Crowdsourcing Meets Ecology: Hemispherewide Spatiotemporal Species Distribution Models

Daniel Fink, Theodoros Damoulas, Nicholas E. Bruns, Frank A. La Sorte, Wesley M. Hochachka, Carla P. Gomes, and Steve Kelling

■ *Ecological systems are inherently complex. The processes that affect the distributions of animals and plants operate at multiple spatial and temporal scales, presenting a unique challenge for the development and coordination of effective conservation strategies, particularly for wide-ranging species. In order to study ecological systems across scales, data must be collected at fine resolutions across broad spatial and temporal extents. Crowdsourcing has emerged as an efficient way to gather these data by engaging large numbers of people to record observations. However, data gathered by crowdsourced projects are often biased due to the opportunistic approach of data collection. In this article, we propose a general class of models called AdaSTEM (for adaptive spatiotemporal exploratory models) that are designed to meet these challenges by adapting to multiple scales while exploiting variation in data density common with crowdsourced data. To illustrate the use of AdaSTEM, we produce intraseasonal distribution estimates of long-distance migrations across the Western Hemisphere using data from eBird, a citizen science project that utilizes volunteers to collect observations of birds. Subsequently, model diagnostics are used to quantify and visualize the scale and quality of distribution estimates. This analysis shows how AdaSTEM can automatically adapt to complex spatiotemporal processes across a range of scales, thus providing essential information for full-life cycle conservation planning of broadly distributed species, communities, and ecosystems.*

> *… there is no single natural scale at which ecological phenomena should be studied — systems generally show characteristic variability on a range of spatial, temporal and organizational scales.*
> *(Levin 1992)*

Broad-scale environmental and ecological systems originate as simultaneous processes operating across a range of spatiotemporal scales. To study and conserve these systems it is crucial to understand the multi-scale structure of underlying processes. For example, consider some of the processes affecting birds during migration. Climatic phenomena, like El Niño southern oscillation and the North Atlantic oscillation (Grosbois et al. 2008) can affect migration timing and direction at hemispheric spatial scales for years at a time. Regional migration pathways are affected by mesoscale spatial processes that define boundaries between major ecosystems like prairies and forests (Fortin and Dale 2005). At a local scale, individual foraging decisions may be based on the availability of specific plants or insects within small habitat patches (Bonter et al. 2009).

 ISSN 0738-4602

An important goal for many conservation applications is spatial prioritization, the identification, delineation, and ranking of regions for management actions (Moilanen, Wilson, and Possingham 2009). For applications with large geographic extents, multiscale spatial prioritization is essential for land managers to identify land parcels for acquisition (Schuster and Arcese 2013) or remediation. For example, with declining populations of long-distance migrating birds, a key question is whether declines are caused by events on breeding grounds, nonbreeding grounds, or during migrations. Answering this question requires the comparison of regional population estimates across continents. Once important large-scale regions are identified, fine-scale information is needed to identify critical habitat patches and individual migration stopover sites.

Multiscale information is also vital to a broad range of related sustainability applications. Scientists need to prioritize regions for disease control management (Ostfeld, Glass, and Keesing 2005). Policy makers need to select sites for human development while trying to minimize ecological costs, for example, when developing wind farms (Drewitt and Langston 2006). In these examples, multiscale information is valuable because it allows managers to inform policy and make objective decisions at the appropriate spatial and temporal scale (Gomes 2009).

One of the fundamental challenges of studying multiscale processes is the collection of data. Consistent sources of fine-resolution data are needed across broad extents. For many types of biodiversity data, the largest collection programs are national in scope. Unfortunately, the variation among national programs hinders ecological study and conservation planning for broadly distributed species. Because of the difficulty and expense of collecting systematic biodiversity data across large extents, many researchers are beginning to use data collected by citizen science projects through crowdsourcing techniques (Dickinson, Zuckerberg, and Bonter 2010).

Crowdsourcing projects that engage the public to collect data have been very successful at collecting data across large areas. However, these data tend to be irregularly and sparsely distributed. When participants opportunistically choose where to report their observations, the data tend to follow patterns of human activity (Hochachka et al. 2012), for example, figure 1. This structure presents a challenge for the analysis of multiscale processes because variation in data density translates into variation in scale at which valid inferences can be made. Intuitively, as data density increases at a particular location, the information available for estimating processes operating there also increases, allowing study of smaller scale processes. In addition to the density of data, the scale structure of analytical models affects the scale at which valid inference can be made (Dungan et al. 2002). Thus, to take full advantage of crowdsourced data, models that can discover multiscale structure by adapting to the varying density of irregularly distributed observations are needed. Additionally, to make full use of these models tools are needed to quantify and communicate the finest scales at which inferences can reliably be made.

The most common approach to account for spatial and spatiotemporal scale has been to model how correlation varies as a function of proximity. This has been an active research area in statistics and machine learning for the past two decades (Cressie 1993; Rasmussen and Williams 2006; Cressie and Wikle 2012). Methodologies such as Kriging (Cressie 1986), Gaussian processes (Paciorek and Schervish 2004), Gaussian Markov random fields (Rue and Held 2005), splines (Pintore, Speckman, and Holmes 2006; Kammann and Wand 2003), and autoregressive models (Huang, Cressie, and Gabrosek 2002; Tzeng, Huang, and Cressie 2005) have been proposed to estimate and account for spatial correlation in stationary settings, where the effects of proximity are assumed to be constant. More recently, research has focused on accounting for nonstationary spatial correlation that allows for varying scales. Nonstationary covariance functions have been proposed for Gaussian processes and Kriging models (Stein 2005, Paciorek and Schervish 2004, Jun and Stein 2008, Pintore and Holmes 2004). Similarly, spline methods have been developed with spatially varying penalties (Pintore, Speckman, and Holmes 2006; Crainiceanu et al. 2007). However, the computational complexity of many of these models is high (Cressie and Johannesson 2008, Gelfand 2012) necessitating trade-offs between computational efficiency and the scale of analysis for large data sets where the number of observations and locations is in the millions.

In this article we present an ensemble model designed to discover scale-dependent, nonstationary predictor-response relationships from large, irregularly distributed observational data (Fink, Damoulas, and Dave 2013). We call this model AdaSTEM, an extension to the spatiotemporal exploratory model (Fink et al. 2010) based on a simple yet effective divide and recombine strategy. The first stage of the model divides the extent of analysis into regional units based on data density using tree data structures. Next, a mixture model is used to organize regional units into a cohesive framework while facilitating discovery of nonstationary patterns of predictor-response associations among regions. Within the regional units, a user-specified model carries out the supervised learning task that associates predictors and responses. AdaSTEM is a highly automated ensemble model with a pleasingly parallel implementation that scales to big data. The experiments described here were conducted on the Lonestar cluster through an allocation on XSEDE (www.xsede.org).

