

# BirdVis: Visualizing and Understanding Bird Populations

Nivan Ferreira, Lauro Lins, Daniel Fink, Steve Kelling, Chris Wood, Juliana Freire, and Cláudio Silva

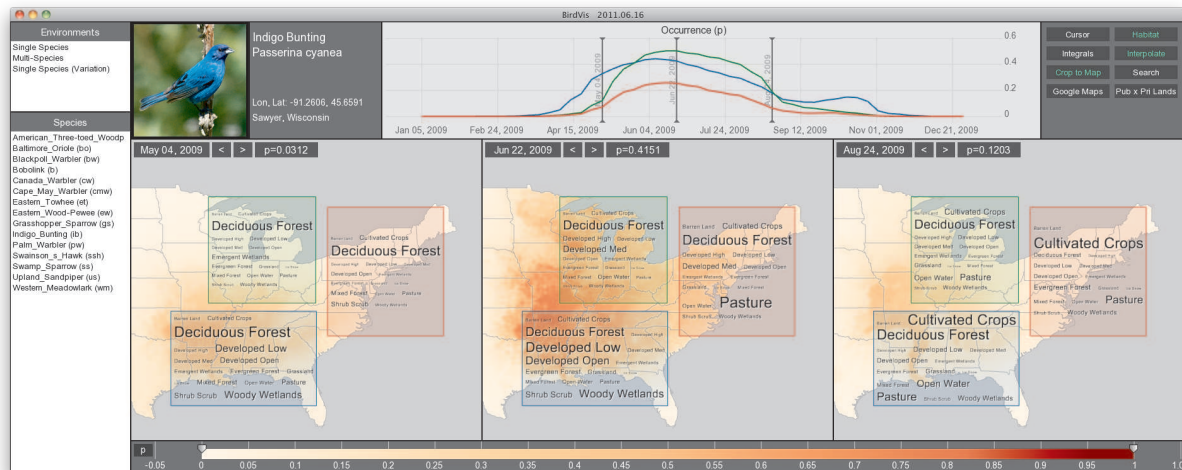


Fig. 1. Tag cloud lenses in BirdVis are used as a tool to explore and understand relative habitat preferences suggested by a species distribution model over space, time, and across species. Here, we show occurrence maps for the Indigo Bunting. The lenses cover three different regions on three different dates of the breeding season of 2009: May 4, June 22, and August 24, 2009. They show important differences in habitat preferences across these regions as well as how the preferences change over time within a region. The ability to interact with these visualizations, by changing regions and dates, and comparing different bird species, provides an unprecedented tool for scientists to understand bird distribution and consequently obtain insights about the environment and how it changes over time. The supplemental video gives an overview of different features and visualizations provided by BirdVis.

**Abstract**—Birds are unrivaled windows into biotic processes at all levels and are proven indicators of ecological well-being. Understanding the determinants of species distributions and their dynamics is an important aspect of ecology and is critical for conservation and management. Through crowdsourcing, since 2002, the eBird project has been collecting bird observation records. These observations, together with local-scale environmental covariates such as climate, habitat, and vegetation phenology have been a valuable resource for a global community of educators, land managers, ornithologists, and conservation biologists. By associating environmental inputs with observed patterns of bird occurrence, predictive models have been developed that provide a statistical framework to harness available data for predicting species distributions and making inferences about species-habitat associations. Understanding these models, however, is challenging because they require scientists to quantify and compare multiscale spatial-temporal patterns. A large series of coordinated or sequential plots must be generated, individually programmed, and manually composed for analysis. This hampers the exploration and is a barrier to making the cross-species comparisons that are essential for coordinating conservation and extracting important ecological information. To address these limitations, as part of a collaboration among computer scientists, statisticians, biologists and ornithologists, we have developed BirdVis, an interactive visualization system that supports the analysis of spatio-temporal bird distribution models. BirdVis leverages visualization techniques and uses them in a novel way to better assist users in the exploration of interdependencies among model parameters. Furthermore, the system allows for comparative visualization through coordinated views, providing an intuitive interface to identify relevant correlations and patterns. We justify our design decisions and present case studies that show how BirdVis has helped scientists obtain new evidence for existing hypotheses, as well as formulate new hypotheses in their domain.

**Index Terms**—Ornithology; species distribution models; multiscale analysis; spatial data; temporal data.

## 1 INTRODUCTION

In the face of global warming and recent environmental disasters, nature conservation has received substantial attention in recent years. The study of birds is an important tool for this purpose, since birds engage in the most spectacular long-distance migrations of any animal

on the planet and demonstrate the biological integration of seemingly disparate ecosystems around the globe. They are unrivaled windows into biotic processes at all levels and are proven indicators of ecological well-being. This is one of the reasons that avian data is one of the richest and most important sources of distributional data for ecology and conservation. However, the vast majority of avian studies have been restricted to a single season, most often the breeding season, and to specific regional spatial scales. As a result, these studies do not fully reflect the limiting factors and important processes that govern these populations, a critical deficiency for conservation planning and ecological understanding.

In order to overcome these limitations and improve the understanding of the broad-scale dynamics of continent-scale bird migrations, scientists from the Cornell Lab of Ornithology (CLO) have devel-

- Nivan Ferreira, Lauro Lins, Juliana Freire and Cláudio Silva are with the SCI Institute, University of Utah
- Daniel Fink, Steve Kelling and Chris Wood are with the Cornell Lab of Ornithology

Manuscript received 31 March 2011; accepted 1 August 2011; posted online 23 October 2011; mailed on 14 October 2011.

For information on obtaining reprints of this article, please send email to: [tvcg@computer.org](mailto:tvcg@computer.org).

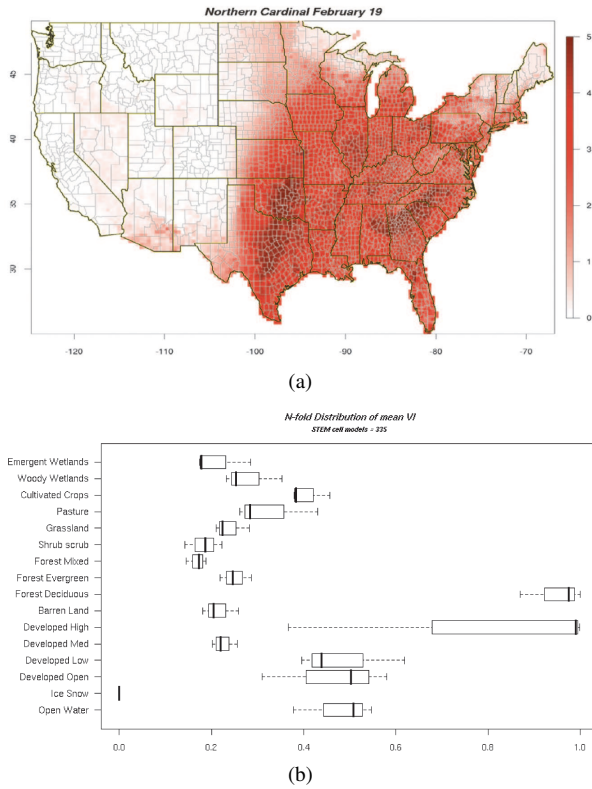


Fig. 2. Examples of visualizations currently used for analyzing bird distribution data. (a) Bagged decision tree predictive model of Northern Cardinal relative abundance in eastern North America on 19 February 2006 [42]; (b) Plot showing the predictors' from statistical modeling analysis of Indigo Bunting in July 2008.

oped a statistical model called Spatio-Temporal Exploratory Model (STEM) [12]. STEM uses bird observation data from the eBird project [10] to predict bird distribution in large spatial and temporal scales. STEM analyses provide the first empirical view of population-level movement on a weekly basis across a range of bird species. However, it is very difficult to directly extract the most important ecological information from STEM model output, since like many machine-learning models, the structure of STEM is itself complex necessitating extensive post-modeling analyses to gain insights. These analyses require several trial-and-error steps, where scientists explore variations in bird populations and the importance of model parameters (*e.g.*, habitat) across multiple space-time windows.

Although scientists in this domain recognize the importance and make use of visualization techniques [21, 42], the detailed information STEM provides about species occurrence variation over space and time poses new challenges. Currently, the use of visualization in this domain is limited to static plots, map views, and jpeg animations created using R [34]. Some examples are shown in Figure 2. In order to answer even simple questions about a single species, several plots and maps are created by the modeling expert, which are then manually composed (*e.g.*, using PowerPoint) for the ornithologists to review. The large number of visualization leads to a cognitive overload, making it hard to identify patterns and correlations among them. This problem is compounded for analyses that involve cross-species comparisons that are essential for coordinating conservation and extracting important ecological information. Furthermore, when the ornithologists formulate new hypotheses, the offline process needs to be repeated: new images are generated and composed so that the hypotheses can be tested. The lack of interactivity makes this process both time-consuming and cumbersome.

