Journal of Applied Ecology

Journal of Applied Ecology 2017, **54**, 618–627 doi: 10.1111/1365-2664.12723

Combining citizen science species distribution models and stable isotopes reveals migratory connectivity in the secretive Virginia rail

Auriel M. V. Fournier[1]†, Alexis R. Sullivan[2], Joseph K. Bump[1]*, Marie Perkins[3]‡, Mark C. Shieldcastle[4] and Sammy L. King[5]

[1]School of Forest Resources and Environmental Sciences, Michigan Technological University, Houghton, MI 49931, USA; [2]Department of Ecology and Environmental Science, Umeå Plant Science Centre, Umeå University, Umeå SE-901 87, Sweden; [3]Louisiana State University AgCenter, Baton Rouge, LA 79803, USA; [4]Black Swamp Bird Observatory, Oak Harbor, OH 43449, USA; and [5]U.S. Geological Survey, Louisiana Cooperative Fish and Wildlife Research Unit, School of Renewable Natural Resources, Louisiana State University, Baton Rouge, LA 70816, USA

Summary

1. Stable hydrogen isotope (δD) methods for tracking animal movement are widely used yet often produce low resolution assignments. Incorporating prior knowledge of abundance, distribution or movement patterns can ameliorate this limitation, but data are lacking for most species. We demonstrate how observations reported by citizen scientists can be used to develop robust estimates of species distributions and to constrain δD assignments.

2. We developed a Bayesian framework to refine isotopic estimates of migrant animal origins conditional on species distribution models constructed from citizen scientist observations. To illustrate this approach, we analysed the migratory connectivity of the Virginia rail *Rallus limicola*, a secretive and declining migratory game bird in North America.

3. Citizen science observations enabled both estimation of sampling bias and construction of bias-corrected species distribution models. Conditioning δD assignments on these species distribution models yielded comparably high-resolution assignments.

4. Most Virginia rails wintering across five Gulf Coast sites spent the previous summer near the Great Lakes, although a considerable minority originated from the Chesapeake Bay watershed or Prairie Pothole region of North Dakota. Conversely, the majority of migrating Virginia rails from a site in the Great Lakes most likely spent the previous winter on the Gulf Coast between Texas and Louisiana.

5. *Synthesis and applications*. In this analysis, Virginia rail migratory connectivity does not fully correspond to the administrative flyways used to manage migratory birds. This example demonstrates that with the increasing availability of citizen science data to create species distribution models, our framework can produce high-resolution estimates of migratory connectivity for many animals, including cryptic species. Empirical evidence of links between seasonal habitats will help enable effective habitat management, hunting quotas and population monitoring and also highlight critical knowledge gaps.

Key-words: Bayesian, citizen science, eBird, feathers, hydrogen isotopes, migration, migratory connectivity, species distribution model, Virginia rail *Rallus limicola*, δD animal origins

Introduction

Understanding migration and dispersal is essential for effective conservation and management. For most organisms, however, these movements are poorly understood due to the size, expense or low return of extrinsic markers. Stable hydrogen isotope ratios (^{2}H:^{1}H, hereafter δD)

---

†Current Address: Arkansas Cooperative Fish and Wildlife Research Unit, Department of Biological Sciences, University of Arkansas, Fayetteville, AR 72701, USA

‡Current Address: Department of Natural Resource Sciences, McGill University, Ste. Anne de Bellevue, QC H9X 3V9, Canada

The first two authors contributed equally to this work.

*Correspondence author. E-mail: jkbump@mtu.edu

of biological tissues (e.g. teeth, hair and feathers) are now widely used to infer animal movements because they require only a single capture event (e.g. Hobson *et al.* 2009; Sullivan *et al.* 2012). This technique exploits the predictable continental variation of δD in precipitation (Bowen, Wassenaar & Hobson 2005), which are transferred up the food chain and leave an isotopic 'signature' that reflects the δD sources ingested during tissue formation (reviewed in Bowen 2010). While δD-based methods are highly tractable, resulting assignments are generally broad because large geographic regions share similar $\delta D_{\mathrm{precipitation}}$ values (Bowen, Wassenaar & Hobson 2005; Bowen 2010).

The resolution of δD-based geographic assignments can be markedly improved by incorporating prior sources of information, such as breeding abundances (Hallworth *et al.* 2013), spatial distribution (Pekarsky *et al.* 2015) or patterns of band recovery data (Hobson *et al.* 2009; Van Wilgenburg & Hobson 2011) into the analyses. However, systematic studies of distribution, abundance and band recovery at an appropriate scale are likely to be lacking for most organisms (e.g. Sullivan *et al.* 2012). For these animals, the abundance of opportunistic observations from citizen scientists for even cryptic organisms can be potentially used to refine δD-based assignments.

Repositories for citizen science observations are increasing in size and scope (e.g. eBird, Sullivan *et al.* 2014; eMammal, Forrester *et al.* 2013; eButterfly, Larrivee *et al.* 2014; Journey North, Miller *et al.* 2011). Although most observations are presently from Europe or North America, global participation is likely to increase with increasing access to technology and outreach from citizen science organizations (Sullivan *et al.* 2014). eBird is presently the largest citizen science repository, with over 275 million observations and growing (Sullivan *et al.* 2014). Although data from eBird and other data bases are not without limitations, they can be deployed in presence-only species distribution models (SDMs; Elith *et al.* 2006; Phillips, Anderson & Schapire 2006; Elith & Leathwick 2009) to develop informative priors for isotope-based geographic assignments.

Species distribution models relate georeferenced occurrence records to a set of environmental covariates to extrapolate the potential distribution of a species to unsampled sites across a landscape (Elith & Leathwick 2009). Although many other techniques exist, the maximum entropy method of species distribution modelling (MaxEnt) is used because of its application of presence-only data and predictive accuracy (Elith *et al.* 2006; Phillips, Anderson & Schapire 2006; Phillips & Dudík 2008). MaxEnt assumes species occupy as close to a uniform distribution across space as possible (i.e. the maximum entropy distribution), subject to environmental constraints (Phillips, Anderson & Schapire 2006). As with other SDM approaches, MaxEnt assumes presence observations are randomly distributed across environmental space, an assumption that requires careful consideration and testing before applying citizen science data (Yackulic *et al.* 2013).

We develop a framework for generating spatially explicit predictions of the origins of migratory animals using combined analysis of SDMs constructed from citizen science observations and stable isotope analyses. Using this method, we determined winter and summer origins of Virginia rails *Rallus limicola*, a declining migratory and secretive marsh bird in North America (Conway 1995). Virginia rails are too small to carry affordable satellite transmitters, and their low site fidelity makes geolocators inefficient (Haramis & Kearns 2007). Band recoveries have also yielded little information: of the 5402 Virginia rails banded in North America in the last 50 years, only six have been recaptured (U.S. Geological Survey Bird Banding Laboratory, https://www.pwrc.usgs.gov/BBL/, accessed 1 April 2016). Due to this paucity of data, Virginia rail populations are managed on the untested assumption that they migrate in a similar way to waterfowl, with whom they often share habitat.

