

LETTER • **OPEN ACCESS**

Analysis of animal-related electric outages using species distribution models and community science data

To cite this article: Mei-Ling E Feng *et al* 2022 *Environ. Res.: Ecology* **1** 011004

View the [article online](#) for updates and enhancements.

You may also like

- [Challenges in Scientific Data Communication from Low-mass Interstellar Probes](#)
David G. Messerschmitt, Philip Lubin and Ian Morrison
- [Collective effects of link failures in linear flow networks](#)
Franz Kaiser, Julius Strake and Dirk Witthaut
- [Anticipating water distribution service outages from increasing temperatures](#)
Emily N Bondank, Mikhail V Chester, Austin Michne et al.

ENVIRONMENTAL RESEARCH ECOLOGY



LETTER

OPEN ACCESS

RECEIVED
25 January 2022

REVISED
27 May 2022

ACCEPTED FOR PUBLICATION
14 June 2022

PUBLISHED
30 August 2022

Original Content from
this work may be used
under the terms of the
[Creative Commons
Attribution 4.0 licence](#).

Any further distribution
of this work must
maintain attribution to
the author(s) and the title
of the work, journal
citation and DOI.



Analysis of animal-related electric outages using species distribution models and community science data

Mei-Ling E Feng^{1,8,*} , Olukunle O Owolabi^{2,8} , Toryn L J Schafer³ , Sanhita Sengupta⁴, Lan Wang⁵, David S Matteson³ , Judy P Che-Castaldo⁶ and Deborah A Sunter^{2,7}

¹ Conservation & Science Department, Lincoln Park Zoo, Chicago, IL, United States of America

² Department of Mechanical Engineering, Tufts University, 419 Boston Ave, Medford, MA 02155, United States of America

³ Department of Statistics and Data Science, Cornell University, Ithaca, New York, United States of America

⁴ School of Statistics, University of Minnesota, Saint Paul, MN, United States of America

⁵ Department of Management Science, University of Miami, Coral Gables, FL, United States of America

⁶ Branch of Species Status Assessment Science Support, U.S. Fish and Wildlife Service, Shirley, New York, United States of America

⁷ Department of Civil and Environmental Engineering, Tufts University, Medford, MA, United States of America

⁸ Authors contributed equally.

* Author to whom any correspondence should be addressed.

E-mail: mefeng7@gmail.com

Keywords: animal activity, birds, community science, electrical distribution system, power outages, reliability, species distribution modeling

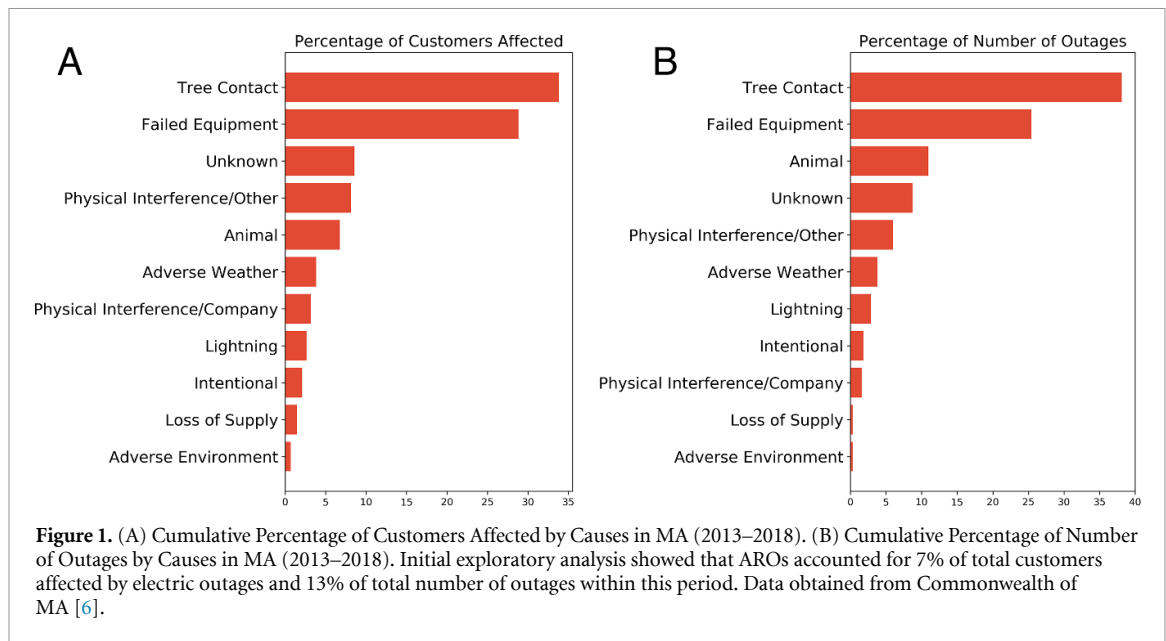
Supplementary material for this article is available [online](#)

Abstract

Animal-related outages (AROs) are a prevalent form of outages in electrical distribution systems. Animal-infrastructure interactions vary across species and regions, underlining the need to study the animal-outage relationship in more species and diverse systems. Animal activity has been an indicator of reliability in the electrical grid system by describing temporal patterns in AROs. However, these ARO models have been limited by a lack of available species activity data, instead approximating activity based on seasonal patterns and weather dependency in ARO records and characteristics of broad taxonomic groups, e.g. squirrels. We highlight available resources to fill the ecological data gap limiting joint analyses between ecology and energy sectors. Species distribution modeling (SDM), a common technique to model the distribution of a species across geographic space and time, paired with community science data, provided us with species-specific estimates of activity to analyze alongside spatio-temporal patterns of ARO severity. We use SDM estimates of activity for multiple outage-prone bird species to examine whether diverse animal activity patterns were important predictors of ARO severity by capturing existing variation within animal-outage relationships. Low dimensional representation and single patterns of bird activity were important predictors of ARO severity in Massachusetts. However, both patterns of summer migrants and overwintering species showed some degree of importance, indicating that multiple biological patterns could be considered in future models of grid reliability. Making the best available resources from quantitative ecology known to outside disciplines can allow for more interdisciplinary data analyses between ecological and non-ecological systems. This can result in further opportunities to examine and validate the relationships between animal activity and grid reliability in diverse systems.

1. Introduction

Power outages result in unwanted disruptions to the functioning of the electric distribution system. These disruptions are a result of a variety of interacting factors including weather, failed equipment, vegetation, and animal activity [1]. Understanding system reliability—the ability to withstand disruptions and minimize supply loss—is a long standing research priority to maintain critical infrastructure and essential public



services [1, 2]. As a result, assessments of distribution reliability can be broken into three research sectors: historical reliability, predicted reliability, and outage causes [2, 3]. Historical reliability is a reactive approach that makes use of historical data on past performance to target less reliable areas of the electric system for improvement. Predictive approaches have developed proactive models that forecast expected outages using a combination of causal factors [4]. The third sector focuses on mechanisms driving outages, identifying specific factors and their interconnected dependencies that disrupt the electrical system [5]. In this study, we focus on the third sector, specifically within the context of animal-related outages (hereafter, AROs), and investigate whether simultaneous activity patterns in multiple species can serve as indicators of reliability in the electric distribution system.