To address this problem, we have designed BirdVis, a new interactive visualization tool that supports the analysis of spatio-temporal bird distribution models. BirdVis was developed as part of a collaboration

among computer scientists, statisticians, ornithologists, and biologists. The system is the first to provide an array of visualizations that can be combined to give insights into the different aspects of large-scale dynamic bird distribution models. An intrinsic aspect of displaying dynamic distributions is the separation of signal and noise. There are two basic problems: 1) Model predictions are imperfect, especially when predicting small occurrence rates. So, in order to focus on ecological signal it is necessary to obscure noisy artifacts. However, doing this is complicated by the second problem; 2) Occurrence rates differ greatly from region to region, through time, and among species. This makes it difficult to visualize the dynamics and make comparisons. Ultimately, this means that to perform comparisons, it is necessary to manually tune two or more images. To support the exploration of these dynamic distributions, BirdVis provides a series of visual representations (and compositions thereof) that aim to simplify the identification of correlations and changes in the presence of noise. These help ecologists and biologists learn from the signals and extract the most relevant pieces of information from the model analysis. They are also instrumental for the statisticians to study the noise, enabling them to better understand and improve the model.

Another important aspect of BirdVis is that it enables the combination of different pieces of information through visualizations that present them in a succinct way. One example is illustrated in Figure 1. In this figure, we use color-mapped maps to show bird occurrence probability. Furthermore, by utilizing a new visualization abstraction, which we call *Tag Cloud Lenses* (TCLs), in the same image, we are able to show the importance of different habitats in different seasons and regions. BirdVis also includes a series of operations that support specific requirements for bird distribution analyses, including the ability drill down into specific regions while exploring correlations among different model parameters. Since searching for patterns requires a trial-and-error process, BirdVis provides different mechanisms for comparing multiple visualizations side-by-side and to interactively manipulate these visualizations.

In this paper, we describe the design and implementation of BirdVis. We discuss the specific visualization requirements for the exploration of dynamic bird distribution models, and how these have shaped up the design of BirdVis. We also describe a series of visual representations that leverage well-known visualization techniques and provide an intuitive and interactive way for scientists to explore bird occurrence predictions. These include the ability to interactively change colormaps and to visualize occurrence probability variations, in addition to the raw probabilities. We also propose the use of tag clouds to represent habitat associations, and introduce a new tool that combines tag clouds and magic lenses, giving scientists the ability to interactively inspect how habitat preferences vary over space and time. BirdVis was used by domain experts to carry out a series of analysis. We report the results of their initial evaluation, the new insights they have obtained as well as limitations they have identified. BirdVis has already been deployed at CLO and is being used by the members of that lab. Not only has BirdVis made it easier (and reduced the amount of time required) to perform common analyses, but it has also enabled new ways to look at the data and compare different scenarios. We should note that although BirdVis was developed for and has been used to explore the results of STEM, its visualizations and features can also be utilized for other bird distribution models and for similar data sets, that require multivariate and multiscale spatio-temporal analyses such as for example, national patterns of air pollution.

## 2 RELATED WORK

STEM data is the first empirical spatio-temporal view of population-level movement on large spatial and temporal scales. BirdVis represents our first attempt at developing an interactive visualization tool for exploring the STEM predictions. Visualizations currently used to analyze bird distribution patterns convey high-level, often static information (see Figure 2)—they do not support interactive exploration. For example, to prevent collision between birds and aircrafts, Shamoun-Baranes *et al.* [40] have developed a Web-based tool to visualize bird avoidance models. Their system uses static images to present model results and it does not support interactive exploration of the data.

There has been substantial work on techniques for spatio-temporal visualization [2, 20]. Although some of the existing systems provide a general framework, because they have evolved from domain-specific applications, they often focus on visualization techniques used to address the requirements of a specific domain. For example, the systems proposed by Guo *et al.* [14] and Chen *et al.* [9] were developed to analyze the market of American technology companies. The main focus was to understand how the market evolved by summarizing the relevant information by state and year. Their methods, however, do not scale for country-level aggregation, and hence, are not applicable for data with high resolution in space and time, such as the STEM results.

A number of systems have been proposed to visualize weather-related data [18, 29, 33], which similar to STEM's, consist of spatio-temporal data that is the average of results of many simulations of a model. Some features from these systems are present in BirdVis, including map views, colormaps, brushing and coordinated views [37]. However, we had to develop new visual representations for the exploration of dynamic distributions. As we discuss in Section 4.2, while analyzing the importance of predictors, the exact values of these variables is not important: scientists need to understand the *relative differences* among them. Consequently, colormaps, contours and glyphs are not sufficient. In BirdVis, we employ tag clouds as a more effective means to display the relative importance of predictors.

We also introduce tag cloud lenses (Section 4.2), a technique that combines ideas from magic lenses [4] and tag maps [19]. Tag cloud lenses provide a novel way to explore multivariate spatio-temporal data, suitable for understanding relative differences among those variables. Magic Lenses have inspired the development of analogous interfaces and interaction methods in many domains: spatio-temporal querying for video data [39], augmented virtual environments [7], exploration of geo-referenced Wikipedia content using mobile devices [17], and Label Lenses [11], to name a few. Since its introduction by Milgram and Jodelet [24], tag clouds have been extensively explored to convey geography-related data. Tag Maps [1, 19] have been used to produce summaries of a collection of geo-referenced photographs. Wood *et al.* [46] and Slingsby *et al.* [41] used a mash-up-based approach to build visualizations using tag maps over the Web. These techniques, however, are used for discrete data and display tags in the exact location of the data entry. As a result, they do not apply in our scenario, where data is defined continuously over space. We use a tag cloud to summarize the information inside a given region—the location of the tags inside a region does not have any spatial meaning. To the best of our knowledge, this is the first work that combines Magic Lenses and tag clouds (tag maps).

### 3 BACKGROUND

**The eBird Project and Bird Observation Data.** eBird [10, 26, 43] is a citizen science project, launched in 2002 by the Cornell Lab of Ornithology at Cornell University and the National Audubon Society, whose goal is to gather and share real-time standardized bird observation data over the Web. The eBird data set contains observations of 9091 species, over 90% of the world's bird species, and submissions from 226 countries. It is unique among broad-scale bird monitoring projects in that it collects observations made throughout the year. The eBird participants follow a protocol where time, location, and counts of birds are all reported in a standardized manner. Furthermore, tied to each observation are covariates that quantify search effort and describe the surrounding environment.

**STEM.** The STEM model [12] was developed to overcome important limitations of previous bird population studies. STEM consists on a semi-parametric machine learning model designed to provide a statistical framework that harnesses available data for predicting species distributions and makes inferences about species-habitat associations. In essence, given the static and noisy eBird data, which is sparsely distributed in space and time, STEM produces bird occurrence estimations in large spatial and temporal scales (continent-wide and inter-annual) using a multiscale strategy to differentiate between local and global-scale spatio-temporal structure. This is achieved by creating a randomized ensemble of overlapping local models, each ap-

plied across a restricted geographic and temporal extent, called spatio-temporal pixel, or stixel for short. A user-specified predictive model accounts for local variation as a function of local predictor values. Decision Trees (DTs) are used as the predictive model [5]. Predictions are made for explicit location-time pairs by taking the mean across all of the overlapping local models that include that location-time. Thus, local patterns are allowed to “scale up” via ensemble averaging to larger scales. This combines the bias-reducing properties of local models with the variance-reducing properties of randomized ensembles.

**The Data Set: STEM Predictions.** For the generation of the data used in this paper, STEM was applied using eBird reports as a source of bird presence as well as of bird absence information. eBird participants indicate when and where they have recorded a species—species that are not present in a set of observations (for a given time and location) are considered to be absent. We considered 622,124 observations reported from 107,295 unique locations, and 22 predictors. Four of the predictors were effort variables included in the analysis to capture the variation in detection rates (the hours spent searching, the kilometer length of the transect, observation time of day, and the number of people in the search party).

To account for habitat selectivity each observation was linked to remote-sensing data from the U.S. 2001 National Land Cover Database [13] (NLCD), where land-cover can be one of 16 types (*i.e.*, Barren Land, Cultivated Crops, Developed High, Developed Low, Developed Med, Emergent Wetlands, Deciduous Forest, Evergreen Forest, Mixed Forest, Grassland, Ice Snow, Pasture, Shrub scrub, Woody Wetlands, Developed Open, Open Water) at a 30-meter cell resolution. From this information, the percentage of coverage and habitat composition was computed for each of the land-cover classes in a 1.5km pixel (square 2.25ha) centered on the location of each observation. The NLCD classes were also used as predictors. Finally, the last predictor was the elevation at the observation site<sup>1</sup>. For each species studied, presence-absence data was analyzed from observations collected during the six-year period 2004-2009 within the conterminous U.S.