Although Virginia rail are difficult to detect, the abundance of citizen science occurrence records allowed construction of robust SDMs for breeding and wintering habitat. Utilizing citizen science observations from other organisms allowed estimation of sampling intensity and correction for unequal sampling effort across the landscape. Bayesian analysis of SDMs and δD data yielded high-resolution assignments previously possible for only conspicuous or well-studied taxa (e.g. Hobson *et al.* 2009; Van Wilgenburg & Hobson 2011). Our results indicate many Virginia rails do not migrate within their administrative management units, which could be revised by extending and explicitly testing the geographic assignments produced by our method.

Materials and methods

We analysed the δD of feathers collected from wintering Virginia rails on the Louisiana and Texas coastal plains and from a spring migration stopover site in the Lake Erie marshes of Ohio, USA (Fig. 1; Table 1). Probable origins of migrants were determined by (i) construction of season-specific maximum entropy SDMs representing the prior probability of Virginia rail occurrence; (ii) calculation of spatially explicit, δD likelihood functions for each individual; and (iii) application of Bayes' rule to incorporate the two probability surfaces at the population level.

SAMPLE COLLECTION

We captured spring migrant Virginia rails at Ottawa National Wildlife Refuge (NWR) in Ohio, USA, and wintering rails at five public wetlands in Louisiana and Texas, USA (Fig. 1; Table 1). Contour feathers are moulted in the winter and incorporate the δD signature of the wintering locality, while primary feathers contain δD signature reflecting the breeding area (Conway 1995). In rare cases, a second primary feather moult may occur on the wintering grounds (Conway 1995).

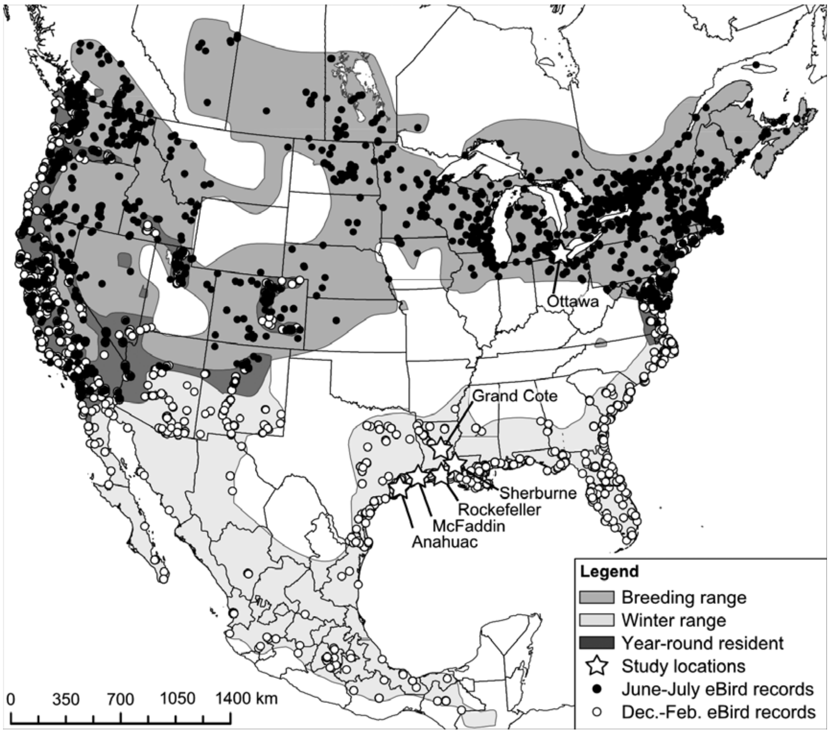

Fig. 1. Breeding, wintering and resident range (shaded areas) of the Virginia rail *Rallus limicola*, a North American marsh bird. Dots represent citizen scientist observations 2002–2012 downloaded from eBird, a global repository for observation records (Sullivan *et al.* 2014). Study locations are denoted by stars. Filled dots (n = 1800) are summer observations; unfilled dots (n = 1832, eBird and n = 15, Global Biodiversity Information Facility) are winter observations.

Table 1. Summary of sampling sites and periods where Virginia Rails *Rallus limicola* were captured and feather samples were taken

Time of year	Site	Sampling period	*N*
Migration	Ottawa NWR, Ohio, USA	15 March 2005 to 31 May 2009	41
Wintering	Anahuac NWR, Texas, USA	February to March 2005	3
		November to March 2005–2006	6
Wintering	McFaddin NWR, Texas, USA	February to March 2005	5
		November to March 2005–2006	13
Wintering	Grand Cote NWR, Louisiana, USA	October to November 2005	26
Wintering	Rockefeller Wildlife Refuge Louisiana, USA	October to March 2004–2005	25
		February to March 2006	9
Wintering	Sherburne Wildlife Management Area, Louisiana, USA	November to January 2005–2006	8

N, number of Virginia Rails.

Accordingly, we collected contour feathers from spring migrants to determine their winter provenance and primary feathers from wintering rails to determine their previous breeding location.

Spring migrants at Ottawa NWR were captured using live traps by Black Swamp Bird Observatory in 2005 and 2009 (Fournier *et al.* 2015). Each bird received a U.S. Fish and Wildlife Service band (permit #20929). Wintering rails were captured by hand or with a dip net from an all-terrain vehicle or airboat (Perkins 2007; Perkins, King & Linscombe 2010). Louisiana State University IACUC approved capture of rails on the wintering grounds (AE04-07), and they were banded under U.S. Fish and Wildlife Service permit #22207.

SPECIES DISTRIBUTION MODELLING

Twenty-one raster layers (1 km^2 resolution) representing land cover (Latifovic *et al.* 2002), wetland presence (Lehner & Döll 2004) and bioclimatic parameters (Hijmans *et al.* 2005; See Table S1, Supporting Information) across North America were used as environmental variables in model training. We constrained SDM predictions to the combined summer and year-round range or winter and year-round range, respectively (BirdLife International & NatureServe 2013), which we expanded to encompass 95% of the verified eBird observations to produce a broader – and more conservative for provenance assignment – assessment of potential Virginia rail distribution.

We downloaded all georeferenced observations of Virginia rails from eBird reported for June and July (n = 4476) and December, January and February (n = 8302) from 2002 to 2012. We also used locality information for Mexico reported in the Global Biodiversity Information Facility (http://data.gbif.org; n = 15 from GBIF; n = 158 from eBird for Mexico). After discarding multiple sightings within 1 km^2, 1800 summer observations and 1832 winter observations remained for SDM construction.

We accounted for sampling bias in the Virginia rail observations using two background point (or 'pseudo-absence') manipulation approaches: (i) the sampling bias grid, which provides an estimate of sampling intensity across the landscape (Elith, Kearney & Phillips 2010; Kramer-Schadt *et al.* 2013; Syfert, Smith & Coomes 2013; Fourcade *et al.* 2014); and (ii) targeted group sampling, which supplies designated background points selected from presence observations of related species collected using similar methods and with the same spatial bias as the focal species (Dudík, Schapire & Phillips 2005; Phillips *et al.* 2009). MaxEnt minimizes the relative entropy between the biased presence and background points to determine environmental constraints (Dudík, Schapire & Phillips 2005; Phillips *et al.* 2009). In other words, if background and presence points share the same bias, then MaxEnt can identify ecological variables that differ between the two, rather than highlighting more heavily sampled regions (Phillips *et al.* 2009).

