linear fixed effects are provided in Fig. 9.

The predicted global distribution of *L. microphyllum* under current climate conditions is illustrated in Fig. 10. These predictions are undertaken while controlling for latent spatial processes at two-levels, sampling bias tied to human population density, and the included climate and edaphic covariates. Fig. 10(A) displays probabilities of occurrence under current climate conditions assuming *L. microphyllum* to be constrained by clustering distance in accordance with the functional relationship illustrated by Fig. 7. The overall distribution indicates a tropical and subtropical species range with the highest probabilities of occurrence in northern Australia and peninsular Florida. Fig. 10(B) also displays predictions under current climate conditions; however, this scenario is modeled with *L. microphyllum* not being constrained by clustering distances. Fig. 11(A) and (B) provide the probability of global occurrence under the projected temperature and precipitation estimates provided by the 8.5 RCP for the year 2070. In a similar manner to

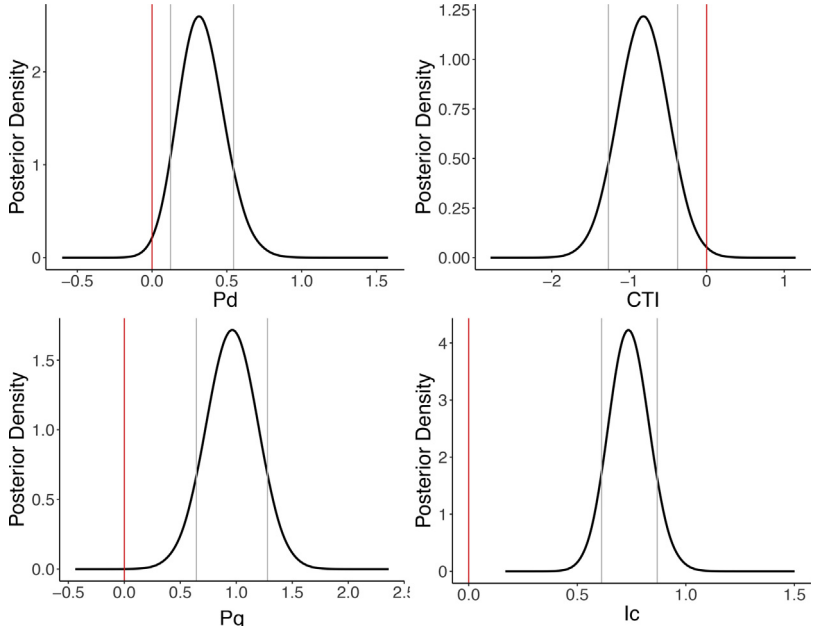

Fig. 9. Posterior densities of Model5 fixed effects. The vertical axis is density and the horizontal axis is on the logit scale. The bold red vertical line demarcates zero and gray vertical lines demarcate the 95% credible interval. Human population density (Pd), Compound Topographic Index (CTI), precipitation of the wettest quarter (Pq), and invasive cohort richness (Ic).

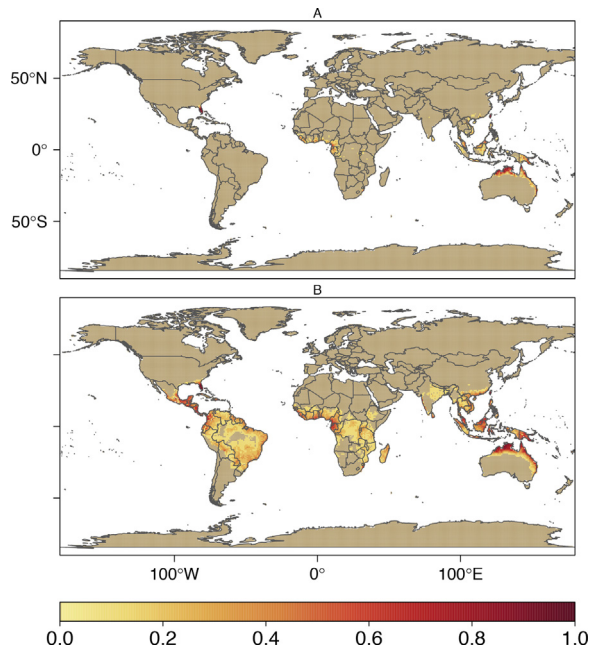

Fig. 10. Probability of *L. microphyllum* occurrence globally as predicted by the selected model (Model5) under current climate conditions (1950–2000). (A) Probability of occurrence with clustering; (B) probability of occurrence without clustering.

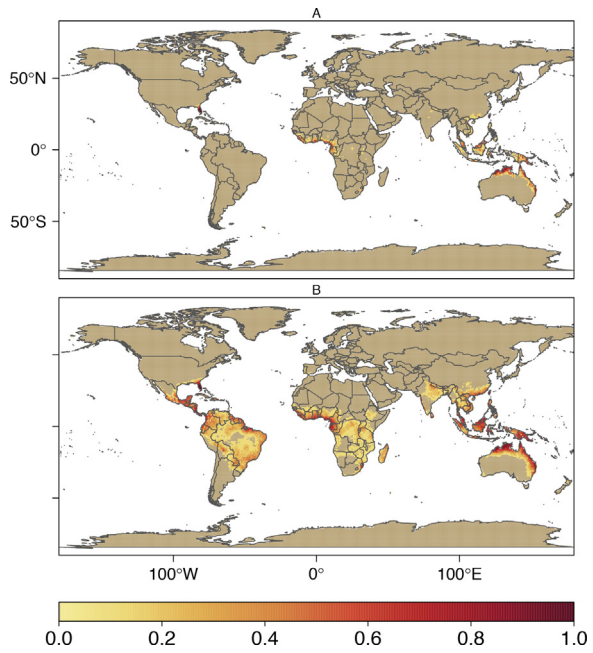

Fig. 11. Probability of *L. microphyllum* occurrence globally as predicted by the selected model (Model5) under future climate conditions (2070, 8.5 RCP). (A) Probability of occurrence with clustering; (B) probability of occurrence without clustering.