This analysis produced one daily *occurrence map* per week, generated for all 52 weeks in 2009. An occurrence map consists of bird *occurrence probability* estimations in 130,769 locations selected from a geographically stratified random design based on a grid of 30 km pixels. Variation in detectability associated with observation effort is controlled by assuming that all effort predictors (search time, transect length, time of day, and number of observers) were constant and additively associated with the true occurrence probability. Thus, occurrence maps show the estimated probability that a single eBird participant will detect the species on a search for a given time range (*e.g.*, from 7-8AM) while traveling a predefined distance (*e.g.*, 1km) on a given day (*e.g.*, 03/04/2009). Another outcome of these predictions was spatio-temporally explicit measures of *predictor importance* (PI). The importance of a Decision Tree predictor is computed as the sum of the empirical improvement in the Decision Tree splitting criterion due to this predictor [5]. The predictor importance over an ensemble of trees is defined as the sum of predictor importance values for each tree in the ensemble [16]. In STEM, roughly speaking, the PI can be intuitively interpreted as the strength of the association between observed species occurrence/absence of a bird and the given predictor variable. This measure has no units. In this work, we use only the PI values for the 16 NLCD predictors, since as explained in Section 4 we are interested in understanding how these values relate to the actual bird habitat usage.

### 4 VISUALIZING BIRD DISTRIBUTION MODELS IN SPACE, TIME AND ACROSS SPECIES

**Desiderata for Visualizations.** When we started this project, our goal was to design a set of interactive visualizations to help scientists (ornithologists, biologists, and statisticians) involved in the generation, validation, and interpretation of bird distribution models, better understand the predictions, tune the models and communicate their results.

<sup>1</sup>Elevation data is available at [http://eros.usgs.gov/#/Find\\_Data/Products\\_and\\_Data\\_Available/gtopo30\\_info](http://eros.usgs.gov/#/Find_Data/Products_and_Data_Available/gtopo30_info)

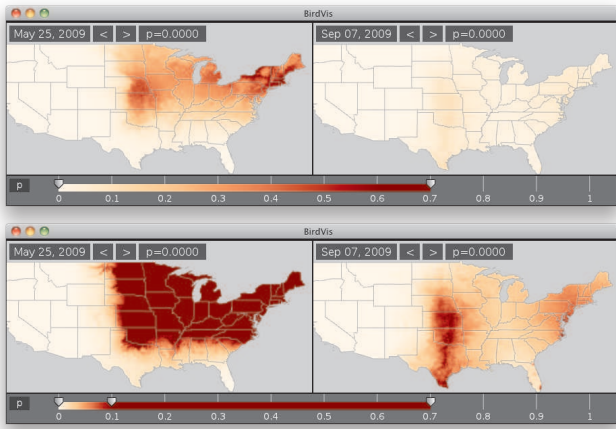


Fig. 3. In general, it is difficult to find a single colormap that is both perceptually good to display bird distribution and that represents well the range of the values in different times. Here, we show occurrence probability maps for Baltimore Oriole in May 25 and September 7 of 2009 using two configurations. The top row uses a continuous sequential colormap from probability 0 to 0.7. While this highlights important features in May 25, the details in September 7 are not clear. The bottom row uses the same colormap but with a range between 0 and 0.1, and dark red is used for values greater than 0.1. This reveals details in September 7, but it is not as effective for May 25.

In order to define the requirements for our design, we have carried out extensive interviews to understand the analysis needs of these scientists and the scientific questions they investigate. These are summarized below.

- Identify and understand migration patterns. Besides visualizing bird movement, they need to explore the timing, direction, speed, and duration of these movements.
- Validate the hypotheses that STEM predictor importance reflects actual species habitat preferences.
- Compare the behavior of different species, both regarding migration patterns and habitat use.
- Perform statistical analyses to validate and improve the prediction model.

Carrying out these tasks is challenging due to the multivariate aspect of the data: multiple dimensions (species, space and time) associated with multiple signals being studied (occurrence probability, predictor importance). For example, to validate hypotheses regarding habitat associations for a species based on habitat predictor importance signals, it is necessary to compare 16 variables, each of which depends on space and time. Figure 2(b) shows a boxplot for the predictor importance distribution of these 16 habitat predictors for a fixed species, region and time period. Often, scientists need to inspect a large number of these plots to explore different space-time windows, as well as multiple species.

To better support these tasks, we need insightful visualizations for different parameter combinations and the ability to compare and interactively manipulate them. In what follows, we present novel visual representations we have designed and discuss our design decisions. We also describe how we combined these visualizations into a new interactive tool.

**Preprocessing and Data Abstraction.** Given a species  $s$  (e.g., Indigo Bunting) and a time  $t$  (e.g., 3rd week of 2009), STEM results, be it the occurrence map for  $s$  at  $t$  or its associated predictor  $\sigma$  importance data (e.g.,  $\sigma = \text{Pasture}$ ), can be described as a function of the form  $f: P \rightarrow \mathbb{R}$ , where  $P$  is a set of points randomly sampled across the spatial domain and  $f(p)$  is a real value associated with each  $p \in P$ . When  $f$  represents an occurrence map,  $f(p)$  is a number between  $[0, 1]$  which indicates the occurrence probability in location  $p$  at time  $t$  for species  $s$ .

Before visualizing the model results, for each function  $f$  generated by STEM, we compute a new function  $f': P' \rightarrow \mathbb{R}$  based on a regular

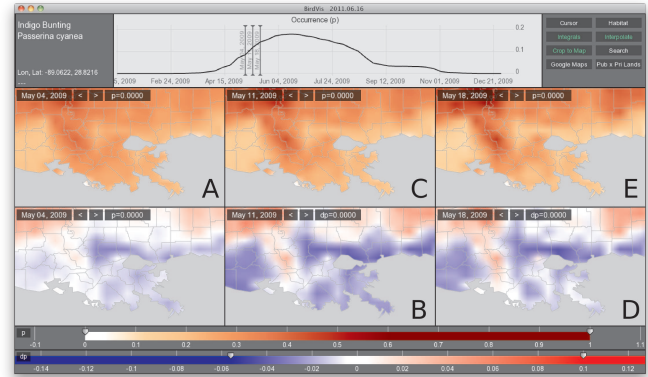


Fig. 4. Occurrence variation maps (bottom row) complement the usual occurrence maps (top row) by highlighting intensity and sign of occurrence probability differences. Let  $A, B, C, D, E$  refer to the underlying scalar fields of the corresponding map-based visualizations. The scalar fields of the occurrence variation maps (bottom row) obey the equations:  $B = C - A$  and  $D = E - C$ . The colors in these views are controlled by an independent divergent colormap controller (bottom colormap). Note that by considering only the top row, it is not clear that the occurrence probability for Indigo Bunting on the Louisiana shore is decreasing. In contrast, this pattern is evident in the bottom row (shown in Blue).

grid: the points in  $P'$  are the center locations of cells in a regular grid containing all points of  $P$ , and  $f'(p')$  is the average of all  $f(p)$  values for the points  $p \in P$  in the cell centered at  $p'$ . Since the points in  $P$  are randomly sampled (a characteristic of STEM), we can always define a regular grid where each cell contains approximately the same number of points in  $P$  ( $\sim 5$  points per cell in our case) and produce an  $f'$  that is a good approximation for  $f$ .

The ability to quickly compute the average value of  $f'$  over rectangular regions is a key requirement to allow interactive responses on some of the visualizations we have designed, including tag cloud lenses (Section 4.2) and temporal trajectory summaries (Section 4.4). Thus, we pre-compute  $f'': P' \rightarrow \mathbb{R}$ , an accumulated version of  $f'$ . For each point  $p' \in P'$  described by grid coordinates  $i_0, j_0$ , we define  $f''(p')$  as the sum of all points described by grid coordinates  $i, j$  where  $i \leq i_0$  and  $j \leq j_0$ . It is easy to check that the sum of a rectangular region in  $f'$  can be computed in constant time, with at most four accesses to  $f''$ .

When we combine all functions  $f''$  corresponding to a given variable, we obtain a new function defined over a 3-dimensional grid. We refer to the set of all variables (occurrence probability and predictor importance) as a *spatio-temporal cube*.

#### 4.1 Interactively Visualizing Bird Occurrence

Occurrence predictions are the basic outcome of bird distribution models. To understand such predictions, domain scientists often use color-coded, map-based static visualizations. Figure 2(a) shows an example, where a sequential colormap is used to represent the predicted occurrence probability of the Northern Cardinal (*Cardinalis cardinalis*). These visualizations have important limitations, notably: the lack of interactivity, the use of fixed colormaps, and the complexity of inferring movement information based only on occurrence maps. As we discuss below, we have designed map-based visualizations that address these limitations.