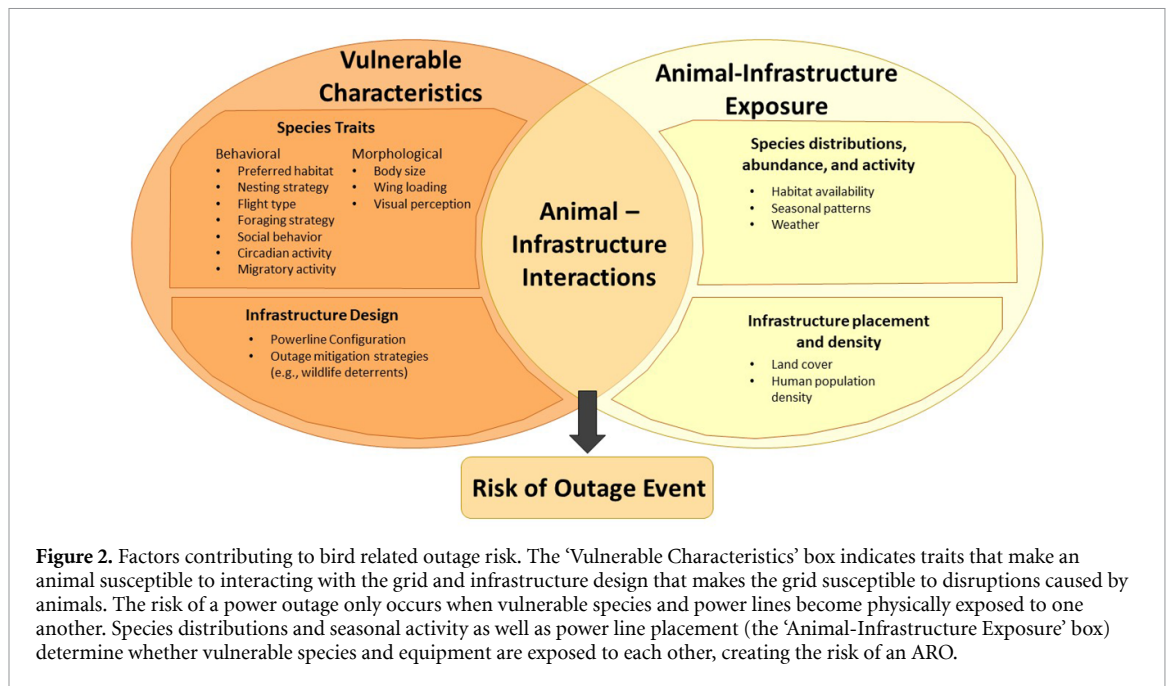
ARO contribute a fair proportion of disruptions to the electrical grid. For example, AROs represented about 7% of the total number of customers affected by electric outages in the state of Massachusetts (hereafter, MA) between 2013 and 2018 (figure 1). While AROs impact comparatively fewer distribution customers than other outage causes, they are often frequently occurring, resulting in considerable impact on the grid [7]. Many taxonomic groups have been identified as causes for AROs including rodents, birds, snakes, ungulates, and medium-sized climbing mammals [8]. The relationship between species populations and grid reliability varies across different species, regions, and habitat types, underlining a need for testing this relationship in more diverse environments and species [9]. However, the inaccessibility of species activity data has forced many ARO models to approximate animal activity levels based on life history characteristics from broad taxonomic groups, e.g. squirrels [4, 10, 11].

Studies in cities across the United States have shown that non-squirrel taxa such as birds are also important for describing distribution reliability [1]. Birds were also listed by the National Rural Electric Cooperative Association as a leading cause of overhead distribution power outages in the United States, second to squirrels [8, 12]. ARO models which have relied on a single proxy of animal activity are limited in their flexibility to be applied to diverse systems. They also lose the ability to account for species-specific mechanisms that drive AROs which allow for more meaningful ARO mitigation strategies [13, 14].

Mechanisms driving AROs are derived from environmental conditions and engineering factors that interact with biological species traits [12]. Behavioral and morphological traits determine a species' likelihood of interacting with electrical equipment and how this interaction manifests figure 2.

In the case of birds, gregarious social behavior encourages multiple individuals to gather near power lines, causing the lines to sag or swing and collide as flocks take off at once [8, 12]. Cavity nesting species build nests in electrical equipment and excavate cavities in wooden utility poles. Nesting in equipment increases their chances of collision while flying in and out of the nest, while the nests and cavities themselves damage pole integrity and attract further animals, i.e. predators, to lines [8, 12, 15]. Perch-hunting strategies make utility poles ideal foraging structures for species of raptors, but also increase their chances of electrocution or collision [12].

Additionally, morphology such as wing shape, eye placement, and body size can influence the likelihood of collision in flight and electrocution on lines. Wing loading, or the ratio of body size to wing span, can indicate flight maneuverability, while binocular vision is better at detecting obstacles than peripheral vision



[16–18]. Species with larger body sizes and wing spans are able to connect conductors and grounding parts on electrical equipment, causing short circuits [12].

The risk of an ARO starts with these vulnerable characteristics; i.e. traits that make an animal susceptible to interacting with the grid as well as infrastructure designs that make the grid susceptible to disruptions caused by animals (e.g. a lack of animal deterrents or guards). However, the realized risk (or likelihood) of a power outage only occurs when vulnerable species and power lines become physically exposed to one another. Species distributions and seasonal activity as well as power line placement determine whether vulnerable species and equipment are exposed to each other, creating the risk of an ARO. Environmental conditions such as land use and habitat cover, as well as seasonality determine this degree of exposure. Different habitats and seasons are suitable for different species and this determines the spatio-temporal exposure of species to infrastructure [13, 14, 19] (figure 2). As a result, there is variable ARO risk across space (species distributions) and time (species seasonal activity) which is largely driven by the traits unique to different outage-prone species.

Considering multiple animal activity patterns in animal-outage models may expose a greater understanding of risk interrelationships between ecological and energy domains. Specifically, it may expose species traits that are directly related to outages (e.g. body size, nesting type, or foraging strategy), help delineate ecological patterns such as migration routes and timing that impact distribution lines, and as a result provide more robust outage prevention and mitigation strategies for AROs. Although case studies examining relationships between animals and electrical infrastructure have shown variation across species and geography due to the variables mentioned above [9], existing ARO models have largely been unable to test this multiple animal activity pattern hypothesis due to limited data for diverse species and taxonomic groups.

The few studies that have used species activity data alongside outage records have shown animal activity to be an important variable in characterizing the reliability of the power grid [1], yet these data remain underutilized in animal-outage studies [4, 10, 11]. Joint analyses of power outage and species activity data have been limited to a few species and coarse resolutions due to the resource demanding nature of biological surveys which limit the accessibility of ecological data [1, 7, 11]. We examined available options to fill the ecological data gap in ARO studies and test our multi-pattern hypothesis by using species distribution models derived from eBird data.

Community science data has the potential to improve the availability of species-specific activity estimates and to provide greater flexibility to explore animal-grid interactions in more diverse systems. Community science databases such as eBird [20] are comprised of opportunistic species observations submitted by the general public. eBird captures year-round bird observations at a global scale and has best practice methodologies in place to predict fine resolution estimates of species distributions across geographic space and time [21, 22]. Wide-scale species-specific population databases such as eBird are key to better understanding and mitigating both bird mortality and ARO disruptions to the electrical distribution system.

eBird data is semi-structured, meaning it collects data on the species observation process. This allows data users to standardize opportunistic data that can be subject to several data biases including taxonomic, spatial, temporal, and detection bias [21]. These biases can become increasingly apparent at finer spatial and temporal resolutions, leading to variable and inconsistent species responses [23], but eBird data has also been used to develop species models that far exceed the spatial and temporal limitations of standardized biological surveys, filling knowledge gaps on the full annual cycles and range-wide distributions of bird species [24].

We demonstrate eBird's ability to increase the availability of animal activity for ARO models by using a case study of outage-prone bird species and AROs in the state of MA from 2013–2018. Using eBird data to develop species distribution models (hereafter, SDMs) for 16 outage-prone bird species in MA, we conducted a data-driven exploration of ARO severity as it relates to spatio-temporal patterns of activity in multiple species. Behavioral and morphological traits differed between our study species, influencing their activity patterns throughout seasons, distributions across the landscape, and their likelihood to cause disruptions to the grid (figure 2). eBird data allowed us to visualize these patterns in animal activity alongside ARO severity at weekly, township resolutions, considered the most accurate temporal resolution at which to estimate AROs [2, 10].

Our results suggest that single animal activity patterns representative of the main biological characteristics exhibited by groups of species were important estimates of ARO severity and grid reliability, but diversifying species within these patterns did not improve the strength of their estimates. However, we found that more than one dominant, temporal bird activity pattern had relatively high importance for estimating ARO severity in MA; these being patterns in summer migrants and overwintering resident species. A multi-pattern approach is important for understanding the biological life cycles and traits, which vary across geographies and species, that influence animal-related grid reliability in different regions. Due to this known variation, we recommend future studies use our approach as a guide to explore this relationship further with different species, taxa, regions, and modeling approaches.