Fig. 10, the top figure [Fig. 11(A)] displays probabilities of occurrence assuming the clustering distance thresholds estimated from the species realized distribution and the bottom figure [Fig. 11(B)] relaxes the clustering assumption to predict occurrences based on habitat suitability alone. To better examine model predictions Figs. 12, 14, and 16 map predictions of *L. microphyllum* occurrence across smaller geographic extents.

Fig. 12(A) displays probabilities of occurrence over the southeastern United States under current climate conditions assuming *L. microphyllum* to be constrained by clustering distance in accordance with the functional relationship illustrated by Fig. 7. Fig. 12(B) also assumes clustering but predicts the probability of occurrence under the projected temperature and precipitation estimates provided by the 8.5 RCP for the year 2070. The bottom row of Fig. 12 likewise displays predictions under current [Fig. 12(C)] and future [Fig. 12(D)] climate conditions; however, these figures do not assume *L. microphyllum* to be constrained to clustering dis-

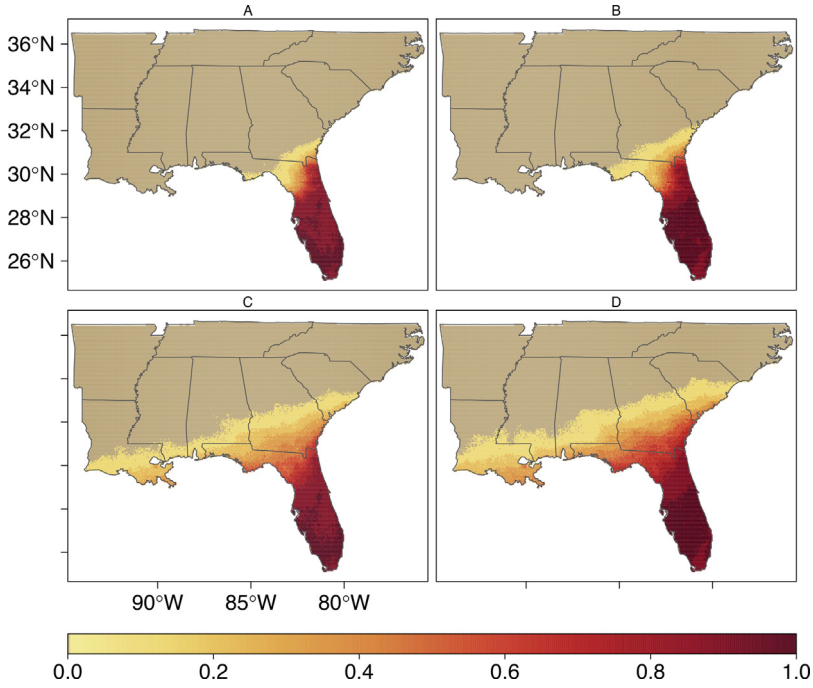

Fig. 12. Probability of *L. microphyllum* occurrence across the southeastern United States as predicted by the selected model (Model5). (A) Probability of occurrence under current climate conditions (1950–2000) with clustering; (B) probability of occurrence under future climate conditions (2070, 8.5 RCP) with clustering; (C) probability of occurrence under current climate conditions without clustering; (D) probability of occurrence under future climate conditions assuming no clustering.

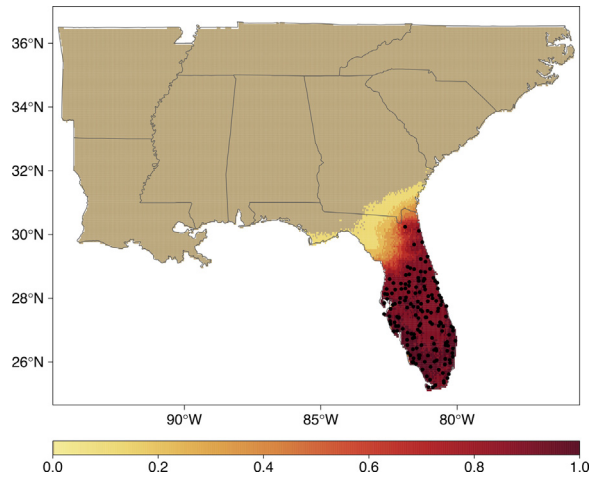

Fig. 13. Comparison of prediction to model points. Color coded map indicates probability of *L. microphyllum* occurrence across the southeastern United States under current climate conditions and clustering. Overlain points display presence locations used in model fitting. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

tances. Fig. 13 compares predictions for the United States to the presence locations used in model fitting.

Fig. 14 displays probabilities of *L. microphyllum* occurrence across Southeast Asia and Oceania. As with predictions for the United States, forecasts are provided for current and future climate conditions assuming clustering Fig. 14(A) and (B) respectively, as well as without being restricted to clustering distances [Fig. 14(C) and (D)]. Fig. 15 compares predictions for region to the presence locations used in model fitting.