##### Adjusting Color Maps to Observe Different Population Thresholds.

Bird occurrence predictions are based on observation data. When a species is migrating, fewer observation records are reported for that species than when it is breeding. This fact is reflected in the range of the predicted occurrence probability values, which are smaller in the migration season and larger in the breeding season. Thus, if a single colormap configuration is used to encode occurrence probability values for a given species throughout the year, important details are likely to be missed. Figure 3 illustrates this problem. It shows occurrence maps for the Baltimore Oriole on two dates where the occurrence

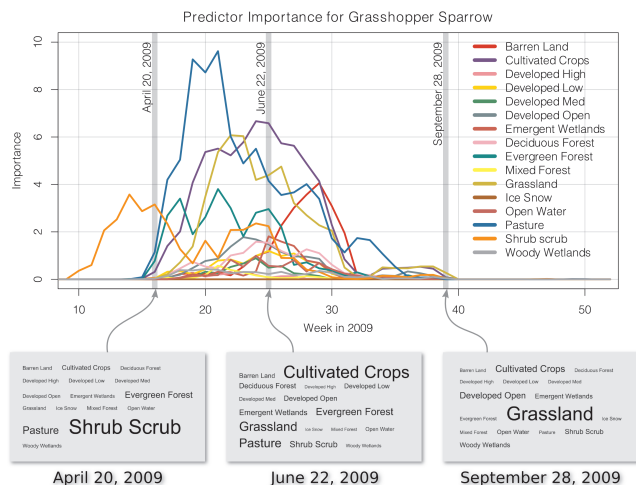


Fig. 5. Top: plot of predictor importance (PI) for 16 habitat variables in the year of 2009 for the species Grasshopper Sparrow. Bottom: tag clouds based on PI values to highlight which habitats are more relevant in three different dates. Tag clouds provide a more effective representation, helping scientists more quickly understand the relative importance of variables in a specific date. The actual tag cloud conveys the information directly and does not require the user to carefully inspect the individual plot curves and the legend. The sequence of tag clouds also clearly shows an important ecological fact: the habitat association for the Grasshopper Sparrow changes throughout the year, from *Shrub-scrub* to *Cultivated Crops* and finally, to *Grassland*.

probability ranges differ substantially. Note that neither colormap configuration is effective for both dates. Although, in this specific case, it is possible to design a single colormap that would simultaneously show the details on both dates (e.g., from yellow at 0.0 to red at 0.1 to gray at 0.2 to cyan at 0.3 to blue at 0.7), this solution yields a less intuitive colormap [38] and does not generalize well if we want to see details simultaneously in more sub-ranges.

To give scientists the flexibility to experiment with different configurations, and adjust the visualizations to focus on specific signal ranges, we have included an *interactive colormap controller* that is always available through the interface. As the user modifies the colormap configurations, the results are immediately reflected in the visualization. As Figure 3 illustrates, the colormap widget (below the maps) serves both as a controller to change the colormap specification and as a legend for the maps.

**Visualizing Variations in Occurrence Probabilities.** Ornithologists are interested in visualizing not only where birds are observed, but also migration patterns, including timing, direction, speed, and duration of movements. While occurrence location is evident in occurrence map visualizations, movement information is not always obvious. The top row of Figure 4 illustrates that it can be hard to perceive occurrence change by using side-by-side occurrence map visualizations. To address this issue, we display *occurrence variation maps* to help scientists see the differences between occurrence maps. The bottom row in Figure 4 shows examples of occurrence variation maps. With an independent (divergent) colormap and interactive controller, occurrence variation maps are useful to highlight the intensity and sign (i.e., negative or positive) of occurrence probability differences in a series of occurrence maps, providing better evidence of movement. For example, with the divergent colormap of Figure 4, a blue-to-white-to-red pattern supports a movement from the blue region to the red region passing through the white region.

## 4.2 Zooming into Habitats using Tag Cloud Lenses

As described in Section 3, for each set of occurrence probability predictions STEM produces, it also derives predictor importance (PI) values for a fixed set of predictors (e.g., *Pasture*, *Barren Land*). A key challenge we faced was to find effective ways to visualize the predictor importance, with the goal of helping scientists obtain insights

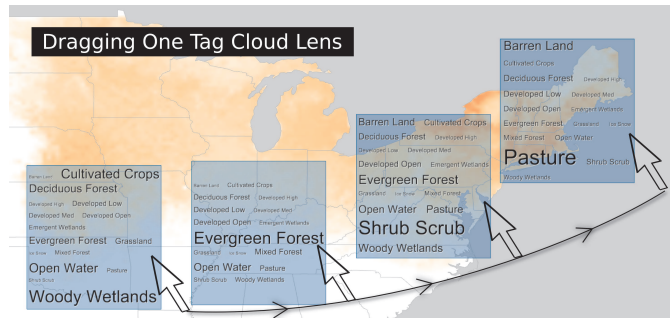


Fig. 6. Tag cloud lenses provide an intuitive tool to pose spatial queries with immediate feedback. The region can be specified by the size of the lens rectangle, and as we drag the rectangle over the map, the query is evaluated and the tag cloud is updated dynamically.

into aspects of the biological behavior the model suggests. Below we discuss how tag clouds provide an intuitive and effective means to visualize predictor importance. We also describe how we have extended tag clouds to be used as an exploratory tool which provides the means to interrogate habitat associations in space, time, and across species.

**Using Tag Clouds to Convey the Relative Importance of Predictors.** If we fix a species and average the predictor importance values for a single predictor (e.g., *Pasture*) in a given spatial-time window (e.g., the whole US territory on April 20, 2009), we get back a number that, by itself, does not have much practical meaning. However if we compute the same average for all other predictors of interest and compare the relative magnitudes of these averages, we can obtain insight into which predictors the model considered more relevant (to describe occurrence) for the species  $s$  in that same spatial-temporal window. Here, we use the 16 habitat predictors considered by STEM. If in a chosen spatial-temporal window  $t$ , the average PI value of one of the 16 habitat predictors is much larger than the others, this would support this predictor as the preferred habitat for  $s$  in  $t$ .

A possible visualization for predictor importance is a line plot. As illustrated in Figure 5, each curve in the plot corresponds to a habitat predictor and follows the average PI values (taken from the whole US territory) for the Grasshopper Sparrow at 52 different moments (one for each week of 2009). While examining PI values, the scientists are primarily interested in identifying the most relevant habitat predictors at a given spatio-temporal window. Although a lot of information can be extracted from the line plot in Figure 5 (e.g., magnitudes, maximum values, trends), their information need is not effectively addressed by this kind of visualization. Suppose the spatio-temporal window is set to be the US territory on April 20, 2009. To discover the most relevant predictors using the line plot in Figure 5, an analyst would need to identify the crossings with largest y-value between the curves and a vertical line on April 20, 2009 (left gray line in the plot), identify the colors of such top crossing curves, and then read the habitat predictor name corresponding to these colors in the plot legend.

Instead of a plot, we decided to represent the relative importance of predictors for a given spatio-temporal window as a tag cloud [41], where each tag is a habitat predictor name, and the tag importance [3, 35] is used to scale the tag font-size. We compute the size as the average PI value in the given spatio-temporal window. To make these tag clouds more perceptually efficient, we sort the tags in alphabetical order [15]. Tag clouds are effective at providing a concise overview about relative magnitude associated to a potentially large number of tags. Note how the three tag clouds on the bottom of Figure 5 clearly convey the habitat preferences for the Grasshopper Sparrow in three different spatio-temporal windows: the preference changes from *Schrub Scrub* in April, to *Cultivated Crops* in June, to *Grassland* in September.

**Tag Cloud Lenses as an Exploratory Tool.** To support the interactive exploration of predictor importance, we introduced tag cloud lenses, a magic lens that displays tag clouds. Tag cloud lenses provide an intuitive mechanism for users to formulate spatial queries on multivariate spatio-temporal data sets and visualize the relative differences among variables. More precisely, given a spatial region of interest (ROI), the

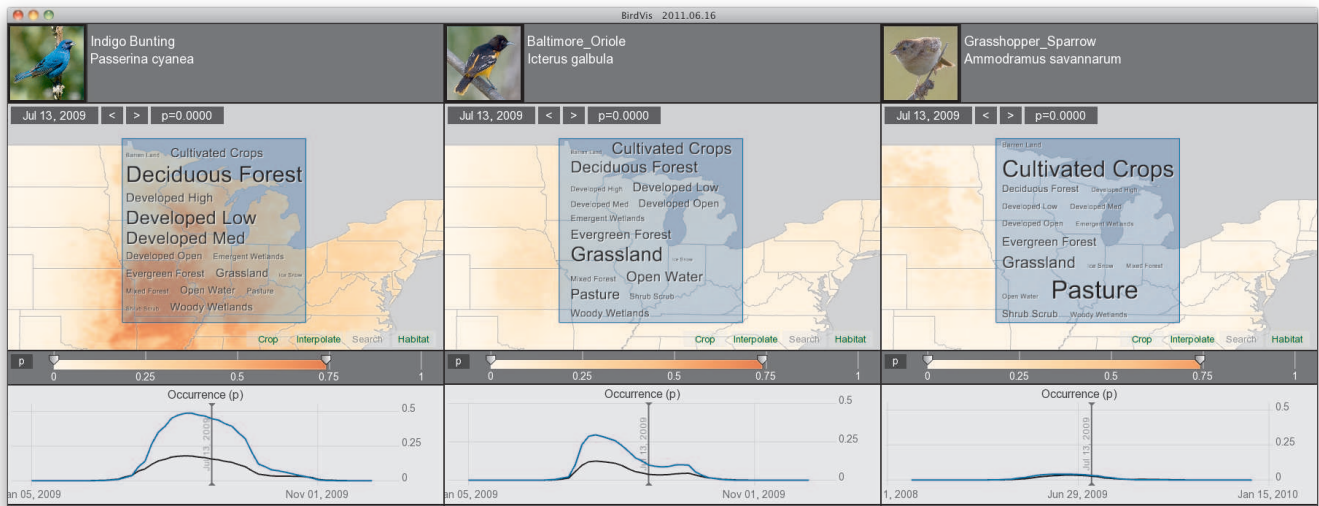


Fig. 7. BirdVis in multi-species mode with three species (one in each column): Indigo Bunting, Baltimore Oriole, and Grasshopper Sparrow. Note that, at July 13 of 2009, the habitat preferences are different for these species in the same region (intersection of the blue rectangle with the US territory).