We submit that SDMs based on community science databases are an improvement for understanding mechanistic drivers of grid reliability as opposed to using animal activity proxies derived from ARO patterns. Community science data can improve the accessibility of ecological data which can then support future joint analyses of ecological and electric distribution systems.

2. Methods

We chose MA as our case study region because of its available high resolution power outage data (street level and minute resolution) which gave us the flexibility to rescale it to match the resolution of our species data. In addition, the outage data in MA had a high degree of completeness, capturing over 90% of all reported distribution level customer outages in the state.

All analyses were performed in R statistical language version 4.0.2 [25].

2.1. Aligning outage data and species activity estimates

2.1.1. Estimating bird activity

We first identified taxonomic groups of birds known to cause outages due to their behavioral and morphological traits [12]. We select 16 species with these outage prone traits (table 1) and who also have their year round or breeding range within MA (referenced from eBird species range maps). To ensure we had sufficient detection data, we did not consider nocturnal species such as owls because eBird observations are predominantly recorded during daylight hours.

The eBird surveys are semi-structured which allows the data to be standardized and adjusted for common sampling biases. As a result, the data have been used to generate estimates of annual and multi-year trends in species populations consistent with structured bird surveys such as the North American Breeding Bird Survey [29]. Detection data (binary data on whether a species was observed) are more readily available from eBird surveys than individual counts. This makes detection rates (hereafter, DR), the probability of observing a species on a standardized survey, a commonly estimated measure of relative abundance. We use DR as a proxy of activity in birds because animal activity can be an underlying driver of DR, i.e. active birds are easier to detect by observers.

To align spatio-temporal resolutions across species' DRs and outage data, we estimate relative bird activity (DR) for our 16 study species at a weekly township-level resolution. We modeled eBird DRs using a random forest approach that appears in an earlier study comparing the performance of multiple modeling methods for relative abundance [23]. We apply the random forest approach because it provides added flexibility to account for the multiple spatial and temporal variables needed to characterize species occurrences. The random forest methodology was derived from the 'encounter rate' methodology in the Best

Table 1. MA bird species selected for their contributions to animal-caused outages. See [23] for the list of focal species (with the exception of Monk Parakeet (*Myiopsittia monachus*) and American Crow (*Corvus brachyrhynchos*)) with scientific names.

Species (Abbreviation)	Disruptive Traits	Outage Cause	References
Brown-headed Cowbird (BHCO) Common Grackle (COGR) Red-winged Blackbird (RWBL) European Starling (EUST) House Sparrow (HOSP) Mourning Dove (MODO) Monk Parakeet (MOPA)	Social behavior, nesting type, flight type	Phase conductor contact, excrement (contamination), nest contact, predator attraction, collision, electrocution	[8, 16, 26–28]
Downy Woodpecker (DOWO) Hairy Woodpecker (HAWO) Northern Flicker (NOFL) Pileated Woodpecker (PIWO) Red-bellied Woodpecker (RBWO)	Nesting type, foraging type	Pole damage	[8, 15]
Osprey (OSPR) Red-tailed Hawk (RTHA) Turkey Vulture (TUVU) American Crow (AMCO)	Nesting type, foraging type, migratory activity, body size	Collision, electrocution, excrement (streamer), nest contact	[8, 18, 28]

Practices for Using eBird Data [21, 22] which outlines how to estimate species distributions across space and time from the eBird Basic Dataset, and addresses common forms of biases found in the data [30].

The random forest model predicted DR in response to temporal variables (year, day of week, week of the year, and starting time of observations) and survey effort (protocol type and list length). The full data preparation and modeling methodology follows the methods in [23]. Here, we add additional spatial predictors in addition to time and survey effort to improve our estimates of species occurrence across both space and time.

MA township names as well as the proportion of land cover types in each township were also included in the model to account for spatial variation in habitat availability and observer density. Land cover proportions were calculated within each township using 1 km resolution land cover types from the National Land Cover Database [31] aggregated into broader habitat types for birds. The final land cover categories used in the RF model were proportion of barren land, cultivated land, developed land, forest, open water, shrub, wetland. Since topography also influences species distributions, the median and standard deviation in elevation within each township were also used as spatial covariates. Elevation data was downloaded from the EarthEnv project [32].

Our species DR models had fair model performance when looking at AUC and Maximum Kappa, with the exception of Monk parakeet. Monk parakeet also had the lowest model performance (AUC = 0.71 and Max Kappa = 0.39) which led us to drop this species from our analysis due to data limitations (see S3 supplementary material for model performance metrics of all study species). Monk parakeet is a fairly recent invasive species to MA, with introductions to the United States beginning with the pet trade in the late 20th century [33]. At the time of this study, eBird observations of Monk parakeet only began in 2009 and were limited to 89 observations on complete eBird checklists. This left us with very few observations to build our DR models. Our two study species with the next fewest observations were Pileated woodpecker (PIWO) ($n = 15\,290$) and Turkey vulture ($n = 34\,932$) for comparison. For the remainder of our study we focus on 15 study species, removing Monk parakeet from the analysis.

2.1.2. Power outage data

We obtained outage records from three electric distribution utilities (Eversource, UNITIL, and National Grid) in MA through the MA Office of Energy and Environmental Affairs [34]. These three utilities account for the majority share of electricity distribution in the state. The data range from 2013–2018 at a township-level spatial resolution. Fields include the date and time of the outage, township of the outage, and the categorical cause of the outage. The details on our outage data pre-processing appears in the supplementary materials (S1 supplementary materials).

For our analysis, we focused on a subset of outages classified as being caused by an ‘animal’. We refer to these outages as AROs throughout this paper. AROs were first obtained as a subset of electric outages with a recorded outage cause classified as ‘Animal’, ‘Animal-other’, ‘Birds’, or ‘Squirrels.’

Previous explorations of animal activity alongside outages have focused on the frequency of ARO events, which is a direct measurement of animal interference with the grid [1, 2, 7]. However, an important measurement of system reliability in the energy sector is the System Average Interruption Duration Index

(SAIDI). SAIDI is used to quantify the amount of time, on average, customers' electricity is disrupted. We recognize that SAIDI incorporates information that can be influenced by other factors outside of animal interactions such as the repair time required by various means of infrastructure design and maintenance strategies [3, 35, 36], which may interfere with its relationship with animal activity. However, we measure the severity of AROs with SAIDI because it is the IEEE standard metric for quantifying electrical grid reliability [37] and is more relevant to grid managers for the purpose of assessing and understanding electrical system reliability. SAIDI was measured in minutes and is calculated as

$$\text{SAIDI}_{i,\ell} = \frac{D_{i,\ell} C_{i,\ell}}{H_\ell}, \quad (1)$$

where

D_i is the duration of the grid disruption in minutes for the grid disruption event i ;

C_i is the number of customers affected during the grid disruption event i ;

H is the total number of households served in the system a location ℓ .

We calculated SAIDI for each outage event before summing SAIDI across all outage records within each township and week to generate our weekly, township-level aggregated outage dataset.

Our aggregated SAIDI data were positively skewed (mean = 0.99, median = 0.12), so we used the natural log transformation of SAIDI (lnSAIDI) for the remainder of the analysis. We removed outliers of these aggregated lnSAIDI values using the interquartile range (IQR) method. We took 1.5 times the IQR and added this value to the 75th quantile and subtracted it from the 25th quantile. We then removed week-township records with lnSAIDI 1.5 IQR below the 25th quantile and 1.5 IQR above the 75th quantile. There were several instances of townships with abnormally high weekly values of SAIDI, so this process removed 2.6% of our week-township outage records, leaving 10 287 observations for the remainder of the analysis.