To estimate sampling intensity across summer and winter ranges, we downloaded the coordinates of all verified complete checklists reported to eBird within the same time period as the Virginia rail data, which yielded 504 023 coordinates for the summer range and 275 704 for winter. By using full-checklists, we can assume observed Virginia rails would have been reported as present and, while the absence of an observation does not necessarily imply the absence of a bird, these checklists reasonably reflect sampling intensity and can serve as appropriate background points. An isotropic Gaussian kernel as implemented in the 'spatstat' package for R (Baddeley, Rubak & Turner 2015) was used to estimate sampling intensity within the summer and winter ranges of Virginia rail. The resulting sampling intensity raster was supplied to MaxEnt via the bias grid option. For the target group sampling, we randomly selected ten sets of 40 000 occurrence points (as determined from initial tuning runs, described below) from within the summer and winter range, respectively, to serve as background points. We extracted the environmental covariate data at these points and supplied them in 'samples-with-data' format.

Our goal was to generate the SDM with the greatest predictive accuracy, as measured by the area under the receiver operating characteristic curve (AUC under ROC; Hanley & McNeil 1982; Phillips, Anderson & Schapire 2006). This distinction is important because our choice of settings is justified for estimating potential species distributions but should not be used to infer underlying environmental drivers (Merow, Smith & Silander 2013). First, we ran a series of tuning runs to systematically determine the optimal number of background samples (10 000–40 000 by 5000), value of the β smoothing parameter (0·0–3·0 by 0·5), number of training iterations (500–1500 by 250) and proportion of training to test data (70/30, 60/40 or 50/50) for the summer and winter SDMs, respectively (Merow, Smith & Silander 2013). Initially, we included all 21 environmental covariates in the analysis and used leave-one-out jackknifing to identify covariates reducing the predictive power of the model which were subsequently removed. For the final SDMs, we randomly assigned the data to 60% training and 40% test sets in 10 bootstrapped replicates to estimate the mean AUC and standard error. Modelling was carried out using the graphical interface of MAXENT v 3.3.3k (Phillips, Anderson & Schapire 2006; Phillips & Dudík 2008).

STABLE ISOTOPE ANALYSIS

We cleaned feathers with phosphate-free detergent and 2:1 chloroform methanol solution, rinsed in deionized water and dried at 50 °C overnight. Approximately 300 μg of tissue from each feather vane was loaded into high-purity silver capsules and prepared for stable isotope mass spectrometry. The Stable Isotope Laboratory at the University of Arkansas analysed migrating rail samples, and the Colorado Plateau Stable Isotope Laboratory at the University of Northern Arizona analysed wintering rails samples. At both facilities, isotope measurements were made on H_2 gas emitted during flash pyrolysis using a continuous-flow isotope ratio mass spectrometer. Because ~20% of the hydrogen in feathers exchanges freely with ambient water vapour (Wassenaar & Hobson 2003), we analysed feathers concurrently with three calibrated keratin standards to allow for comparison across laboratories (Wassenaar & Hobson 2003). We report the non-exchangeable δD fraction in parts per mil (‰) normalized to the Vienna Standard Mean Ocean Water–Standard Light Antarctic Precipitation (VSMOW-SLAP) standard (Wassenaar & Hobson 2003). Two to eight technical replicates were in each run, which yielded a reproducibility of ±1·87 ‰.

MIGRANT RAIL ORIGINS

A strong relationship between the δD of precipitation and the tissue is a prerequisite for stable isotope-derived geographic assignments. As only a few adult Virginia rails (n = 10) were captured during the moult period and because keratin synthesis is conserved across similar species, we also included samples from a sympatric congener, the king rail *R. elegans* (Perkins 2007) to determine the relationship between $\delta D_{feathers}$ and $\delta D_{precipitation}$. The calibration set comprised a total of 54 flight feathers from: (i) 10 Virginia rails captured at a single location; (ii) 13 king rail museum specimens <100 years old at the time of analysis from 11 different localities; and (iii) 31 king rails captured live on the breeding ground, albeit 28 were from a single site (see Appendix S1 for full details). We regressed the δD of flight feathers to mean annual growing season $\delta D_{precipitation}$ at the site of sample collection and used this relationship to convert a GIS-based model of $\delta D_{precipitation}$ across North America to a spatially explicit raster depicting mean expected $\delta D_{feather}$ values (~37 × 37 km resolution; Bowen, Wassenaar & Hobson 2005).

Due to sources of variance inherent in isotope mass spectrometry and interindividual physiology and behaviour, the $\delta D_{feather}$ values expected from any given site are best characterized as a distribution of potential values (e.g. Sullivan *et al.* 2012). We represented the origins of a given bird as a normal density function (Royle & Rubenstein 2004):

$$f(y*|\mu_c,\sigma_c) = \left(\frac{1}{\sqrt{2\pi}\sigma_c}\right)\exp\left[\frac{1}{2\sigma_c^2}(y*-\mu_c)^2\right]$$

where $f(y*|\mu_c,\sigma_c)$ is the probability a $\delta D_{feather}$ (y*) value originates from a pixel on the raster, given the expected mean of the pixel μ_c, and σ_c, the standard deviation of the residuals from the $\delta D_{feathers} - \delta D_{precipitation}$ regression. Stable isotope assignments were made using the 'raster' and 'rgdal' packages in R v. 3.0.2 (Hijmans & van Etten 2012; Bivand, Keitt & Rowlingson 2014).

We incorporated the SDM as prior information for the stable isotope assignment using Bayes' rule:

$$f(b|y) = \frac{f(y|b)f(b)}{\sum_{b=1}^{B} f(y|b)f(b)}$$

where $f(b|y)$ is the posterior probability a given pixel on the raster represents the origins of a bird, $f(b)$ is the probability of Virginia rail occurrence based on the SDM, and $f(y|b)$ is the δD assignment. Bayes' Rule was applied using the Raster Calculator tool in the Spatial Analyst extension in ARCMAP 10.2 (ESRI, Denver, CO, USA). To depict the likely origins of each population, we assigned each individual separately and converted the raster to a binary surface using a 9:1 odds ratio, which is three times more conservative than the commonly employed 3:1 odds ratio (Van Wilgenburg & Hobson 2011; Hallworth *et al.* 2013). The resulting assignments were summed to yield population-level assignments in units of the number of rails with origins consistent with a given pixel and converted to proportions to enable comparisons across sites. We generated one assignment map per study site location, with the exception of Rockefeller NWR, because we sampled there 2 years (Table 1) and mean δD values varied significantly (Mann–Whitney *U*-test, $P < 0{\cdot}002$). Although rails were also collected at McFaddin NWR and Ottawa NWR in two different years, we found no evidence for a difference in δD values (Mann–Whitney *U*-test, both $P > 0{\cdot}50$).