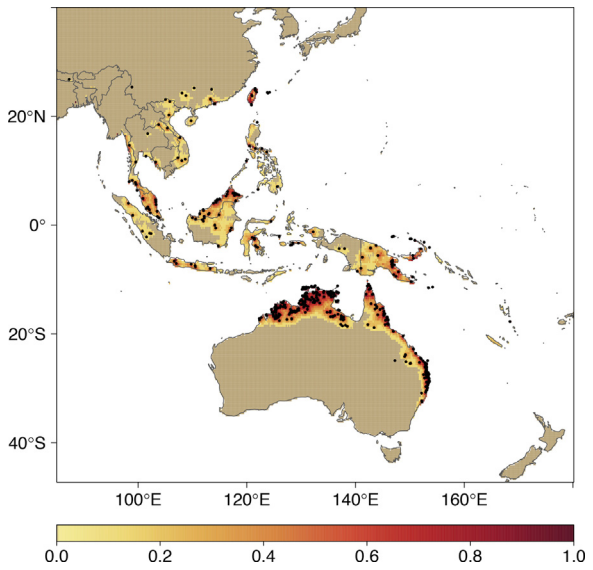

Fig. 15. Comparison of prediction to model points. Color coded map indicates probability of *L. microphyllum* occurrence across the Southeast Asia and Oceania under current climate conditions and clustering. Overlain points display presence locations used in model fitting. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Fig. 16 displays probabilities of *L. microphyllum* occurrence over southern Mexico, Central America, and the northern countries of South America. Predictions are provided for current climate and future conditions assuming clustering limitation Fig. 16(A) and (B) respectively, as well as no clustering constraint [Fig. 16(C) and (D)].

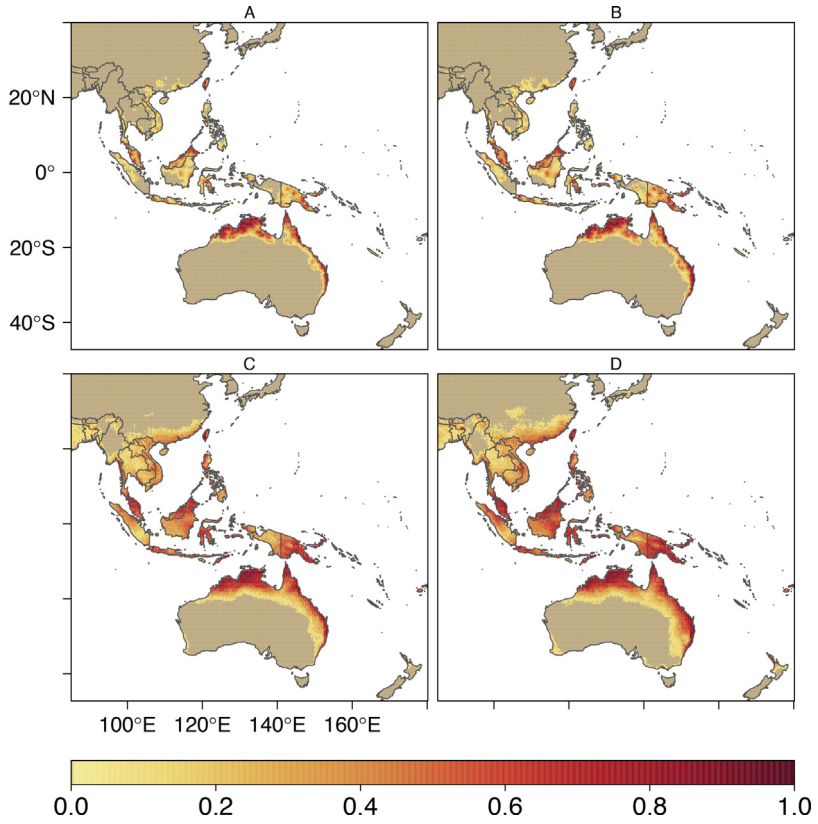

Fig. 14. Probability of *L. microphyllum* occurrence across Southeast Asia and Oceania as predicted by the selected model (Model5). (A) Probability of occurrence under current climate conditions (1950–2000) with clustering; (B) probability of occurrence under future climate conditions (2070, 8.5 RCP) with clustering; (C) probability of occurrence under current climate conditions assuming no clustering; (D) probability of occurrence under future climate conditions assuming no clustering.

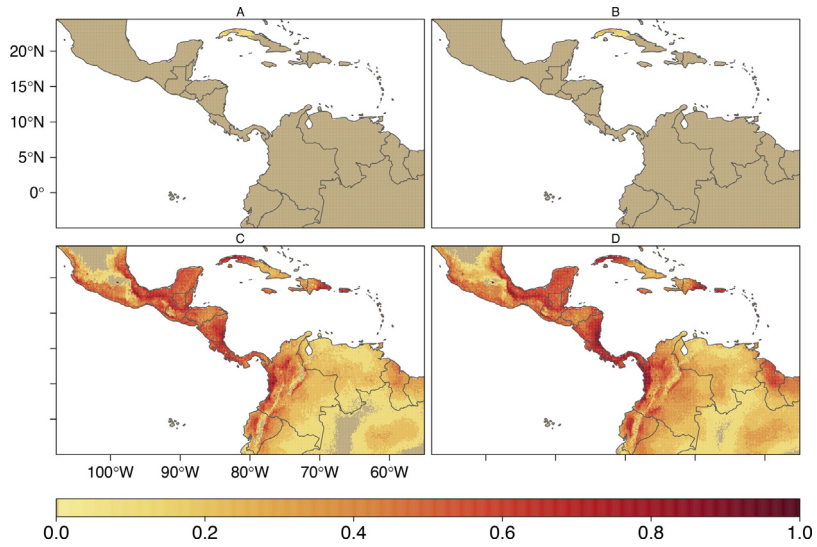

Fig. 16. Probability of *L. microphyllum* occurrence over southern Mexico, Central America, and northern portions of South America as predicted by the selected model (Model5). (A) Probability of occurrence under current climate conditions (1950–2000) with clustering; (B) probability of occurrence under future climate conditions (2070, 8.5 RCP) with clustering; (C) probability of occurrence under current climate conditions assuming no clustering; (D) probability of occurrence under future climate conditions assuming no clustering.