We illustrate the use of AdaSTEM with an analysis

of crowdsourced data from eBird (www.ebird.org) (Sullivan et al. 2009). The goal was to estimate the daily distributions of long-distance migratory birds across the Western Hemisphere (figure 4) with the finest spatial resolution possible. Using AdaSTEM, we produced the first hemispherewide population-level distribution estimates for three species of long-distance migratory birds. To facilitate the interpretation and application of the distribution estimates we then used model diagnostics to quantify and visualize the spatial variation in scale, bias, and uncertainty. Together, these results provide the information necessary to make statistical comparisons and ecological interpretations across a range of spatial scales.

In the sections following we will describe the fixed-scale ensemble framework STEM and then discuss the AdaSTEM extension. We will then describe the analysis of the eBird data in detail and assess the scale and quality of the AdaSTEM estimates. Finally, we will follow up with a brief discussion of our findings and methodology.

STEM: Spatiotemporal Exploratory Models

STEM (Fink et al. 2010) is a mixture model designed to adapt to nonstationary, scale-dependent processes. This is achieved by creating a dense mixture of local learning models with compact overlapping support. A user-specified supervised learning model, the base model, accounts for variation as a function of predictor values within its support set, which we call a stixel (for spatiotemporal pixel). Because the stixels are compact sets, the learning model can adapt to local predictor-response associations while limiting long-range extrapolation. Utilizing the fact that stixels overlap, predictions at a specified location, s, are made by taking an average across all base models whose stixels include that location. This combines the bias-reducing properties of local models (for example, decision trees [Breiman et al. 1984]) with the variance-reducing properties of randomized ensembles (for example, bagging [Breiman 1996]). In this article we consider two classes of base modes — linear models fit through least squares for the synthetic experiments and logistic generalized additive models (GAM) (Wood 2006) for the binary classification of eBird data.

The Mixture Model

The approach described here is based on ensemble or mixture modeling (Kuncheva and Whitaker 2003; Hastie, Tibshirani, and Friedman 2009) with a focus on prediction for large data sets. To this end we treat the estimation of the base models independently. The ensemble response is computed as the weighted average taken across base models with shared support, that is, within overlapping stixels; see figure 1 (center). For simplicity, all supporting base models are weighed equally. STEM can be considered as a spatiotemporal wrapper for any user-specified base model.

Formally, let $\{y_n(s), \mathbf{x}_n(s)\}^N_{n=1}$ be the set of observed responses and predictors $\mathbf{x}_i(s) = [x_{i,1}(s),\cdots, x_{i,d}(s)]$ indexed by locations[1] $s \in R^k$ within the study area $D \subset R^k$. $y(s)$ is modeled as the ensemble response:

$$y_e(s) = \sum_{m=1}^{M} \alpha_m(s) f_m(\mathbf{x}(s), D_m, s) \quad (1)$$

with M base models f_m each defined on its own stixel $D_m \subset D$ with mixture weights $\alpha_m(s)$. Each base model f_m is independently fit to N_m observations falling within D_m. The mixture weights at coordinates s are

$$\alpha_m(s) = n^{-1}(s) I(s \in D_m) \quad (2)$$

where the indicator function $I(s \in D_m)$ indicates membership of s in the support set D_m. The ensemble support $n(s)$ is the number of base models that support coordinate

$$s : n(s) = \sum I(s \in D_m) \quad (3)$$

with the sum taken over the M base models.

STEM uses a simple ensemble design with fixed size stixels. The ensemble is created by partitioning the study extent D into a regular set of M square stixels D_m with sides of length λ. Second, P such partitions are sampled, randomizing the position of each left corner π to form an ensemble of overlapping stixels. Within each stixel D_m we require that the number of observations N_m meet a minimum sample size, γ, to fit a base model. Stixels where $N_m < \gamma$ are omitted from the ensemble. Thus, the maximum ensemble support $n(s)$ at location s is P the number of partitions. The algorithm is given in algorithm 1. Note that estimates of parameters are indicated by placing a hat over the corresponding symbol.

For mixture models the smallest scale signal that can reliably be estimated at a given location s is determined by characteristics of both the stixels supporting s and the choice of base model f_m. In general, for a given class of base model, the smallest scale signal that can be estimated will increase with the size of the stixels. For a single base model, variation that cannot be explained will tend to be averaged out across larger areas as stixel size increases. Similarly, at the ensemble level, base model estimates will be averaged across larger areas as stixel size increases. Thus, for the mixture model, the range over which information is shared and averaged increases for larger stixels. As a result of this, the minimum scale over which inferences can be made will also increase with larger stixels.

Adaptive Multiscale Modeling with AdaSTEM

The parameter λ controls the size of the stixels, which indirectly controls the minimum scale of the signal

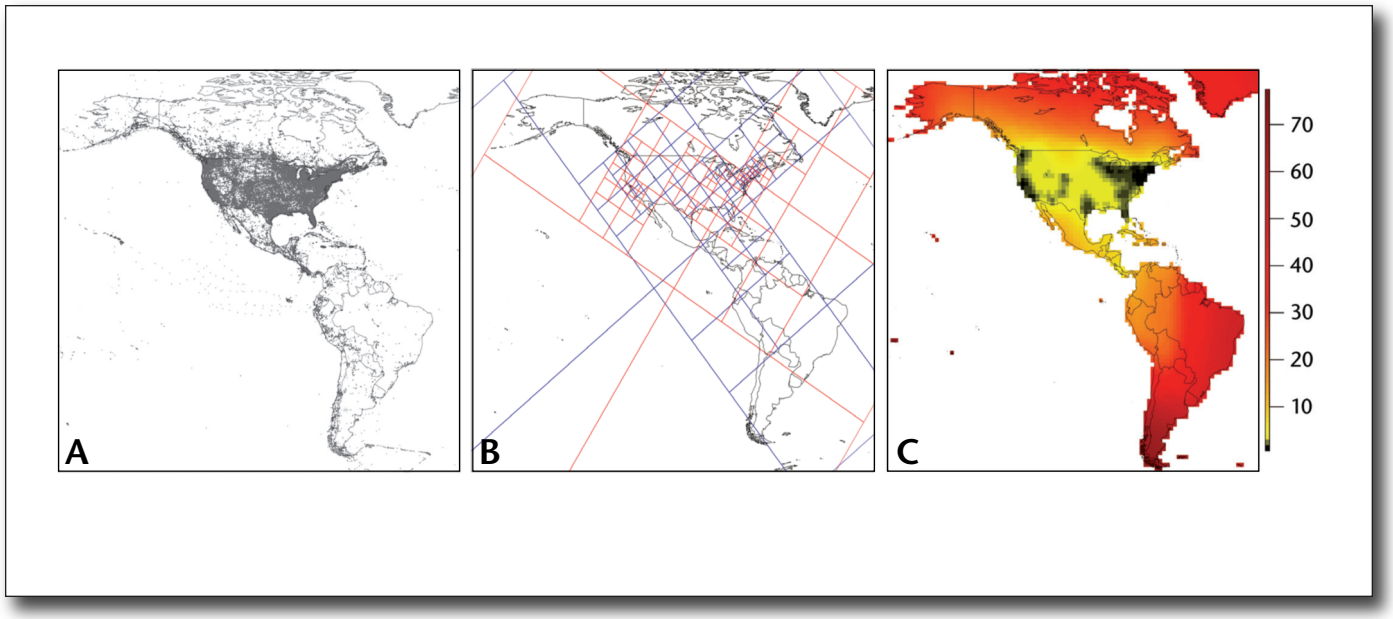

Figure 1. Quadtree.