2.1.3. Aligning outage severity and bird activity

The MA outage dataset had certain weeks and townships without any reported AROs. ARO reporting can be subjected to false negatives (unreported occurrences) from both reporting procedures and the mobility of the animals themselves. Federal and state agencies require utilities to comply with conservation laws and regulations (e.g. the Migratory Bird Treaty Act and Endangered Species Act) when engaging in activities related to species protected by these laws (such as nest removal and installing animal deterrents) [12]. Additional time, effort, and costs are associated with the actions needed to address any impacts on these protected species by power lines, which may deter ARO reporting. In addition, scavenging of animal carcasses can vary between geographic areas and species, with smaller animal carcasses being removed by scavengers within 24–48 h. Similarly, not all animal-grid interactions result in mortality, resulting in the animal(s) moving away from the scene of the outage before the cause can be reported [12].

Due to the likelihood of these false 0's in our power outage data and lack of capacity to monitor the above mentioned reporting biases, we interpret our results as conditional on an outage having occurred. By using only non-zero values of SAIDI at the weekly-township resolution, our study focuses on the impact of bird activity on the severity of AROs. We merged our weekly, township-level outage and bird datasets, dropping 56 MA townships that did not have any reported AROs within our study period. Out of the 313 weeks from 2013–2018 across the remaining 351 MA townships (109 863 possible observations), we were left with 10 287 weekly-township observations that had both species DRs and SAIDI records (9.3% of all possible week-township observations). The combined dataset consisted of 16 variables, with lnSAIDI being our measure of ARO severity, and 15 species DR estimates that represented the activity of each of our 15 remaining study species.

2.2. Spatio-temporal patterns in outage-prone bird species

We first applied a principal component analysis (PCA) to examine the spatial and temporal patterns in bird activity (DR) across all species. PCA allows discovery of clear subgroups of species with shared spatio-temporal patterns in their weekly-township observations of DR. We used the *prcomp* R-function to run the PCA on all 15 bird species DRs, accounting for complete weekly observations between 2013 and 2018 for all 351 MA townships. The resulting DR principal components (PCs) were linear combinations of DRs that explained the most variation in activity across the MA populations of these species. Bi-plots of these PC loading vectors then group species with distinct spatial and temporal patterns of activity.

We grouped species with similar spatio-temporal variation determined by the first two PCs, i.e. species with positive PC values, species with negative PC values, and species with values close to zero (figure 3), and plotted their state average DRs to visualize distinct temporal activity patterns in our study species. We also looked at their distinct spatial patterns by mapping mean species DR averaged across time for each MA township.

2.3. Accounting for several activity patterns in models of grid reliability

We developed a set of models to test our main hypothesis that multiple animal activity patterns, specifically bird activity patterns, improve our ability to describe changes in outage severity than using single activity patterns or not accounting for bird activity at all in ARO models. Our model set was grouped into three main categories to test this hypothesis: ‘null models’ that only used time and habitat as predictors, ‘single-pattern’ models that used activity patterns of a single species that represented one of four main spatio-temporal trends found within our study species (see section 3.1), and ‘multi-pattern’ models that used multiple activity patterns found within our study species. We estimated multiple versions of outage severity models within each of these three categories, using lnSAIDI as our dependent response variable.

For our predictors, we used species DRs to represent bird activity, the proportion of land cover occupied by distinct habitat types in each township (see section 2.1.2), and temporal variables of month and year as factors. We began with five habitat types and our 15 study species DRs as potential predictors, but removed habitats and species that were highly correlated to minimize collinearity among our predictors. We removed the developed land habitat type because it was highly anti-correlated with forested habitat and retained only one of our three blackbird species (Red-winged blackbird (RWBL)) and three of our five woodpecker species (Red-bellied woodpecker (RBWO), Northern flicker (NOFL), and PIWO) because activity patterns between species within each of these families were highly correlated. This left four of our five habitat predictors and 11 of our 15 bird species which were used for our final SAIDI models.

In our null models, we did not include any measures of bird activity and only relied on changes in habitat, month, and year to account for spatio-temporal trends in SAIDI. We used four habitat types (forest, barren land, open water, and grassland) to represent spatial patterns. We include one model that only includes habitat types, a second model that only accounts for time, and a third that accounts for time and habitat together. Next we added different combinations of bird activity patterns to this time and habitat model to test if bird activity improved model performance.

We estimated three multi-pattern models that used different approaches to incorporate multiple bird activity patterns into SAIDI models. We estimated one model that used individual detection rates of each of our study species ($n = 11$ after removing those with collinearity; hereafter the ‘All Species’ model), a second model that used detection rates from only four species that represented the dominant spatio-temporal activity patterns within our study species (the ‘All Representative Species’ model), and a third model that used the first two PCs of all our study species’ detection rates (the ‘All PC’ model; see section 2.2).

From the All Representative Species model, we created four single-pattern models, each using detection rates from a single species representing a dominant bird activity pattern. These were temporal patterns of overwintering resident species, temporal patterns of summer migrant species, spatial patterns of forest dependent species, and spatial patterns of urban generalist species. Our ‘Resident Species’ model, based on the seasonal activity patterns of species (figure 4(A)), used RBWO detection rates to represent seasonal activity patterns in overwintering resident species and our ‘Migration Species’ model (figure 4(C)) used RWBL to address seasonal activity patterns unique to summer bird species migrants. For these temporal trend representative species, we chose species that also had broad spatial distributions (found fairly ubiquitously across the state) in order to focus on temporal patterns in activity (S2 supplementary materials). Our ‘Forest Species’ model used PIWO activity to describe spatial patterns of forest dependent species (S2 supplementary materials, figure 2), and our ‘Urban Species’ model used European starling activity to address spatial distributions unique to urban specialist species (S2 supplementary materials, figure 4). Similar to our temporal trend species, the spatial trend representative species had weak temporal patterns (activity generally stayed constant throughout the year) to focus on spatial patterns in activity (figure 4(B)).

For each of our models that included bird activity (single- and multi-activity models), we included a model that only accounted for species activity without time and habitat, and two more models that accounted for species, habitat, and time together, one as additive predictors (‘Additive’ models) and another using interactions between activity and habitat and activity and month (‘Interaction’ models). This resulted in a total of 24 models in our model set.

Since our data is structured as weekly observations per MA township, we had replicate observations within each township and year. Spatial relatedness within each township may be due to town specific characteristics associated with ARO severity such as customer population size, number of households, and infrastructure. We also expect observations within years to be influenced by inter-annual changes in variables such as climate which impact timing of periodic ecological events in biological life cycles as well as customer demand on the electrical grid for heating and cooling. Therefore, we considered linear mixed effects (LMM) models with additive random intercepts for townships and year [38]. We used an analysis of variance to test the significance of these random effects by comparing a multiple linear regression and a LMM version of each candidate model. Finding the random effects to be significant for each candidate model, we chose to use LMMs as our final models.