MIGRATION STRATEGY

We tested for significant differences in the δD values of feathers – and therefore migratory origin – according to sex and wing chord length (a proxy for size, e.g. Hobson *et al.* 2009) in general linear models constructed separately for migrating and wintering habitats. The sex of sampled rails was determined with morphometric discriminate analysis of tarsus and culmen (Fournier *et al.* 2013). We included an interaction term between sex and sample location to test for differences in δD within each of the winter sites. For rails sampled during migration at Ottawa NWR, we tested for differences in the winter origin of rails according to their capture date.

Results

SPECIES DISTRIBUTION MODELS

Strong spatial bias was detected in eBird observations reported concurrently with the Virginia rail observations used as SDM construction (June–July and December–February, 2002–2012; see Materials and methods). Within the winter and summer ranges, sampling intensity was highest on the east and west coasts of the USA, with potential habitat in central Canada and Mexico less sampled. This spatial bias in observations translated into environmental bias: supplying the bias grid as an 'environmental covariate' to MaxEnt when constructing the summer and winter SDMs indicated sampling intensity as the single strongest predictor of Virginal rail presence. Correcting for spatial and corresponding bias in the SDM, using bias grids and target group sampling, yielded qualitatively similar results, but the bias grid option consistently produced more discriminative models (Fig. S2).

PROVENANCE ASSIGNMENTS

Variation in the modelled mean annual δD of growing season precipitation explained 75·6% of the variance of δD of rail feathers of known provenance ($\delta D_{feathers} = 1{\cdot}156 \times \delta D_{precipitation} - 43{\cdot}82$, $\sigma = 8{\cdot}60$, $P < 10^{-15}$). Variation in $\delta D_{feather}$ values within a site was relatively small (SD = 5·60), and most uncertainty in the relationship arose from deviation in the expected $\delta D_{feather}$ among sites (Fig. S1). Examination of the regression residuals for constant variance yielded no indication of heteroscedasticity.

To illustrate our method, we depict the probable origins of a single Virginia rail captured at a migratory stopover site at Ottawa NWR, Ohio (Fig. 2a–c) and one captured at a wintering site at Grand Cote NWR, Louisiana (Fig. 2d–f). Model uncertainty exacerbates the natural lack of resolution in stable hydrogen based assignments (Fig. 2a,d). Poor resolution is most conspicuous when determining the summer origins of wintering rails, as probable isotopic origins span from coast-to-coast (Fig. 2d). Although limiting the assignments to the known range of the Virginia rail removes biologically implausible origins in northern Canada, much of the USA would still have $\delta D_{precipitation}$ values consistent with the origins of wintering rails (Fig. 2d). Setting the isotope assignment conditional on a summer-specific SDM produced assignments with markedly greater resolution, but most individual assignments of wintering rails still had small areas (i.e. several 37 km^2 pixels) identified as plausible summer habitat on the basis of combined $\delta D_{feather}$ and SDM analysis (Fig. 2e,f). Isotopic provenance assignment of migrating rails produced narrower origins, as $\delta D_{precipitation}$ rapidly loses heavy hydrogen with increasing distance inland (Fig. 2a). Only a relatively narrow band along the Gulf Coast is consistent with the $\delta D_{feathers}$, but applying the SDM as a prior in the Bayesian framework further limited the possible origins to the coasts of Texas, western Louisiana and southernmost Florida (Fig. 2f).

Combined δD and SDMs at the population level revealed most rails wintering along the Gulf Coast are predicted to breed in eastern North America across a broad, wetland-rich region spanning from the western shores of Lake Michigan to the southern border of Ontario and Québec (Fig. 3a–f). While the catchment areas of all five sites overlapped, they varied in the proportion of rails originating from the north-eastern and south-western Great Lakes region (Fig. 3). At each site, a majority (≥50%) of Virginia rails were assigned to the Great Lakes region, while fewer (7–35%) were assigned to the Prairie Pothole Region of the Dakotas and the Chesapeake Bay region. Interannual variation in $\delta D_{feathers}$ at Rockefeller WR translated into markedly different summer catchment areas: in 2004–2005, a large minority of rails (35%) originated from the Prairie Potholes Regions, while no rails were exclusively assigned to this area in 2006 and instead originated from further east across the Great Lakes. At the individual assignment level, however, many rails (21%) could not be unambiguously assigned to either the Great Lakes or Chesapeake watershed region. More overwintering rails in Texas (Anahuac and McFaddin) were exclusively

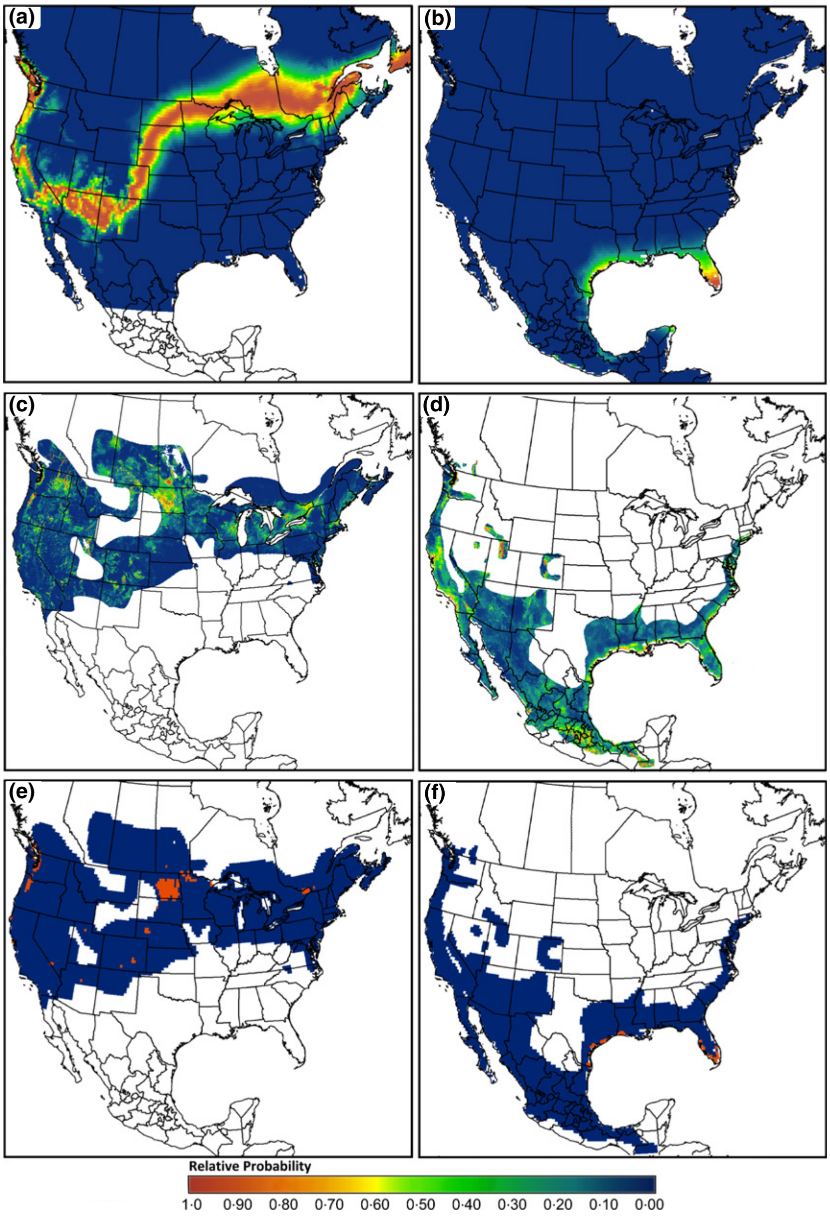

Fig. 2. Probable origins of a Virginia rail *Rallus limicola* captured at Ottawa National Wildlife Refuge, Ohio and one captured at Sherburne Wildlife Management Area Louisiana (USA) as determined by δD values of feathers (a, b) and season-specific species distribution models (c, d). The probability each cell on the map represents the origin of the rail (e, f) is determined by applying Bayes' Theorem to make the isotope-inferred origins (a, b) conditional on the species distribution models (c, d). Binary assignment surfaces were created by applying a 9:1 odds ratio, whereby all pixels with a probability >14·29% were considered 'likely' and all others as 'unlikely' (e, f). [Colour figure can be viewed at wileyonlinelibrary.com].

