Inconsistent shifts in warming and temperature variability negatively impact avian fitness.

Conor C. Taff   J. Ryan Shipley

CCT: Department of Ecology & Evolutionary Biology and Lab of Ornithology, Cornell University

JRS: Swiss Federal Institute for Forest, Snow and Landscape Research WSL

Both authors contributed equally to this study.

Correspondance: Conor C. Taff; cct63@cornell.edu

ABSTRACT

As the climate has warmed many birds have advanced their breeding timing. However, climate change also changes temperature distributions; thus breeding earlier in a hotter world might increase nestling exposure to either extreme heat or cold. We combined >300,000 breeding records from 24 North American birds with historical temperature data to understand how exposure to extreme temperatures has changed. Average spring temperature increased since 1950, but change in timing of extremes was inconsistent in direction and magnitude; thus, populations could not track both average and extreme temperatures. Relative fitness was reduced following heat-waves and cold-snaps in 11 and 16 of 24 species, respectively. Latitudinal variation in sensitivity in three widespread species suggests that vulnerability to extremes at range limits may contribute to range shifts. Our results add to evidence demonstrating that understanding individual sensitivity and its links to population level processes is critical for predicting vulnerability to changing climates.

Keywords: climate change, cold-snaps, heat-waves, reproductive success, phenology, extreme temperature, community science

1 INTRODUCTION

Human induced climate change has resulted in an accelerating increase in the average temperatures that animals experience around the globe.[1] A rich literature now documents the consequences that span from the scale of individuals to entire populations and species[2]. The most universal observed responses to warming are reductions in body size,[3–5] shifts in species ranges,[2] and changes in the timing of annual events (i.e., phenological shifts),[6–8] which have the potential to impact fitness via increased interspecific competition[9] or phenological mismatches between predators and prey.[10–13] However, mean temperatures alone fail to capture the complexity of climate change, as the rate and magnitude of warming

2

often differs between day and night as well as seasonally. Another component of climate change is increasing temperature variability, which is predicted to have greater impacts on species survival than increases in mean temperature.[14] As the pace, magnitude, and variability of climate change determines the risk of exposure to conditions outside those historically experienced, there is a critical need to understand how and why the magnitude of these consequences differs between species to predict and mitigate the effects of climate change for wild animals.[15]

Recently, Rosenberg et al.[16] estimated that the total population of North American birds has declined by almost 3 billion since 1970. While these declines arise from a combination of several factors (e.g., introduced predators;[17] noise and light pollution;[18] land use change[19]; disease[20]), climate change is generally considered to be one of the primary threats and drivers.[21,22] However, the declines in overall abundance are far from uniform, with some groups, such as aerial insectivores, showing steep declines while others, such as many wetland birds, have increased in numbers.[16] These differences suggest that life history and habitat requirements many play an important role in determining the sensitivity of different species to climate change.

One of the major goals of many climate studies has been to assess whether wild birds are shifting their ranges or advancing breeding phenology fast enough to track changes in mean temperatures[23-25], with several recent studies suggesting that observed changes are generally inadequate to keep pace.[26-28] However, it is also increasingly apparent that changes in mean temperatures are insufficient to understand the response to climate change, as the mean only provides a coarse approximation of changes in minimum (nighttime) and maximum (daytime) temperatures which can change independently in pace, magnitude, and variability. Different organismal traits such as timing of reproduction or when a species is active likely shape the response because animals may need to balance changing temperatures simultaneously with other abiotic gradients, such as day length, UV radiation, or water

availability.[29,30] Even when only considering mean temperatures, perfectly tracking average daily temperatures may result in a more variable thermal environment when breeding activities are occurring earlier in the year[31] or when climate warming is associated with increased climate variability.[32–34] When predictable conditions are linked to survival, increases in either the magnitude or the frequency of conditions outside historical norms can result in reduced fitness and declines at the population or species level.

The relative importance of different abiotic gradients for each species and their link to fitness, coupled with the pace of plasticity or evolutionary change, likely dictates which species persist and which will perish.[22] The strength of this relationship is determined by how efficiently organisms transfer heat and energy in different environmental conditions and how these traits influence individual performance as well as the ability to survive, develop, grow, and reproduce.[35] For example, several studies demonstrate that life history traits predict different aspects of the response to climate change.[19,36,37] Among birds, foraging mode is one key life history characteristic that might influence the relative challenge posed by average temperature versus variation in temperature. The vast majority of birds are reliant on insects to provide energy and nutrients to their rapidly developing offspring. For example, obligate aerial insectivores depend on flying insects, even modest decreases in daytime temperatures can reduce insect activity and thereby overall foraging success[38] (hereafter "cold-snaps"). If these reductions in insect activity occur during critical periods when nestlings are most vulnerable, they can trigger mass mortality events.[31] Accordingly, species that do not rely on insect activity while foraging are likely to be less sensitive to cold-snaps. In contrast, heat-waves would be expected to impact a broader range of species as they may approach physiological upper temperature limits.[39,40] A more thorough understanding of relative sensitivity to average temperature, cold-snaps, and heat-waves during breeding attempts requires data that spans multiple species across a wide range of conditions paired with detailed records of reproductive performance.

The most detailed studies of climate change on populations tend to focus on single species with a long time series of detailed breeding data.[3,31,32,41] These studies provide evidence for the mechanisms by which climate change influences populations, but they typically do not address variation in responses between species or different populations. Another approach is to use large scale observations to model spatial, temporal, and cross species variation in population phenology,[7,8] abundance,[16] body size,[4,42] or ranges[43,44] in response to climate change. However, these studies typically cannot link the patterns observed at coarse scales to specific processes and mechanisms that occur at the level of individual animals.[25] Linking individual fitness to population level processes with large scale and spatially variable outcomes is needed to accurately predict variation in sensitivity to climate change across species.

In this study, we used a database of community-contributed observations on the timing of breeding and reproductive success in wild North American birds compiled by Project MartinWatch, by the NestWatch program at the Cornell Lab of Ornithology, and by Project Nestwatch from Birds Canada between 1995 and 2020. We identified >300,000 breeding records from 24 common species with nesting records spanning most of the United States and Canada. By documenting the reproductive performance of individuals, this database allows for detailed exploration of spatial, temporal, and interspecific variation in temperature exposure and vulnerability at near continental scales.

We first asked how the change in timing of cold-snaps and heat-waves compares with changes in overall average spring temperature across the range of locations from which breeding data were available. If these abiotic gradients have advanced at different paces or in different directions, then populations will be unable to track the same breeding conditions regardless of phenological advancement. Next, we determined the temperatures experienced during every nesting attempt and asked whether each species showed evidence of performance declines during cold-snaps or heat-waves during the breeding attempt (i.e., whether these

[111] events qualify as extreme climatic events 'ECEs' with biological consequences[45,46]). We

[112] predicted that performance declines would be most pronounced for cold-snaps experienced

[113] by aerial insectivores because of the direct impact on flying insect availability. In contrast,

[114] we predicted that all species in the dataset would be vulnerable to heat-waves.

2 MATERIALS & METHODS

2.1 Breeding records

[117] We combined breeding records from three different databases. First, we requested raw

[118] breeding records from Nestwatch, a project managed by the Cornell Lab of Ornithology

[119] (www.nestwatch.org). Second, we requested raw breeding records from Project Nestwatch, a

[120] program managed by Birds Canada (www.birdscanada.org/bird-science/project-nestwatch/).

[121] Finally, we obtained publicly archived records of breeding purple martins (*Progne subis*)

[122] from Project MartinWatch (www.purplemartin.org[47]). We initially manipulated each data

[123] source into a common format to allow us to merge records together. Beginning with the

[124] combined dataset, we then proceeded through a series of filtering and quality control steps to

[125] arrive at a final dataset for analysis.

[126] While the Nestwatch database includes some older observations, most nest records are from

[127] 1995 or later (>98%); we therefore removed older records to focus on the period from 1995

[128] to 2020. We also removed species that had fewer than 300 records. This minimum sample

[129] size was somewhat arbitrary, but given the complexity of models that we planned to fit, the

[130] desire for spatial and temporal sampling, and the need to model ECEs which, by definition,

[131] only occur in a small percentage of nesting attempts, we chose a 300 nest minimum to ensure

[132] that we could reliably fit models for each species included. For a species with exactly 300

[133] nests, we would only expect to observe an average of 7.5 nestling attempts that experienced

[134] a 2 SD cold-snap or heat-wave. We also removed European starling (*Sturnus vulgaris*) and

house sparrow (*Passer domesticus*) records from the dataset, because these two invasive species are often considered nuisance pests and it was unclear whether a high reporting rate of failed nests reflected human eviction from nest boxes. This resulted in 24 remaining species that we considered further (Table S1).

Next, we removed any records that had incomplete location data, that were outside of the study extent, were flagged by the data providers as having errors, or that had impossible values reported (e.g., more young than eggs or negative clutch sizes). We also removed nests that did not include information on the date of laying or hatching. In cases where only a laying date was provided, we inferred approximate hatching dates by using the typical incubation length as reported in species accounts from the Birds of the World online.[48] While fledging success was reported, exact fledging date was generally unknown and we estimated fledging dates using the typical fledging age for a species from Birds of the World online.[48]

At this point, we also roughly categorized the included species into four foraging modes based on the primary way that food is obtained during the nestling provisioning period using the species accounts in Birds of the World Online. These included aerial capture (both fly-catching and aerial insectivores), sally foraging (flying to ground from a perch after visually detecting arthropod movement), gleaning (a broad category of strategies involving active search for arthropods), and seed eating. We did not have enough total species or phylogenetic variation in foraging mode to formally analyze diet, but we discuss the possible role that these foraging styles may have had on susceptibility to cold-snaps and heat-waves.

Finally, we plotted distributions of clutch sizes and hatching dates for each of the 24 species and removed records that had values that were likely due to data entry error (i.e., outside of the possible range for each species). This filtering step was somewhat subjective, but we were conservative in setting limits so that only records well outside the expected ranges were excluded. The end result of these filtering steps was a final dataset that included 301,514 breeding records from 24 species. The spatial extent of the dataset is illustrated in Figure S1

161 and the number of nests included for each species is shown in Table S1. Using the final set of

162 filtered breeding records, we determined the elevation of each nest by accessing a digital

163 elevation model raster spanning the records using the `elevatr` package in R.[49] Temperature

164 data were added to each nest record as described below.

2.2 Spatial and temperature data

166 **Spatial data.** We downloaded a base map for the region encompassing the breeding records

167 included using the package `rnaturalearth`[50] in R version 4.0.2.[51] For some analyses (see

168 below) we wanted to summarize breeding records and historical temperature by grouping

169 records and weather stations that were recorded close together. To facilitate this grouping,

170 we created a grid of hexagons each with an area of 40,000 km^2 across the land area of North

171 America with the `st_make_grid` function from package `sf`.[52] We then clipped this grid to

172 include only hexagons that contained breeding records from the dataset described above.

173 **Temperature records.** We accessed historical temperature data from meteorological

174 stations in the United States using the package `rnoaa`[53] and in Canada using the package

175 `weathercan`.[54] In both cases, we filtered stations to include those that were reported to have

176 at least 50 years of data between 1920 and 2020 (not all stations actually yielded 50 years of

177 data because some were active but missing the data we required). Using these criteria, we

178 identified 2,608 stations from the National Oceanic and Atmospheric Administration

179 (NOAA) and 1,125 stations from Environment and Climate Change Canada (ECCC) that

180 covered the spatial extent of our breeding records (Figure S2). From these stations we

181 downloaded all available daily high temperature records between 1920 and 2020. When

182 summarizing temperature data for the hexagonal grid, we averaged all stations that were

183 within the bounds of each hexagonal grid cell.

184 **Temperature anomalies.** We downloaded a monthly temperature anomaly raster from

185 the Berkeley Earth project (www.berkeleyearth.org/data/). This data product compares the

8

monthly average temperature within each 1° square grid to the average temperature over the period 1951-1980. Monthly anomalies in each year are expressed relative to that 30 year baseline period. Because we were interested in conditions during the breeding season, we extracted the anomaly for April, May, and June for each year from 1920-2020. We next clipped the square anomaly grid to the boundaries of each equal area hexagon and calculated a weighted average representing the anomaly within each hexagon and then merged these readings with the breeding records described above. We hereafter refer to this averaged value as the spring temperature anomaly.

Definitions of extreme climate events. In order to usefully summarize and analyze long-term climatic data, we had to make decisions about what counts as an extreme event. These decisions included choices about how to handle the severity, duration, and timing of cold-snaps and heat-waves. We followed two connected, but distinct strategies for the historical analysis of the timing of extreme events and for the analysis of the biological effects of extreme events on each species. For historical analyses, we needed to select absolute temperature values and we defined these based on the distribution of temperatures actually experienced by the 24 species we studied (see details below).

For the analyses focused on fitness effects for individual species, we were interested in the evidence for ECEs resulting from extreme temperatures. The definition of an ECE is inconsistent,[45,46] but typically these events refer to a combination of both extreme climatic conditions defined based on a distribution (e.g., events occurring <5% of the time) and a negative biological response, sometimes requiring a long recovery period.[45] No single definition is universally applicable across studies[46], but for the purposes of our study, we considered our results to indicate evidence for an ECE if 3-day cold or heat events two standard deviations from the species mean were associated with reduced reproductive success for each species.