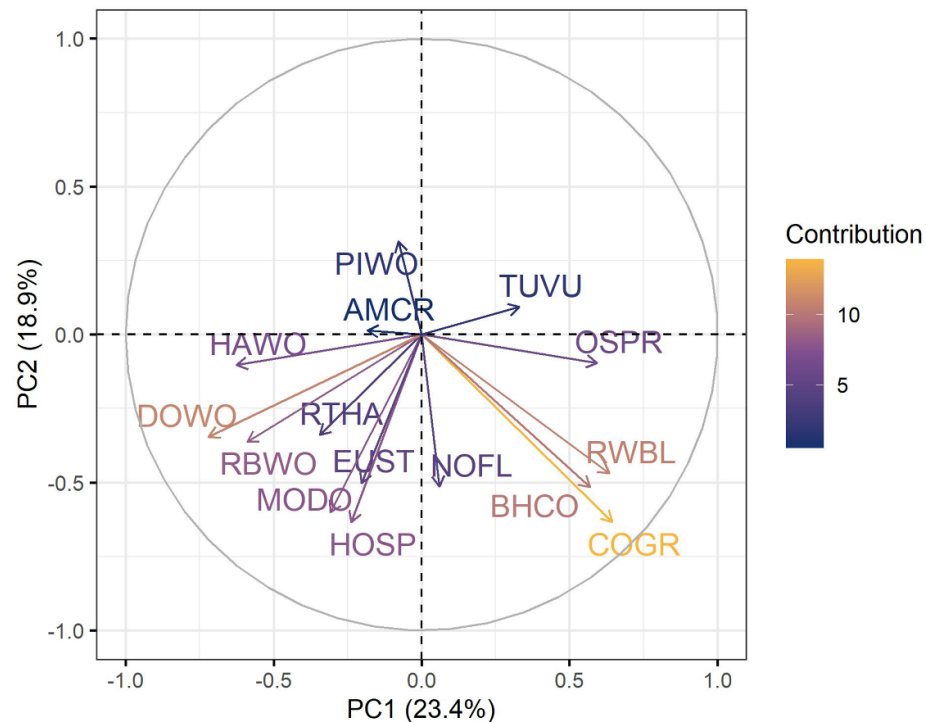


Figure 3. The first two PCs of weekly detection rates across 15 outage-prone bird species in MA townships. Each species is placed along each axis by its loading vector in the first PC and the second PC. These two axes divide species into their unique patterns in activity throughout seasons and spatial distributions across the landscape. The color and length of each vector indicates each species' contribution to the variance in the PCs. Species in the bottom right quadrant are summer migrants limited to urban environments, while the upper left quadrant are resident species occupying more rural areas of the state. The upper right quadrant contains migratory species occupying rural regions and the bottom left quadrant contains resident species concentrated in urban regions.

We used three metrics to evaluate and rank the performance of each LMM in our comparison set using R-package 'Performance' [39]. Akaike's Information Criteria (AIC) weight is used for estimating the relative fit and parsimony between alternative models, and we supplement these indicators which compare model performance relative between models with conditional and marginal R^2 , which describe the variance explained by both the fixed and random effects (conditional) and just the fixed effects (marginal) of each model [40]. The difference between these two indices reflects the variation captured by random effects. Our primary goal was to explore the relationship between the response and features; therefore, discussion based on multiple measures provides a more holistic picture of the limitations in modeling ARO severity in MA.

We expected that specific activity patterns would be important for describing ARO severity in different seasons and habitat types. Since the temporal activity patterns of summer migrant and overwintering resident bird species complement each other throughout the year and habitat preferences determine spatial distributions of species, this may result in different groups of species causing outages at different times of the year and locations within the state. We expected this phenomenon to result in multi-pattern models or more than one single-pattern model to be competitive in our model comparison.

3. Results and findings

3.1. Spatio-temporal patterns in bird activity

The first two PCs of our bird DR PCA explained 40% of the variance across species DRs and highlighted the main spatio-temporal patterns of activity seen in our 15 focal species (figure 3). By visualizing time series and maps of DRs for species grouped by their loading vectors along the first two PCs, we found the first PC largely grouped species based on their seasonal activity patterns, i.e. heightened summer activity in migratory species such as RWBL (figure 4(C)) versus heightened winter activity in resident species such as RBWO (figure 4(A)).

Given the first grouping, the second PC further grouped species based on their spatial distributions, i.e. PIWO which occupies continuous forested areas versus NOFL which can occupy forest patches in more urbanized environments (figure 5). See our supplementary materials for maps of all study species' distributions and land cover types in MA (S2 supplementary materials). Together these PCs indicate four

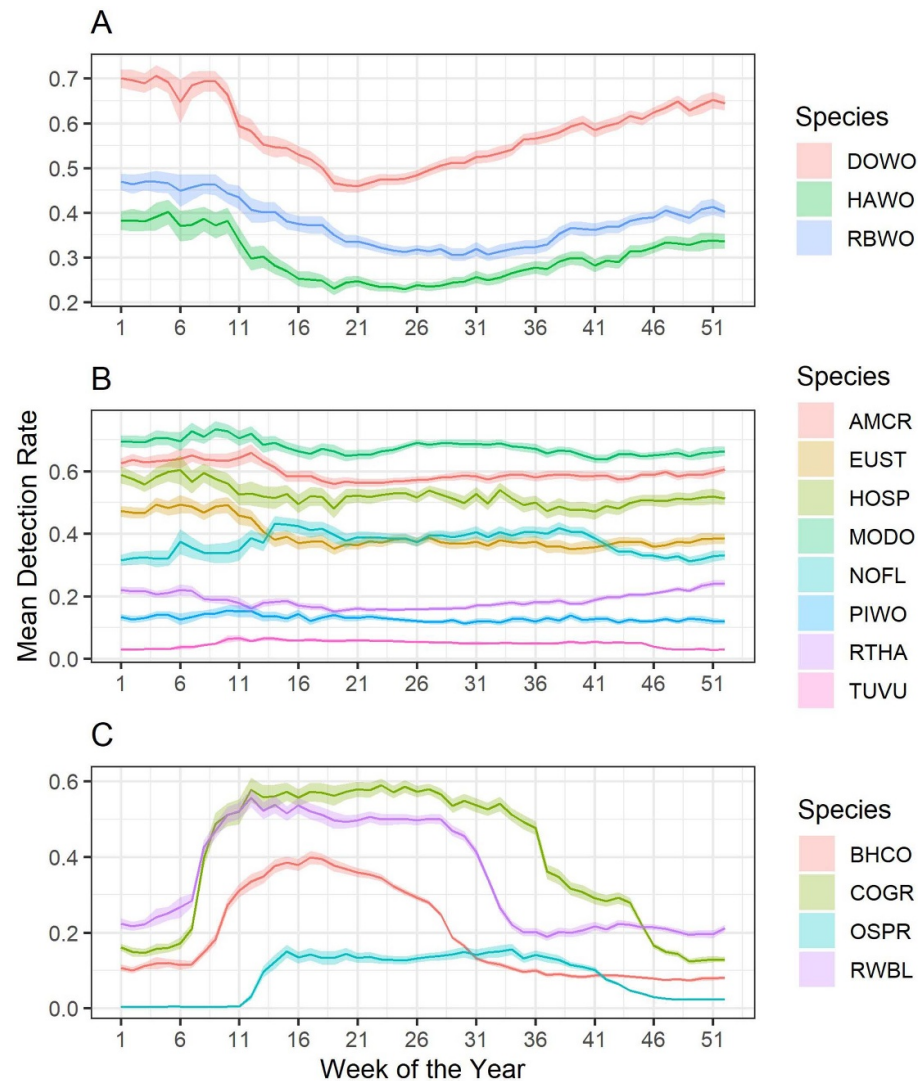


Figure 4. Mean and 95th percentiles of species detection rates, averaged across years and townships, are plotted for each week of the year. Resident species in group (A) have higher winter activity, species in (B) did not express strong seasonal activity patterns, and migratory species in (C) have higher summer activity in MA. These species were grouped by their PC1 loading vector values.

distinct spatio-temporal patterns of bird activity, i.e. rural resident species, urban resident species, rural migratory species, and urban migratory species. Plotting species by their loading vectors along these two PCs grouped our study species into these four distinct activity patterns (figure 3).

3.2. Grid reliability vs species activity patterns

The additive Resident Species model was ranked the highest performing model by AIC weight (0.624, ~62% across all models in the set). In particular, proportion of forest cover and barren land were statistically significant spatial covariates and the activity of overwintering resident birds in addition to the months of June, July, and November were statistically significant temporal covariates, all of which saw positive correlations with SAIDI. However, the additive All PC model also carried some AIC weight (0.105, ~10%), along with the null model using only time and habitat (AIC weight = 0.089, ~9%). Overall we found single-pattern models had better performance than multi-pattern models, with the majority AIC weight attributed to the Resident Species, single-pattern model and the remaining AIC weight split relatively evenly between non-activity and multi-activity models. In addition, additive models of bird activity, habitat, and time had higher AIC weight than models with interaction terms (table 2).