assigned to breeding habitat in the Chesapeake Bay watershed on the east coast (14% vs. 1%), but Louisiana rails (Rockefeller, Grand Cote and Sherburne) were more likely to have spent the breeding season west of the Great Plains (17% vs. 3%).

There was less variability observed in the probable origins of the migrating rails sampled in Ohio. All 41 rails most likely spent the winter along the coast of Texas or Louisiana (Fig. 4). We found no significant interannual variation among migrating rails captured at Ottawa NWR between 2005 and 2009, in contrast to the variation observed among wintering rails at Rockefeller NWR, Louisiana.

MIGRATION STRATEGY

Individual sex was determined using morphometric measurements for 111 of the 136 rails (81·2%) with ≥0·70 posterior probability. The proportion of males:females varied across sites, from 0·24 at Grand Cote NWR in Louisiana to 3·86 at Ottawa NWR in the breeding range in Ohio (Table S2). We found no indication that provenance (i.e. δD values) varies with wing chord at any of the six sites ($P > 0{\cdot}05$). At Ottawa NWR, δD did not vary according to passage date ($P > 0{\cdot}05$).

Mean δD (±SD) values were similar for breeding male and female rails at Ottawa NWR (−40·42 ± 14·93, −38·46 ± 14·87, respectively) and across the wintering range (−102·75 ± 20·13, −104·75 ± 21·74, respectively). However, the interaction between sex and location was significant at Sherburne WMA ($P < 0{\cdot}01$) and Grand Cote NWR ($P < 0{\cdot}05$) and was marginally significant ($P = 0{\cdot}05$) among the rails sampled during 2006 at Rockefeller WR (Fig. 5). We observed a similar trend at

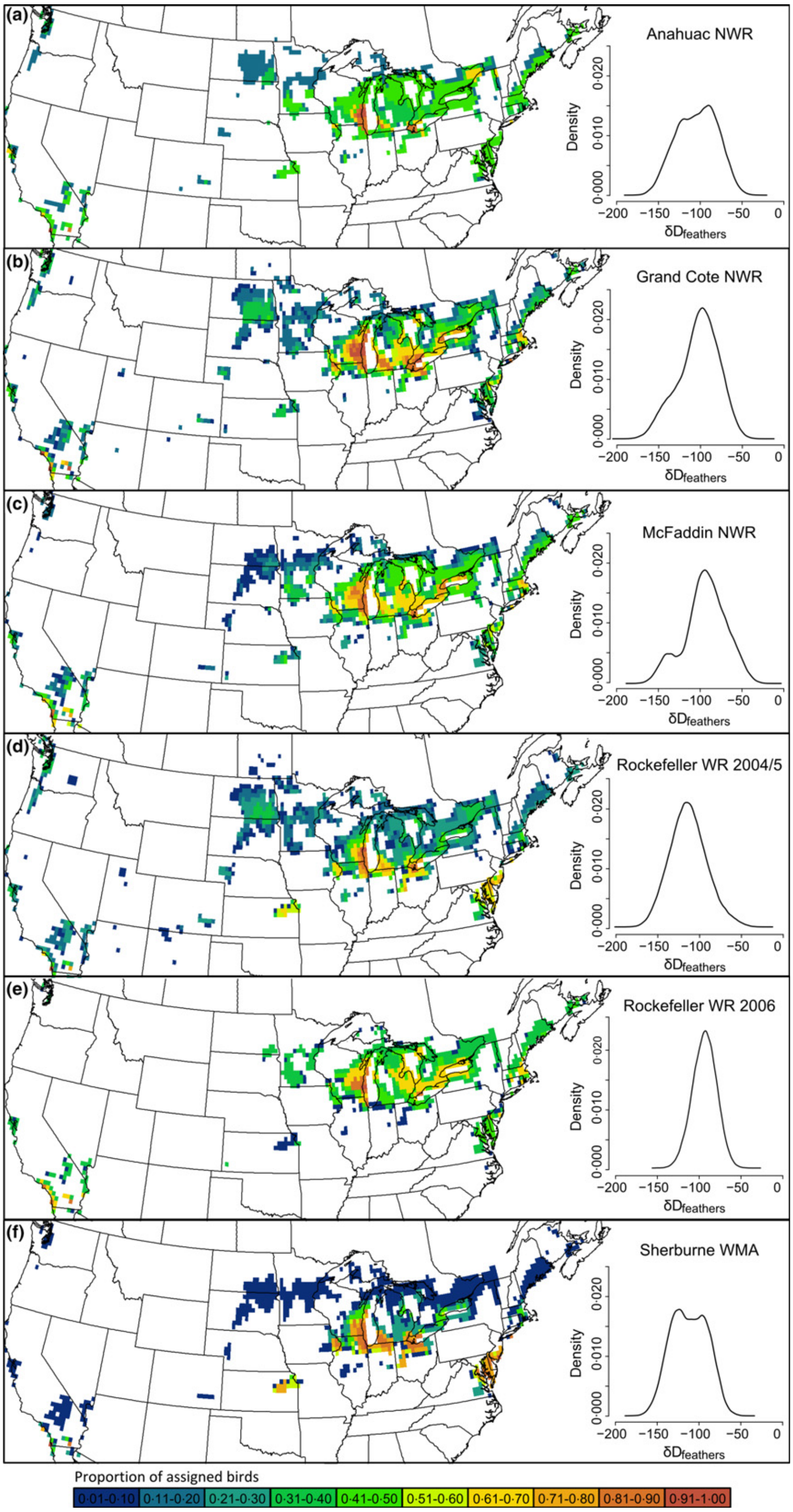

Fig. 3. Probable breeding ranges of wintering Virginia rail *Rallus limicola* inferred from feather δD and species distributions models. Wintering rails were captured on the wintering grounds at five sites along the Gulf Coast, USA: (a) Anahuac National Wildlife Refuge, Texas; (b) Grand Cote National Wildlife Refuge, Louisiana; (c) McFaddin National Wildlife Refuge, Texas; (d) Rockefeller Wildlife Refuge 2004–05, Louisiana; (e) Rockefeller Wildlife Refuge 2006; and (f) Sherburne Wildlife Management Area, Louisiana. Panel insets depict the derived probability density function used to incorporate isotopic variance into assignments of origin. Population origins are depicted as the proportion of the individual assignments in a population consistent with a probable origin at a given pixel. [Colour figure can be viewed at wileyonlinelibrary.com].