The exact choice of a 3-day time period for considering cold-snaps and heat-waves was also

somewhat arbitrary, but previous work suggests that multi-day poor weather events tend to have a greater impact on offspring survival than single day events. This pattern has been observed in raptors,[55–57] in swallows,[38,58] and in other passerines.[34] We acknowledge that longer cold-snaps and heat-waves likely have even more severe impacts, but they will also be correspondingly rarer to observe when matched with breeding records and 3 days represents a compromise between a multi-day challenge and enough observation of extreme temperatures to allow us to model fitness effects.

Categorizing patterns of historical cold-snap and heat-wave occurrence. Using the averaged daily high temperature for each grid cell, we determined the date of the latest 3-day cold-snap and the date of the earliest 3-day heat-wave for each year and hexagon cell between 1920 and 2020. For this analysis, it was necessary to use absolute temperature values to define a cold-snap and heat-wave so that we could ask how the annual timing of the same temperature conditions may have changed over the past 100 years. We selected the threshold temperatures used by examining the distribution of temperatures recorded during nest attempts for the 24 species included in our study.

For each species, we determined the 5th, 10th, and 20th percentile of 3-day temperatures (for cold-snaps) along with the 80th, 90th, and 95th percentile of 3-day temperatures (for heat-waves) experienced during all the nesting attempts in our cleaned dataset (see below for details on nest level temperature). This resulted in 24 temperatures for each percentile value. We then averaged these species-specific estimates to arrive at a single threshold value for each percentile (Figure S3).

This summary resulted in cold-snap thresholds of 13.9°, 15.5°, and 17.4° C along with heat-wave thresholds of 32.1°, 33.6°, and 34.8° C. In each year, we only considered cold-snaps and heat-waves that occurred after the 60th and before the 240th day of the year (approximately March 1st to August 28th), because we were interested in the timing of these events in relation to breeding activities. We present all three thresholds in most analyses and

10

₂₃₈ in all cases the patterns were qualitatively similar regardless of threshold, but for some

₂₃₉ summary plots at the continental scale we used the milder thresholds (17.4° and 32.1°)

₂₄₀ because these allowed for the inclusion of a wider geographic area since the most extreme

₂₄₁ temperatures for cold and heat were rarely recorded at low and high latitudes, respectively.

₂₄₂ **Historical trends in cold-snap and heat-wave anomalies.** We derived a measure of

₂₄₃ yearly cold-snap and heat-wave timing anomaly for each grid cell to compare with the

₂₄₄ average temperature anomaly data described above. To do this we averaged the last

₂₄₅ cold-snap and earliest heat-wave date for the years 1951-1980 in each grid and then for each

₂₄₆ year and grid combination calculated the deviation from that average date. Negative values

₂₄₇ for the anomaly indicate years in which the last cold-snap or earliest heat-wave date

₂₄₈ occurred earlier than the historical average and positive values indicate years in which the

₂₄₉ last cold-snap or earliest heat-wave occurred later than the historical average. These values

₂₅₀ were calculated to be comparable to the temperature anomaly data described above.

₂₅₁ **Temperatures experienced by individuals during breeding.** Using the temperature

₂₅₂ data described above, we matched each breeding record to temperature from the most

₂₅₃ similar station using a two step process. We first matched records to the closest station. We

₂₅₄ next compared the difference in elevation between the breeding record and the station. If the

₂₅₅ closest station differed in elevation by >300 m, we searched for the station at the most

₂₅₆ similar elevation within 50 km. Using this approach, we paired records to weather stations

₂₅₇ that were 18.6 ± 10.5 km (SD) from the nest and within 47.2 ± 67.1 m elevation. While this

₂₅₈ approach resulted in the best matched station, we also note that standardized temperature

₂₅₉ records are usually recorded ~1.5m above the ground, whereas the species included in our

₂₆₀ dataset could have nested at varying heights and in open cups or nest boxes. Thus, the exact

₂₆₁ thermal environment for nests likely differed from the temperature records we used, but we

₂₆₂ could not account for this detailed level of variation.

₂₆₃ Using the identified station, we determined the coldest and hottest 3-day period experienced

11

during each nesting attempt separately for the incubation and nestling provisioning period. To accomplish this, we found the sequence of 3 days with the lowest combined daytime high temperature and considered this value as the lowest short term temperature exposure. For the highest temperature, we first determined the high temperature for each day in a string of 3-days; for each group of 3 days, we considered the lowest daytime high temperature and we then selected the string of 3-days that maximized this value (i.e., the hottest 3-day period experienced). It is important to note that these temperature determinations were continuous and, unlike the historical analysis above, they did not rely on any choice of threshold values. We used species specific timing for incubation and nestling stages to ensure that the temperatures we recorded would actually have been experienced during the reproductive attempt.

2.3 Data analysis

Spring temperature and cold-snap timing. We first modeled the change over the last 100 years in spring temperature anomalies and in the timing of cold-snaps and heat-waves across the spatial range of the breeding records included in our dataset. We were interested in determining how these temperature variables have changed over time, how much regional variation there is in those patterns, and the extent to which average temperature anomalies were correlated with cold-snap and heat-wave timing.

We initially plotted the average temperature, cold-snap, and heat-wave anomalies over the past 25 years (the time period covering our breeding records) in each of the hexagon grids described above for illustration purposes. We also fit a generalized additive model (GAM) for each anomaly measure using the entire time series with the anomaly in each grid-year combination as the response variable along with a basis smoothed predictor variable for year and a spatial smooth for latitude and longitude to account for spatial autocorrelation. We used these models to describe the overall change in spring temperatures and cold-snap or

12

heat-wave timings and the degree of spatial variation in those changes.

Next, we fit a GAM with the average timing of last cold-snap or earliest heat-wave anomaly for each grid over the past 25 years as the response variable and with the corresponding average spring temperature anomaly, and a spatial smooth for latitude and longitude as predictor variables. These models were used to infer whether spring temperature anomaly and the timing of the latest cold-snap or earliest heat-wave covaried, while accounting for spatial autocorrelation in the dataset.

Impact of cold-snaps and heat-waves on reproductive success. To model the impact of cold-snaps and heat-waves on reproductive success, we used a two-step approach to fit GAMs for each species separately for the incubation and nestling provisioning phase. First, as weather measurements were indicative of fine-scale regional conditions rather than those measured at specific nest locations, we calculated the response to weather conditions by averaging the number of chicks fledged grouped by hatch date and nearest weather station. We then created a model using these averaged records and 3-day temperature extremes experienced during the incubation period with mean number of chicks fledged as the response variable.

Next, we fit the same model for each species, but restricted records to those that hatched at least a single chick successfully and included only temperature records during the nestling provisioning stage. This allowed us to minimize the direct effects of temperature extremes during incubation on our models investigating the nestling phase. We separated between these two life history stages because we suspected that temperature extremes experienced during the nestling provisioning phase could have a stronger direct effect on fledging success, as the ability of the parent to buffer against extreme temperatures varies.[31] We implemented a Gaussian model for both the incubation and nestling provisioning models as they have lower false positive Type I error rates when compared to a Poisson due to overdispersion, and thus tend to be more conservative.[59]

13

In both models, the structure of the GAM was identical. Predictors included a basis smooth for the coldest and hottest 3-day period during either reproductive stage (incubation stage or provisioning) as well as a spatial smooth for latitude and longitude to account for spatial autocorrelation. The models also included smoothed predictors for date and a random effect for year. The basis dimension value of k for the spatial smooth was chosen iteratively by comparing the effective degrees of freedom (edf) with the k-index as per Wood 2017[60]. Predictors were checked for whether their smooths contributed unique information to the model (concurvity). We fit each model separately for each of the 24 species included in our analyses. Not all species had records in every year, so the number of years included in each model varies by species, and the basis dimension value was set to this in each model. To facilitate comparison between species, we standardized the coldest and hottest 3-day period within species so that the mean was zero and standard deviation was one for each species.

We summarized sensitivity to cold-snaps and heat-waves from these models by calculating pointwise estimates for number of offspring for each species during a 2 standard deviation cold-snap or heat-wave that occurred during either incubation or the nestling provisioning phase while controlling for the other model parameters. We then converted the estimates to model predicted relative fitness by taking the predicted number of offspring fledged divided by the average number of offspring fledged for each species. We considered point estimates with confidence intervals that did not overlap one (the average value of relative fitness) to indicate significant sensitivity to temperature extremes.

Species that experienced these reductions in fitness would be considered to have experienced an ECE based on the combined climatic extremity and biological response definition outlined by Smith, 2011.[45] However, Bailey & van de Pol (2016) argue that using an arbitrary distribution cutoff might miss important biological responses at different values or when responses differ between species[46] (in our case using 2 SD results in point estimates at approximately the 2.5th and 97.5th percentiles). In an online appendix, we include the full

14

relative fitness surface for each species across the range of coldest and hottest 3-day periods. These comparisons do not rely on any arbitrary choice of a single point estimate from the climatic distribution.

Latitudinal variation in cold-snap and heat-wave susceptibility.

After fitting the global models described above, we investigated whether the susceptibility to cold-snaps and heat-waves varied from the northern to southern limits of the breeding range in three species: purple martins (*Progne subis*), eastern bluebirds (*Sialia sialis*), and tree swallows (*Tachycineta bicolor*). We chose these three species because they had the largest sample sizes and because all three also have a wide latitudinal distribution and the global analysis above indicated susceptibility to both cold-snaps and heat-waves. For each species, we split the records into 5 latitude bands with an equal number of nest records per band and then repeated the global analyses exactly as described above within each band. The number of records was sufficient for these species that even after splitting into five datasets each latitude band included >9,000 nesting records for each species. For tree swallows, we limited this analysis to records east of the Rocky Mountains because mountain and western populations have very different breeding timing from eastern populations at similar latitudes.

2.4 Data and code availability

All data used in the analyses here are available from publicly accessible databases. Breeding records for individual nesting attempts can be accessed from each organization by filing a request and completing a data sharing agreement. Historical temperature data was accessed from the Berkeley Earth project (www.berkeleyearth.org/data/) or from meteorological station records maintained by NOAA in the United States and by ECCC in Canada.

All data processing, analyses, and figures were created in R version 4.0.2.[51] A complete set of annotated code to reproduce the full analysis, manuscript, and supplemental materials is

permanently archived on Zenodo (DOI: 10.5281/zenodo.6011738). The archived repository also includes an appendix with complete model summary tables for every GAM described in the results. Figure S4 provides a conceptual overview of the complete analysis pipeline with reference to each script in the code repository.
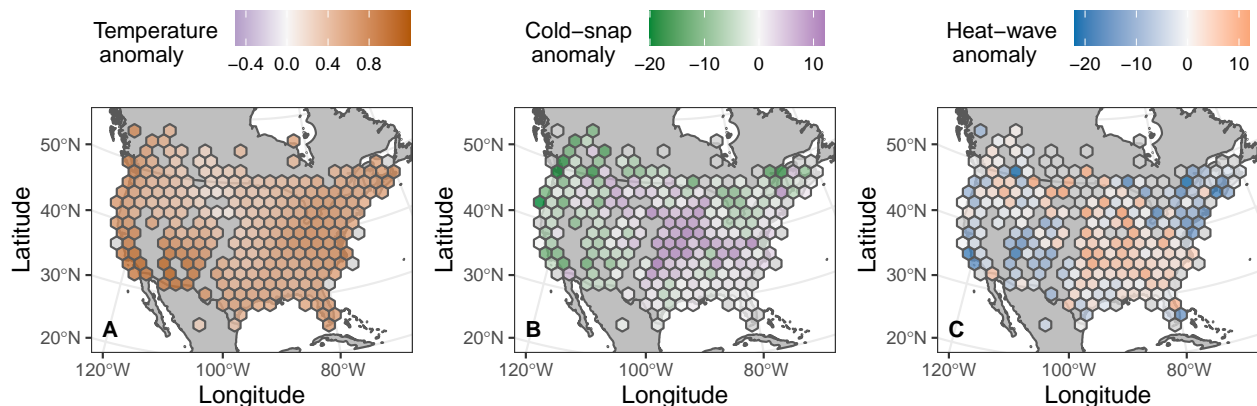
3 RESULTS

Figure 1: Spring temperature anomaly (A), cold-snap anomaly (B), and heat-wave anomaly (C) for the North American spatial extent studied over the years 1995-2020. Spring temperature anomaly is defined as the average of April, May, and June using data from the Berkeley Earth Project measured in degrees C compared to the baseline period of 1950-1980. Cold-snap anomaly is the difference in average timing of the latest three day period in which temperatures did not rise above 17.4 C from 1995-2020 compared to the average date of the latest cold-snap from 1950-1980, measured in days. Heat-wave anomaly is the difference in timing of the earliest three day period in which temperatures always rose above 32.1 C from 1995-2020 compared to the average date of the earliest heat-wave from 1950-1980, measured in days. Gray cells indicate regions with missing data.

3.1 Spring temperature and cold-snap or heat-wave timing

Across the spatial range that we studied, the average spring temperature anomaly from 1995 to 2020 was universally positive, although there was spatial variation in the magnitude of this increase in average spring temperature (Figure 1A). In contrast, change in the date of the latest 3-day cold-snap or the earliest 3-day heat-wave was inconsistent in both sign and magnitude (Figure 1B-C).