All of our models explained moderate variance in SAIDI, with conditional R^2 ranging from 0.146 to 0.171 [41]. We found that the random variation among townships and years explained the majority of the variation in ARO severity (R^2 conditional greater than R^2 marginal for all models, table 2). Many towns had low sample sizes (72 of 292 towns had ≤ 10 observations) and it is likely the model may overfit in these groups. The 11 models with the highest conditional R^2 values were all within close range of each other (all

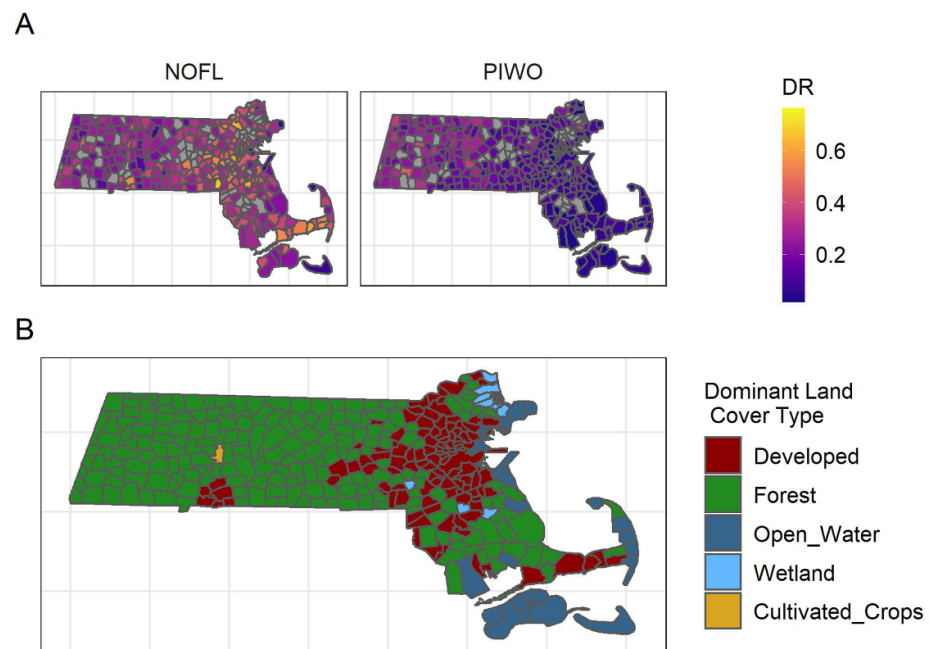


Figure 5. (A) Maps of the average weekly detection rates for PIWO and NOFL in MA townships. These species display two distinct spatial distributions across the state. PIWO are more frequently observed in continuous forested habitat, while NOFL can occupy forest patches in more urbanized environments surrounding the cities of Springfield and Boston. (B) A map of dominant land cover types in each MA township is provided for referencing these spatial patterns in habitat. See our supplementary materials for maps of all 15 study species (S2 supplementary materials).

>0.169 , with less than a 0.002 difference between them). The interaction Migrant Species model had the highest conditional R^2 of these models, followed by the null model with habitat and time. The majority of models with the highest conditional R^2 were single-pattern models with habitat and time. Of these single-pattern models, temporal patterns (Migratory and Resident Species models) had higher conditional R^2 for interaction models, and spatial patterns (Urban and Forest Species models) had higher conditional R^2 for additive models. Multi-pattern models generally explained less variance in their random effect (lower conditional R^2), but exhibited higher relative marginal R^2 (variance from fixed effects only) (table 2). Note, although marginal R^2 may decrease with an increase in the number of fixed parameters, it is unlikely as the variance explained by the fixed effects will increase [40].

4. Discussion

4.1. Improving models of ARO severity using multiple animal activity patterns

Overall, we found that modeling ARO severity using diverse bird species and all of their exhibited activity patterns did not greatly improve model performance over models using low dimensional, single-patterns or only time and habitat cover as predictors of SAIDI. However, the dominant temporal activity patterns belonging to species with different biological characteristics, i.e. migratory and overwintering behavior, were important covariates of ARO severity. Although model performance varied across different model performance metrics, SAIDI models using single bird activity patterns had the highest AIC weight, which could have been due to their greater parsimony, and had a relatively high conditional R^2 compared to models that included multiple bird activity patterns from diverse species covariates.

Outage-prone bird species in MA either had little variation in activity across the state or over time, or exhibited a combination of four main spatio-temporal patterns (summer peaks in migrant species activity, winter peaks in overwintering residents, concentrations of urban generalist species activity in developed areas, and concentrations of forest specialists in rural areas) resulting from distinct behavioral traits and life histories. Species that had similar traits also shared complementary activity patterns. Migratory species such as Osprey (OSPR) and RWBL had the strongest seasonal fluctuations in activity and are nearly inactive in the state during the winter (figure 4(C)), while late spring into fall had the highest levels of activity for these species during breeding and migration. Resident woodpecker species have more stable, year-long activity and increase in activity during the winter likely due to northern populations moving south and overwintering in the state (figure 4(A)). Species such as PIWO, that occupy large, intact mature forests are distributed across

Table 2. Model set comparing whether including multiple bird activity patterns improved models of SAIDI. The model set was broken into three hypotheses: 1. Non-activity model with habitat and time (null model), 2. the null model with a single bird activity pattern (Single-Pattern Model), and 3. the null model with multiple bird activity patterns (Multi-Pattern Models). The multi-pattern models used all study species detection rates (All Species), the PCs of all study species detection rates (All PC), and a model that uses one representative species from four dominant spatio-temporal trends in bird activity (All Representative Species). The single-pattern models used only one of these four representative species at a time. Model performance is compared with conditional R^2 , marginal R^2 , AIC weight (AIC; measure of model parsimony), and the number of fixed effects included in each model. Models are ordered by AIC weight.

Model hypothesis	$\ln(\text{SAIDI}) \sim \text{model}$	R^2 conditional	R^2 marginal	AIC weight	Fixed effects
Single-Pattern Model	Resident Species, Habitat, and Time (Additive)	0.1693	0.0551	0.6247	17
Multi-Pattern Model	All PC, Habitat, and Time (Additive)	0.1695	0.0549	0.1057	18
Null Model (No Bird Activity)	Time and Habitat	0.1701	0.0542	0.0895	16
Multi-Pattern Model	All Representative Species, Habitat, and Time (Additive)	0.1683	0.0548	0.0492	20
Single-Pattern Model	Migrant Species, Habitat, and Time (Additive)	0.1696	0.0543	0.0465	17
Single-Pattern Model	Forest Species, Habitat, and Time (Additive)	0.1696	0.0540	0.0385	17
Single-Pattern Model	Urban Species, Habitat, and Time (Additive)	0.1698	0.0541	0.0362	17
Single-Pattern Model	Migrant Species, Habitat, and Time (Interaction)	0.1709	0.0557	0.0074	29
Multi-Pattern Model	All Species, Habitat, and Time (Additive)	0.1694	0.0572	0.0015	27
Single-Pattern Model	Resident Species, Habitat, and Time (Interaction)	0.1700	0.0559	0.0005	29
Single-Pattern Model	Forest Species, Habitat, and Time (Interaction)	0.1648	0.0523	0.0002	29
Single-Pattern Model	Urban Species, Habitat, and Time (Interaction)	0.1697	0.0550	0.0001	29
Multi-Pattern Model	All Species, Habitat, and Time (Interaction)	0.1692	0.0665	0.0000	159
Null Model (No Bird Activity)	Time	0.1688	0.0113	0.0000	12
Multi-Pattern Model	All PC, Habitat, and Time (Interaction)	0.1673	0.0554	0.0000	42
Multi-Pattern Model	All Representative Species, Habitat, and Time (Interaction)	0.1652	0.0567	0.0000	68
Multi-Pattern Model	All PC	0.1601	0.0051	0.0000	3
Single-Pattern Model	Resident Species	0.1602	0.0008	0.0000	2
Multi-Pattern Model	All Species	0.1565	0.0215	0.0000	12
Single-Pattern Model	Migrant Species	0.1576	0.0026	0.0000	2
Null Model (No Bird Activity)	Habitat	0.1583	0.0430	0.0000	5
Single-Pattern Model	Urban Species	0.1557	0.0043	0.0000	2
Multi-Pattern Model	All Representative Species	0.1495	0.0105	0.0000	5
Single-Pattern Model	Forest Species	0.1463	0.0043	0.0000	2

forested areas in rural sections of MA (figure 5). In contrast, opportunistic species such as the NOFL can take advantage of smaller forest patches within the suburbs of urbanized areas in the state.