Anahuac NWR in Texas, but only two of the seven sexed rails were female (Fig. 5). Together, these results indicate individual wintering sites may attract male and female rails from different breeding habitats.

Discussion

Understanding the timing and connectivity of animal movements is a persistent challenge in ecology, especially for rare or difficult to detect species. To address this challenge, we created a Bayesian framework combining SDMs derived from citizen science observations with δD assignments to produce high-resolution estimates of the origins of migratory animals. Incorporating SDMs with stable isotopes provides an empirical method to remove biologically implausible origins without relying on presumed knowledge or surrogate species.

Our method applies to any migratory species where large-scale citizen science data are available. Of the growing number of citizen science data bases, eBird is presently the largest, comprising over 275 million observations and representing 95% of the world's known avifauna (Sullivan *et al*. 2014). Data bases focused on other taxa, such as mammals (eMammal, Forrester *et al*. 2013) and butterflies (eButterfly, Larrivee *et al*. 2014), are also in development, in addition to general initiatives targeting target citizen scientists from diverse geographic areas (e.g. iSpot, Clow & Makriyannis 2011 and iNaturalist, iNauturalist.org). Such data bases provide a rich but largely untapped resource for studies of migratory connectivity (but see Fink *et al*. 2010; Miller *et al*. 2011).

Presence data sourced from citizen scientists are invaluable for species that are difficult to detect with systematic inventories (e.g. BioBlitz; Lundmark 2003; Kindberg, Ericsson & Swenson 2009). However, citizen science data are generally not evenly distributed across the landscape, which can confound SDMs and requires bias correction. To account for sampling bias, we employed two separate background point manipulations, but the optimal bias-corrected method will likely depend on the species, spatial and temporal focus, utilized data base and modelling method (e.g. Kramer-Schadt *et al*. 2013; Fourcade *et al*. 2014). Bias correction methods are an area of active development (e.g. Boria *et al*. 2013; Kramer-Schadt *et al*. 2013; Syfert, Smith & Coomes 2013; Fourcade *et al*. 2014), and large data bases, such as eBird, deserve detailed study. However, several studies have indicated simple spatial filtering can also significantly reduce sampling bias and may in fact be the optimal correction method if presence data are abundant (Kramer-Schadt *et al*. 2013; Fourcade *et al*. 2014). Importantly, the summer SDM after bias correction indicated areas of high habitat suitability concordant with Virginia rail abundance estimates from the systematic Breeding Bird Survey (Sauer *et al*. 2014), even in regions poorly represented by citizen science observations. Error in geographic assignments may be introduced through a biased prior, but there are also limitations and caveats inherent to δD-based assignments. For example, we still could not unambiguously discriminate between breeding origins on the Great Lakes or Chesapeake Bay for about 20% of the birds. These two regions are geographically distant but share similar $\delta D_{precipitation}$ values and provide suitable breeding habitat, and therefore, our model still could not fully distinguish them. This limitation could potentially be overcome by using additional stable isotopes. Stable oxygen in particular shows predictable, continental variation and can be used to complement δD assignments (Pekarsky *et al*. 2015), and continental-level predictions for strontium isotope variation may soon be available (Chesson *et al*. 2012). An important caveat – we emphasize that stable isotope-based assignments are model predictions and should be verified with independent data (e.g. Hallworth *et al*. 2013) before widespread deployment in management programmes.

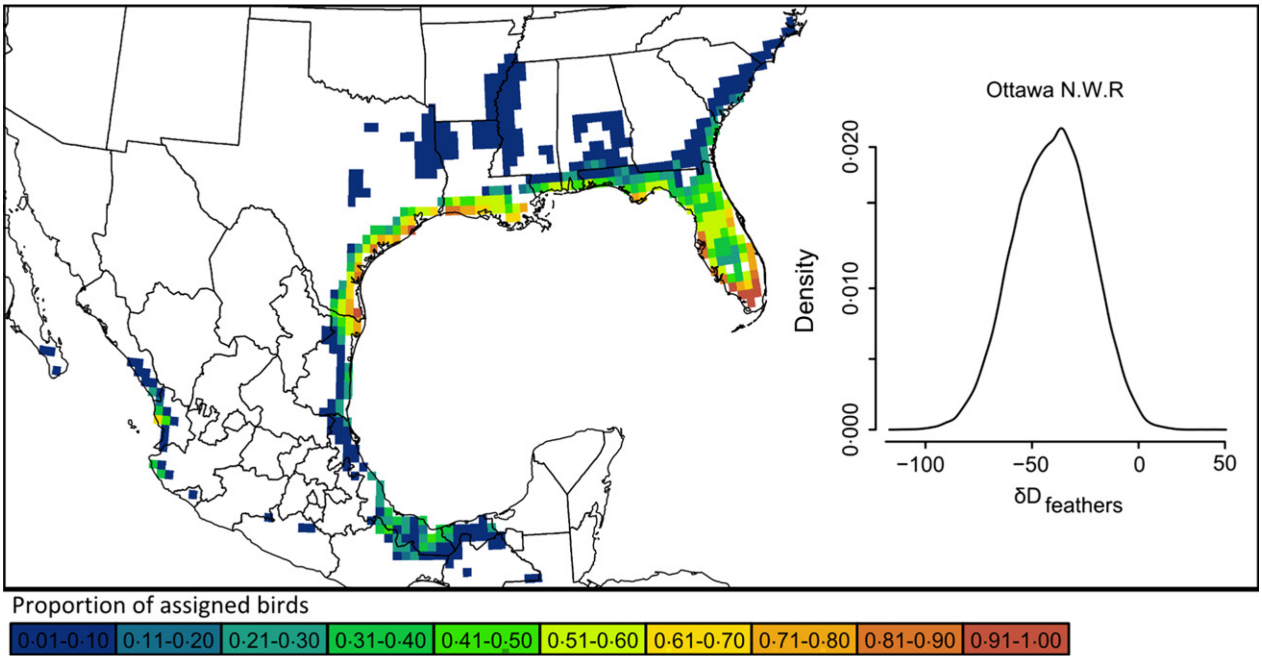

Fig. 4. Winter provenance of migrating Virginia rails *Rallus limicola* captured at Ottawa National Wildlife Refuge (Ohio, USA) and inferred from probability distributions based on stable hydrogen isotope analysis of feathers and species distributions models. Panel insets depict the derived probability density function used to incorporate isotopic variance into assignments of origin. [Colour figure can be viewed at wileyonlinelibrary.com].