Left: eBird data locations showing the varying density of observations. *Center:* Two realizations of quadtree generated stixels, red and blue. *Right:* Average quadtree stixel size (in degrees) follows the density of observations. (Color version of figure presented in electronic version of *AI Magazine*).

Input: Spatial dataset with extent D
Output: Ensemble model estimator $\hat{y}_e(s)$
Set λ by cross-validation
for p=*1* **to** P **do**
  Randomize partition corner $\pi_p \sim \mathcal{U}(0, \lambda)$
  Partition D into M_{p} stixels each with length λ
  for m=*1* to M_{p} **do**
    if $N_{\mathrm{m}} \geq \gamma$ **then**
      Fit base model f_{m} in D_{m}, get estimator $\hat{f}_m$
$\hat{y}_e(s) = \sum_{m=1}^{M} a_m(s)\hat{f}_m(s)$ (Eq. 1)

Algorithm 1. STEM.

that can be modeled by the mixture. In STEM λ is a fixed, universal parameter that does not vary with location. It can be estimated through cross-validation to identify the scale of analysis best supported by data.

AdaSTEM proposes an adaptive scheme based on tree data structures (Samet 2006) where stixel size λ(*s*) varies with location *s* as a function of data density (Fink, Damoulas, and Dave 2013). Letting the stixel size λ vary with data density allows the mixture better to exploit unevenly distributed data in the presence of a multiscale signal. In densely sampled regions λ will be small and the base models can adapt to fine-scale signals producing low bias estimators. In sparsely sampled regions λ will be large and base models are forced to adapt to large-scale signals producing low variance estimators.

The center panel of figure 1 shows two partitions

Input: Spatial dataset with extent D
Output: Ensemble model estimator $\hat{y}_e(s)$
Set λ by cross-validation
for p=*1* **to** P **do**
    Randomize partition corner $\pi_p \sim \mathcal{U}(0, \lambda)$
    Partition D into M_p stixels each with length λ
    for m=*1* to M_p **do**
        if $N_\mathrm{m} \geq \gamma$ **then**
            Fit base model f_m in D_m, get estimator $\hat{f}_m$
$\hat{y}_e(s) = \sum_{m=1}^{M} a_m(s)\hat{f}_m(s)$ (Eq. 1)

Algorithm 2. AdaSTEM.

of the study extent generated using quadtrees where each stixel corresponds to a leaf node in a tree. The right panel of figure 1 shows the quadtree stixel size averaged across 100 partitions. Note how the distribution of data-driven stixel sizes $\lambda(s)$ follows the pattern of data density shown in the left panel of figure 1.

Variance between base models is controlled by ensemble averaging (Breiman 1996) and lower covariance between the base models in a neighborhood $\{m|s \in D_\mathrm{m}\}$ is encouraged by bootstrapping the data and randomizing the angle $\theta \sim \mathcal{U}(0, 360]$ and center $c \sim \mathcal{U}(D)$ of each tree partition P. The algorithm is given in algorithm 2.

To investigate the value of letting the stixel size $\lambda(s)$ vary with data density we conducted a synthetic experiment comparing the performance of STEM and AdaSTEM. Both models were repeatedly trained with realizations of noisy data and then evaluated against the true regression function. This was done for two separate two-dimensional regression functions, one single-scale and one multiscale function, each fit with observations sampled from two different densities, one where observations were uniformly distributed and another where observations came from a nonuniform multiscale density. STEM and AdaSTEM were specificed as spatial mixtures of linear regression base models for these experiments.

Figure 2 shows results from a typical realization of the four synthetic experiment scenarios. These results show how AdaSTEM can adapt to multiscale structure when it is present and data density is sufficient, while retaining predictive performance when there is only coarse-scale signal or data are sparse. For the single-scale function and when data are too sparse to detect multiscale signal, performance between the models is comparable.

When multiscale signal is present and there is sufficient data density, the AdaSTEM surface estimate correctly captures both fine- and coarse-scale patterns while the STEM surface estimate does not. We refer the reader to Fink, Damoulas, and Dave (2013) for a more detailed description of the experimental setup and analysis.

eBird

The ecological goal for developing AdaSTEM was to estimate the daily distribution of terrestrial, diurnal bird species across the Western Hemisphere, excluding Greenland, throughout their annual cycle. The bird observation data come from the citizen science project, eBird (Sullivan et al. 2009). eBird is a broad-scale bird monitoring project that collects observations made throughout the year. Participants follow a protocol in which they collect observations of the bird species that they see — checklists — along with ancillary information about the time, location, and search effort. By asking participants to indicate when they have contributed complete checklists of all the species they detect on a search, we assume that lack of detection conveys partial information about absence. Together, the reports of absence and effort add information that we use to capture and control for sources of variation associated with the detection process.

The results from analyses shown in this article used presence-absence data from complete checklists collected with effort data from January 1, 1900, to December 31, 2011, within the Western Hemisphere. The data set, figure 1 (left panel), consisted of