The redundancy of these four main species activity patterns within our study species aligns with our model results. Using one representative species from each of these patterns was sufficient for estimating ARO severity and adding species that exhibited the same patterns did not add additional information. We found including migratory bird activity with habitat and time explained the greatest variance in SAIDI

(conditional R^2) and including resident bird activity with habitat and time was ranked the top performing model for AIC. The importance of both these models could also be indicative that different groups of species contribute more to outages at different times of the year, depending on their life history traits and timing for biological cycles such as migration and breeding. In this case, considering more than one animal activity pattern (as opposed to simply diversifying species) could still be important to account for the effects of biological patterns and mechanics on ARO severity.

4.2. Data considerations for future studies of AROs

Despite diverse bird activity showing less importance for ARO severity compared to single activity patterns or simply measures of habitat and time, this finding could be specific to our case study location and species. One explanation could be that birds contribute less to outages in MA compared to other outage-prone taxa such as rodents or small mammals. Out of our 15 062 ARO records, only 30 records were specifically recorded as attributed to birds. Our case study also used only one taxonomic group in a relatively small geographic area and animal-outage interactions are known to vary greatly across geographic locations depending on local species communities and the engineering design of distribution lines [8]. For instance, bird species played a significant role in modeling outages in Arizona [1], while squirrel biology was used to reliably model AROs in Kansas [2]. Better data are needed to begin incorporating activity for more diverse taxa into analyses of grid reliability, especially for small mammals. Existing species occurrence databases for these taxa such as i-Naturalist and the Global Biodiversity Information Facility (GBIF) [42] do not compile complete survey data. As a result, they cannot account for sampling bias and true absences, a benefit that allows eBird to produce more accurate SDMs.

Animal activity is not the only factor that drives animal-infrastructure interactions. In some cases it is the increasing susceptibility of existing populations that plays a greater role in determining grid interactions. Stochastic variables such as weather events can further drive bird-infrastructure interactions by increasing bird susceptibility to line collisions, while data on infrastructure composition can improve knowledge of species-infrastructure overlap and exposure. In these cases, the spatio-temporal distribution of species may be important for explaining animal activity, while species distributions in conjunction with stochastic environmental data such as extreme weather events and information on infrastructure may add further understanding to the relationship between animals and electrical grid reliability (figure 2). Therefore, the use of animal activity as an indicator of grid reliability could be strengthened by including additional environmental and infrastructure data alongside estimates of species activity.

A further data limitation in our study was the inconsistent reporting in the outage data we obtained from utility companies. The reported ARO causes in our data consisted of mostly non-descriptive classifications, i.e. ‘animal’ or ‘animal-other,’ with taxon specific causes identifying between birds and squirrels limited to a few hundred records. In many cases AROs can be difficult to report, with animals moving away from the site after causing an outage, or predators carrying away carcasses [12]. Legal processes can also deter reporting of species protected by conservation laws. As a result, accurate ARO data and their specific causes can be difficult to extract from reported outage records. In order to improve our understanding of AROs, there continues to be a need for better data in both the ecology and energy sectors.

5. Conclusions

We highly recommend further collaborations between energy experts and ecologists to study the importance of diverse animal activity patterns in estimating grid reliability, as there is currently a lack of research on this topic. Future studies could use our framework applying community science data to supplement sparse ecological data for studying AROs in other geographic locations. Additionally, future work should apply activity patterns of non-bird taxa to ARO models as standardized observation data and fine scale species distribution models for these other taxa become more widely available. Furthermore, collaboration may lead to development of better modeling strategies. We explored many linear mixed models, but a mechanistic or non-linear model may better explain the patterns in ARO severity. Additionally, improvements in data collection and study design may be explored.

Although further studies with different taxa, species, and geographic locations should be considered to validate our findings, our conclusion that multiple dominant animal activity patterns were important predictors of grid reliability indicates that more than one life history characteristic should be considered in studies of AROs. However, obtaining activity data for multiple species may not be necessary as long as these dominant patterns are captured. Making the best available resources from quantitative ecology known to outside disciplines can allow for more successful interdisciplinary data analyses between ecological and non-ecological disciplines. Species distribution modeling is used by ecologists to estimate the distribution of species across continuous spatio-temporal resolutions from observational data collected by limited biological

surveys. We suggest that SDMs in combination with more widely available community science databases can be used to improve the availability of animal activity estimates in models of AROs and allow these models to be applied to more diverse regions.

Data availability statement

The data that support the findings of this study are openly available at the following URL/DOI: <https://doi.org/10.5281/zenodo.6550833> [43].

Acknowledgments

This work was made possible by the Massachusetts Office of Energy and Environmental Affairs [34], through which we obtained outage datasets for Eversource Energy, National Grid, and Unitil Corporation as used in this work, as well as the Cornell Lab of Ornithology for their open source bird data [20]. We thank the volunteers and community scientists who contributed to the eBird database as well as the eBird project team. Funding for this research was provided by the NSF Harnessing the Data Revolution (HDR) program (Award Numbers 1940276, 1940176, 1940160, 2023755).

Data accessibility

The R-code used in the analysis of this study are available on Github at https://github.com/mefeng7/Bird_outages_MA_Script. The version of the analysis code at the time of publication is archived on Zenodo (<https://doi.org/10.5281/zenodo.6550833>). The accompanying data and outputs with this analysis code are available on Figshare (https://figshare.com/articles/dataset/Bird_outages_MA_data/19602142).

The compiled power outage data and estimated detection rates for the state of Massachusetts are also publicly available in Columbia University's International Research Institute for Climate and Society Data Library [44].

Links to these Data Library datasets: <http://iridl.ldeo.columbia.edu/SOURCES/.PRISM/.eBird/.derived/.detectionProbability/> <http://iridl.ldeo.columbia.edu/SOURCES/.EOEEA/>.

Recommended citation for the Data Library datasets: Resnick, D, Cousin, R, Kaplan, A, Chourio, X M, Vadillo, A, Hall, K, Graff, K, and Khomyakov, I (2021), Suite of Tools to Disseminate Information Derived from Multiple Domains Science Results, [IN54B-06] presented at 2021 Fall Meeting, AGU, 13–17 December.

Conflict of interest

The authors of this manuscript have no conflicts of interest to declare.

Authors contribution

All authors contributed to the study conception and design; M-L E F and O O O collected the data; M-L E F, O O O, T L J S, and S S analyzed the data; M-L E F and O O O led the writing of the manuscript. All authors reviewed, contributed critically, and approved the final manuscript.

ORCID iDs

Mei-Ling E Feng  <https://orcid.org/0000-0001-5979-240X>
Olukunle O Owolabi  <https://orcid.org/0000-0002-4379-5718>
Toryn L J Schafer  <https://orcid.org/0000-0001-5594-7697>
David S Matteson  <https://orcid.org/0000-0002-2674-0387>
Judy P Che-Castaldo  <https://orcid.org/0000-0002-9118-9202>
Deborah A Sunter  <https://orcid.org/0000-0003-2024-9543>

References

- [1] Maliszewski P J, Larson E K and Perrings C 2012 Environmental determinants of unscheduled residential outages in the electrical power distribution of Phoenix, Arizona *Reliab. Eng. Syst. Saf.* **99** 161–71
- [2] Gui M 2009 Advanced methods for prediction of animal-related outages in overhead distribution systems *PhD Thesis* Kansas State University
- [3] Sekhar P C, Deshpande R A and Sankar V 2016 Evaluation and improvement of reliability indices of electrical power distribution system *2016 National Power Systems Conf. (NPSC)* pp 1–6