As shown by Fig. 2, this region is not associated with any documented presence locations.

4. Discussion

Comparison metrics (Table 2) reveal that models combining environmental and sptial covariates (Model1–Model5) out-perform the non-spatial model (Model6) in all cases. The metrics also demonstrate that specifying the spatial clustering term as a non-linear effect (Model1, Model2, and Model5) produces better results (e.g., increased accuracy and improved parsimony) than does approaching the covariate as a linear fixed effect (Model3) or excluding it altogether (Model4). These outcomes highlight the need to address spatial latencies, possibly even at multiple scales, when undertaking SDM.

Model results also point to the need to account for data collection bias. As shown in Table 3 and quantified in the Results section, a positive relationship exists between the frequency of human observed *L. microphyllum* occurrences and human population density. We interpret this relationship to mean that the observed species distribution reflects both ecological patterns and human accessibility; therefore, studies failing to account for collection bias may confound understanding of the underlying ecological processes. In brief, researchers need consider both observations and the observer.

Modeling of *L. microphyllum* temperature tolerance suggests that plant occurrences are likely to be limited to regions where the minimum temperature of the coldest month remains above 3.85 °C [Fig. 8]. This estimated temperature threshold is intermediate to the −2.2 °C tolerance investigated by Hutchinson and Langeland (2014) in a laboratory setting and a 8.0 °C limit approximated by Goolsby (2004) through examination of known occurrence locations. The model derived estimate also corresponds to the onset of plant tissue necrosis as anecdotally observed in the field [4.5 °C via Hutchinson and Langeland (2014)]. The rapid increase in probability of occurrence (from 0% to 73.1%, see Results) between the limiting temperature of 3.85 °C and 10.00 °C indicates a strong role for temperature in forecasting the plant's distribution under future climate warming.

Invasive cohort richness serves as a important predictor of the fern's presence and improves model performance with other fixed and random effects held constant (Model1 versus Model5). It is also possible that some portion of the variance explained by the biotic covariate is due to factors unrelated to biotic interaction. For example, it may be that co-occurrence of *L. microphyllum* with the invasive cohorts is due to a common, but un-modeled, soil or nutrient requirement, rather than a competitive or mutualistic interaction between species. That is, invasive cohorts may serve as a good *L. microphyllum* indicator species simply because they share a similar niche.

As with temperature and invasive cohort richness, the Compound Topographic Index and precipitation variables are found to be significant as judged by the credible interval. The inference drawn from the negative relationship between the CTI index and probability of *L. microphyllum* occurrence is that the plant avoids habitats that are subject to permanent inundation. Lakes, rivers, and other permanently inundated areas exhibit CTI values at and above the domain mean; however, this study and others (Volin et al., 2010; Gandiaga et al., 2009) suggest that *L. microphyllum* has preference for saturated or infrequently inundated habitats. As scaled by the CTI covariate, saturated and infrequently inundated areas fall just below the mean response. The precipitation covariate appears to support this interpretation; *L. microphyllum* shows a positive relationship to total precipitation of the wettest quarter, which obtains its highest levels across the wet tropics.

Although we cannot directly infer a mechanistic process from a correlative approach, to the extent that the model reproduces the statistical characteristics of the observations, we believe that the parameterizations used in the model can be said to adequately represent the salient underlying mechanisms (e.g., dispersal limitation). With this caveat, we speculate from the results that *L. microphyllum* is dispersal limited and has not yet achieved equilibrium over its introduced range in the southeastern United States. This conclusion is drawn primarily from the fine-scale clustering term modeled as a function of nearest neighbor distance [Fig. 7]. Although the tendency for *L. microphyllum* to cluster has been noted by others (Nauman and Austin, 1978; Volin et al., 2004; Wu et al., 2006; Fujisaki et al., 2010), model outcomes demonstrate that probability of presence increases non-linearly with proximity to known observations and is an important influence at distances of 205 km or less. In concordance with Volin et al. (2004) and Wu et al. (2006), we postulate that observed clustering by sporophytes is related to

spore dispersal by the plants. Further, we speculate that *L. microphyllum* exhibits stratified dispersal and that the model's clustering term provides a gross approximation for the decay rate in propagule pressure as distance from *L. microphyllum* increases. That is, as distance from an established plant increases, the chance of occurrence decreases. Stratified dispersal is a combination of gradual outward diffusion punctuated by occasional long-distance dispersal events (Schooler et al., 2009; Ferrarese and Garono, 2010). During a dispersal event, *L. microphyllum* may bypass or "jump" areas of unsuitable habitat to establish nascent clusters at distance that in turn spread by way of simple diffusion. Stratified dispersal explains why *L. microphyllum* shows a diffuse northward advance at the regional scale while simultaneously exhibiting local clustering.