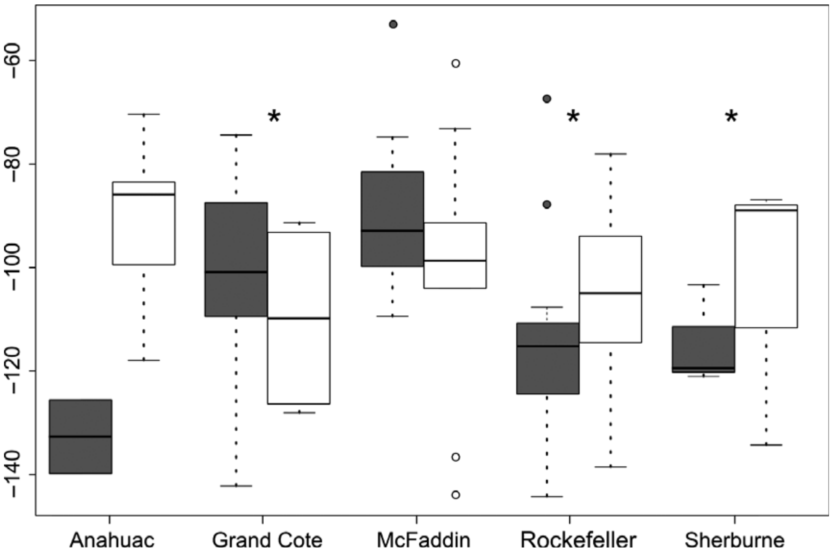

Fig. 5. Feather δD values by sex and site (see Table 1) for wintering Virginia rails *Rallus limicola* along the Gulf Coast, USA. Boxplots for females are denoted in grey. Asterisks indicate a significant difference in δD values ($\alpha = 0{\cdot}05$) between males and females at a site.

Combined analysis of citizen science data and δD produced high-resolution estimates of the origins of migrant Virginia rails and also revealed uncertainties in the current management strategy. Virginia rails are managed in administrative flyways delineated from surrogate species, but we found limited congruence between the movements of Virginia rails and their administrative flyways (Louisiana = Mississippi Flyway; Texas = Central Flyway). Less than 20% of the individuals from Texas had probable breeding origins located within the Central Flyway, while nearly 45% Louisianan birds originated either further east (30%) or west (14%) of the Mississippi Flyway. These assignments show the need for organism-specific studies of migratory connectivity, especially for game species like the Virginia rail whose numbers may be declining across their range (Conway 1995). Detailed knowledge of Virginia rail migration enabled by this, and future studies will inform the implementation of science-based wetland management, funding strategies and harvest quotas (Conway 1995; Case & McCool 2009). As Virginia rail is a cryptic species, we suggest our framework can also yield estimates of migratory connectivity of sufficient resolution to guide conservation and management decisions for many other migratory animals, especially those with low detection probability.

Acknowledgements

We are grateful to Shawn O'Neil and an anonymous reviewer for constructive comments on earlier drafts of the manuscript. Thanks to the Ecosystem Science Center at Michigan Technological University for funding part of this project through their undergraduate research programme. We thank the following museums for access to specimens: The Sam Noble Oklahoma Museum of Natural History; The Field Museum of Natural History, Chicago; University of Kansas Natural History Museum; Ornithology Department, Museum of Comparative Zoology, Harvard. The Louisiana Department of Wildlife and Fisheries provided support for the Gulf Coast samples through the Rockefeller Wildlife Refuge. This research was funded in part by the National Science Foundation (DEB Proposal 1545611 to JKB). Any use of trade, firm or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

Data accessibility

Data available from the Dryad Digital Repository http://dx.doi.org/10.5061/dryad.r4847 (Fournier *et al.* 2016).

References

Baddeley, A., Rubak, E. & Turner, R. (2015) *Spatial Point Patterns: Methodology and Applications with R*. Chapman and Hall/CRC Press, London, UK. http://www.crcpress.com/Spatial-Point-Patterns-Methodology-and-Applications-with-R/Baddeley-Rubak-Turner/9781482210200/

BirdLife International & NatureServe (2013) *Bird Species Distribution Maps of the World*, Version 3.0. BirdLife International, Cambridge, UK and NatureServe, Arlington, USA. http://www.birdlife.org/datazone/info/spcdownload

Bivand, R., Keitt, T. & Rowlingson, B. (2014) rgdal: bindings for the geospatial data abstraction library. R Package, version 0.9-1. Available online http://cran.r-project.org/web/packages/rgdal/

Boria, R.A., Olson, L.E., Goodman, S.M. & Anderson, R.P. (2013) Spatial filtering to reduce sampling bias can improve the performance of ecological niche models. *Ecological Modelling*, **275**, 73–77.

Bowen, G.J. (2010) Isoscapes: spatial pattern in isotopic biogeochemistry. *Annual Review of Earth and Planetary Sciences*, **38**, 161–187.

Bowen, G.J., Wassenaar, L.I. & Hobson, K.A. (2005) Global application of stable hydrogen and oxygen isotopes to wildlife forensics. *Oecologia*, **143**, 337–348.

Case, D. & McCool, D. (2009) Priority information needs for rails and snipe. Developed by the Association of Fish and Wildlife Agencies' Migratory Shore and Upland Game Bird Support Task Force. https://www.fws.gov/migratorybirds/pdf/surveys-and-data/Info-Needs-Rails-Snipe.pdf

Chesson, L.E., Tipple, B.J., Mackey, G.N., Hynek, S.A., Fernandez, D.P. & Ehleringer, J.R. (2012) Strontium isotopes in tap water from the coterminous USA. *Ecosphere*, **3**, 1–17.

Clow, D. & Makrivannis, E. (2011) iSpot analysed: participatory learning and reputation. *Proceedings of the 1st International Conference on Learning Analytics and Knowledge*, **1**, 34–43.

Conway, C.J. (1995) Virginia rail. *The Birds of North America*, No. 173 (eds A. Poole & F. Gill). The Academy of Natural Sciences and The American Ornithologists' Union, Philadelphia, PA, USA.

Dudík, M., Schapire, R.E. & Phillips, S.J. (2005) Correcting sample selection bias in maximum entropy density estimation. *Advances in Neural Information Processing Systems*, **17**, 323–330.

Elith, J., Kearney, M. & Phillips, S. (2010) The art of modelling range-shifting species. *Methods in Ecology and Evolution*, **1**, 330–342.

Elith, J. & Leathwick, J.R. (2009) Species distribution models: ecological explanation and prediction across space and time. *Annual Review of Ecology, Evolution, and Systematics*, **40**, 677–697.

Elith, J., Graham, C.H., Anderson, R.P., Dudík, M., Ferrier, S., Guisan, A. *et al.* (2006) Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, **29**, 129–151.

Fink, D., Hochachka, W.M., Zuckerberg, B., Winkler, D.W., Shaby, B., Munson, M.A. *et al.* (2010) Spatiotemporal exploratory models for broad-scale survey data. *Ecological Applications*, **20**, 2131–2147.

Forrester, T., McShea, W.J., Kays, R.W., Costello, R., Baker, M. & Parsons, A. (2013) eMammal – citizen science camera trapping as a solution for broad-scale, long-term monitoring of wildlife populations. Ecological Society of American 98th Annual Convention.

Fourcade, Y., Engler, J.O., Rödder, D. & Secondi, J. (2014) Mapping species distributions with MAXENT using a geographically biased sample of presence data: a performance assessment of methods for correcting sampling bias. *PLoS One*, **9**, e97122.