- [4] Gui M, Pahwa A and Das S 2011 Bayesian network model with Monte Carlo simulations for analysis of animal-related outages in overhead distribution systems *IEEE Trans. Power Syst.* **26** 1618–24
- [5] Che-Castaldo J P *et al* 2021 Critical Risk Indicators (CRIs) for the electric power grid: a survey and discussion of interconnected effects *Environ. Syst. Decis.* **41** 594–615
- [6] Commonwealth of Massachusetts 2020 Historic power outages (available at: www.mass.gov/info-details/power-outages#historic-power-outages-)
- [7] Doostan M and Chowdhury B 2019 Statistical analysis of animal-related outages in power distribution systems—a case study 2019 *IEEE Power & Energy Society General Meeting (PESGM)* (IEEE) pp 1–5
- [8] Frazier S D and Bonham C 1996 Suggested practices for reducing animal-caused outages *IEEE Ind. Appl. Mag.* **2** 25–31
- [9] Loss S R, Will T and Marra P P 2014 Refining estimates of bird collision and electrocution mortality at power lines in the United States *PLoS One* **9** e101565
- [10] Das S, Kankanala P and Pahwa A 2021 Outage estimation in electric power distribution systems using a neural network ensemble *Energies* **14** 4797
- [11] Sahai S and Pahwa A 2006 A probabilistic approach for animal-caused outages in overhead distribution systems *Int. Conf. on Probabilistic Methods Applied to Power Systems, 2006* pp 1–7
- [12] Avian Power Line Interaction Committee (APLIC) 2006 Suggested practices for avian protection on power lines: the state of the art in 2006 Washington, DC; Sacramento, CA: Edison Electric Institute, APLIC; California Energy Commission
- [13] Burgio K, Rubega M and Sustaita D 2014 Nest-building behavior of Monk Parakeets and insights into potential mechanisms for reducing damage to utility poles *PeerJ* **2** e601
- [14] Burnham J *et al* 2004 Preventive measures to reduce bird-related power outages-part I: electrocution and collision *IEEE Trans. Power Deliv.* **19** 1843–7
- [15] Polat A, Yumak K, Atilla N and Baayriyanik M 2016 An overview of bird related issues in electrical power systems *IOP Conf. Ser.: Mater. Sci. Eng.* **161** 012091
- [16] D'Amico M, Martins R C, Álvarez-Martínez J M, Porto M, Barrientos R and Moreira F 2019 Bird collisions with power lines: prioritizing species and areas by estimating potential population-level impacts *Diver. Distribut.* **25** 975–82
- [17] Bernardino J, Bevanger K, Barrientos R, Dwyer J F, Marques A T, Martins R C, Shaw J M, Silva J P and Moreira F 2018 Bird collisions with power lines: state of the art and priority areas for research *Biol. Conserv.* **222** 1–13
- [18] Martin G and Shaw J 2010 Bird collisions with power lines: failing to see the way ahead? *Biol. Conserv.* **143** 2695–702
- [19] Rollan A, Real J, Bosch R, Tintà A and Hernandez-Matias A 2010 Modeling the risk of collision with power lines in Bonelli's Eagle *Hieraetus fasciatus* and its conservation implications *Bird Conserv. Int.* **20** 279–94
- [20] Sullivan B L, Wood C L, Liff M J, Bonney R E, Fink D and Kelling S 2009 eBird: a citizen-based bird observation network in the biological sciences *Biol. Conserv.* **142** 2282–92
- [21] Strimas-Mackey M, Hochachka W M, Ruiz-Gutierrez V, Robinson O J, Miller E T, Auer T, Kelling S, Fink D and Johnston A 2020 Best practices for using eBird data *Version 1.0*. (Ithaca, NY: Cornell Lab of Ornithology)
- [22] Johnston A, Hochachka W, Strimas-Mackey M, Ruiz Gutierrez V, Robinson O, Miller E, Auer T, Kelling S and Fink D 2020 Analytical guidelines to increase the value of citizen science data: using eBird data to estimate species occurrence (Cold Spring Harbor Laboratory) *bioRxiv Preprint* (available at: www.biorxiv.org/content/early/2020/06/23/574392.full.pdf)
- [23] Feng M-L E and Che-Castaldo J 2021 Comparing the reliability of relative bird abundance indices from standardized surveys and community science data at finer resolutions *PLoS One* **16** e0257226
- [24] Fink D, Auer T, Johnston A, Ruiz-Gutierrez V, Hochachka W M and Kelling S 2020 Modeling avian full annual cycle distribution and population trends with citizen science data *Ecol. Appl.* **30** e02056
- [25] R Core Team 2020 R: a language and environment for statistical computing (Vienna: R Foundation for Statistical Computing) (available at: www.R-project.org/)
- [26] Alonso J A and Alonso J C 1999 Collision of birds with overhead transmission lines in Spain *Birds and Power Lines: Collision, Electrocution and Breeding* (Madrid: Quercus)
- [27] Pettersson J 2005 The impact of offshore wind farms on bird life in Southern Kalmar Sound Sweden: final report based on studies 1999–2003 *Report to the Swedish Energy Agency, Lund: Department Animal Ecology, Lund University* OCLC: 500542969
- [28] Sundararajan R *et al* 2004 Preventive measures to reduce bird related power outages-part II: streamers and contamination *IEEE Trans. Power Deliv.* **19** 1848–53
- [29] Walker J and Taylor P 2017 Using eBird data to model population change of migratory bird species *Avian Conserv. Ecol.* **12** 4
- [30] eBird, Cornell Lab of Ornithology 2020 eBird basic dataset Version: EBD_relmay-2020
- [31] Homer C *et al* 2020 Conterminous united states land cover change patterns 2001–2016 from the 2016 national land cover database *ISPRS J. Photogramm. Remote Sens.* **162** 184–99
- [32] Amatulli G, Domisch S, Tuanmu M-N, Parmentier B, Ranipeta A, Malczyk J and Jetz W 2018 A suite of global, cross-scale topographic variables for environmental and biodiversity modeling *Sci. Data* **5** 180040
- [33] Russello M A, Avery M L and Wright T F 2008 Genetic evidence links invasive monk parakeet populations in the United States to the international pet trade *BMC Evol. Biol.* **8** 217
- [34] Department of Public Utilities: Energy and Environmental Affairs 2013 Fileroom (available at: <https://eeonline.eea.state.ma.us/DPU/Fileroom/dockets/bynumber>) (Accessed 4 May 2020)
- [35] Heidari A, Agelidis V G and Naderi M S 2013 Effects of switch type and location on the reliability of power distribution systems considering distributed generation 2013 *IEEE Int. Conf. on Industrial Technology (ICIT)* pp 710–5
- [36] Hanser F M 2015 Control strategies for demand-side management in smart grids part 1: assessment methodology based on reliability 2015 *IEEE Innovative Smart Grid Technologies - Asia (ISGT ASIA)* pp 1–6
- [37] Warren C A and Saint R 2005 IEEE reliability indices standards *IEEE Ind. Appl. Mag.* **11** 16–22
- [38] Hurlbert S 1984 Pseudoreplication and the design of ecological field experiments *Ecol. Monogr.* **54** 187–211
- [39] Lüdtke D, Ben-Shachar M S, Patil I, Waggoner P and Makowski D 2021 performance: an R package for assessment, comparison and testing of statistical models *J. Open Source Softw.* **6** 3139
- [40] Nakagawa S and Schielzeth H 2013 A general and simple method for obtaining r^2 from generalized linear mixed-effects models *Methods Ecol. Evol.* **4** 133–42
- [41] Cohen J 1988 *Statistical Power Analysis for the Behavioral Sciences* 2nd ed. (New York: Lawrence Erlbaum Associates) (<https://doi.org/10.4324/9780203771587>)
- [42] GBIF: the global biodiversity information facility 2021 what is GBIF? (available at: www.gbif.org/what-is-gbif) (Accessed 29 November 2021)

- [43] Feng M-L *et al* 2022 Analysis of animal-related electric outages using species distribution models and community science data *Zenodo* (available at: <https://zenodo.org/record/6550833#.Ys4XE3bMK3B>)
- [44] Blumenthal M B, Bell M, del Corral J, Cousin R and Khomyakov I 2014 IRI Data Library: enhancing accessibility of climate knowledge *Earth Perspec.* **1** 19