Importantly, having constructed the clustering covariate we can undertake prediction presupposing unlimited ecological opportunity and simultaneously controlling for temperature, precipitation, preference for wetland habitats, invasive cohorts, sampling bias, and spatial latencies. Through forecasting occurrence without clustering (e.g., dispersal limitation), we can better assess the availability of suitable habitat outside of *L. microphyllum*'s realized distribution. Fig. 12(A) displays probability of *L. microphyllum* occurrence over the southeastern United States under current climatic conditions with documented observations held static. Stated differently, Fig. 12(A) provides predictions for when the nearest neighbor distances applied to the clustering function [Fig. 7] are measured from realized fern locations [Fig. 13]. The figure shows that probabilities of occurrence greater than 50% (red areas) either fall within Florida, or are in close proximity to the Florida state boundary. By comparison, Fig. 12(C) also displays the probability of occurrence under current climate conditions, but predictions here are made specifying that the nearest neighbor distance is zero for all mapped locations. By setting nearest neighbor distances to zero, we can identify locations of suitable habitat assuming observed clustering patterns to be artifacts of the realized distribution with no relevance to dispersal. Fig. 12(C) differs from Fig. 12(A) in that probabilities of occurrence greater than 50% spread outward beyond Florida to coastal South Carolina, the lower third of Georgia, and westward to southern Louisiana. Having changed the clustering term while holding all other effects constant and accounting for spatial latencies, the interpretation is that suitable environmental conditions are available beyond *L. microphyllum*'s documented range; however, the fern is constrained by time or (clustering) distance and not yet able to colonize them. As a more dramatic example of *L. microphyllum* apparent dispersal limitation, Fig. 16(C) illustrates high probabilities of occurrence across majority portions of southern Mexico, Central America, and South America given unlimited access by *L. microphyllum*, but adding the clustering constraint to the prediction reduces the probability of occurrence substantially [Fig. 16(A)]. Once again, the implication is that suitable habitat exists in these regions as evidenced by temperature, precipitation, wetness, and the presence of invasive cohorts, but these regions have yet to be invaded.

Extrapolating model predictions to the year 2070 under the 8.5 RCP indicates that *L. microphyllum* will expand its current distribution globally to inhabit regions currently inaccessible due to temperature and dispersal limitation. Fig. 12(B) and (D) compare projected species distributions in the southeastern United States for constrained and unconstrained clustering, respectively. Unless the plant is introduced to the neotropics, the probability of *L. microphyllum* occurrence is minimal under the 2070 scenario [Fig. 16(B)]. However, should the fern invade or be introduced, it is likely to spread rapidly throughout Mexico, Central America, and South America [Fig. 16(D)]. Some spreading is also predicted for the fern's native habitat in Southeast Asia and Australia [Fig. 14]; however, expansion in the plant's native habitat does not appear as dramatic as those from the introduced range considering the number of documented occurrences throughout the region [Fig. 15]. This is not a surprising finding as *L. microphyllum* would already have had greater time and opportunity to fill available habitat throughout its native range.

5. Conclusion

Species distribution modeling aimed at forecasting the distribution of invasive species under projected global warming offers land managers an important tool for assessing future ecological risk and for prioritizing management actions. The current study applies Bayesian inference and newly accessible geostatistical tools to forecast global range expansion for the ecosystem altering invasive climbing fern *L. microphyllum*. Results indicate that *L. microphyllum* will undergo global range expansion in concert with anthropogenic global warming and continued dispersal. In applying ecological modeling techniques to invasive species, we emphasize the importance of considering species equilibrium assumptions, the potential for non-linear responses to environmental gradients, the explanatory power of biotic covariates, accounting for spatial errors, and to the extent possible correcting for sampling bias inherent to species occurrence data.

Acknowledgement

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

References

Aiello-Lammens, M.E., Boria, R.A., Radosavljevic, A., Vilela, B., Anderson, R.P., 2015. spThin: an R package for spatial thinning of species occurrence records for use in ecological niche models. Ecography 38 (5), 541–545.
Allen, J.M., Bradley, B.A., 2016. Out of the weeds? Reduced plant invasion risk with climate change in the continental United States. Biol. Conserv. 203, 306–312, http://dx.doi.org/10.1016/j.biocon.2016.09.015.
Austin, M., 2007. Species distribution models and ecological theory: a critical assessment and some possible new approaches. Ecol. Model. 200 (1–2), 1–19.
Austin, M.P., Belbin, L., Meyers, J.A., Doherty, M.D., Luoto, M., 2006. Evaluation of statistical models used for predicting plant species distributions: role of artificial data and theory. Ecol. Model. 199 (2), 197–216.
Austin, M.P., Nicholls, A.O., Margules, C.R., 1990. Measurement of the realized qualitative niche: environmental niches of five eucalyptus species. Ecol. Monogr. 60 (2), 161–177.
Ayllón, D., Nicola, G.G., Elvira, B., Parra, I., Almodóvar, A., 2013. Thermal carrying capacity for a thermally-sensitive species at the warmest edge of its range. PLOS ONE 8 (11).
Baddeley, A., Turner, R., 2005. spatstat: an R package for analyzing spatial point patterns. J. Stat. Softw. 12 (6), 1–42 http://www.jstatsoft.org/v12/i06/.
Bakka, H., Vanhatalo, J., Illian, J., Simpson, D., Rue, H., 2016. Accounting for Physical Barriers in Species Distribution Modeling With Non-Stationary Spatial Random Effects., pp. 1–21, ArXiv, http://arxiv.org/abs/1608.03787.
Beckner, J., 1968. *Lygodium microphyllum*, another fern escaped in Florida. Am. Fern J. 58, 93–94.
Belsley, D.A., Kuh, E., Welsch, R.E., 1980. Identifying Influential Data and Sources of Collinearity, Book Chapter.
Boria, R.A., Olson, L.E., Goodman, S.M., Anderson, R.P., 2014. Spatial filtering to reduce sampling bias can improve the performance of ecological niche models. Ecol. Model. 275, 73–77, http://dx.doi.org/10.1016/j.ecolmodel.2013.12.012.
Bradley, B.A., Wilcove, D.S., Oppenheimer, M., 2010. Climate change increases risk of plant invasion in the Eastern United States. Biol. Invasions 12 (6), 1855–1872.
Brown, C., Law, R., Illian, J.B., Burslem, D.F.R.P., 2011. Linking ecological processes with spatial and non-spatial patterns in plant communities. J. Ecol. 99 (6), 1402–1414.
Buckley, Y.M., 2008. The role of research for integrated management of invasive species, invaded landscapes and communities. J. Appl. Ecol. 45 (2), 397–402.
CIESIN, 2005. Gridded Population of the World, Version 3 (GPWv3): Population Density Grid., http://dx.doi.org/10.7927/H4XK8CG2.
Dennis, R., Thomas, C., 2000. Bias in butterfly distribution maps the influence of hot spots and recorder's home range. J. Insect Conserv. 4 (2), 73–77.
Dormann, C.F., 2007. Effects of incorporating spatial autocorrelation into the analysis of species distribution data. Global Ecol. Biogeogr. 16 (2), 129–138.
EDDMAPS, 2016. (EDDMAPS) Early Detection and Distribution Mapping System. http://www.eddmaps.org.