lenses summarize the values of each variable inside the ROI and display a tag cloud showing the summary superimposed on the ROI. By making the lens transparent, we can see through it and maintain the context of the underlying map. Furthermore, multiple lenses can be displayed that correspond to different variables, and by using multiple coordinated views (over the temporal dimension), we can see how the summaries provided by the lenses change over time. As illustrated in Figure 6, the scientists can use tag cloud lenses to seamlessly move the tag clouds over the map. They can also scale the tag cloud and adjust it to the size of the ROI.

One challenge we faced was performance: each TCL, when moved onto a region, issues a query that needs to be evaluated over the underlying data. We use a preprocessing step (see Section 4) to obtain interactive rates for computing the TCLs. As Figure 1 shows, multiple TCLs can be displayed simultaneously on a single map allowing different regions to be inspected concurrently. Figure 1 also illustrates the use of coordinated views to inspect the same regions in different time steps. This is achieved by using coordinated lenses across the different views, *i.e.*, if a user creates, drags, resizes, or deletes a lens in one of the views, the result of the operation is synchronized in all other views. The only difference is that the tag clouds rendered on top of corresponding lenses (*e.g.*, the blue lenses in the three views) summarize a different time window. It is worth stressing that each TCL is a local summary and should be used to understand the relative importance of its tags in the corresponding spatio-temporal window. One should not compare sizes of tags across different TCLs.

### 4.3 Comparing Species

A key challenge in visualizing occurrence maps arises when multiple species need to be compared, since spatial dynamics and overall prevalence levels vary dramatically across species. Practically, this means that maps must be tuned for each species individually. This was done by our collaborators using R scripts—a process that is both tedious and error prone. To support this task, we have adopted a spreadsheet-based layout [37], which allows multiple visualizations (*e.g.*, occurrence maps, predictor importance) to be displayed side-by-side for comparison (see Figure 7). The cells in the spreadsheet can be synchronized in space, time and colormap: changes to one of these view parameters will affect the display of the synchronized cells.

### 4.4 Putting it all Together: The BirdVis Tool

We have combined the visual representations described in the previous sections into a new tool, which we call BirdVis. The user interface

of BirdVis consists of a single window partitioned into three regions: the top left contains the *Environment List View*, the bottom left contains a *Species Data Set List View* and, on the right is the *Display Area* (see Figure 1). At start up, the Display Area has no components. The idea is that a user, before exploring a data set, must first choose an *environment* for the Display Area. An environment is a pre-defined set of coordinated components [37] designed for a set of tasks. The current BirdVis prototype has three environments: *Single Species*, *Single Species with Occurrence Variation*, and *Multi-Species*. These environments consist of different coordination schemes for a set of basic components: *Map View*, *Temporal Trajectory Summary View*, *Colormap Widget*, *Information Panel*, *Tools Panel*. The environments and components of BirdVis are illustrated in Figure 8 and described below.

#### 4.4.1 Environments

**Single Species Environment.** To use this environment, a user drags and drops its name from the Environments List View component into the display area. On the top of the window, BirdVis shows the Temporal Trajectory Summary View, an Information Panel, and a Toolbar (see *e.g.*, Figure 1). The spreadsheet which shows a series of occurrence maps (Map Views) sits right below these components; the number of rows and columns in the spreadsheet are specified by the user. To allow the user to control the colormap for the different Map Views, a colormap widget linked to these views is available at the bottom of the window.

Once the environment fills the Display Area with its components, the user can drag any species in the Species List View into the Map Views. As soon as a species is dropped into a view, a summary trajectory curve is computed and displayed on the Temporal Trajectory Summary View. Occurrence maps visualizations are then computed for the species across all Map Views. When initialized, the Map Views are assigned different time offset signals. As soon as the visualization appears, the user can interact with the views by panning and zooming; modify the colormaps by dragging handles to change color assignment rules; move forward and backward in time, by either using the buttons associated with each Map View (which change only the data in of the corresponding Map View), or using arrow keys on the keyboard, which change the dates for all Map Views simultaneously. For the sizes of the data sets that our collaborators are currently exploring, all the updates happen smoothly for any numbers of rows and columns that are of practical use in terms of screen space. Another interaction available on Map Views is to create rectangular selections.

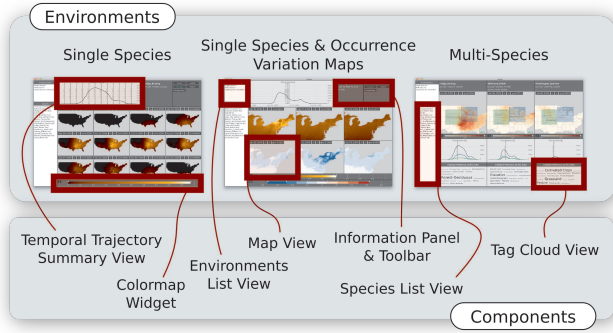


Fig. 8. BirdVis uses the notion of environments as predefined combinations and coordination of basic components. Here we show the three environments implemented in BirdVis and its basic components: a small set of components rewired in different ways to provide different functionality.

These selections are synchronized with all Map Views and once it is drawn in one of the views it appears in all other Map Views. A summary trajectory curve for the selected area also appears on the Temporal Trajectory Summary View with the same color as the selection rectangle on the map.

**Single Species Environment with Occurrence Variation Maps.** This environment is the same as the Single Species Environment with one extra element. For each occurrence map generated in the first environment, there is an associated occurrence variation map positioned below the original one. This is illustrated in Figure 4. These occurrence variations maps, described in Section 4.1, help in the identification of migration patterns. An independent Colormap Widget is linked to the occurrence variation maps.

**Multi-Species Environment.** While in the two previous environments, dropping of one species in a cell replaces all visualizations to reflect the selected species, in the multi-species environment different species can be dropped and visualized in different cells. Cell coordination in the Multi-Species Environment is similar to the other environments. The Map Views are synchronized: by zooming and panning in one of them the same action is applied to the other views (which are associated with different species). Selections are also synchronized, and by using the same commands as in the previous environments, the user can move back and forward in time. One difference in this environment is that the Temporal Trajectory Summary View is not shared by the cells—there is one per cell, *i.e.*, one for each species (see Figure 7).

#### 4.4.2 Basic Components

**Map View.** The *Map View* widget displays spatial faces, *i.e.*, points in a given time step, for the variables in the spatio-temporal cube whose values are represented using a color map. Given a variable  $v$  in a spatio-temporal cube, a Map View displays the vector function  $v_t$  for a given time  $t$ . Following the observations discussed in Section 4.1, besides the occurrence probability, the user can also see the probability variation for any variable—the occurrence variation maps display the scalar fields  $\Delta v = v_t - v_{t-1}$ .

In a Map View, a user can pan, zoom, and brush. Colors in a map are controlled by the *Colormap Widget* (described below). Using the brushing capability, the user can define a rectangular ROI that will be shown on the map. An ROI can also be dragged and re-scaled, and when this happens, the visualizations associated with it—curves in the Temporal Trajectory Summary View and the tag cloud lenses that are overlaid on them—are updated dynamically.

**Temporal Trajectory Summary View.** Together with a map, a user may also display the associated *Temporal Trajectory Summary View*. This view shows the temporal behavior of the data shown in the Map View, *i.e.*, the curve that represents the average occurrence value of that variable over an rectangular spatial region over time. By default, when data is initially loaded into the Map View, its temporal behavior

curve over the entire map is shown in the Temporal Trajectory Summary View; when the user defines a ROI, the view is linked to the ROI and it is dynamically updated as the ROI rectangle is dragged around the map. The user can pan, zoom and scale this view—it is possible to change the scales in both axis, as well as zoom and pan over the graph. **Colormap Widget.** The *Colormap Widget* controls the colors used by the Map View. The user can interactively change the colormap value range by dragging the handles, and the Map View is updated accordingly. The user can also add and hide intermediate handles. The intermediate handles allow colors to be distributed over the range of values in a non-uniform way (see *e.g.*, Figure 3).