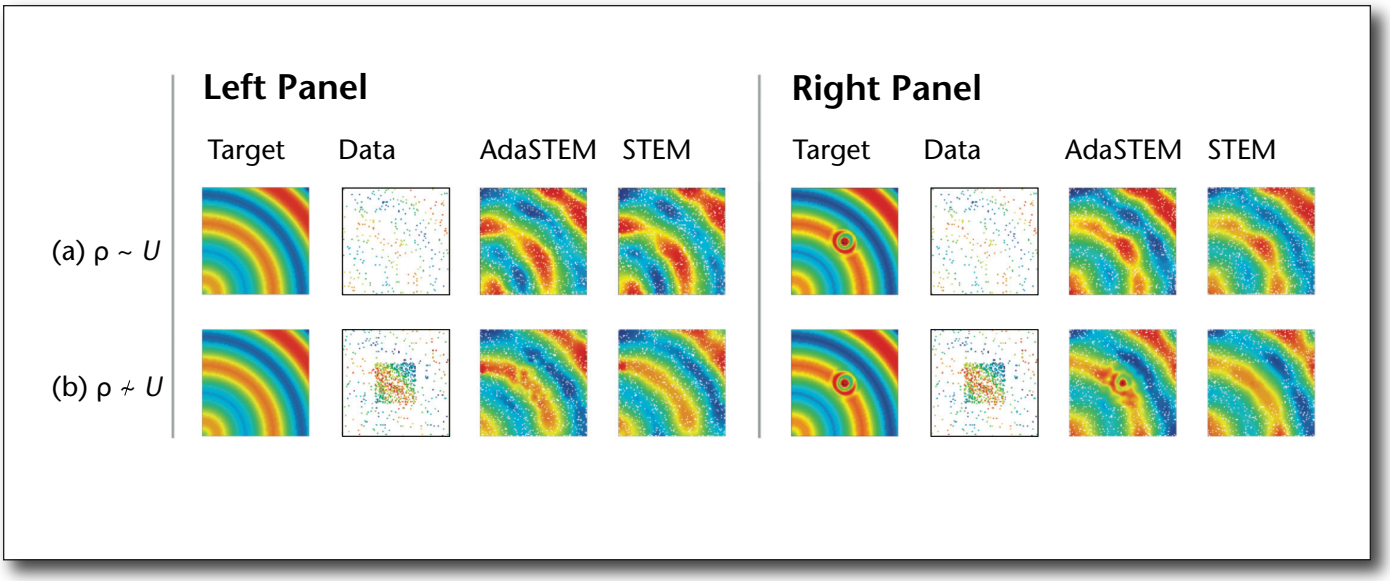

Figure 2. AdaSTEM Versus STEM Synthetic Data Experiment for Uniform or Nonuniform Density of Observations and Single- or Multiscale Signal.

Left Panel: For the single-scale function both models perform comparably for both uniform (row a) and nonuniform (row b) data density. *Right Panel:* In the presence of multiscale signal AdaSTEM clearly outperforms STEM when the density of observations is sufficient (row b) to capture the small scale correlation. (Color version of figure presented in electronic version of *AI Magazine*).

approximately 2.5 million checklists collected across 385 thousand unique locations. All models were trained with 2.25 million checklists made across 360 thousand unique locations, with the remaining held out for model evaluation. At small spatial scales, the data density can be seen to correlate with human population and travel patterns. At larger spatial scales the observations are seen to be most densely distributed in the United States where eBird originated, and sparser in Central and South America even in regions of high human density (see the left panel, figure 1).

The spatiotemporal distributions presented here were modeled as a spatial mixture of local temporal trajectories. We estimated the trajectory within each stixel using a binary response GAM as the class of base model. The binary response indicates the presence or absence of a single bird species recorded on a given search. The logit of the probability of occurrence was modeled as an additive function of the day of the year and several other factors describing the effort spent searching for birds. Seasonal variation is captured by a smooth function of the day-of-year covariate and fit with a penalized cyclic spline basis. To account for variation in detection rates we included effort covariates for the amount of time spent on a search, the distance traveled while searching, and the number of observers in the search party. The time of the day was used to account for diurnal variation in behavior, suchas higher detectability of birds during their participation in the dawn chorus (Diefenbach et al. 2007), which make species more or less conspicuous. The ensemble was created by partitioning the study extent into square stixels measured in units of degrees latitude and longitude with $P = 200$. The minimum sample size per base model was $\gamma = 500$.

The predictive performance of STEM and AdaSTEM were compared using distribution estimates for Barn Swallow (*Hirundo rustica*) in (Fink, Damoulas, and Dave 2013). In these tests, AdaSTEM outperformed STEM for all measures of predictive performance for all 12 months of the year. These results demonstrated the ability of AdaSTEM to take advantage of the varying eBird observation density by reducing bias in regions with high data density and controlling variance in regions with low data density.

Autumn Migration Estimates

To demonstrate how AdaSTEM can adapt to different distributional dynamics across a range of extents and scales we estimated the distributions for Barn Swallow (*Hirundo rustica*), Blackpoll Warbler (*Setophaga striata*), and Black-throated Blue Warbler (*Setophaga caerulescens*). These three species are all broadly distributed migratory birds with very different autumn migration strategies — different distribution locations, distribution extents, and timing of movement.

To develop rangewide estimates of species' distributions we selected the smallest stixel size necessary to achieve at least half the maximum ensemble support, *P,* across 90 percent of the Western Hemisphere. Then we used this model to estimate one daily dis-

tribution surface per week for 52 weeks of the year. The surface is the probability of occurrence on the given day estimated at 130 thousand locations from a geographically stratified random design. All effort predictors were held constant to remove variation in detectability. The precise quantity estimated is the relative probability that a typical eBird participant will detect the species on a search at a given location from 7 to 8 AM while traveling 1 kilometer on the given day of the year.

Figure 3 shows the distribution estimates for Barn Swallow (top), Blackpoll Warbler (middle), and Black-throated Blue Warbler (bottom) on June 28 (left), October 11 (center), and December 20 (right). For the three species these dates fall in the breeding season, during autumn migration, and in the nonbreeding season, respectively. To control for seasonal variation in detectability when comparing distributions across dates and species, we standardized the predicted probability of occurrence across distributions.

Across their annual cycle, Barn Swallow occur throughout most of the longitudinal extent of the terrestrial Western Hemisphere with broad and complex movements during autumn migration (figure 3, top and center), which occur over several months. This contrasts with the Blackpoll Warbler, a neotropical migrant that breeds in large numbers in the boreal forests of North America and undertakes one of the longest migrations of any North American warbler (DeLuca et al. 2013). During autumn migration, it travels first eastward to the North American coast from which the majority of individuals make a transatlantic flight through the eastern Caribbean to Northern South America. Figure 3 (center) shows how the Blackpoll population occurrence is concentrated along the northeastern coast of the United States in October, with much lower rates of occurrence along the southeast coast and a second high concentration in the Western Caribbean islands. These distributional patterns are in agreement with studies of Blackpoll Warbler migration that have relied on a variety of different data sources, based primarily on observations of individual birds (DeLuca et al. 2013). Finally, the Black-throated Blue Warbler, one of the most extensively studied passerine species in North America, migrates south from the eastern deciduous forest of North America along a broad front from the eastern seaboard to the Appalachians where it reaches its wintering grounds in the Caribbean (Holmes, Rodenhouse, and Sillett 2013).