Ehrenfeld, J.G., 2010. Ecosystem consequences of biological invasions. Annu. Rev. Ecol. Evol. Syst. 41 (1), 59–80, http://dx.doi.org/10.1146/annurev-ecolsys-102209-144650.
Elith, J., 2015. Predicting Distributions of Invasive Species., pp. 1–28, eprint arXiv:1312.0851, http://arxiv.org/abs/1312.0851.
Elith, J., Kearney, M., Phillips, S., 2010. The art of modelling range-shifting species. Methods Ecol. Evol. 1 (4), 330–342.
Elith, J., Leathwick, J., 2009. Species distribution models: ecological explanation and prediction across space and time. Annu. Rev. Ecol. Evol. Syst. 40 (2009), 677–697.
Elsner, J.B., Fricker, T., Widen, H.M., Castillo, C.M., Humphreys, J., Jung, J., Gredzens, C., 2016. The relationship between elevation roughness and tornado activity: a spatial statistical model fit to data from the central great plains. J. Appl. Meteorol. Climatol. 55 (4), 849–859.
Ferrarese, E., Garono, R.J., 2010. Dispersal of *Galerucella pusilla* and *G. calmariensis* via passive water transport in the Columbia River Estuary. Biol. Control 52 (2), 115–122, http://dx.doi.org/10.1016/j.biocontrol.2009.10.009.
Fithian, W., Elith, J., Hastie, T., Keith, D.A., 2015. Bias correction in species distribution models: pooling survey and collection data for multiple species. Methods Ecol. Evol. 6 (4), 424–438.
Florida Exotic Pest Plant Council, 2015. Florida Exotic Pest Plant Council's 2015 List of Invasive Plant Species. http://www.fleppc.org/list/2015FLEPPCLIST-LARGEFORMAT-FINAL.pdf.
Fuglstad, G.-A., Simpson, D., Lindgren, F., Rue, H., 2015. Constructing Priors that Penalize the Complexity of Gaussian Random Fields., pp. 1–33, arXiv.org, http://arxiv.org/abs/1503.00256.
Fujisaki, I., Brandt, L.A., Chen, H., Mazzotti, F.J., 2010. Colonization, spread, and growth of *Lygodium microphyllum* on tree islands in a wetland in Florida. Invasive Plant Sci. Manag. 3 (4), 412–420.
Gandiaga, S., Volin, J.C., Kruger, E.L., Kitajima, K., 2009. Effects of hydrology on the growth and physiology of an invasive exotic, *Lygodium microphyllum* (Old World climbing fern). Weed Res. 49 (3), 283–290.
Gelman, A., Hwang, J., Vehtari, A., 2014. Understanding predictive information criteria for Bayesian models. Stat. Comput. 24 (6), 997–1016.
Golding, N., Purse, B.V., 2016. Fast and flexible Bayesian species distribution modelling using Gaussian processes. Methods Ecol. Evol. 7, 598–608.
Goolsby, J.A., 2004. Potential distribution of the invasive Old World Climbing Fern, *Lygodium microphyllum* in North and South America. Nat. Areas J. 24, 351–353.
Gordon, D.R., 1998. Effects of invasive, non-indigenous plant species on ecosystem processes: lessons from Florida. Ecol. Appl. 8 (4), 975–989.
Guisan, A., Thuiller, W., 2005. Predicting species distribution: offering more than simple habitat models. Ecol. Lett. 8 (9), 993–1009.
Hendricks, J., Pelzer, B., 2004. Collinearity involving ordered and unordered categorical variables. In: RC33 conference in Amsterdam (1991)., pp. 1–17.
Hijmans, R.J., Cameron, S., Parra, J., Jones, P., Jarvis, A., 2005. Very high resolution interpolated climate surfaces for global land areas. Int. J. Climatol. 25, 1965–1978 http://www.worldclim.org/.
Hijmans, R.J., Elith, J., 2016. Species Distribution Modeling With R Introduction, ftp://cran.r-project.org/pub/R/web/packages/dismo/vignettes/sdm.pdf.
Hobbs, R.J., Arico, S., Aronson, J., Baron, J.S., Bridgewater, P., Cramer, V.A., Epstein, P.R., Ewel, J.J., Klink, C.A., Lugo, A.E., Norton, D., Ojima, D., Richardson, D.M., Sanderson, E.W., Valladares, F., Vilà, M., Zamora, R., Zobel, M., 2006. Novel ecosystems: theoretical and management aspects of the new ecological world order. Global Ecol. Biogeogr. 15 (1), 1–7.
Hoeting, J.A., 2009. The importance of accounting for spatial and temporal correlation in analyses of ecological data. Ecol. Appl. 19 (3), 574–577 http://www.jstor.org/stable/27645996.
Hooten, M.B., Hobbs, N.T., 2015. A guide to Bayesian model selection for ecologists. Ecol. Monogr. 85 (1), 3–28.
Hulme, P.E., 2016. Climate change and biological invasions: evidence, expectations, and response options. Biol. Rev., http://dx.doi.org/10.1111/brv.12282.
Humphreys, J.M., Elsner, J.B., Jagger, T.H., Mahjoor, A., 2016. Disaggregating the patchwork. Wetlands, 1–15, http://dx.doi.org/10.1007/s13157-016-0859-z.
Huntley, B., Berry, P.M., Cramer, W., Mcdonald, A.P., 1995. Modelling present and potential future ranges of some European higher plants using climate response surfaces. J. Biogeogr. 22 (6), 967–1001.
Hurtt, G.C., Pacala, S.W., 1995. The consequences of recruitment limitation: reconciling chance, history and competitive differences between plants. J. Theor. Biol. 176 (1), 1–12 http://www.sciencedirect.com/science/article/pii/S0022519385701701.
Hutchinson, J.T., Langeland, K.A., 2014. Tolerance of *Lygodium microphyllum* and *L. japonicum* spores and gametophytes to freezing temperature. Invasive Plant Sci. Manag. 7 (2), 328–335, http://dx.doi.org/10.1614/IPSM-D-13-00074.1.
Illian, J., Martino, S., Sorbye, S.H., Gallego-Fern??ndez, J.B., Zunzunegui, M., Esquivias, M.P., Travis, J.M.J., 2013. Fitting complex ecological point process models with integrated nested Laplace approximation. Methods Ecol. Evol. 4 (4), 305–315.
IMapInvasives, 2015. iMapInvasives: An Online Data System Supporting Strategic Invasive Species Management. http://imapinvasives.org.
Jablonski, D., 2008. Biotic interactions and macroevolution: extensions and mismatches across scales and levels. Evolution 62 (4), 715–739.