In a spatial GAM averaging across the entire range studied, spring temperature anomaly was consistently positive from 1995-2020 and the last cold-snap tended to occur 3 to 5 days earlier than the reference period regardless of the cold-snap threshold used (Figure 2A-B). However, the first heat-wave was not consistently different from the reference period regardless of the temperature threshold used (Figure 2C).
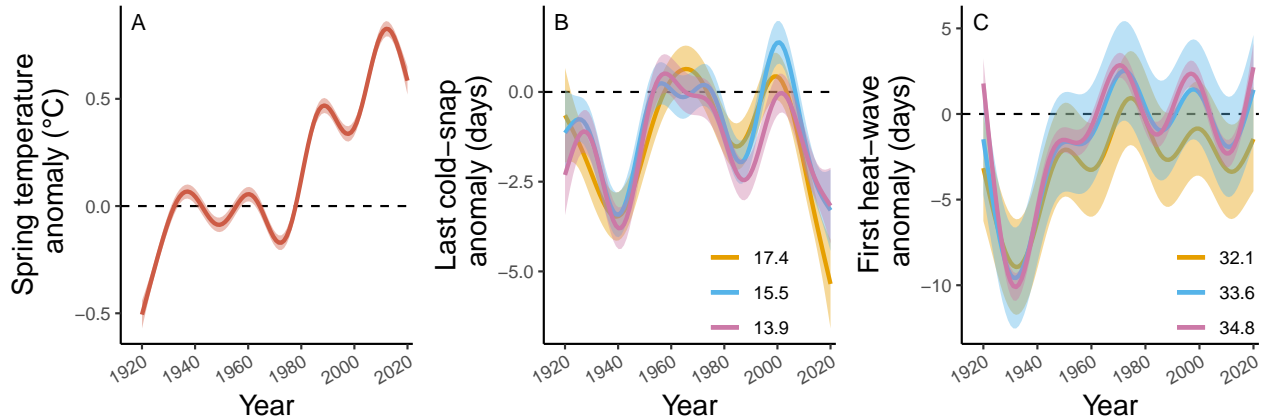

Figure 2: The historical change across the North American spatial extent studied in average spring temperature anomaly (A), timing of the last cold-snap (B), or timing of the first heat-wave (C) over the last 100 years. Lines and 95% confidence intervals are from spatial GAMs that account for latitude and longitude of each grid cell. For the cold-snap and heat-wave panels, three different threshold values are shown to illustrate increasingly more severe cold-snaps or heat-waves.

When considering average anomalies over the last 25 years, there was no evidence that a larger overall spring temperature anomaly was associated with any consistent difference in the change in cold-snap timing (Figure 3A) or heat-wave timing (Figure 3B). When comparing the temperature anomaly in each individual year at the level of hexagonal grids to the cold-snap or heat-wave anomaly in each individual grid-year, there was an association such that grid-years with higher spring temperature anomalies tended to have both earlier last cold-snaps and earlier heat-waves (Figure S5), but there was wide variation for individual grid-years.
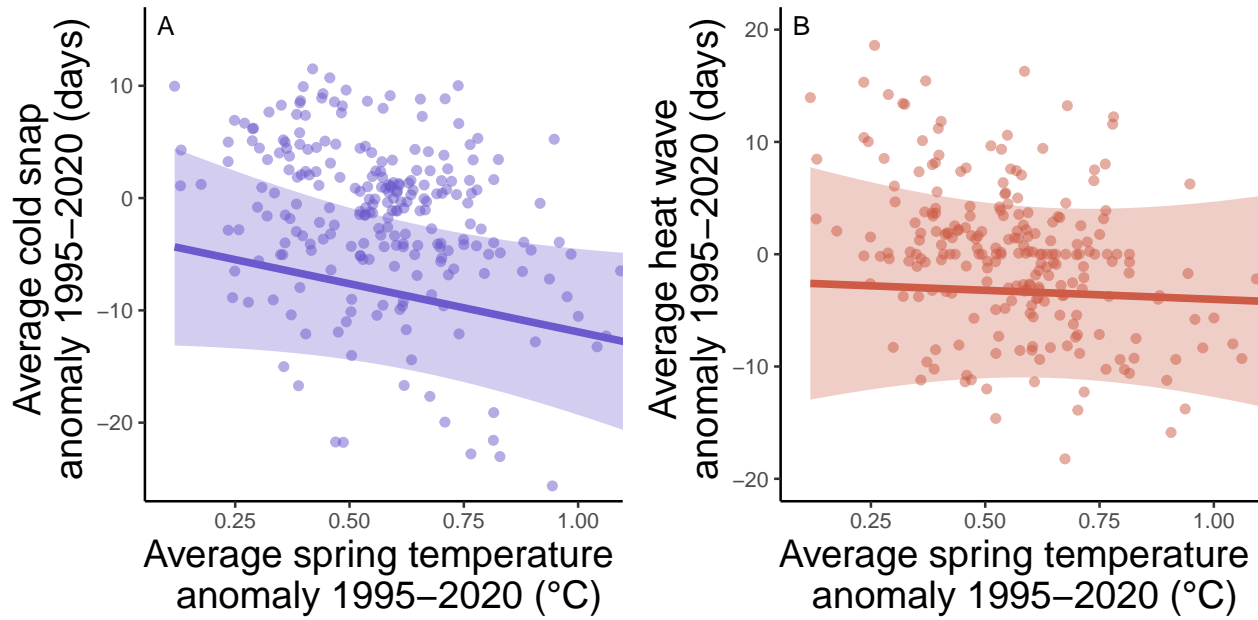
Figure 3: Relationship between average spring temperature anomaly for the North American spatial extent studied for each grid cell from 1995-2020 and (A) average cold-snap anomaly or (B) average heat-wave anomaly over the same period. Points are the average values for each hexagon grid cell. Lines and 95% confidence intervals are derived from GAMs that include a spatial smooth for latitude and longitude to account for spatial autocorrelation.

3.2 Impact of cold-snaps and heat-waves on reproductive success

Because of differences in range and breeding timing, the species in our dataset varied widely in their exposure to temperature during breeding (Figure 4A). During incubation, point estimates from GAMs that controlled for date, year, and location, indicated that a two standard deviation cold-snap reduced relative fitness in 8 out of 24 species (Figure 4B). The species that were sensitive to cold-snaps in this period included purple martins, tree swallows, eastern bluebirds, western bluebirds (*Sialia mexicana*), mountain bluebirds (*Sialia currocoides*), prothonotary warblers (*Protonotaria citrea*), Carolina chickadee (*Poecile carolinensis*), and mountain chickadees (*Poecile gambeli*). Relative fitness estimates ranged from 0.64 to 0.93 for these species; all values and confidence intervals for point estimates are included in Table S2). Only two of these same species also showed evidence of reduced fitness during a two SD incubation heat-wave (purple martin and eastern bluebird; Figure

18

4B, Table S2). No species showed evidence of significantly increased fitness from either cold-snap or heat-wave exposure during incubation.

During the nestling provisioning period, 75% of the species (6 of the 8) that were sensitive to cold-snaps in incubation also showed reduced relative fitness from two SD cold-snaps occurring after hatching (prothonotary warbler and mountain chickadee were no longer susceptible; Figure 4C). In addition to these six species, eastern phoebes (*Sayornis phoebe*), violet-green swallows (*Tachycineta thalassina*), and six other species had reduced fitness when a two SD cold-snap occurred during the provisioning period (Figure 4C). Overall, the effect sizes for reduction in relative fitness during provisioning were comparable to those seen during incubation (range of estimates for susceptible species was 0.74 to 0.94; Table S2). Nine of the 14 species susceptible to cold-snaps were also susceptible to heat-waves during provisioning (Figure 4C). Two additional species (brown-headed nuthatch: *Sitta pusilla* and chestnut-backed chickadee: *Poecile rufescens*) were susceptible to heat-waves but not cold-snaps during provisioning (Figure 4C). Similar to the incubation phase, no species showed evidence of significantly increased fitness from either cold-snap or heat-wave exposure during the nestling phase.
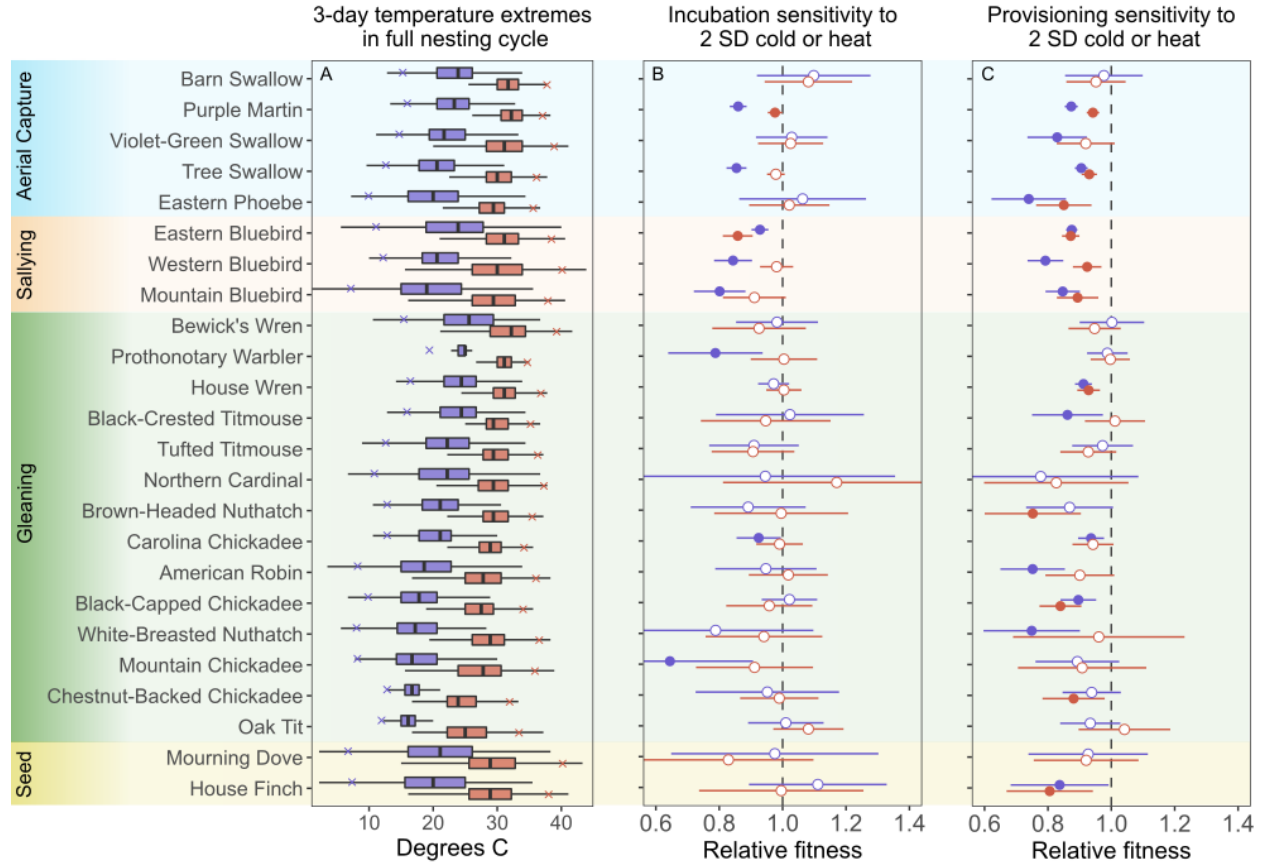
Figure 4: Exposure and sensitivity to cold-snaps and heat-waves across 24 species. (A) Distribution of the coldest and hottest 3-day high temperatures experienced during the full nesting cycle for all species sorted by the coldest average temperature within each foraging style. Boxplots indicate median, IQR, and 1.5 times IQR for the coldest (blue) and hottest (red) 3-day period. Blue and red x marks indicate points that are 2 SD below (cold) or above (hot) the mean exposure. (B and C) Predicted relative fitness for breeding attempts that experienced a 2 SD cold-snap (blue) or 2 SD heat-wave (red) during incubation (B) or provisioning (C). Point estimates are derived from GAMs that control for spatial patterns, year, and date. Circles show point estimate, lines show 95% CI; open circles indicate no difference from mean reproductive success while closed circles indicate a significant effect of temperature on relative reproductive success. Color bands show the primary foraging strategy that each species uses to provision nestlings.

3.3 Latitudinal variation in temperature exposure and susceptibility

We examined latitudinal trends in susceptibility to cold-snaps and heat-waves for eastern bluebirds, purple martins, and tree swallows. In all three species, breeding date was later farther north, but breeding attempts from northern areas still experienced lower 3-day

20

coldest and 3-day hottest temperatures on average (Figure 5A-B, Figure S6A-B, and Figure S7A-B). While the overall patterns of susceptibility to cold-snaps and heat-waves were largely similar across these species (Figure 4B-C), the latitudinal patterns differed somewhat for each species. During both incubation and provisioning, eastern bluebirds had reduced fitness from cold-snaps only near the northern edge of the range (Figure 5C-D; full details on point estimates in Table S3). In contrast, relative fitness was reduced from heat-waves over a wider, but somewhat inconsistent latitudinal extent (Figure 5C-D).