Assessing the Scale and Quality of AdaSTEM

High-quality species distribution and movement information is useful for a variety of ecological and conservation applications across a range of scales. However, distribution estimates by themselves are not sufficient for most applications because they do not convey information about the scale or quality of the estimates. More often than not, distribution estimates from spatially explicit models are computed at arbitrarily fine resolutions. Visualizations and maps generated from these products risk communicating the existence of fine-scale patterns where none may be supported by the data. Without understanding the spatial resolution of an estimate it is easy to overinterpret the results and make inference about fine-scale patterns where this is not warranted.

In this section we present a set of model diagnostics to assess and visualize spatial patterns of scale, bias, and uncertainty. First, we determine the spatial scale of the distribution estimates so that spurious inferences about fine-scale patterns can be avoided. Understanding the spatial scale of estimates is also necessary for constructing statistical comparisons between regions. Second, we present an analysis of regional bias.

Understanding spatial patterns of bias is especially useful when using crowdsourced data. Finally, we provide a quantitative assessment of the uncertainty attached to distribution estimates. This information is essential for making decisions in the face of uncertainty.

Together these three diagnostics provide useful information to interpret and apply summaries of distribution estimates to real-world sustainability problems. All of the diagnostics discussed here are for the June 28 Barn Swallow AdaSTEM distribution estimate (figure 3, top, left).

Assessing Spatial Scale

We formalize the notion of scale as the effective range, the shortest distance at which the correlation between pairs of measurements within a neighborhood becomes negligibly small (Banerjee, Carlin, and Gelfand 2004). To assess scale of a distribution estimate we measure the effective range of the residuals. The spatial variation in scale is visualized by computing effective ranges across a half degree grid and interpolating.

Figure 4 (left) shows the interpolated effective range for the distribution of Barn Swallow based on residuals from June 24–July 1. The portion of the study area with insufficient residual density to estimate the effective range is shown in grey and the effective range of this Barn Swallow distribution estimate can be seen to vary from less than 10 kilometers to over 100 kilometers depending on location. For example, in regions with very short ranges, like Ithaca, New York, in the northeastern United States, estimates of occurrence separated by as little as 5 kilometers are independent. At the same time, in the state of Montana, located in the northcentral United States, occurrence estimates must be separated by at least 60 kilometers to be independent of each other.

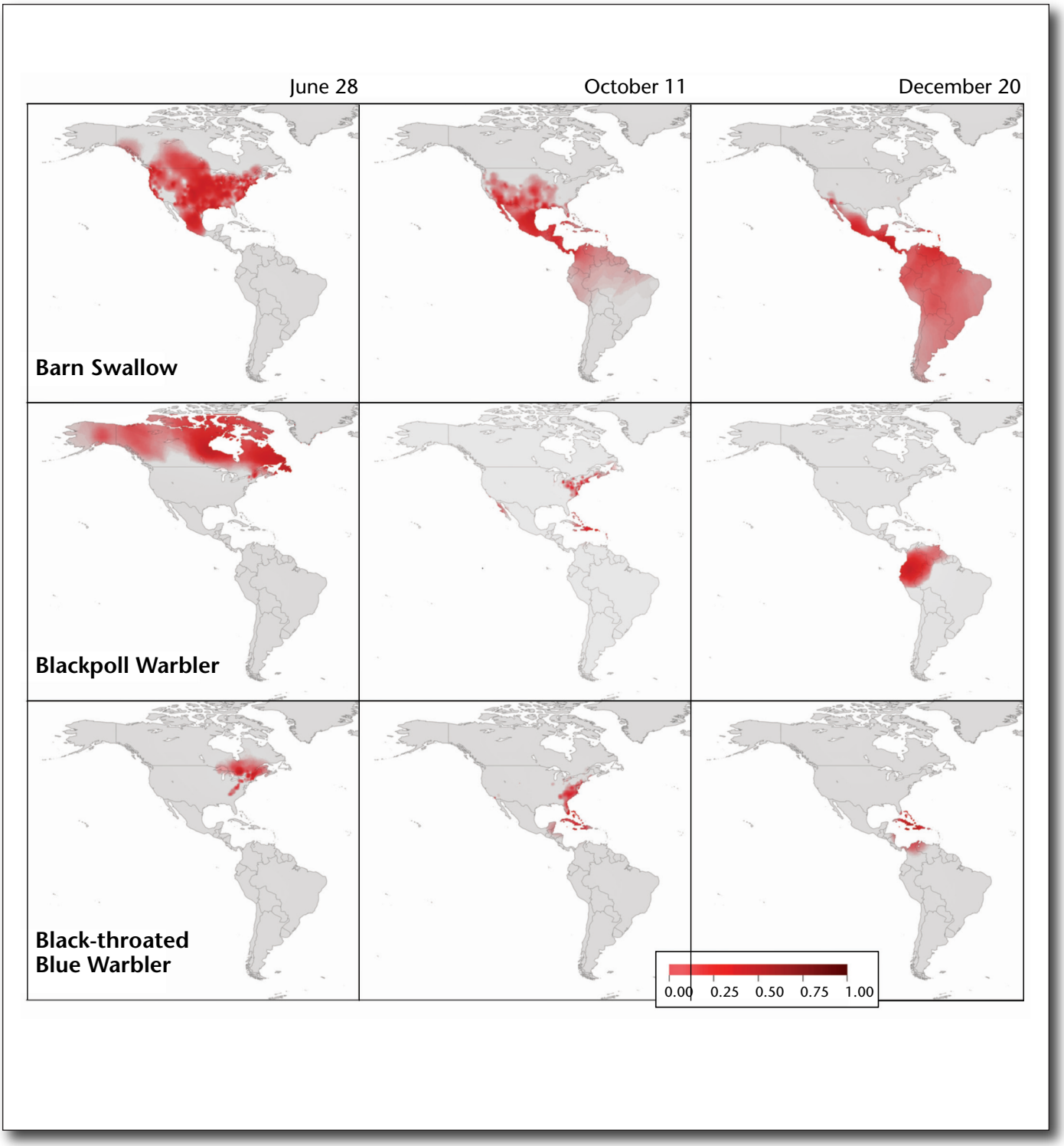

Figure 3. AdaSTEM Distribution Estimates for Barn Swallow, Blackpoll Warbler, and Black-throated Blue Warbler.

Breeding Season (June 28, left), autumn migration (October 11, center), and the nonbreeding season (December 20, right). Darker shades of red indicate higher relative probability of occurrence for each distribution. These three species span the rich variation in avian distributional dynamics that characterize bird species' annual cycles. The quality of these distribution estimates highlights how AdaSTEM adapts to a variety of complex spatiotemporal processes across a range of scales. (Color version of figure presented in electronic version of *AI Magazine*).