BirdVis uses the sequential and divergent colormaps from Color Brewer [6] and Moreland [25] (see Figure 4). The sequential colormaps are used to display occurrence probabilities, while the divergent ones are used in the variation maps. The Colormap Widget uses the labeling algorithm proposed by Talbot *et al.* [44] to show reference intermediate values as legends in the bottom.

**Tools Panel.** This panel contains a set of tools that aid in exploratory analyses of the predictions. The *Habitat button* turns the tag cloud lenses on and off. When this is turned on, the ROIs present in the Map View are used as canvases to tag clouds that represent the predictor importance in those areas. The *Integral button* allows the user to select whether the Temporal Trajectory Summary View is used to represent the whole map or just a region. To search for a region in the map by its name, the user can type the name of the region in the *Search Tool*. The *Google Maps Tool* shows the currently displayed Map View on Google Maps. This tool is especially useful to correlate the displayed colormap with other features, *e.g.*, the cities in the region.

**Information Area.** To provide the user context information about the active task, the *Information Area* shows the species being displayed together with coordinates and the geographical location where the mouse is currently placed.

**Tag Cloud View.** This is an independent tag cloud view area. We use it in the cells of the Multi-Species environment to show the tag cloud for the whole spatial domain at the time instant associated with the cell.

## 5 IMPLEMENTATION

BirdVis was implemented in C++, and its graphical user interface was written in OpenGL and Qt. BirdVis is tightly integrated with R [34]. It uses R interface packages to C++: Rcpp and RInside. The spatio-temporal cubes used by BirdVis are stored as R data files and we use R for all data handling tasks. An important benefit of this approach is that we can leverage the R infrastructure and algorithms (*e.g.*, to deal with matrix data). Furthermore, because our collaborators use R for their current analyses, it makes it easier for us to integrate them into BirdVis.

Both the spatial and temporal resolution of the data sets can be configured in the input file, consequently, results of model evaluation in different scales can be visualized and compared. Furthermore, in the data file, the user can define two kinds of variables: variables that are depicted using colormaps and variables that are queried by the tag cloud lenses. The number of variables of the latter type and the tags used by BirdVis can also be configured in the data files, allowing the user to include different sets of predictors, and thus, enabling the application of BirdVis in other domains.

## 6 CASE STUDIES

In this section we present two case studies where BirdVis was used to analyze STEM data. The first one shows how BirdVis enabled scientists to put forward new hypotheses about the Palm Warbler migration. The second case study shows how tag clouds and tag cloud lenses were used to study the Indigo Bunting habitat preferences, and illustrates the usefulness of BirdVis for the interactive visualization and analysis of spatial-temporal data.

### 6.1 Understanding the Migration of Different Warbler Taxa

**Background.** The Palm Warbler (*Dendroica palmarum*) is an eastern Canadian-breeding warbler that winters primarily in the southeastern

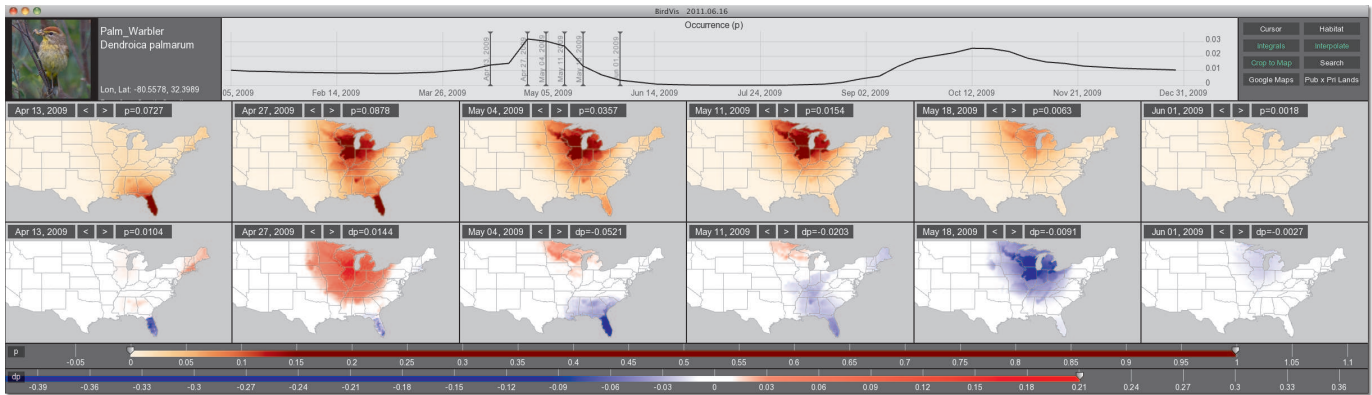


Fig. 9. Series of occurrence maps (top row) and occurrence variation maps (bottom row) for the 2009 Spring Migration of Palm Warbler.

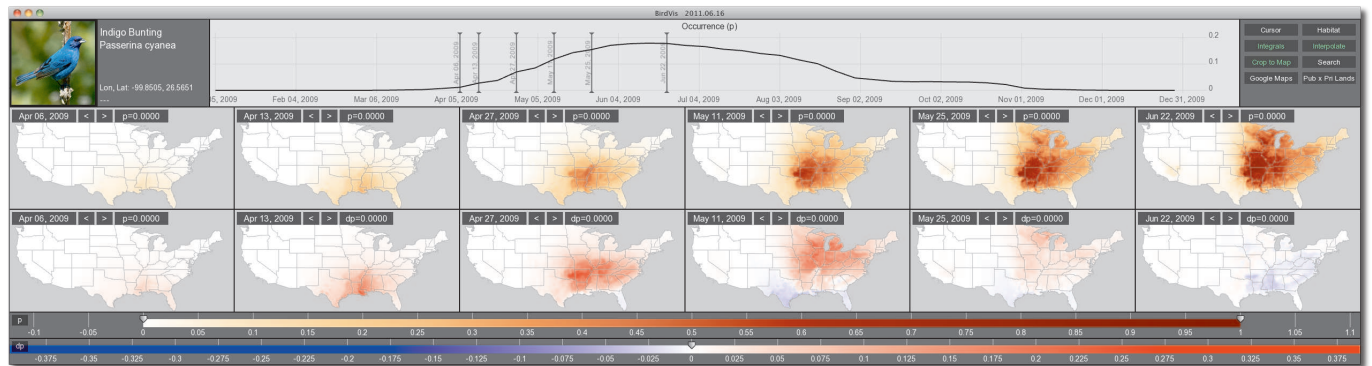


Fig. 10. Indigo Bunting migration: The first column shows the earliest arrivals of Indigo Bunting along the Gulf Coast during the last week of March/first week of April. As expected, the predicted occurrence distribution (shown on the top row) and predicted occurrence variation (on the bottom row) are very similar. From the first to the second week of April (column 2), the Indigo Bunting population greatly increases along the Gulf Coast. Indigo Buntings continue moving north the week of April 15 (column 3) and expand into much of the Ohio River Valley; predicted occurrence and change are still quite similar. By early May (column 4), dramatic differences appear between the two images. Indigo Buntings occur in the highest frequencies from south MO, south IL, south IN, south to east TX and west MS, but the greatest change in distribution is in the Upper Midwest, where Indigo Buntings now arrive in large numbers. By mid/end-May (column 5), buntings are largely present throughout their breeding range. Higher probabilities of detection are centered around the Upper Midwest where the bulk of Indigo Buntings are arriving. By early about 10 June, Indigo Buntings are on breeding territories and there is no change from the previous week.

United States and western Caribbean. Like other warblers that winter in the United States, it is one of the earlier species to move north in spring, with the primary northward push occurring in April. In the northern United States their migration winds down in May, while most other species of northern-nesting warblers are approaching peak occurrence. The Palm Warbler is especially interesting in that it has two populations. A subspecies of eastern breeders (*D. p. hypochrysea*; Yellow Palm Warbler) nests in Atlantic Canada and winters from northern Florida to east Texas, while another (*D. p. palmarum*; Western Palm Warbler) nests further west in Canada but winters further east, from the southeastern U.S. to south Florida and the western Caribbean. The migration trajectories of these two subspecies thus form an X, and Yellow Palm moves north earlier in spring and south later in fall than its western counterpart. STEM has provided the first data set showing the complete annual cycle for this species [12,32].

**The Analysis.** The scientists produced occurrence maps and occurrence variation maps to visualize the Palm Warbler spring migration. In Figure 9, we can see that, on April 13 (first column), the migration starts: the southeastern wintering range shows well on the top occurrence map, with peak occurrence in Florida; declines in Florida and increases in New England indicate that migration is underway with Yellow Palms leaving the southeast and heading to Atlantic Canada. On April 27, the migration of Yellow Palm is already on the decline, with lower occurrence in New England as compared to the previous week, but an obvious push to the north-northwest as Western Palms move from Florida towards the Great Lakes region. Subsequent fig-

ures show a clear picture of northward passage, with Yellow Palm migration ending well before that of the western population. By May 18, Palm Warblers are less common everywhere than they were the previous week, indicating that spring migration for the species as a whole is drawing to a close.