Purple martins susceptibility to heat-waves was only apparent near the southern edge of the range and was more pronounced in provisioning than during the incubation period (Figure S6C-D). Unlike eastern bluebirds, purple martins had reduced relative fitness from cold-snaps during incubation and provisioning at nearly every latitude band. Only the southernmost two bands during provisioning showed no impact on fitness associated with cold-snaps (Figure S6C-D; Table S4).

Like purple martins, tree swallows had consistently reduced relative fitness when cold-snaps occurred during incubation (Figure S7C). Despite an aggregate effect of heat-waves during incubation (Figure 4C), there was no signal for heat-wave effects during incubation or provisioning in any individual latitude band (Figure S7C-D). During provisioning, tree swallows only showed a sensitivity to cold-snaps at the northern range limit and no clear susceptibility to heat-waves (Figure S7D; Table S5).
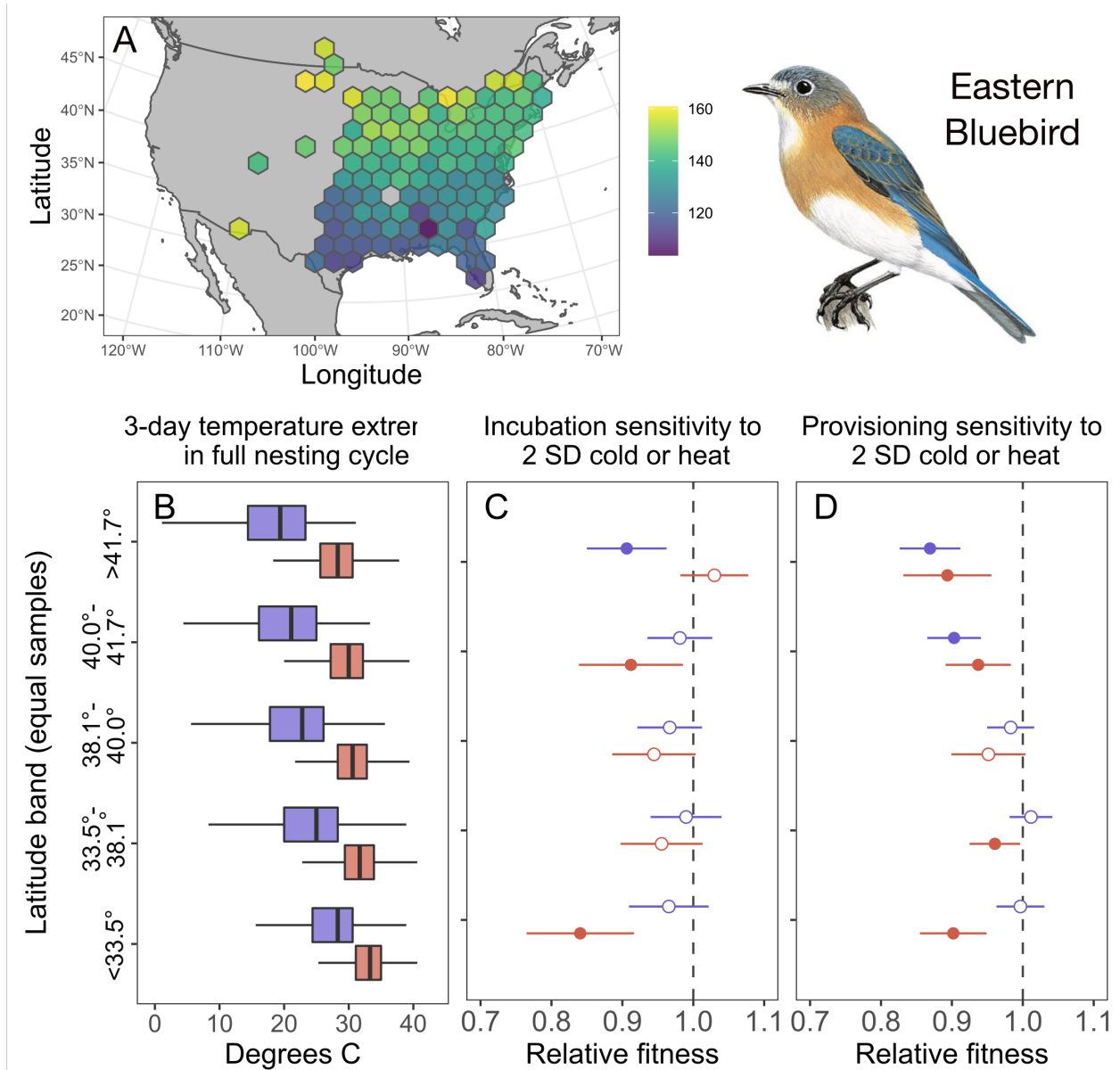
Figure 5: Breeding timing, temperature exposure, and extreme temperature sensitivity in eastern bluebirds. (A) Average date of first egg laying across the records included. (B) Distribution of the coldest and hottest 3-day high temperatures in the full nesting cycles in five latitude bands with an equal number of samples per band. The top of the panel is the northernmost band and bottom is the southernmost. Boxplots indicate median, IQR, and 1.5 times IQR for the coldest (blue) and hottest (red) 3-day period. (C and D) Predicted relative fitness for breeding attempts that experienced a 2 SD cold-snap (blue) or 2 SD heat-wave (red) during incubation (C) or provisioning (D) for each latitude band. Point estimates are derived from GAMs that control for spatial patterns, year, and date. Circles show point estimate, lines show 95 percent CI; open circles indicate no difference from mean reproductive success while closed circles indicate a significant effect of temperature on relative reproductive success. Illustration from Handbook of the Birds of the World by Ian Willis, copyright Lynx Edicions.

4 DISCUSSION

Using community scientist collected breeding records, we show that exposure to three-day cold-snaps during a breeding attempt is associated with reduced relative fitness in 16 of 24 common species and that exposure to three-day heat-waves is associated with reduced fitness in 11 of 24 species. In three widespread species, sensitivity to cold-snaps and heat-waves was spatially heterogeneous, indicating a mechanism by which climate change might generate population trends that differ across a species range. Historical temperature records for the same area showed that the date of the latest cold-snap is only slightly earlier and the date of the earliest heat-wave has not changed consistently over the past 70 years despite warmer springs overall. Taken together, our results demonstrate that the likelihood of encountering cold-snaps or heat-waves during breeding might differ as a consequence of climate change and that exposure to these events often results in reduced reproductive success for many common and widespread species . Moreover, in many cases it is impossible to track both average temperature and temperature variability through phenological shifts in the timing of breeding. As a consequence, species that are most sensitive to extreme temperature events may be less able to adapt to changes in average temperature. Within species, variation in sensitivity at northern and southern range limits might contribute to range shifts and local population declines. Understanding the mechanisms that result in different sensitivity to temperature variability within and between species should help accurately predict which populations are most vulnerable to climate change.

Across the range of breeding records that we studied, average spring temperatures have increased over the last 70 years, but the average timing of latest cold-snaps and earliest heat-waves during the breeding season has not changed consistently in sign or magnitude. As a consequence, the historical association between average temperature and temperature variability differs from the association observed over the last 25 years. Thus, cues that wild birds use to time seasonal events may not contain the same information about average

23

temperature and temperature variability that they did historically. Changes in breeding phenology have been identified as one of the 'universal responses' to climate change.[61,62] In some cases, these shifts can minimize the effects of phenological mismatch with food resources,[6,63–65] but in many cases the degree of change in breeding timing seems inadequate to keep pace with average temperatures.[23,26,27] These inadequate shifts may be attributable to a lack of ability for rapid evolutionary change or incomplete phenotypic plasticity.[66,67] However, average breeding temperature and food availability are not the only gradients that could be disrupted by climate change.[29] For species that are especially susceptible to cold-snaps or heat-waves during breeding, changes in the timing of breeding events to track average temperature might result in increased exposure to temperature variability.[31,34] Our results highlight the consequences of temperature variation and suggest that in some cases incomplete plasticity and a degree of phenological mismatch may represent an adaptive response that balances competing selection pressures.[68]

We predicted that the consequences of cold-snaps might be especially severe during the nestling provisioning period and for species that rely on insect activity for foraging. A recent meta-analysis demonstrated that environmental insect food abundance is a strong predictor of nestling body condition and fledging success, especially in species in which insects make up a large component of the diet.[69] Daily total flying insect biomass and emergence rates are strongly influenced by temperature,[38,70] so cold-snaps can have immediate and direct effects on food availability. These impacts on food availability could compound the increased thermoregulatory challenges and associated increase in energy use during cold-snaps. We did find that more species were sensitive to cold-snaps during nestling provisioning, although several species were sensitive in both stages (6 species in both stages, 2 species in incubation only, 8 species in provisioning only). The 14 species that were sensitive to cold during provisioning included four of the five obligate aerial insectivores (purple martin, tree swallow, violet-green-swallow, and eastern phoebe) and all three species that rely extensively on insect movement for sally-capture foraging (eastern, western, and mountain bluebird). However,

white-breasted nuthatches (*Sitta carolinensis*) and American robins (*Turdus migratorius*) also showed declines despite a less obvious link to insect availability and the other aerial insectivore did not exhibit sensitivity (barn swallow, *Hirundo rustica*). Thus our results suggest the possibility of foraging mode as a mechanism generating susceptibility to cold-snaps during provisioning, but more data for a broader distribution of phylogenetically independent species with different foraging styles would be needed to evaluate this prediction convincingly.

In contrast to cold-snaps, we predicted that the primary challenge from heat-waves would be their impact on thermoregulation, rather than indirect effects of food availability. For example, several studies demonstrate that heat dissipation rate may constrain reproductive investment[71,72] and hot conditions increase cooling costs.[73] Therefore, we expected to find more widespread evidence for heat-wave impacts during both incubation and provisioning. While we did find fairly widespread effects of heat-waves, fewer species were sensitive to heat-waves than to cold-snaps (11 versus 18 species across both stages). For nine of these species the sensitivity was only observed in one breeding stage. Moreover, the effect sizes for reduction in relative fitness from heat-waves were generally smaller than those observed for cold-snaps. While natural heat-waves during breeding can clearly drive reduced fitness in some cases,[57,74] over the range and set of species we studied, cold-snaps seemed to generate more consistent and stronger reductions in breeding success even though spring breeding conditions are warmer overall than they were historically. One reason for these smaller effects might be that the species in our study were able to mitigate the risks of hyperthermia by changing behaviors or by choosing nesting sites that were cooler than the overall temperature recorded from nearby weather stations. It is also possible that a different characterization of heat-waves, such as one using average rather than daily high temperature or a longer time period than three days, might have detected more vulnerability.

In one sense, it is surprising that we did not find universal evidence for cold-snap and

heat-wave consequences, because ultimately it is clear that a sufficiently severe temperature extreme would reduce fitness. Several non-mutually exclusive factors likely explain the lack of evidence for sensitivity that we observed in some species. First, species ranges are shaped, in part, by direct and indirect effects of temperature.[75] Because we took the approach of generating point estimates that were 2 SD from the mean of the observed breeding records for each species, the absolute value of temperature challenges that we modeled differed across species. While our approach explores the realized fitness declines from actual temperature exposure, manipulations or extensions beyond these temperature extremes would eventually uncover sensitivities in all species. Many experimental studies have demonstrated physiological or fitness costs from cooling[76,77] or heating[78,79] nest boxes during breeding, but these often employ more sustained or extreme temperatures than the natural variation we modeled. On a related note, some species with relatively restricted ranges limited by temperature exposure might be highly sensitive to temperature extremes, but rarely experience those conditions during breeding, preventing us from detecting effects of temperature challenges. Pigot et al. 2010 argued that widespread species range limits are more likely to be defined by temperature, so it may be more common to detect sensitivities to temperature extremes in those same widespread species[75]. Finally, the community science datasets that we used yielded highly variable sample sizes and do not necessarily include representative observations across the range of most species. For widespread and easily observed species it is likely easier to detect effects both because of larger sample sizes and because of more dispersed sampling across the extent of each species range. Samples near the range limits might be especially important for detecting sensitivity to temperature extremes and are not equally available for all species in our dataset (e.g., bluebird trails and nest boxes make monitoring especially easy and widespread for these species).

The results from our more detailed analysis of records from eastern bluebirds, purple martins, and tree swallows do indeed suggest that within-species spatial dynamics in sensitivity to temperature variability may be highly relevant. Eastern bluebirds were

26

vulnerable to heat-waves over most of their range especially during provisioning, but were only sensitive to cold-snaps near the northern range limit. Purple martins were vulnerable to cold-snaps across their range during incubation and in the northern half of the range during provisioning, but were only sensitive to heat-waves near their southern range limit. Tree swallows had no clear sensitivity to heat-waves within narrow latitude bands; they were sensitive to cold-snaps in incubation across most of their range, but only sensitive to cold-snaps during provisioning near their northern range limit. We did not quantify overall abundance changes in our study, but it is interesting to note that since 2007 purple martins have declined near their southern range limit and tree swallows have declined most precipitously near their northern range limit while expanding their range southward.[80] These patterns qualitatively match with the range limits near which we found differences in sensitivity for each species. In contrast, eastern bluebirds have declined west of the Appalachian Mountains and increased on the Atlantic seaboard over the same period, so it is less clear how the patterns might relate to the sensitivity differences we found.[80] Nevertheless, the fact that all three species show at least some evidence of increased sensitivity to temperature near a range limit suggests that exposure to increased temperature variability might be an important contributor to range shifts with climate change. However, despite the fact that all three species have similar body sizes, breeding behavior, and distributions, with two being fairly close relatives that are very similar ecologically, there was no common latitudinal pattern. Presumably these idiosyncratic responses are driven by subtle differences, such as exposure to multiple additive stressors, life history details, or habitat requirements. Thus, predicting differential sensitivity to temperature variability will likely require a detailed understanding of the ecology for each species.