Kadmon, R., Farber, O., Danin, A., 2003. A systematic analysis of factors affecting the performance of climatic envelope models. Ecol. Appl. 13 (3), 853–867.
Kadmon, R., Farber, O., Danin, A., 2004. Effect of roadside bias on the accuracy of predictive maps produced by bioclimatic models. Ecol. Appl. 14 (2), 401–413, http://dx.doi.org/10.1890/02-5364.
Leach, K., Montgomery, W.I., Reid, N., 2016. Modelling the influence of biotic factors on species distribution patterns. Ecol. Model. 337, 96–106, http://dx.doi.org/10.1016/j.ecolmodel.2016.06.008.
Leung, B., Lodge, D.M., Finnoff, D., Shogren, J.F., Lewis, M.A., Lamberti, G., 2002. An ounce of prevention or a pound of cure: bioeconomic risk analysis of invasive species. Proc. Biol. Sci./R. Soc. 269 (1508), 2407–2413.
Lindgren, F., Rue, H., 2015. Bayesian spatial modelling with R-INLA. J. Stat. Softw. 63 (19), 1–25 http://www.jstatsoft.org/v63/i19/.
Lindgren, F., Rue, H., Lindström, J., 2011. An explicit link between Gaussian fields and Gaussian Markov random field: the stochastic partial differential equations approach. J. R. Stat. Soc. Ser. B: Stat. Methodol. 73, 423–498.
McKinney, M.L., Lockwood, J.L., 1999. Biotic homogenization: a few winners replacing many losers in the next mass extinction. Trends Ecol. Evol. 14, 450–453.
Moore, et al., 1993. Terrain attributes: estimation methods and scale effects. In: Jakeman, J.A., Beck, M.B., Wiley, M.M. (Eds.), Modeling Change in Environmental Systems, London. , pp. 189–214.
Nauman, C.E., Austin, D.F., 1978. Spread of the exotic fern *Lygodium microphyllum* in Florida. Am. Fern J. 68 (4), 65–66.
Olden, J.D., Poff, N.L., Douglas, M.R., Douglas, M.E., Fausch, K.D., 2004. Ecological and evolutionary consequences of biotic homogenization. Trends Ecol. Evol. 19 (1), 18–24.
Pachauri, R., Meyer, L., 2014. IPCC, 2014: Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Tech. Rep.
Pearson, R.G., Dawson, T.P., Pearson, R.G., Dawson, T.P., 2003. Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful? Global Ecol. Biogeogr. 12 (5), 361–371.
Pellissier, L., Anne Bråthen, K., Pottier, J., Randin, C.F., Vittoz, P., Dubuis, A., Yoccoz, N.G., Alm, T., Zimmermann, N.E., Guisan, A., 2010. Species distribution models reveal apparent competitive and facilitative effects of a dominant species on the distribution of tundra plants. Ecography 33 (6), 1004–1014.
Pemberton, R., Ferriter, A.P., 1998. Old world climbing fern (*Lygodium microphyllum*), a dangerous invasive weed in Florida. Am. Fern J. 88 (4), 165–175.
Pysek, P., Jarok, V., Hulme, P.E., Pergl, J., Hejda, M., Schaffner, U., Vila, M., 2012. A global assessment of invasive plant impacts on resident species, communities and ecosystems: the interaction of impact measures, invading species' traits and environment. Global Change Biol. 18 (5), 1725–1737.
R Core Team, 2016. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria https://www.R-project.org/.
Renner, I.W., Elith, J., Baddeley, A., Fithian, W., Hastie, T., Phillips, S.J., Popovic, G., Warton, D.I., 2015. Point process models for presence-only analysis. Methods Ecol. Evol. 6 (4), 366–379.
Roberts, R.E., 1996. Climbing Fern Wreaks Wetland Havoc., Tech. Rep. Florida Department of Environmental Protection Resource Management Notes 8:13.
Rodgers, L., Pernas, T., Hill, S.D., 2014. Mapping invasive plant distributions in the Florida everglades using the digital aerial sketch mapping technique. Invasive Plant Sci. Manag. 7 (2), 360–374, http://dx.doi.org/10.1614/IPSM-D-12-00092.1.
Rogers, C.E., McCarty, J.P., 2000. Climate change and ecosystems of the Mid-Atlantic Region. Climate Res. 14 (3 SPECIAL 7), 235–244.
Rue, H., Martino, S., Chopin, N., 2009. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. J. R. Stat. Soc. Ser. B: Stat. Methodol. 71 (2), 319–392.
Rydgren, K., Økland, R.H., Økland, T., 2003. Species response curves along environmental gradients. A case study from SE Norwegian swamp forests. J. Veg. Sci. 14 (August 2014), 869–880, http://dx.doi.org/10.1111/j.1654-1103.2003.tb02220.x.
Schooler, S.S., McEvoy, P.B., Hammond, P., Coombs, E.M., 2009. Negative per capita effects of two invasive plants, *Lythrum salicaria* and *Phalaris arundinacea*, on the moth diversity of wetland communities. Bull. Entomol. Res. 99 (3), 229–243.
Simpson, D., Illian, J., Lindgren, F., Sørbye, S.H., Rue, H., 2015. Going off grid: computationally efficient inference for log-Gaussian Cox processes. Biometrika 103 (1), 49–70.
Soberón, J., Nakamura, M., 2009. Niches and distributional areas: concepts, methods, and assumptions. Proc. Natl. Acad. Sci. 106 (Supplement 2), 19644–19650.
USDA, APHIS, PPQ, 2012. Federal Noxious Weed List. Tech. Rep. USDA, Animal and Plant Health Inspection Service (APHIS), Plant Protection and Quarantine (PPQ).
Václavík, T., Meentemeyer, R.K., 2012. Equilibrium or not? Modelling potential distribution of invasive species in different stages of invasion. Divers. Distrib. 18 (1), 73–83.
Velázquez, E., Martínez, I., Getzin, S., Moloney, K.A., Wiegand, T., 2016. An evaluation of the state of spatial point pattern analysis in ecology. Ecography (December 2015), 1–14.
Velázquez, E., Paine, C.E.T., May, F., Wiegand, T., 2015. Linking trait similarity to interspecific spatial associations in a moist tropical forest. J. Veg. Sci. 26 (6), 1068–1079.
Verbruggen, H., Tyberghein, L., Belton, G.S., Mineur, F., Jueterbock, A., Hoarau, G., Gurgel, C.F.D., De Clerck, O., 2013. Improving transferability of introduced species' distribution models: new tools to forecast the spread of a highly invasive seaweed. PLOS ONE 8 (6), 1–13.