This visualization of bird occurrence patterns, showing multiple dates and both the occurrence and occurrence variation, has provided new insights and hypotheses. Past authors have discussed the migration of the Yellow Palm Warbler as heading northeastward up the Atlantic Coast (e.g., [8]). However, Figure 9 (on April 27 and May 4) shows high Palm Warbler occurrence New England and the Southeastern U.S., but the species never seems to reach high occurrence in the mid-Atlantic states, which shows up as a wedge of low occurrence on Figure 9, April 27. The pattern actually suggests that a significant proportion of the population may shortcut across the Atlantic Ocean. While intentional autumn movement over the western Atlantic Ocean is well-known in several species of shorebirds and some passerines such as Blackpoll Warbler [27, 28], it has not been suggested to be a significant migration path in spring. STEM models do suggest that this may occur in spring in both Yellow Palm and Blackpoll Warblers, and anecdotal observations from field birders (A. Farnsworth, M. Iliff pers. obs.) provide further support for the possibility of an offshore migration path in spring.

In this case, BirdVis not only highlighted migration differences for these two taxa, but generated an intriguing new hypothesis that some species may be migrating directly to the Northeast without making

landfall in the MidAtlantic. Confirmation of this hypothesis would contribute significantly to our understanding of migration pathways in eastern North America.

## 6.2 Analyzing the Habitat Preference for Indigo Bunting

**Background.** Indigo Bunting (*Passerina cyanea*) is a migratory song-bird that breeds primarily in the eastern United States and winters in Mexico, Central America and the Caribbean. While general patterns of migration are known in broad terms, most publications focus on first arrival dates, band recoveries of individual birds, or some combination of the two (e.g., [23, 30, 36]).

**The Analysis.** First, the scientists used BirdVis to obtain an overview of the Indigo Bunting distribution (Figure 10). In addition to map views, the scientists used the tag cloud lenses to visualize how the habitat preferences change on the maps. As Figure 1 illustrates, by creating multiple tag cloud lenses for different regions and examining different dates, they were able to identify interesting regions where the local importance of the predictors differs from the importance computed over the entire map.

Figure 11 highlights three relationships: Indigo Bunting distribution, date, and the important of habitat (NLCD classes) predictors in relation to predicted species occurrence. On the left, we see the species' core breeding range (dark orange), with the most important factor affecting their predicted occurrence being an association with Deciduous Forest. At this season, Indigo Buntings are strongly associated with forest edge, power-line cuts, and regenerating clear-cuts, where they feed primarily on insects. But during fall migration (shown on the right), we visualize a different story. The core population has shifted southward (dark orange), and now the most powerful association for predicted occurrence is Cultivated Crops, with Deciduous Forest and Pasture weighing heavily in the mix. This apparent habitat shift from shrubby thickets to more open, grassy, agricultural areas fits well with the species known biology [31]. During the breeding season, Indigo Buntings require protein-rich insects to raising young. But after breeding, the buntings switch to seeds that allow them to gain fat reserves necessary to fuel their journey over the Gulf of Mexico. BirdVis is able to clearly visualize not only the changes in distribution, but the changes in habitats use at different times of the year. By using tag cloud lenses, scientists were able to identify interesting features about how habitat preference changes over space and time. The tag cloud lenses superimposed on the maps in Figure 11 exemplify this use.

We should note that the data we analyzed is the first to show fine spatial-scale movements and habitat-associations across the entire breeding range throughout the year. While the dietary shift of Indigo Bunting was known, this knowledge was based on observations from a very small number of small study plots. Before this analysis, there was no empirical evidence that these relationships hold across large regions, or knew when exactly these shifts in diet occurred.

## 7 DISCUSSION

The combination of the different visual components provided by BirdVis gives scientists the ability to navigate through and explore all dimensions of their data sets: space, time, species, probability occurrences and predictor importance (habitat association). These components were designed to support a set of queries and tasks identified as critical by the ornithologists and biologists who analyze the STEM predictions, and the statisticians who have developed and calibrated the model. An initial evaluation of BirdVis by these experts has shown that the unprecedented flexibility afforded by the tool to explore and visualize bird population-level movement on a weekly basis across a range of bird species is promising. They were able to both confirm existing hypotheses as well as formulate new ones. The BirdVis prototype has been deployed at the Cornell Ornithology lab and is already being used, on a regular basis, by the scientists in that lab.

The initial evaluation has also uncovered some limitations. Since the TCLs use tag clouds as a basic component, they inherit both the advantages and limitations of tag clouds. As we have discussed above, the ability to intuitively convey relative importance for predictors is beneficial, and this was the main motivation for us to use tag clouds.

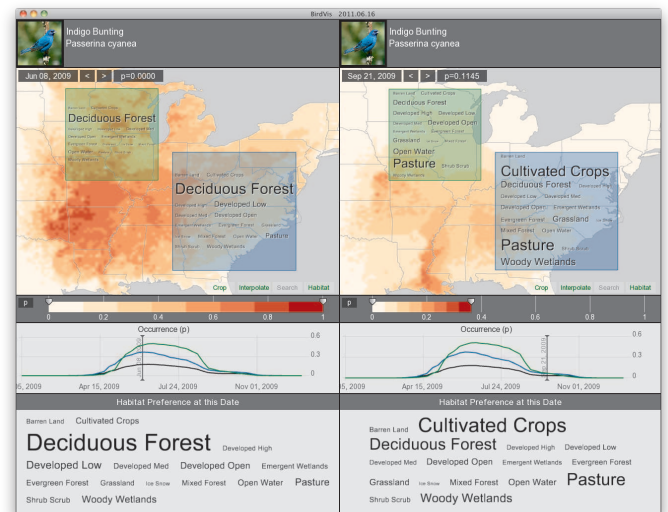


Fig. 11. The predicted breeding distribution of Indigo Bunting on June 8, 2009 is shown on the left. Note the strong association with NLCD class Deciduous Forest, indicating this species edge-nesting habits. On the right is the predicted distribution during fall migration, 21 September 2009. We can observe a shift in the prevalence of NLCD class Cultivated Crops as an occurrence predictor, the retention of Deciduous Forest, and Pasture becoming important for the species. These changes reflect dietary shifts, as the species moves from a primarily insect diet during breeding, to one based mostly on seeds during migration and winter. By using tag cloud lenses synchronized across the two views, it is possible to analyze how habitat preference changes over time.

However, tag clouds have well-known problems. Notably, the length of the tags can influence their perceived sizes [45]. In this paper, we have focused on rectangular lenses. One interesting direction we would like to pursue in future work is to use more general lenses shape and understand how to respond to the more general underlying queries at interactive rates.

Another interesting direction we plan to explore is how to represent trends. The ability to visualize how habitat associations vary over time is key to discovering regions when and where these associations are stationary. Stationary regions are contiguous pieces of land where all habitat predictor importance throughout a year follow similar curves when compared to curves of neighboring areas. Ecologists studying climate change are now realizing how important it is to understand where and when these habitat-occurrence relationships are stationary. When modeling bird distributions, if a model developed for one area is transferred to another stationary area, the results become spurious extrapolations. With the tag cloud lenses in BirdVis, it is possible to get evidence of stationary regions: by moving the lenses in a near by region, if the tag relative sizes do not change much, we are able to identify a region as stationary. But while TCLs offer a means to continuously explore the spatial dimensions of the data, they only support exploration over discrete moments in the temporal dimension. We would like to investigate the use of techniques such as the ones proposed by Lee *et al.* [22] to represent trends in tag cloud lenses.

**Acknowledgments.** We thank the eBird participants for providing the data used in our case studies and the staff at the Cornell Laboratory of Ornithology for managing these data. The bird pictures displayed in BirdVis were provided by Bill Schmoker (schmoker.org). This work has been supported by the Leon Levy Foundation; the Wolf Creek Foundation; the National Science Foundation through DataONE (0830944), IIS-0905385, IIS-0844546, IIS-0746500, CNS-0751152; the Institute for Computational Sustainability (0832782), research grant (1017793); TeraGrid computing resources provided under grant numbers TG-DEB100009 and DEB110008.