One important limitation on the conclusions of our study is that our data only allowed us to model overall fledging success, but it is likely that temperature variability also generates sub-lethal effects that could have long term consequences on population demographics. Early developmental conditions, including temperature, are well known to have a wide variety of

27

long lasting effects on wild birds, such as changes in body size and morphology, physiology, immunology, and survival to recruitment as a breeding adult.[81] Indeed, long term declines in body size is another widespread consequence of global warming.[5] While changes in body size may sometimes result from rapid evolutionary responses,[4] many of the observed changes in body size could also be explained by changes in developmental temperature.[82] For example, in tree swallows fledgling body size is positively correlated with average developmental temperature and predicts both adult body size and likelihood of recruiting.[82] Thus, cold-snaps or heat-waves that are not severe enough to reduce fledging success could still have important consequences on the demographics of bird populations through their long term effects on nestlings. Our results should therefore be considered as a minimum measure of the fitness costs from experiencing extreme temperature conditions.

A great deal of research has focused on the effects of climate change on wild bird populations over the past several decades. While the large-scale patterns of climate change on ranges, phenology, and morphology have been described for many species, studies that focus on these patterns are often unable to characterize the organismal and ecological processes that are operating. At the same time, targeted studies that focus on individually marked birds or experimental manipulations often isolate mechanisms but cannot directly link their results to population and range wide consequences. Our results add to a growing number of studies that seek to make connections from individual-to-population level effects[25] in an effort to understand how the sensitivity of populations to specific conditions ultimately creates larger patterns that may differ between species. Moreover, we highlight the fact that average temperature and temperature variability are both important and that they-along with other gradients-must be considered simultaneously in order to understand the impacts of climate change. Because many climate change models predict increasingly extreme weather in addition to overall warming, one of the challenges in predicting vulnerability for bird species moving forward is to understand the extent to which average conditions versus extreme events drive demographic consequences for populations as conditions change.

AUTHOR CONTRIBUTIONS

CCT & JRS contributed equally to conception and analysis of the manuscript. CCT drafted the paper with input from JRS and both authors contributed to revisions.

ACKNOWLEDGMENTS

REFERENCES

1.

NOAA. NOAA national centers for environmental information, state of the climate: Global climate report for october 2021, published online november 2021, retrieved on november 24, 2021 from https://www.ncdc.noaa.gov/sotc/global/202110. (2021).

2.

Parmesan, C. & Yohe, G. A globally coherent fingerprint of climate change impacts across natural systems. *Nature* **421**, 37–42 (2003).

3.

Ozgul, A. *et al.* The dynamics of phenotypic change and the shrinking sheep of st. kilda. *Science* **325**, 464–467 (2009).

4.

Weeks, B. C. *et al.* Shared morphological consequences of global warming in north american migratory birds. *Ecology Letters* **23**, 316–325 (2020).

5.

Gardner, J. L., Peters, A., Kearney, M. R., Joseph, L. & Heinsohn, R. Declining body size: A third universal response to warming? *Trends in Ecology & Evolution* **26**, 285–291 (2011).

6.

Charmantier, A. *et al.* Adaptive phenotypic plasticity in response to climate change in a wild bird population. *Science* **320**, 800–803 (2008).

7.

Youngflesh, C. *et al.* Migratory strategy drives species-level variation in bird sensitivity to vegetation green-up. *Nature Ecology & Evolution* **5**, 1–8 (2021).

8.

Horton, K. G. *et al.* Phenology of nocturnal avian migration has shifted at the continental scale. *Nature Climate Change* **10**, 63–68 (2020).

9.

Blackford, C., Germain, R. M. & Gilbert, B. Species differences in phenology shape coexistence. *The American Naturalist* **195**, E168–E180 (2020).

10.

Samplonius, J. M. *et al.* Strengthening the evidence base for temperature-mediated phenological asynchrony and its impacts. *Nature Ecology & Evolution* **5**, 155–164 (2021).

11.

Cohen, J. M., Lajeunesse, M. J. & Rohr, J. R. A global synthesis of animal phenological responses to climate change. *Nature Climate Change* **8**, 224–228 (2018).

12.

Both, C., Van Asch, M., Bijlsma, R. G., Van Den Burg, A. B. & Visser, M. E. Climate change and unequal phenological changes across four trophic levels: Constraints or adaptations? *Journal of Animal Ecology* **78**, 73–83 (2009).

13.

Shipley, J. R. *et al.* Climate change shifts the timing of nutritional flux from aquatic insects. *Current Biology* **32**, 1342–1349 (2022).

14.

Vasseur, D. A. *et al.* Increased temperature variation poses a greater risk to species than climate warming. *Proceedings of the Royal Society B: Biological Sciences* **281**, 20132612 (2014).

15.

Thackeray, S. J. *et al.* Phenological sensitivity to climate across taxa and trophic levels. *Nature* **535**, 241–245 (2016).

16.

Rosenberg, K. V. *et al.* Decline of the north american avifauna. *Science* **366**, 120–124 (2019).

17.

Loss, S. R., Will, T. & Marra, P. P. The impact of free-ranging domestic cats on wildlife of the united states. *Nature communications* **4**, 1–8 (2013).

18.

Senzaki, M. *et al.* Sensory pollutants alter bird phenology and fitness across a continent. *Nature* **587**, 605–609 (2020).

19.

Julliard, R., Jiguet, F. & Couvet, D. Common birds facing global changes: What makes a species at risk? *Global Change Biology* **10**, 148–154 (2004).

20.

Robinson, R. A. *et al.* Emerging infectious disease leads to rapid population declines of common british birds. *PLoS One* **5**, e12215 (2010).

21.

Møller, A. P., Rubolini, D. & Lehikoinen, E. Populations of migratory bird species that did not show a phenological response to climate change are declining. *Proceedings of the National Academy of Sciences* **105**, 16195–16200 (2008).

22.

Franks, S. E. *et al.* The sensitivity of breeding songbirds to changes in seasonal timing is linked to population change but cannot be directly attributed to the effects of trophic asynchrony on productivity. *Global Change Biology* **24**, 957–971 (2018).

23.

Devictor, V., Julliard, R., Couvet, D. & Jiguet, F. Birds are tracking climate warming, but not fast enough. *Proceedings of the Royal Society B: Biological Sciences* **275**, 2743–2748 (2008).

24.

Both, C., Bouwhuis, S., Lessells, C. & Visser, M. E. Climate change and population declines in a long-distance migratory bird. *Nature* **441**, 81–83 (2006).

25.

Socolar, J. B., Epanchin, P. N., Beissinger, S. R. & Tingley, M. W. Phenological shifts conserve thermal niches in north american birds and reshape expectations for climate-driven range shifts. *Proceedings of the National Academy of Sciences* **114**, 12976–12981 (2017).

26.

Simmonds, E. G., Cole, E. F., Sheldon, B. C. & Coulson, T. Phenological asynchrony: A ticking time-bomb for seemingly stable populations? *Ecology Letters* **23**, 1766–1775 (2020).

27.

Radchuk, V. *et al.* Adaptive responses of animals to climate change are most likely insufficient. *Nature Communications* **10**, 1–14 (2019).

28.

Gienapp, P. *et al.* Predicting demographically sustainable rates of adaptation: Can great tit breeding time keep pace with climate change? *Philosophical Transactions of the Royal Society B: Biological Sciences* **368**, 20120289 (2013).

29.

Spence, A. R. & Tingley, M. W. The challenge of novel abiotic conditions for species undergoing climate-induced range shifts. *Ecography* **43**, 1571–1590 (2020).

30.

DeGregorio, B. A., Westervelt, J. D., Weatherhead, P. J. & Sperry, J. H. Indirect effect of climate change: Shifts in ratsnake behavior alter intensity and timing of avian nest predation. *Ecological Modelling* **312**, 239–246 (2015).

31.

Shipley, J. R. *et al.* Birds advancing lay dates with warming springs face greater risk of chick mortality. *Proceedings of the National Academy of Sciences* **117**, 25590–25594 (2020).

32.

Marrot, P., Garant, D. & Charmantier, A. Multiple extreme climatic events strengthen selection for earlier breeding in a wild passerine. *Philosophical Transactions of the Royal Society B: Biological Sciences* **372**, 20160372 (2017).

33.

Townsend, A. K. *et al.* The interacting effects of food, spring temperature, and global climate cycles on population dynamics of a migratory songbird. *Global Change Biology* **22**, 544–555 (2016).

34.

Regan, C. E. & Sheldon, B. C. Phenotypic plasticity increases exposure to extreme climatic events that reduce individual fitness. *Global Change Biology* **29**, 2968–2980 (2023).

35.

Kearney, M. R., Jusup, M., McGeoch, M. A., Kooijman, S. A. & Chown, S. L. Where do functional traits come from? The role of theory and models. *Functional Ecology* **35**, 1385–1396 (2021).

36.

Belitz, M. *et al.* Climate drivers of adult insect activity are conditioned by life history traits. *Ecology Letters* **24**, 2687–2699 (2021).

37.

Riddell, E. *et al.* Exposure to climate change drives stability or collapse of desert mammal and bird communities. *Science* **371**, 633–636 (2021).

38.

Winkler, D. W., Luo, M. K. & Rakhimberdiev, E. Temperature effects on food supply and chick mortality in tree swallows (tachycineta bicolor). *Oecologia* **173**, 129–138 (2013).

39.

Albright, T. P. *et al.* Combined effects of heat waves and droughts on avian communities across the conterminous united states. *Ecosphere* **1**, 1–22 (2010).

40.

McKechnie, A. E. & Wolf, B. O. Climate change increases the likelihood of catastrophic avian mortality events during extreme heat waves. *Biology Letters* **6**, 253–256 (2010).

41. Visser, M. E., Lindner, M., Gienapp, P., Long, M. C. & Jenouvrier, S. Recent natural variability in global warming weakened phenological mismatch and selection on seasonal timing in great tits (parus major). *Proceedings of the Royal Society B* **288**, 20211337 (2021).

42. Jirinec, V. *et al.* Morphological consequences of climate change for resident birds in intact amazonian rainforest. *Science Advances* **7**, eabk1743 (2021).

43. Freeman, B. G., Scholer, M. N., Ruiz-Gutierrez, V. & Fitzpatrick, J. W. Climate change causes upslope shifts and mountaintop extirpations in a tropical bird community. *Proceedings of the National Academy of Sciences* **115**, 11982–11987 (2018).

44. Thomas, C. D. & Lennon, J. J. Birds extend their ranges northwards. *Nature* **399**, 213–213 (1999).

45. Smith, M. D. An ecological perspective on extreme climatic events: A synthetic definition and framework to guide future research. *Journal of Ecology* **99**, 656–663 (2011).

46. Bailey, L. D. & Pol, M. van de. Tackling extremes: Challenges for ecological and evolutionary research on extreme climatic events. *Journal of Animal Ecology* **85**, 85–96 (2016).

47.

Williams, H. M., Siegrist, J. & Wilson, A. M. Support for a relationship between demography and modeled habitat suitability is scale dependent for the purple martin progne subis. *Journal of Animal Ecology* **90**, 356–366 (2021).

48.

Billerman, S. M., Keeney, B. K., Rodewald, P. G. & Schulenberg, T. S. *Birds of the world.* (https://birdsoftheworld-org.proxy.library.cornell.edu/bow/home; Cornell Laboratory of Ornithology, Ithaca, NY USA, 2020).

49.

Hollister, J., Shah, T., Robitaille, A. L., Beck, M. W. & Johnson, M. *Elevatr: Access elevation data from various APIs.* (2020).

50.

South, A. *Rnaturalearth: World map data from natural earth.* (2017).

51.

R Core Team. *R: A language and environment for statistical computing.* (R Foundation for Statistical Computing, 2020).

52.

Pebesma, E. Simple Features for R: Standardized Support for Spatial Vector Data. *The R Journal* **10**, 439–446 (2018).

53.

Chamberlain, S. *Rnoaa: 'NOAA' weather data from r.* (2020).

54.

LaZerte, S. E. & Albers, S. weathercan: Download and format weather data from environment and climate change canada. *The Journal of Open Source Software* **3**, 571 (2018).

55.

Fisher, R. J. *et al.* Extreme precipitation reduces reproductive output of an endangered raptor. *Journal of Applied Ecology* **52**, 1500–1508 (2015).

56.

G. MCDonald, P., Olsen, P. D. & Cockburn, A. Weather dictates reproductive success and survival in the australian brown falcon falco berigora. *Journal of Animal Ecology* **73**, 683–692 (2004).

57.

Corregidor-Castro, A. *et al.* Experimental nest cooling reveals dramatic effects of heatwaves on reproduction in a mediterranean bird of prey. *Global Change Biology* (2023).

58.

Brown, C. R. & Brown, M. B. Intense natural selection on body size and wing and tail asymmetry in cliff swallows during severe weather. *Evolution* **52**, 1461–1475 (1998).

59.