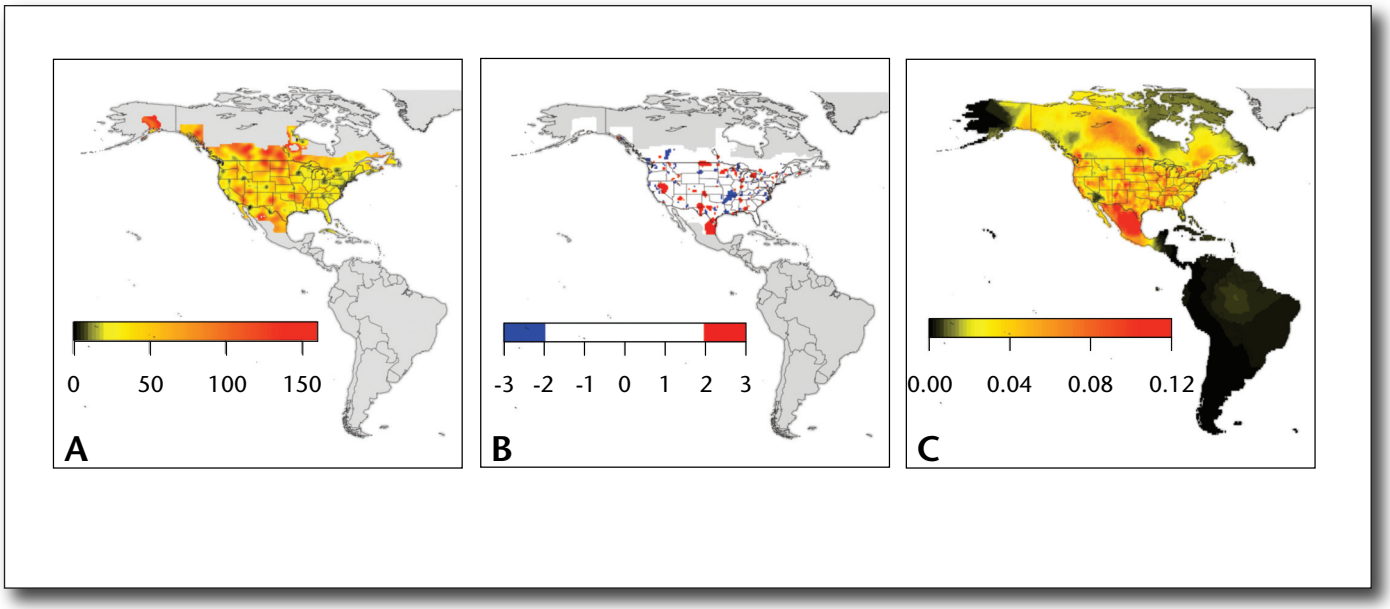

Figure 4. Assessing the Scale and Quality of the June 28 Barn Swallow AdaSTEM Distribution.

Left: Effective range of spatial correlation in kilometers. Center: Interpolated bias estimates shown in units of standard errors. *Right:* Pointwise standard errors computed across 50 individual bootstrap replicates. These diagnostics capture the interplay of the ecological process (that is, species occurrence), the data density, and the scale structure of AdaSTEM to affect the scale and quality of the estimated distribution. (Color version of figure presented in electronic version of *AI Magazine*).

In general, the effective range varies as a joint function of data density and the prevalence of the species. For example, in regions where the species is correctly predicted to be present with very low probability, residuals are uniformly small and the range of residual correlation tends to be larger.

Assessing Bias

For observational data, especially crowdsourced data, bias assessment is important because biases incurred during the data collection process may produce regions where the estimatedprobabilities of occurrence are systematically high or low compared to the observed rates of occurrence. It is important to know where biased regions occur, how big the biased regions are, and the strength of the bias.

To identify biased regions we interpolated the residuals from June 24–July 1 across the same extent as that used to visualize the effective range. Then we looked for areas where the interpolated residuals were substantially larger or smaller than expected by chance alone. This was done by standardizing the residuals and plotting only those regions where the standardized residuals were more than twice as large as their associated standard errors. Figure 4 (center) shows regions with systematically large residuals. Regions where the estimated occurrence rates are too low are shown in red and regions where the estimated occurrence rates are too high are shown in blue. Most of the contiguous regions of bias shown in figure 4b are relatively small, with larger regions in Montana, Nevada, Texas, and Arkansas.

Assessing Uncertainty

Uncertainty estimates are required when making statistical inference about distributional summaries. For spatial prioritization we may want to evaluate whether the difference in expected occurrence rates between regions is larger than that expected by chance. To do this we need uncertainty estimates for the AdaSTEM occurrence rates. These uncertainties can be approximated based on the variation across bootstrap replicates. However, because the AdaSTEM estimator $y_e(s)$ is computed as an average across bootstrap replicates, the standard errors $\sigma y_e(s)$ will be smaller than the variance across the bootstrap replicates. If we assume that the bootstrap replicates are independent, then $\sigma y_e(s)$ will be smaller by a factor of $n(s)^{-1/2}$. For example, if $n(s) = 50$ the standard error of the ensemble estimate will be approximately 15 percent of the standard error of estimates across individual bootstrap replicates.

Figure 4 (right) shows the pointwise standard errors computed across 50 bootstrap replicates. These standard errors are conservative, that is, larger than the actual standard errors for $y_e(s)$. Like the spatial scale and bias diagnostics, the uncertainty estimates vary jointly with data density and species prevalence. For example, many data dense regions have relatively high uncertainties. One reason for this follows

from the binary classification problem itself — variability is greatest where predicted probabilities are near 0.5 and smallest when the predicted probabilities are closest to 0 or 1. Another reason for relatively large standard errors in data dense regions is the fact that AdaSTEM's adaptive partitioning tends to minimize bias in data dense regions, potentially leading to higher variance.

Discussion

Using AdaSTEM, we have produced the first hemispherewide population-level distribution estimates of long-distance migrations using crowdsourced data from eBird. These estimates demonstrate how AdaSTEM can automatically adapt to patterns across several orders of magnitude. While the hemispheric extent of analysis extends over 10,000 kilometers north to south, we found that the spatial resolution of the distribution estimates was less than 100 kilometers within most of the continental United States and Southern Canada. In several data rich regions of North America, the spatial resolution was found to be less than 10 kilometers.

The simple adaptive divide and recombine strategy employed by AdaSTEM provides sufficient flexibility to model complex spatiotemporal processes across a range of scales. AdaSTEM as a class of models will be useful in other spatial and spatiotemporal domains where data are irregularly and sparsely distributed, such as applications based on geographic surveys and geolocated data collected by volunteers through crowdsourcing platforms.