## REFERENCES

- [1] S. Ahern, M. Naaman, R. Nair, and J. Yang. World Explorer: Visualizing Aggregate Data from Unstructured Text in Geo-Referenced Collections. In *Proc. of the 7th ACM/IEEE-CS joint Conf. on Digital libraries*, pages 1–10, 2007.
- [2] N. Andrienko and G. Andrienko. *Exploratory analysis of spatial and temporal data: a systematic approach*. Springer Verlag, 2006.
- [3] S. Bateman, C. Gutwin, and M. Nacenta. Seeing things in the clouds: the effect of visual features on tag cloud selections. In *Proc. of the ACM Conf. on Hypertext and Hypermedia*, pages 193–202, 2008.
- [4] E. Bier, M. Stone, K. Pier, W. Buxton, and T. DeRose. Toolglass and magic lenses: the see-through interface. In *Proc. of the Conf. on Computer graphics and interactive techniques*, pages 73–80, 1993.
- [5] L. Breiman, J. Friedman, C. J. Stone, and R. Olshen. *Classification and regression trees*. Wadsworth International Group, 1984.
- [6] Colorbrewer: Color advice for maps. <http://colorbrewer2.org>.
- [7] L. Brown and H. Hua. Magic lenses for augmented virtual environments. *IEEE Computer Graphics and Applications*, 26(4):64–73, 2006.
- [8] J. L. Bull. *Birds of New York State*. Doubleday, Garden City, N.Y., 1974.
- [9] J. Chen, A. MacEachren, and D. Guo. Supporting the Process of Exploring and Interpreting Space-Time Multivariate Patterns: The Visual Inquiry Toolkit. *Cartography and geographic information science*, 35(1):33, 2008.
- [10] The eBird Project. <http://www.ebird.org>.
- [11] N. Elmquist, P. Dragicevic, and J. Fekete. Rolling the dice: Multidimensional visual exploration using scatterplot matrix navigation. *IEEE Transactions on Visualization and Computer Graphics*, pages 1141–1148, 2008.
- [12] D. Fink, W. Hochachka, B. Zuckerberg, D. Winkler, B. Shaby, M. Munson, G. Hooker, M. Riedewald, D. Sheldon, and S. Kelling. Spatiotemporal exploratory models for broad-scale survey data. *Ecological Applications*, 20(8):2131–2147, 2010.
- [13] J. Fry, M. Coan, C. Homer, D. Meyer, and J. Wickham. Completion of the national land cover database (nlcd) 19922001 land cover change retrofit product. Technical Report 20081379, Geological Survey Open-File Report, 2009.
- [14] D. Guo, J. Chen, A. MacEachren, and K. Liao. A visualization system for space-time and multivariate patterns (vis-stamp). *IEEE Transactions on Visualization and Computer Graphics*, pages 1461–1474, 2006.
- [15] M. Halvey and M. Keane. An assessment of tag presentation techniques. In *Proc. of the Int. Conf. on World Wide Web*, pages 1313–1314, 2007.
- [16] T. Hastie, R. Tibshirani, and J. Friedman. *The elements of statistical learning: data mining, inference, and prediction*. Springer Verlag, 2009.
- [17] B. Hecht, M. Rohs, J. Schöning, and A. Krüger. Wikikey—using magic lenses to explore georeferenced Wikipedia content. In *Proc. of the Int. Workshop on Pervasive Mobile Interaction Devices (PERMID)*, 2007.
- [18] B. Hibbard and D. Santek. The VIS-5D system for easy interactive visualization. In *Proc. of the 1st Conf. on Visualization'90*, pages 28–35, 1990.
- [19] A. Jaffe, M. Naaman, T. Tassa, and M. Davis. Generating summaries and visualization for large collections of geo-referenced photographs. In *Proc. of the 8th ACM Int. workshop on Multimedia information retrieval*, pages 89–98, 2006.
- [20] D. A. Keim, C. Panse, and M. Sips. Information Visualization: Scope, Techniques and Opportunities for Geovisualization. In J. Dykes, A. MacEachren, and M.-J. Kraak, editors, *Exploring Geovisualization*. Oxford: Elsevier, 2004.
- [21] S. Kelling, W. Hochachka, D. Fink, M. Riedewald, R. Caruana, G. Ballard, and G. Hooker. Data-intensive science: a new paradigm for biodiversity studies. *BioScience*, 59(7):613–620, 2009.
- [22] B. Lee, N. Riche, A. Karlson, and S. Carpendale. SparkClouds: Visualizing Trends in Tag Clouds. *IEEE Transactions on Visualization and Computer Graphics*, 16(6):1182–1189, 2010.
- [23] G. H. Lowery, L. W. Life, and F. Commission. *Louisiana birds*. Published for the Louisiana Wild Life and Fisheries Commission by Louisiana State University Press, [Baton Rouge], 1974.
- [24] S. Milgram and D. Jodelet. Psychological maps of Paris. *Environmental psychology*, pages 104–124, 1976.
- [25] K. Moreland. Diverging Color Maps for Scientific Visualization. *Advances in Visual Computing*, pages 92–103, 2009.
- [26] M. A. Munson, S. D. K. Webb, D. Fink, W. M. Hochachka, M. Iliff, M. Riedewald, D. Sorokina, C. Sullivan, B. and Wood, and S. Kelling. The ebird reference dataset 2.0. [http://www.avianknowledge.net/content/features/archive/eBird\\_Ref](http://www.avianknowledge.net/content/features/archive/eBird_Ref), 2010.
- [27] I. C. T. Nisbet. Autumn Migration of the Blackpoll Warbler: Evidence for Long Flight Provided by Regional Survey. *Bird-Banding*, 41(3):207–240, 1970.
- [28] I. C. T. Nisbet, D. B. McNair, W. Post, and T. C. Williams. Transoceanic migration of the Blackpoll Warbler: Summary of scientific evidence and response to criticisms by Murray. *Journal of Field Ornithology*, 66(4):612, 1995.
- [29] T. Nocke, M. Flechsig, and U. Böhm. Visual exploration and evaluation of climate-related simulation data. In *Proc. of the Conf. on Winter simulation: 40 years! The best is yet to come*, pages 703–711, 2007.
- [30] R. B. Payne. *Indigo Bunting (Passerina cyanea)*. Birds of North America Online. Cornell Lab of Ornithology; Retrieved from the Birds of North America Online: <http://bna.birds.cornell.edu/bna/species/004>, oct 2006.
- [31] R. B. Payne. Indigo bunting (*passerina cyanea*), the birds of north america online. <http://bna.birds.cornell.edu/bna/species/004>, 2006.
- [32] R. B. Payne. Palm warbler (*dendroica palmarum*), the birds of north america online. <http://bna.birds.cornell.edu/bna/species/238>, 2006.
- [33] K. Potter, A. Wilson, P. Bremer, D. Williams, C. Doutriaux, V. Pascucci, and C. Johnson. Ensemble-vis: A framework for the statistical visualization of ensemble data. In *IEEE Int. Conf. on Data Mining Workshops*, pages 233–240, 2009.
- [34] R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2011. ISBN 3-900051-07-0.
- [35] A. Rivadeneira, D. Gruen, M. Muller, and D. Millen. Getting our head in the clouds: toward evaluation studies of tagclouds. In *Proc. of the SIGCHI Conf. on Human factors in computing systems*, pages 995–998, 2007.
- [36] S. D. Robbins. *Wisconsin birdlife : population & distribution past & present*. University of Wisconsin Press, 1991.
- [37] J. Roberts. State of the art: Coordinated & multiple views in exploratory visualization. In *Coordinated and Multiple Views in Exploratory Visualization (CMV)*, 2007, pages 61–71, 2007.
- [38] B. Rogowitz and A. Kalvin. The which blair project: A quick visual method for evaluating perceptual color maps. In *Proc. of the Conf. on Visualization'01*, pages 183–190, 2001.
- [39] K. Ryall, Q. Li, and A. Esenther. Temporal magic lens: Combined spatial and temporal query and presentation. *Human-Computer Interaction-INTERACT*, pages 809–822, 2005.
- [40] J. Shamoun-Baranes, W. Bouten, L. Buurma, R. DeFusco, A. Dekker, H. Sierdsema, F. Sluiter, J. Van Belle, H. Van Gasteren, and E. Van Loon. Avian information systems: Developing web-based bird avoidance models. *Ecology and Society*, 13(2):38, 2008.
- [41] A. Slingsby, J. Dykes, J. Wood, and K. Clarke. Interactive tag maps and tag clouds for the multiscale exploration of large spatio-temporal datasets. In *Proc. of the Int. Conf. Information Visualization*, pages 497–504, 2007.
- [42] B. Sullivan, S. Kelling, C. Wood, M. Iliff, D. Fink, M. Herzog, D. Moody, and G. Ballard. Data Exploration Through Visualization Tools. In *Proc. of the Int. Partners in Flight Conference: Tundra to Tropics*, pages 415–418, Feb. 2008.
- [43] B. Sullivan, C. Wood, M. Iliff, R. Bonney, D. Fink, and S. Kelling. eBird: A citizen-based bird observation network in the biological sciences. *Biological Conservation*, 142(10):2282–2292, 2009.
- [44] J. Talbot, S. Lin, and P. Hanrahan. An Extension of Wilkinson's Algorithm for Positioning Tick Labels on Axes. *IEEE Transactions on Visualization and Computer Graphics*, 16(6):1036–1043, 2010.
- [45] F. Viégas and M. Wattenberg. Timelines tag clouds and the case for vernacular visualization. *interactions*, 15(4):49–52, 2008.
- [46] J. Wood, J. Dykes, A. Slingsby, and K. Clarke. Interactive Visual Exploration of a Large Spatio-Temporal Dataset: Reflections on a Geovisualization Mashup. *IEEE Transactions on Visualization and Computer Graphics*, pages 1176–1183, 2007.