Knief, U. & Forstmeier, W. Violating the normality assumption may be the lesser of two evils. *Behavior Research Methods* **53**, 2576–2590 (2021).

60.

Wood, S. N. *Generalized additive models: An introduction with r.* (CRC press, 2017).

61.

Kharouba, H. M. *et al.* Global shifts in the phenological synchrony of species interactions over recent decades. *Proceedings of the National Academy of Sciences* **115**, 5211–5216 (2018).

62.

Horton, K. G., Morris, S. R., Van Doren, B. M. & Covino, K. M. Six decades of north american bird banding records reveal plasticity in migration phenology. *Journal of Animal Ecology* **92**, 738–750 (2023).

63.

Iler, A. M. *et al.* Maintenance of temporal synchrony between syrphid flies and floral resources despite differential phenological responses to climate. *Global Change Biology* **19**, 2348–2359 (2013).

64.

Vatka, E., Orell, M. & Rytkönen, S. Warming climate advances breeding and improves synchrony of food demand and food availability in a boreal passerine. *Global Change Biology* **17**, 3002–3009 (2011).

65.

Twining, C. W., Shipley, J. R. & Matthews, B. Climate change creates nutritional phenological mismatches. *Trends in Ecology & Evolution* **37**, 736–739 (2022).

66.

Both, C. & Visser, M. E. Adjustment to climate change is constrained by arrival date in a long-distance migrant bird. *Nature* **411**, 296–298 (2001).

67.

Visser, M. E., Both, C. & Lambrechts, M. M. Global climate change leads to mistimed avian reproduction. *Advances in Ecological Research* **35**, 89–110 (2004).

68.

Visser, M. E., Marvelde, L. te & Lof, M. E. Adaptive phenological mismatches of birds and their food in a warming world. *Journal of Ornithology* **153**, 75–84 (2012).

69.

Grames, E. M., Montgomery, G. A., Youngflesh, C., Tingley, M. W. & Elphick, C. S. The effect of insect food availability on songbird reproductive success and chick body condition: Evidence from a systematic review and meta-analysis. *Ecology Letters* **26**, 658–673 (2023).

70.

Dunn, P. O. *et al.* Extensive regional variation in the phenology of insects and their response to temperature across n orth a merica. *Ecology* **104**, e4036 (2023).

71.

Nord, A. & Nilsson, J.-Å. Heat dissipation rate constrains reproductive investment in a wild bird. *Functional Ecology* **33**, 250–259 (2019).

72.

Tapper, S., Nocera, J. J. & Burness, G. Heat dissipation capacity influences reproductive performance in an aerial insectivore. *Journal of Experimental Biology* **223**, jeb222232 (2020).

73.

Riddell, E., Iknayan, K., Wolf, B., Sinervo, B. & Beissinger, S. Cooling requirements fueled the collapse of a desert bird community from climate change. *Proceedings of the National Academy of Sciences* **116**, 21609–21615 (2019).

74.

Piatt, J. F. *et al.* Extreme mortality and reproductive failure of common murres resulting from the northeast pacific marine heatwave of 2014-2016. *PloS One* **15**, e0226087 (2020).

75.

Pigot, A. L., Owens, I. P. & Orme, C. D. L. The environmental limits to geographic range expansion in birds. *Ecology Letters* **13**, 705–715 (2010).

76.

Ardia, D. R., Pérez, J. H. & Clotfelter, E. D. Experimental cooling during incubation leads to reduced innate immunity and body condition in nestling tree swallows. *Proceedings of the Royal Society B: Biological Sciences* **277**, 1881–1888 (2010).

77. Nilsson, J. F., Stjernman, M. & Nilsson, J.-Ä. Experimental reduction of incubation temperature affects both nestling and adult blue tits cyanistes caeruleus. *Journal of Avian Biology* **39**, 553–559 (2008).

78. Pérez, J. H., Ardia, D. R., Chad, E. K. & Clotfelter, E. D. Experimental heating reveals nest temperature affects nestling condition in tree swallows (Tachycineta bicolor). *Biology letters* **4**, 468–471 (2008).

79. Andreasson, F., Nord, A. & Nilsson, J.-Ä. Experimentally increased nest temperature affects body temperature, growth and apparent survival in blue tit nestlings. *Journal of Avian Biology* **49**, jav–01620 (2018).

80. Fink, D. *et al.* eBird status and trends, data version: 2021; released: 2022. *Cornell Lab of Ornithology, Ithaca, New York* (2022) doi:https://doi.org/10.2173/ebirdst.2021.

81. Lindström, J. Early development and fitness in birds and mammals. *Trends in ecology & evolution* **14**, 343–348 (1999).

82. Shipley, J. R., Twining, C. W., Taff, C. C., Vitousek, M. N. & Winkler, D. W. Selection counteracts developmental plasticity in body-size responses to climate change. *Nature Climate Change* **12**, 863–868 (2022).

SUPPLEMENTAL TABLES AND FIGURES

Table S1: Total number of breeding records for each included species.

| Common name	Scientific name	Records
American Robin	*Turdus migratorius*	2101
Barn Swallow	*Hirundo rustica*	1003
Bewick's Wren	*Thryomanes bewickii*	1421
Black-Capped Chickadee	*Poecile atricapillus*	2993
Black-Crested Titmouse	*Baeolophus atricristatus*	527
Brown-Headed Nuthatch	*Sitta pusilla*	375
Carolina Chickadee	*Poecile carolinensis*	5287
Chestnut-Backed Chickadee	*Poecile rufescens*	951
Eastern Bluebird	*Sialia sialis*	80447
Eastern Phoebe	*Sayornis phoebe*	998
House Finch	*Haemorphous mexicanus*	627
House Wren	*Troglodytes aedon*	16946
Mountain Bluebird	*Sialia currocoides*	9051
Mountain Chickadee	*Poecile gambeli*	526
Mourning Dove	*Zenaida macroura*	524
Northern Cardinal	*Cardinalis cardinalis*	512
Oak Tit	*Baeolophus inornatus*	1066
Prothonotary Warbler	*Protonotaria citrea*	2121
Purple Martin	*Progne subis*	85368
Tree Swallow	*Tachycineta bicolor*	64222
Tufted Titmouse	*Baeolophus bicolor*	1224
Violet-Green Swallow	*Tachycineta thalassina*	2410
Western Bluebird	*Sialia mexicana*	20335
White-Breasted Nuthatch	*Sitta carolinensis*	479
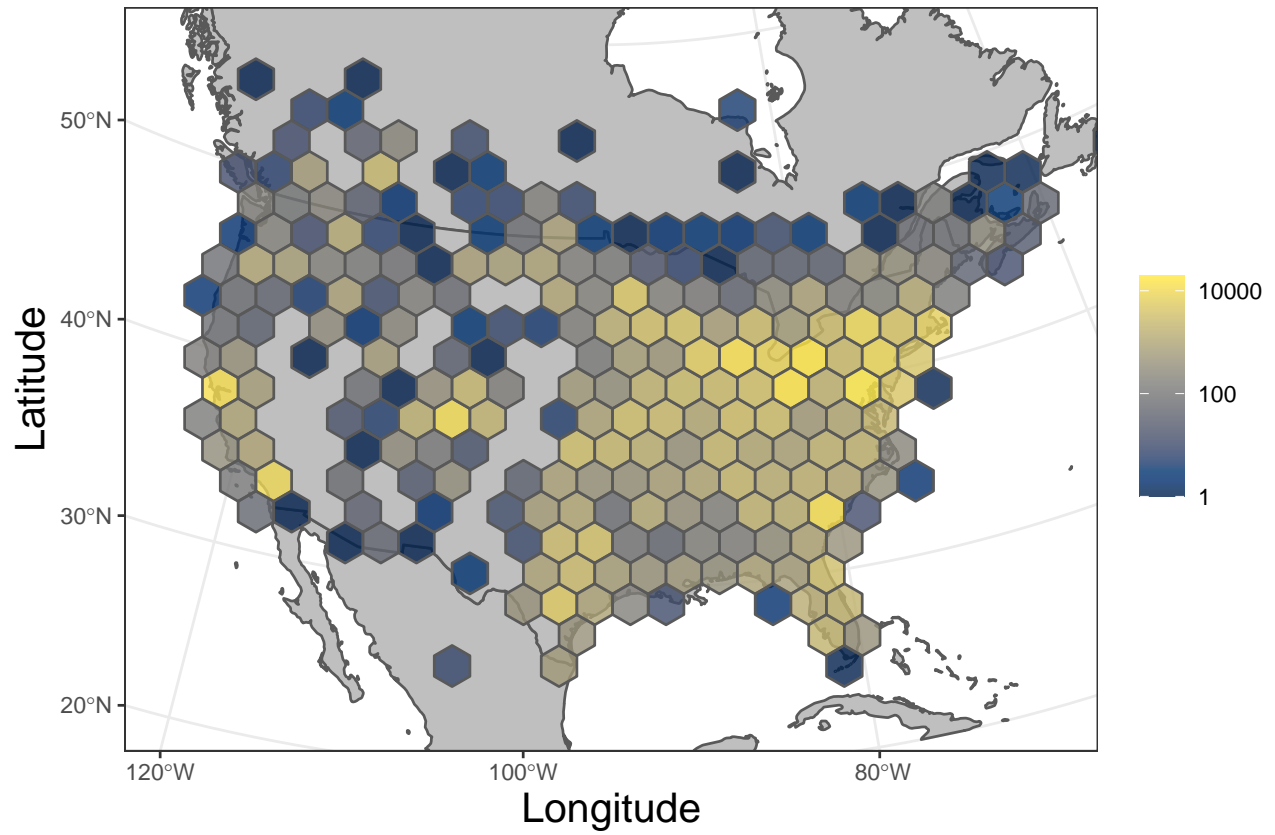
Figure S1: Number of breeding records included within each grid cell for all species and years from the combined datasets after filtering. A total of 301,514 records from 24 species are included from 1995 to 2020.
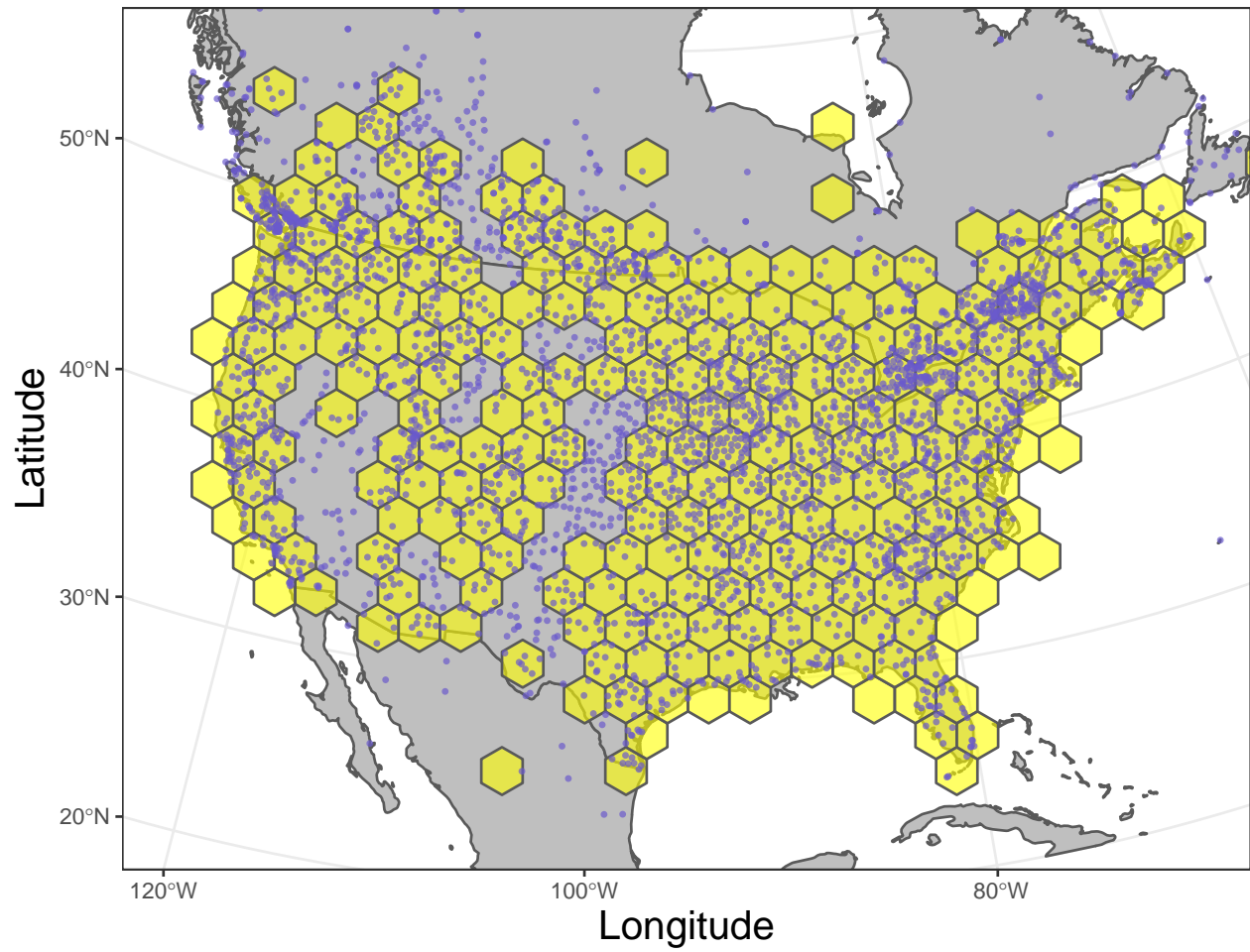
Figure S2: Weather stations used to access historical temperature data. Individual points show station locations with data downloaded from NOAA (n = 2608) or Environment and Climate Change Canada (n = 1125). Stations that had at least 50 years of available data were included. Yellow hexagons show the spatial extent of the avian breeding records used in the analysis to illustrate how coverage of temperature data corresponds to the main dataset.
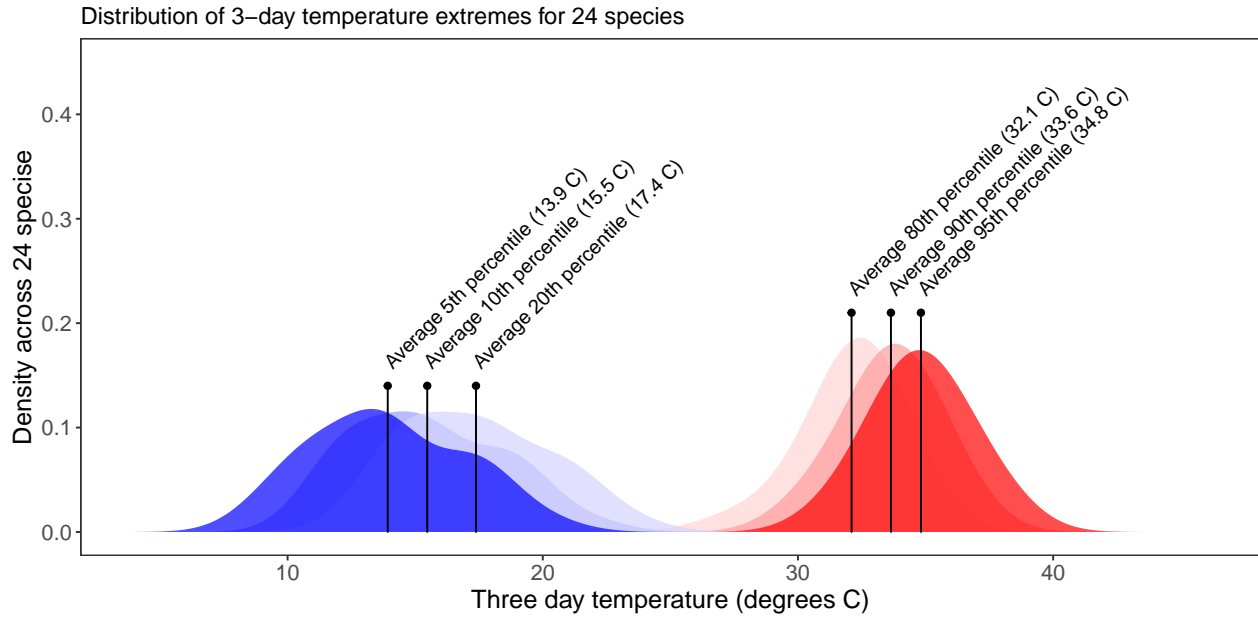
Figure S3: Distribution of extreme climatic events for the 24 species included in this study used to determine threshold temperatures for analysis of historical cold-snap and heat-wave timing. Six distrubutions are shown, illustrating the 5th, 10th, and 20th percentile of the coldest three day period during breeding attempts and the 80th, 90th, and 95th percentile of hottest three day period during breeding attempts. Black lines indicate the raw averages across 24 species for each distribution and these values were used in the historical analysis described in the text.
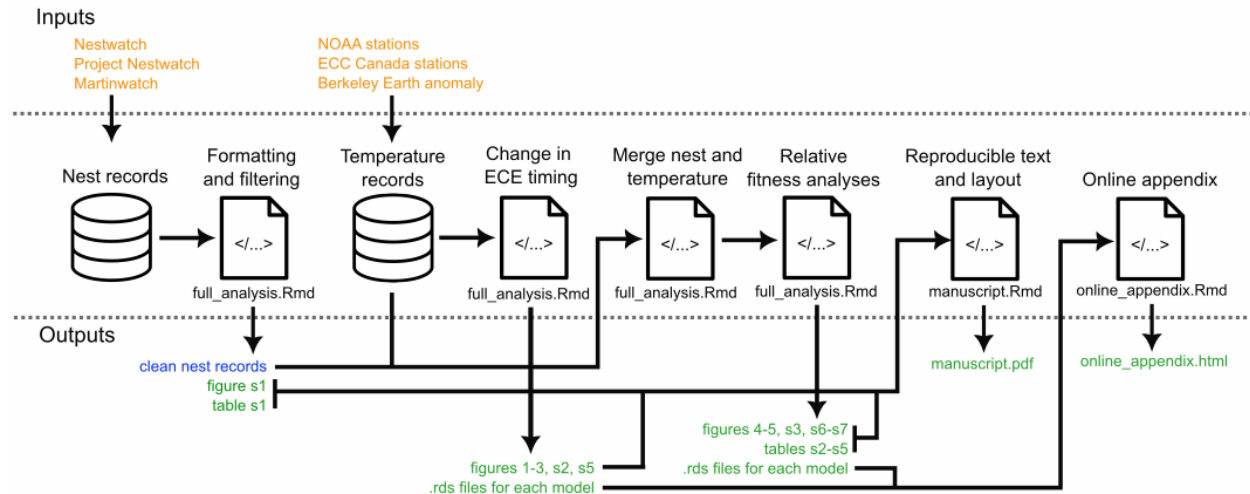