The three species whose data were analyzed in this article span the rich variation in avian distributional dynamics that characterize bird species' annual cycles. For long-distance migrants, these dynamics extend well beyond the conterminous USA, where research and conservation efforts are often focused, including our own efforts based on eBird data (La Sorte et al. [2013], for example). By modeling occurrences across the entire Western Hemisphere, AdaSTEM provides novel information on how these dynamics are structured for entire populations of multiple species across their entire annual cycles, even for species that are panhemispheric migrants. This information has tremendous potential to generate novel inferences in avian ecology and evolution, and to benefit national and international efforts in avian conservation. For example, we can now gain a more detailed understanding of theprocess of migration — how fast birds travel, which routes they take, whether the same routes are followed northward and southward, and whether there are discrete collections of species that travel along the same flyways — that have previously only been studied over smaller region (La Sorte et al. 2013) or for very small numbers of individual birds (Stutchbury et al. 2009). By expanding and improving our existing knowledge base, conservation efforts can be more effective and efficient, with implications not only for protecting current avian populations but for providing the basis for their long-term sustainability under global environmental change.

Acknowledgments

We thank the thousands of eBird participants, K. Rosenberg, S. Sukhnanand, and the Lab of Ornithology IS eBird team — K. Webb, M. Iliff, J. Gerbracht, T. Lenz, W. Morris, B. Sullivan, and C. Wood. This work was supported by the Leon Levy Foundation, the Wolf Creek Foundation, and the National Science Foundation (OCI-0830944, CCF-0832782, ITR-0427914, DBI-1049363, DBI-0542868, DUE-0734857, IIS-0748626, IIS-0844546, IIS-0612031, IIS-1050422, IIS-0905385, IIS-0746500, IIS-1209589, AGS-0835821, CNS-0751152, CNS-0855167, IIS-1017793, CDI-1125098) with computing support from CNS-1059284, OCI-1053575 and DEB-110008.

References

Banerjee, S.; Carlin, B.; and Gelfand, A. 2004. Hierarchical Modeling and Analysis for Spatial Data, vol. 101. London: Chapman & Hall/CRC.

Bonter, D. N.; Donovan, T. M.; Brooks, E. W.; and Hobson, K. A. 2009. Daily Mass Changes in Landbirds During Migration Stopover on the South Shore of Lake Ontario. *The Auk* 124(1): 122–133. dx.doi.org/10.1642/0004-8038(2007)124[122:DMCILD]2.0.CO;2

Breiman, L. 1996. Bagging Predictors. *Machine Learning* 24(2): 123–140. dx.doi.org/10.1007/BF00058655

Breiman, L.; Friedman, J.; Stone, C.; and Olshen, R. 1984. Classification and Regression Trees. London: Chapman & Hall/CRC.

Crainiceanu, C.; Ruppert, D.; Carroll, R.; Joshi, A.; and Goodner, B. 2007. Spatially Adaptive Bayesian Penalized Splines with Heteroscedastic Errors. *Journal of Computational and Graphical Statistics* 16(2): 265–288. dx.doi.org/10.1198/106186007X208768

Cressie, N. 1993. *Statistics for Spatial Data.* New York: John Wiley, 2nd ed.

Cressie, N. 1986. Kriging Nonstationary Data. *Journal of the American Statistical Association* 81(395): 625–634. dx.doi.org/10.1080/01621459.1986.10478315

Cressie, N., and Johannesson, G. 2008. Fixed Rank Kriging for Very Large Spatial Data Sets. *Journal of the Royal Statistical. Society B* 70(1): 209–226. dx.doi.org/10.1111/j.1467-9868.2007.00633.x

Cressie, N., and Wikle, C. K. 2012. *Statistics for Spatio-Temporal Data.* New York: John Wiley.

DeLuca, W.; Holberton, R.; Hunt, P. D.; and Eliason, B. C. 2013. Blackpoll Warbler (*Setophaga striata*). In The Birds of North America Online, ed. A. Poole. Ithaca, NY: Cornell Laboratory of Ornithology.

Dickinson, J. L.; Zuckerberg, B.; and Bonter, D. N. 2010. Citizen Science as an Ecological Research Tool: Challenges and Benefits. *Annual Review of Ecology, Evolution, and Systematics*

41(1): 149–172. dx.doi.org/10.1146/annurev-ecolsys-102209-144636

Diefenbach, D.; Marshall, M.; Mattice, J.; Brauning, D.; and Johnson, D. 2007. Incorporating Availability for Detection in Estimates of Bird Abundance. *The Auk* 124(1): 96–106. dx.doi.org/10.1642/0004-8038(2007)124[96:IAFDIE]2.0.CO;2

Drewitt, A. L., and Langston, R. H. 2006. Assessing the Impacts of Wind Farms on Birds. *Ibis* 148(s1): 29–42. dx.doi.org/10.1111/j.1474-919X.2006.00516.x

Dungan, J. L.; Perry, J.; Dale, M.; Legendre, P.; Citron-Pousty, S.; Fortin, M.-J.; Jakomulska, A.; Miriti, M.; and Rosenberg, M. 2002. A Balanced View of Scale in Spatial Statistical Analysis. *Ecography* 25(5): 626–640. dx.doi.org/10.1034/j.1600-0587.2002.250510.x

Fink, D.; Damoulas, T.; and Dave, J. 2013. Adaptive Spatio-Temporal Exploratory Models: Hemisphere-Wide Species Distributions From Massively Crowdsourced eBird Data. In Proceedings of the Twenty-Seventh AAAI Conference on Artificial Intelligence. Palo Alto, CA: AAAI Press.

Fink, D.; Hochachka, W.; Zuckerberg, B.; Winkler, D.; Shaby, B.; Munson, M.; Hooker, G.; Riedewald, M.; Sheldon, D.; and Kelling, S. 2010. Spatiotemporal Exploratory Models for Broad-Scale Survey Data. *Ecological Applications* 20(8): 2131–2147. dx.doi.org/10.1890/09-1340.1

Fortin, M., and Dale, M. 2005. Spatial Analysis: A Guide for Ecologists. Cambridge, UK: Cambridge University Press.

Gelfand, A. 2012. Hierarchical Modeling for Spatial Data Problems. *Spatial Statistics* 1: 30–39. dx.doi.org/10.1016/j.spasta.2012.02.005

Gomes, C. P. 2009. Computational Sustainability: Computational Methods for a Sustainable Environment, Economy, and Society. *The Bridge* 39(4): 5–13.

Grosbois, V.; Gimenez, O.; Gaillard, J.; Pradel, R.; Barbraud, C.; Clobert, J.; Møller, A.; and Weimerskirch, H. 2008. Assessing the Impact of Climate Variation on Survival in Vertebrate Populations. *Biological Reviews* 83(3): 357–399. dx.doi.org/10.1111/j.1469-185X.2008.00047.x

Hastie, T. J.; Tibshirani, R.; and Friedman, J. 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Berlin: Springer. dx.doi.org/10.1007/978-0-387-84858-7

Hochachka, W. M.; Fink, D.; Hutchinson, R. A.; Sheldon, D.; Wong, W.-K.; and Kelling, S. 2012. Data-Intensive Science Applied to Broad-Scale Citizen Science. *Trends in Ecology and Evolution* 27(2): 130–137. dx.doi.org/10.1016/j.tree.2011.11.006

Holmes, R. T.; Rodenhouse, N.; and Sillett, T. 2013. Black-throated Blue Warbler (*Setophaga caerulescens*). In The Birds of North America Online, ed. A. Poole. Ithaca, NY: Cornell Laboratory of Ornithology.