Fournier, A.M.V., Sheildcastle, M.C., Fries, A.C. & Bump, J.K. (2013) A morphometric model to predict the sex of Virginia Rails (*Rallus limicola*). *Wildlife Society Bulletin*, **37**, 881–886.

Fournier, A.M.V., Shieldcastle, M.C., Kashmer, T. & Mylecraine, K.A. (2015) Comparison of arrival dates of spring rail migration in the southwest Lake Erie marshes, Ohio, USA. *Waterbirds*, **38**, 312–314.

Fournier, A.M.V., Sullivan, A.R., Bump, J.K., Perkins, M., Shieldcastle, M.C. & King, S.L. (2016) Data from: Combining citizen science species distribution models and stable isotopes reveals migratory connectivity in the secretive Virginia rail. *Dryad Digital Repository*, http://dx.doi.org/10.5061/dryad.r4847

Hallworth, M.T., Studds, C.E., Sillett, T.S. & Marra, P.P. (2013) Do archival light-level geolocators and stable hydrogen isotopes provide comparable estimates of breeding-ground origin? *The Auk*, **130**, 273–282.

Hanley, J.A. & McNeil, B.J. (1982) The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, **143**, 29–36.

Haramis, G.M. & Kearns, G.D. (2007) Soras in tidal marsh: banding and telemetry studies on the Patuxent River, Maryland. *Waterbirds*, **30**, 105–121.

Hijmans, R.J. & van Etten, J. (2012) raster: Geographic analysis and modeling with raster data. g. R package, version 2.3-12.

Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G. & Jarvis, A. (2005) Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, **25**, 1965–1978.

Hobson, K.A., Lormée, H., Van Wilgenburg, S.L., Wassenaar, L.I. & Boutin, J.M. (2009) Stable isotopes (δD) delineate the origins and migratory connectivity of harvested animals: the case of European woodpigeons. *Journal of Applied Ecology*, **46**, 572–581.

Kindberg, J., Ericsson, G. & Swenson, J.E. (2009) Monitoring rare or elusive large mammals using effort-corrected voluntary observers. *Biological Conservation*, **142**, 159–165.

Kramer-Schadt, S., Niedballa, J., Pilgrim, J.D., Schröder, B., Lindenborn, J., Reinfelder, V. *et al.* (2013) The importance of correcting for sampling bias in MaxEnt species distribution models. *Diversity and Distributions*, **19**, 1366–1379.

Larrivee, M., Prudic, K.L., McFarland, K. & Kerr, J. (2014) eButterfly: a citizen-based butterfly database in the biological sciences. http://www.e-butterfly.org

Latifovic, R., Zhu, Z.-L., Cihlar, J. & Giri, C. (2002) *Land Cover of North America 2000*. Natural Resources Canada, Canada Center for Remote Sensing, US Geological Service EROS Data Center, South Dakota, SD, USA.

Lehner, B. & Döll, P. (2004) Development and validation of a global database of lakes, reservoirs and wetlands. *Journal of Hydrology*, **296**, 1–22.

Lundmark, C. (2003) BioBlitz: getting into backyard biodiversity. *BioScience*, **53**, 329.

Merow, C., Smith, M.J. & Silander, J.A. Jr (2013) A practical guide to MaxEnt for modeling species' distributions: what it does, and why inputs and settings matter. *Ecography*, **36**, 1058–1069.

Miller, N.G., Wassenaar, L.I., Hobson, K.A. & Norris, D.R. (2011) Monarch butterflies cross the Appalachians from the west to recolonize the east coast of North America. *Biology Letters*, **7**, 43–46.

Pekarsky, S., Angert, A., Haese, B., Werner, M., Hobson, K.A. & Nathan, R. (2015) Enriching the isotopic toolbox for migratory connectivity analysis: a new approach for migratory species breeding in remote or unexplored areas. *Diversity and Distributions*, **21**, 416–427.

Perkins, M. (2007) *The use of stable isotopes to determine the ratio of resident of migrant king rails in southern Louisiana and Texas*. Master's Thesis, Louisiana State University.

Perkins, M., King, S.L. & Linscombe, J. (2010) Effectiveness of capture techniques for rails in emergent marsh and agricultural wetlands. *Waterbirds*, **33**, 376–380.

Phillips, S.J., Anderson, R.P. & Schapire, R.E. (2006) Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, **190**, 231–259.

Phillips, S.J. & Dudík, M. (2008) Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography*, **31**, 161–175.

Phillips, S.J., Dudík, M., Elith, J., Graham, C.H., Lehmann, A., Leathwick, J. & Ferrier, S. (2009) Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. *Ecological Applications*, **19**, 181–197.

Royle, A.J. & Rubenstein, D.R. (2004) The role of species abundance in determining breeding origins of migratory birds with stable isotopes. *Ecological Applications*, **14**, 1780–1788.

Sauer, J.R., Hines, J.E., Fallon, J.E., Pardieck, K.L., Ziolkowski, D.J. Jr & Link, W.A. (2014) *The North American Breeding Bird Survey, Results and Analysis 1966–2013*, Version 01.30.2015. USGS Patuxent Wildlife Research Center, Laurel, MD, USA.

Sullivan, A.R., Bump, J.K., Kruger, L.A. & Peterson, R.O. (2012) Bat-cave catchment areas: using stable isotopes (δD) to determine the probable origins of hibernating bats. *Ecological Applications*, **22**, 1428–1434.

Sullivan, B.L., Aycrigg, J.L., Barry, J.H., Bonney, R.E., Bruns, N., Cooper, C.B. *et al.* (2014) The eBird enterprise: an integrated approach to development and application of citizen science. *Biological Conservation*, **169**, 31–40.

Syfert, M.M., Smith, M.J. & Coomes, D.A. (2013) The effects of sampling bias and model complexity on the predictive performance of MaxEnt species distribution models. *PLoS One*, **8**, e55158.

Van Wilgenburg, S.L. & Hobson, K.A. (2011) Combining stable-isotope (δD) and band recovery data to improve probabilistic assignment of migratory birds to origin. *Ecological Applications*, **21**, 1340–1351.

Wassenaar, L.I. & Hobson, K.A. (2003) Comparative equilibration and online technique for determination of non-exchangeable hydrogen of keratins for use in animal migration studies. *Isotopes in Environmental Health Studies*, **39**, 211–217.

Yackulic, C.B., Chandler, R., Zipkin, E.F., Royle, J.A., James, D., Grant, E.H.C. & Veran, S. (2013) Presence-only modeling using MAXENT: when can we trust the inferences? *Methods in Ecology and Evolution*, **4**, 236–243.

Received 21 December 2015; accepted 16 June 2016
Handling Editor: Des Thompson

Supporting Information

Additional Supporting Information may be found in the online version of this article.

Appendix S1. Individuals used in the $\delta D_{feather} - \delta D_{precipitation}$ regression, including species, collection date, coordinates, museum accession number, and δD value.

Table S1. Bioclimatic variables used in maximum entropy (MaxEnt) species distribution models.

Table S2. Sex of Virginia rails as determined by morphometric analysis.

Fig. S1. Regression of δD of *Rallus* spp. feathers of known provenance on δD of mean annual growing-season precipitation.

Fig. S2. Comparison between uncorrected and bias-corrected species distribution models.