Figure S4: Schematic illustration of the data analysis workflow from raw data inputs to data outputs, manuscript file, and online supplement. To reproduce the analyses, raw data shown in orange will need to be accessed from each appropriate repository in order to produce the output shown in blue. Downstream output files shown in green are available in the archived data and code repository along with all of the scripts referenced in the figure.
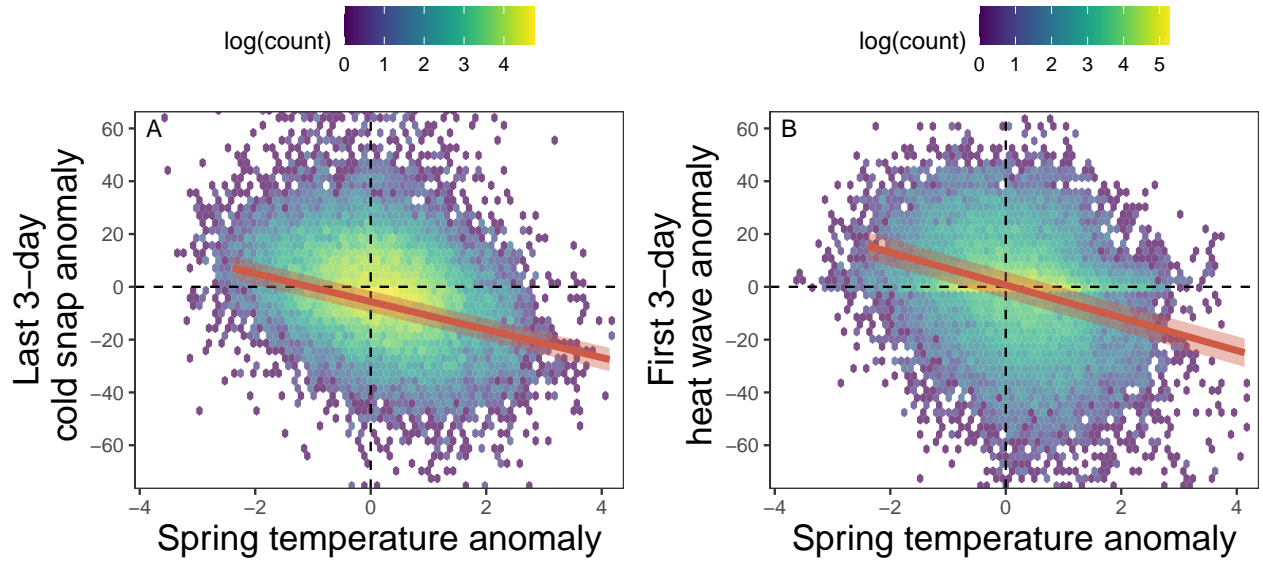
Figure S5: Spring average temperature anomaly in relation to the date of the last three day cold-snap (A) or earliest heat-wave (B). Grid shows the count of total cell level observations at each combination with all years counted. Red line is a fit from a GAM that includes a basis spline for year and tensor smooth for latitude and longitude to account for spatial and temporal autocorrelation. Shaded red region shows the 95% confidence interval of the fit GAM.

Table S2: Point estimates for the expected relative fitness of each species during a two standard deviation cold-snap or heat-wave based on GAMs described in text. Numbers in brackets show 95-percent confidence intervals.

	Incubation cold-snap	Incubation heat-wave	Provisioning cold-snap	Provisioning heat-wave
American robin	0.95 [0.79, 1.11]	1.02 [0.89, 1.14]	0.75 [0.65, 0.85]	0.90 [0.79, 1.01]
Barn swallow	1.10 [0.92, 1.28]	1.08 [0.94, 1.22]	0.98 [0.85, 1.10]	0.95 [0.86, 1.05]
Bewick's wren	0.98 [0.85, 1.11]	0.93 [0.78, 1.07]	1.00 [0.90, 1.10]	0.95 [0.86, 1.03]
Black-capped chickadee	1.02 [0.93, 1.11]	0.96 [0.82, 1.09]	0.90 [0.84, 0.95]	0.84 [0.77, 0.91]
Black-crested titmouse	1.02 [0.79, 1.26]	0.95 [0.74, 1.15]	0.86 [0.75, 0.97]	1.01 [0.92, 1.11]
Brown-headed nuthatch	0.89 [0.71, 1.07]	1.00 [0.78, 1.21]	0.87 [0.73, 1.01]	0.75 [0.60, 0.90]
Carolina chickadee	0.92 [0.85, 0.99]	0.99 [0.92, 1.06]	0.94 [0.90, 0.98]	0.94 [0.88, 1.01]
Chestnut-backed chickadee	0.95 [0.73, 1.18]	0.99 [0.87, 1.11]	0.94 [0.85, 1.03]	0.88 [0.78, 0.98]
Eastern bluebird	0.93 [0.90, 0.96]	0.86 [0.81, 0.91]	0.88 [0.86, 0.89]	0.87 [0.84, 0.90]
Eastern phoebe	1.06 [0.86, 1.26]	1.02 [0.89, 1.15]	0.74 [0.62, 0.86]	0.85 [0.76, 0.94]
House finch	1.11 [0.89, 1.33]	1.00 [0.74, 1.26]	0.84 [0.68, 0.99]	0.81 [0.67, 0.94]
House wren	0.97 [0.92, 1.02]	1.00 [0.95, 1.06]	0.91 [0.89, 0.94]	0.93 [0.89, 0.96]
Mountain bluebird	0.80 [0.72, 0.88]	0.91 [0.81, 1.01]	0.85 [0.79, 0.90]	0.89 [0.83, 0.96]
Mountain chickadee	0.64 [0.38, 0.91]	0.91 [0.73, 1.10]	0.89 [0.76, 1.03]	0.91 [0.71, 1.11]
Mourning dove	0.98 [0.65, 1.30]	0.83 [0.56, 1.10]	0.93 [0.74, 1.12]	0.92 [0.76, 1.09]
Northern cardinal	0.95 [0.54, 1.35]	1.17 [0.81, 1.53]	0.78 [0.47, 1.09]	0.83 [0.60, 1.05]
Oak titmouse	1.01 [0.89, 1.13]	1.08 [0.97, 1.19]	0.93 [0.84, 1.03]	1.04 [0.90, 1.19]
Prothonotary warbler	0.79 [0.64, 0.94]	1.00 [0.90, 1.11]	0.99 [0.92, 1.05]	1.00 [0.93, 1.06]
Purple martin	0.86 [0.83, 0.89]	0.98 [0.95, 1.00]	0.87 [0.85, 0.89]	0.94 [0.92, 0.96]
Tree swallow	0.85 [0.82, 0.89]	0.98 [0.95, 1.01]	0.91 [0.89, 0.93]	0.93 [0.91, 0.95]
Tufted titmouse	0.91 [0.77, 1.05]	0.91 [0.78, 1.04]	0.97 [0.88, 1.07]	0.93 [0.84, 1.02]
Violet-green swallow	1.03 [0.92, 1.14]	1.03 [0.92, 1.13]	0.83 [0.74, 0.92]	0.92 [0.83, 1.01]
Western bluebird	0.84 [0.78, 0.90]	0.98 [0.93, 1.03]	0.79 [0.74, 0.85]	0.92 [0.88, 0.97]
White-breasted nuthatch	0.79 [0.48, 1.10]	0.94 [0.76, 1.13]	0.75 [0.60, 0.90]	0.96 [0.69, 1.23]

Table S3: Point estimates for the expected relative fitness of eastern bluebirds during a two standard deviation cold-snap or heat-wave based on GAMs described in text. Breeding records are split into five latitude bands with an equal number of records per band; the southernmost band includes nests below 33.5 degrees latitude and is followed by 33.5 to 38.1 degrees, 38.1 to 40.0 degrees, 40.0 to 41.7 degrees, and finally nests above 41.7 degrees in the northernmost band. Numbers in brackets show 95-percent confidence intervals.