Huang, H.-C.; Cressie, N.; and Gabrosek, J. 2002. Resolution-Consistent Spatial Prediction of Global Processes from Satellite Data. *Journal of Computational and Graphical Statistics* 11(1): 63–88. dx.doi.org/10.1198/106186002317375622

Jun, M., and Stein, M. L. 2008. Nonstationary Covariance Models for Global Data. *Annals of Applied Statistics* 2(4): 1271–1289. dx.doi.org/10.1214/08-AOAS183

Kammann, E., and Wand, M. 2003. Geoadditive Models. *Journal of the Royal Statistical Society: Series C* (Applied Statistics) 52(1): 1–18. dx.doi.org/10.1111/1467-9876.00385

Kuncheva, L., and Whitaker, C. 2003. Measures of Diversity in Classifier Ensembles and Their Relationship with the Ensemble Accuracy. *Machine Learning* 51(2): 181–207. dx.doi.org/10.1023/A:1022859003006

La Sorte, F. A.; Fink, D.; Hochachka, W. M.; DeLong, J. P.; and Kelling, S. 2013. Population-Level Scaling of Avian Migration Speed with Body Size and Migration Distance for Powered Fliers. *Ecology* 94(8): 1839–1847. dx.doi.org/10.1890/12-1768.1

Levin, S. A. 1992. The Problem of Pattern and Scale in Ecology. The Robert H. MacArthur Award Lecture. *Ecology* 73(6): 1943–1967. dx.doi.org/10.2307/1941447

Moilanen, A.; Wilson, K. A.; and Possingham, H. P. 2009. Spatial Conservation Prioritization: Quantitative Methods and Computational Tools. Oxford, UK: Oxford University Press.

Ostfeld, R.; Glass, G.; and Keesing, F. 2005. Nonparametric Spatial Covariance Functions: Estimation and Testing. *Trends in Ecology and Evolution* 20(6): 328–336. dx.doi.org/10.1016/j.tree.2005.03.009

Paciorek, C., and Schervish, M. 2004. Nonstationary Covariance Functions for Gaussian Process Regression. Advances in Neural *Information Processing Systems* 16: 273–280.

Pintore, A., and Holmes, C. C. 2004. Spatially Adaptive Non-Stationary Covariance Functions Via Spatially Adaptive Spectra. Technical Report, Department of Statistics, Oxford University, Oxford, UK. (www.stats.ox.ac.uk/?cholmes/Reports/spectral tempering.pdf).

Pintore, A.; Speckman, P.; and Holmes, C. C. 2006. Spatially Adaptive Smoothing Splines. *Biometrika* 93(1): 113–125. dx.doi.org/10.1093/biomet/93.1.113

Rasmussen, C., and Williams, C. 2006. *Gaussian Processes for Machine Learning*. Cambridge, MA: The MIT Press.

Rue, H., and Held, L. 2005. *Gaussian Markov Random Fields: Theory and Applications, Monographs on Statistics and Applied Probability,* volume 104. London: Chapman & Hall. dx.doi.org/10.1201/9780203492024

Samet, H. 2006. *Foundations of Multidimensional and Metric Data Structures*. San Francisco: Morgan Kaufmann.

Schuster, R., and Arcese, P. 2013. Using Bird Species Community Occurrence to Prioritize Forests for Old Growth Restoration. *Ecography* 36(4): 499–507. dx.doi.org/10.1111/j.1600-0587.2012.07681.x

Stein, M. L. 2005. Space-Time Covariance Functions. *Journal of the American Statistical Association* 100: 310–321. dx.doi.org/10.1198/016214504000000854

Stutchbury, B. J.; Tarof, S. A.; Done, T.; Gow, E.; Kramer, P. M.; Tautin, J.; Fox, J. W.; and Afanasyev, V. 2009. Tracking Long-Distance Songbird Migration by Using Geolocators. *Science* 323(5916): 896–896. dx.doi.org/10.1126/science.1166664

Sullivan, B. L.; Wood, C. L.; Iliff, M. J.; Bonney, R. E.; Fink, D.; and Kelling, S. 2009. eBird: A Citizen-Based Bird Observation Network in the Biological Sciences. *Biological Conservation* 142(10): 2282–2292. dx.doi.org/10.1016/j.biocon.2009.05.006

Tzeng, S.; Huang, H.-C.; and Cressie, N. 2005. A Fast, Optimal Spatial-Prediction Method for Massive Datasets. *Journal of the American Statistical Association* 100(472): 1343–1357. dx.doi.org/10.1198/016214505000000420

Wood, S. 2006. *Generalized Additive Models: An Introduction with R,* volume 66. London: Chapman & Hall/CRC.

Daniel Fink is a research associate in the information science program at the Cornell Lab of Ornithology. His research interests are in statistics and machine learning with applications to large-scale spatiotemporal problems in environmental and ecological sciences.

Theodoros (Theo) Damoulas is a research assistant professor at New York University working at the New York University Center for Urban Science and Progress and the New York University Polytechnic School of Engineering in the area of Urban Informatics. His research interests are in statistical machine learning and artificial intelligence with applications to biology, sustainability and urban informatics.

Nicholas E. Bruns is a developer on the information science team at the Cornell Lab of Ornithology. He implements large-scale analyses of the data from the citizen observation network, eBird.

Frank A. La Sorte is a research associate at the Cornell Lab of Ornithology. His research interests are in the ecology, biogeography, and conservation of migratory birds within the context of global environmental change.

Wesley M. Hochachka is a senior research associate at Cornell University, working at the Cornell Lab of Ornithology where he is the assistant director of the Bird Population Studies program. The two main focuses of his research are the development and use of predictive models of species distributions, and investigations into ecological and evolutionary aspects of host-pathogen dynamics.

Carla P. Gomes is a professor of computer science and the director of the Institute for Computational Sustainability at Cornell University. Her research themes include constraint reasoning, mathematical programming, and machine learning, for large scale combinatorial problems. Recently, Gomes has helped found the new field of computational sustainability, which is her current main research focus.

Steve Kelling, director of information science at the Cornell Lab of Ornithology, leads a team of ornithologists, computer scientists, statisticians, application developers, and data managers to develop programs, tools, and analyses to gather, understand, and disseminate information on birds and the environments they inhabit.