Volin, J.C., Kruger, E.L., Volin, V.C., Tobin, M.F., Kitajima, K., 2010. Does release from natural belowground enemies help explain the invasiveness of *Lygodium microphyllum*? A cross-continental comparison. Plant Ecol. 208 (2), 223–234.
Volin, J.C., Lott, M.S., Muss, J.D., Owen, D., 2004. Predicting rapid invasion of the florida everglades by old world climbing fern author. Divers. Distrib. 10 (5), 439–446.
Watanabe, S., 2013. A widely applicable bayesian information criterion. Mach. Learn. Res. 14 (1), 867–897.
Wickham, J., Barnes, C., Wade, T., 2015. Combining NLCD and MODIS to create a land cover-albedo database for the continental United States. Remote Sens. Environ. 170, 143–152.
Wiens, J.J., 2011. The niche, biogeography and species interactions. Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci. 366, 2336–2350 http://rstb.royalsocietypublishing.org/content/366/1576/2336.
Wiens, J.J.A., 1989. Spatial scaling in ecology. Funct. Ecol. 3 (4), 385–397.
Wilcove, D.S., Rothstein, D., Dubow, J., Phillips, A., Losos, E., 1998. Quantifying threats to imperiled species in the United States. BioScience 48, 607–615.
Wu, Y., Rutchey, K., Wang, N., Godin, J., 2006. The spatial pattern and dispersion of Lygodium microphyllum in the Everglades wetland ecosystem. Biol. Invasions 8 (7), 1483–1493.
Zurek, et al., 2008. Towards New Scenarios for Analysis of Emissions, Climate Change, Impacts, and Response Strategies. Intergovernmental Panel on Climate Change, Geneva.