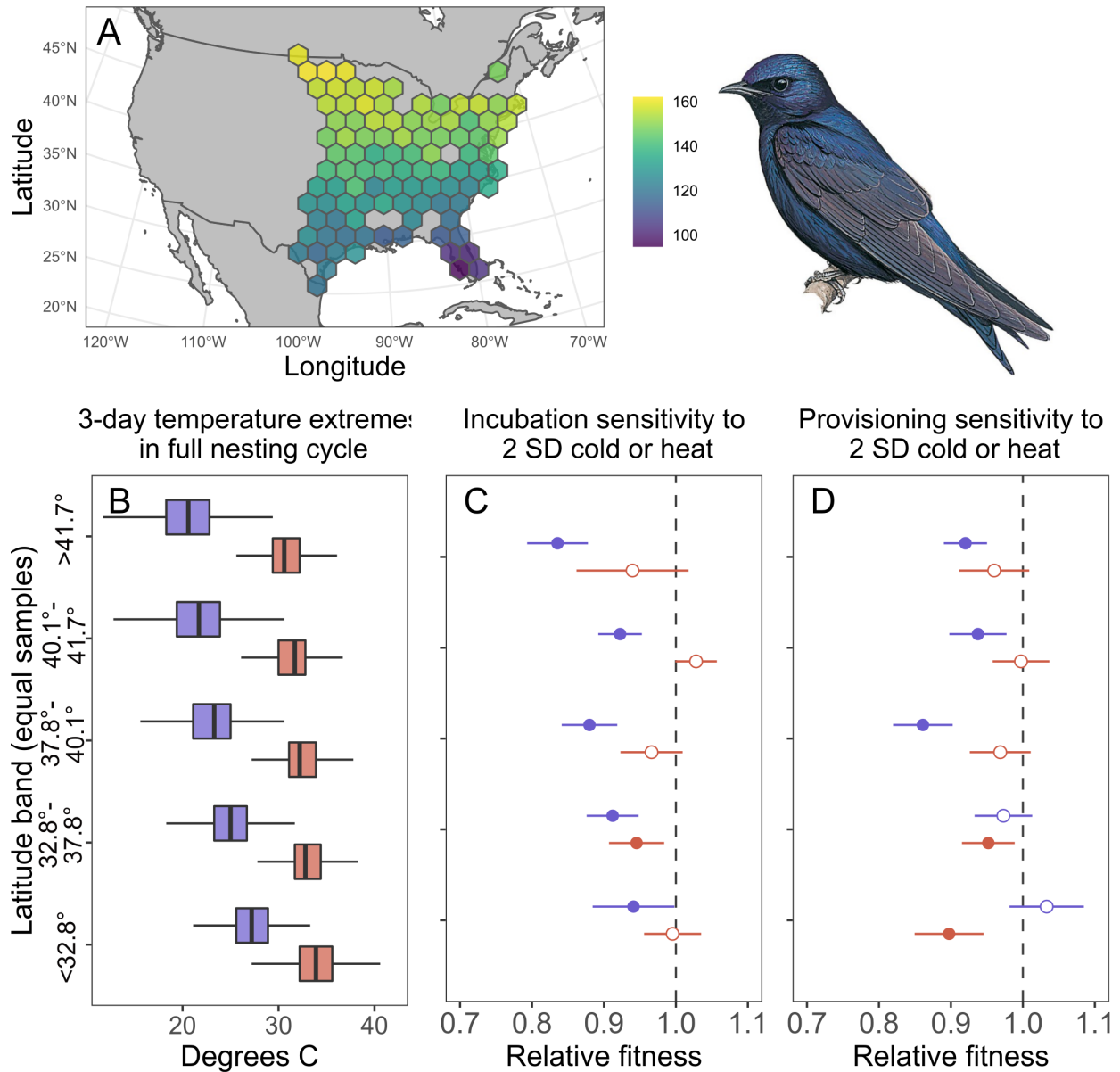|	Incubation cold-snap	Incubation heat-wave	Provisioning cold-snap	Provisioning heat-wave
Northernmost	0.91 [0.85, 0.96]	1.03 [0.98, 1.08]	0.87 [0.83, 0.91]	0.89 [0.83, 0.96]
North	0.98 [0.94, 1.03]	0.91 [0.84, 0.99]	0.90 [0.87, 0.94]	0.94 [0.89, 0.98]
Middle	0.97 [0.92, 1.01]	0.94 [0.89, 1.00]	0.98 [0.95, 1.02]	0.95 [0.90, 1.00]
South	0.99 [0.94, 1.04]	0.96 [0.90, 1.01]	1.01 [0.98, 1.04]	0.96 [0.92, 1.00]
Southernmost	0.97 [0.91, 1.02]	0.84 [0.76, 0.92]	1.00 [0.96, 1.03]	0.90 [0.85, 0.95]

Figure S6: Breeding timing, temperature exposure, and extreme temperature sensitivity in purple martins. (A) Average date of first egg laying across the records included. (B) Distribution of the coldest and hottest 3-day high temperatures in the full nesting cycles in five latitude bands with an equal number of samples per band. The top of the panel is the northernmost band and bottom is the southernmost. Boxplots indicate median, IQR, and 1.5 times IQR for the coldest (blue) and hottest (red) 3-day period. (C and D) Predicted relative fitness for breeding attempts that experienced a 2 SD cold-snap (blue) or 2 SD heat-wave (red) during incubation (C) or provisioning (D) for each latitude band. Point estimates are derived from GAMs that control for spatial patterns, year, and date. Circles show point estimate, lines show 95 percent CI; open circles indicate no difference from mean reproductive success while closed circles indicate a significant effect of temperature on relative reproductive success. Illustration from Handbook of the Birds of the World by Hilary Bum, copyright Lynx Edicions.

Table S4: Point estimates for the expected relative fitness of purple martins during a two standard deviation cold-snap or heat-wave based on GAMs described in text. Breeding records are split into five latitude bands with an equal number of records per band; the southernmost band includes nests below 32.8 degrees latitude and is followed by 32.8 to 37.8 degrees, 37.8 to 40.1 degrees, 40.1 to 41.7 degrees, and finally nests above 41.7 degrees in the northernmost band. Numbers in brackets show 95-percent confidence intervals.

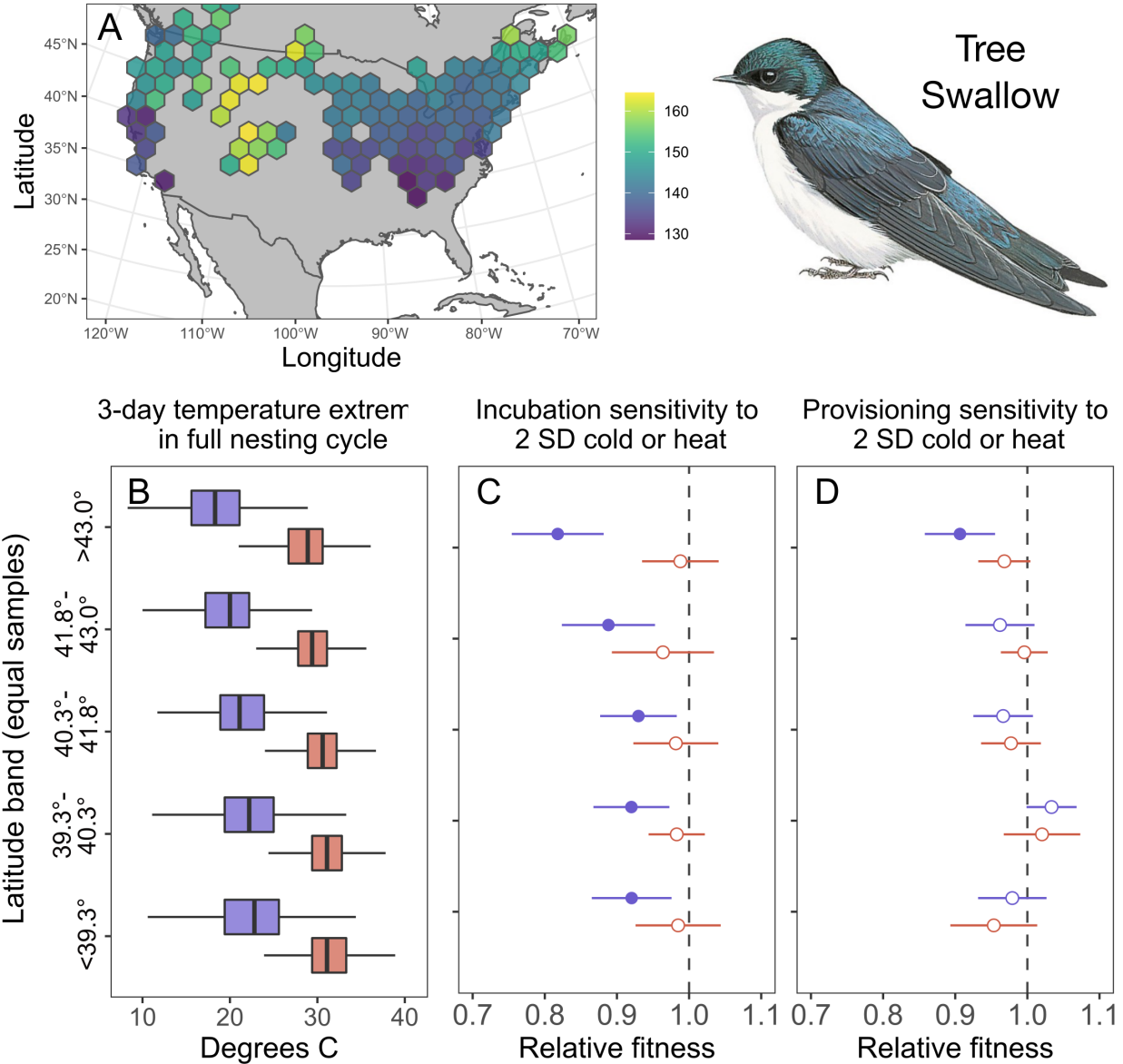|	Incubation cold-snap	Incubation heat-wave	Provisioning cold-snap	Provisioning heat-wave
Northernmost	0.84 [0.79, 0.88]	0.94 [0.86, 1.02]	0.92 [0.89, 0.95]	0.96 [0.91, 1.01]
North	0.92 [0.89, 0.95]	1.03 [1.00, 1.06]	0.94 [0.90, 0.98]	1.00 [0.96, 1.04]
Middle	0.88 [0.84, 0.92]	0.97 [0.92, 1.01]	0.86 [0.82, 0.90]	0.97 [0.93, 1.01]
South	0.91 [0.88, 0.95]	0.95 [0.91, 0.98]	0.97 [0.93, 1.01]	0.95 [0.92, 0.99]
Southernmost	0.94 [0.88, 1.00]	1.00 [0.96, 1.04]	1.03 [0.98, 1.08]	0.90 [0.85, 0.95]

Figure S7: Breeding timing, temperature exposure, and extreme temperature sensitivity in tree swallows. (A) Average date of first egg laying across the records included. (B) Distribution of the coldest and hottest 3-day high temperatures in the full nesting cycles in five latitude bands with an equal number of samples per band. The top of the panel is the northernmost band and bottom is the southernmost. Boxplots indicate median, IQR, and 1.5 times IQR for the coldest (blue) and hottest (red) 3-day period. (C and D) Predicted relative fitness for breeding attempts that experienced a 2 SD cold-snap (blue) or 2 SD heat-wave (red) during incubation (C) or provisioning (D) for each latitude band. Point estimates are derived from GAMs that control for spatial patterns, year, and date. Circles show point estimate, lines show 95 percent CI; open circles indicate no difference from mean reproductive success while closed circles indicate a significant effect of temperature on relative reproductive success. Analysis by latitude band included only species in the eastern range, see methods for details. Illustration from Handbook of the Birds of the World by Hilary Bum, copyright Lynx Edicions.

Table S5: Point estimates for the expected relative fitness of tree swallows during a two standard deviation cold-snap or heat-wave based on GAMs described in text. Breeding records are split into five latitude bands with an equal number of records per band; the southernmost band includes nests below 39.3 degrees latitude and is followed by 39.3 to 40.3 degrees, 40.3 to 41.8 degrees, 41.8 to 43 degrees, and finally nests above 43 degrees in the northernmost band. Numbers in brackets show 95-percent confidence intervals.

	Incubation cold-snap	Incubation heat-wave	Provisioning cold-snap	Provisioning heat-wave
Northernmost	0.82 [0.75, 0.88]	0.99 [0.93, 1.04]	0.91 [0.86, 0.96]	0.97 [0.93, 1.00]
North	0.89 [0.82, 0.95]	0.96 [0.89, 1.03]	0.96 [0.91, 1.01]	1.00 [0.96, 1.03]
Middle	0.93 [0.88, 0.98]	0.98 [0.92, 1.04]	0.97 [0.93, 1.01]	0.98 [0.94, 1.02]
South	0.92 [0.87, 0.97]	0.98 [0.94, 1.02]	1.03 [1.00, 1.07]	1.02 [0.97, 1.07]
Southernmost	0.92 [0.87, 0.98]	0.98 [0.93, 1.04]	0.98 [0.93, 1.03]	0.95 [0.89, 1.01]