

A continental system for forecasting bird migration

Authors: Benjamin M. Van Doren^{1*}, Kyle G. Horton²

Affiliations:

¹Edward Grey Institute, Department of Zoology, University of Oxford, Oxford OX1 3PS, UK

²Cornell Lab of Ornithology, Cornell University, Ithaca, NY 14850, USA

*Correspondence to: benjamin.vandoren@zoo.ox.ac.uk

Abstract: Billions of animals cross the globe each year during seasonal migrations, but efforts to monitor them are hampered by the unpredictability of their movements. We developed a bird migration forecast system at a continental scale by leveraging 23 years of spring observations to identify associations between atmospheric conditions and bird migration intensity. Our models explained up to 81% of variation in migration intensity across the United States at altitudes of 0-3000 m, and performance remained high when forecasting events 1-7 days in advance (62-76% variation explained). Avian migratory movements across the United States likely exceed 500 million individuals per night during peak passage. Bird migration forecasts will reduce collisions with buildings, airplanes, and wind turbines, inform a variety of monitoring efforts, and engage the public.

One Sentence Summary: Forecasting bird migration across the United States can reduce avian mortality.

Main Text: Billions of birds migrate between distant breeding and wintering sites each year, through landscapes and airspaces increasingly transformed by humans. Hundreds of millions die annually from collisions with buildings, automobiles, and energy installations (1), and light pollution exacerbates these effects (2). Dramatic pulses of intense migration interspersed with periods of low activity characterize birds' movements aloft (3, 4), and efforts to reduce negative impacts on migrants (e.g. turning off lights and wind turbines at strategic times, 5) would be most effective if they targeted the few nights with intense migratory pulses. However, bird movements are challenging to predict days or even hours in advance.

For decades, scientists have studied the drivers of avian migration. Winds, temperature, barometric pressure, and precipitation play key roles (6–8). However, such general relationships have not produced migration forecasts accurate at both broad continental extents and fine spatial and temporal resolutions (9, 10). Local topography, regional geography, and time of season modify relationships between conditions and migration intensity, and hundreds of species with diverse behaviors frequently pass over a single location during migration. The complex interactions between environmental conditions and animal behavior make predicting bird migration at the assemblage level a grand challenge.

One major difficulty has been amassing behavioral data that appropriately characterize bird migration at a continental scale. Radar, used globally as a tool to study animal migration (3, 11–14), offers a realistic solution to monitor hundreds of species (15). In the continental United States, the Next-Generation Radar (NEXRAD) network comprises 143 weather surveillance radars (16) and an archive with over two decades of data. Although designed for meteorological applications, these radars measure energy reflected by a diversity of aerial targets, including birds. Only recently have advances in computational methods (e.g. 17) facilitated use of the entire radar archive for longitudinal studies of bird migration at continental scales.

Using the NEXRAD archive, we quantified 23 years (1995-2017) of spring nocturnal bird migration across the US (Fig. 1). We developed a classifier to eliminate radar scans contaminated with precipitation. We then trained gradient boosted trees (18) to predict bird migration intensity from atmospheric conditions reported by the North American Regional Reanalysis (NARR) (19). Our model used 12 predictors, including winds, air temperature, barometric pressure, and relative humidity (Fig. S1), which we used to predict a cube-root transformed index of migration intensity ($\text{cm}^2 \text{km}^{-3}$). The cube root transform reduces skewness but is less extreme than a log-transformation, which would have given considerable weight to biologically unimportant differences between small values. We measured migration intensity in 100-m altitude bins up to 3 km to model the three-dimensional distribution of migrating birds over the continent. To express migration intensities in numbers of birds, we assumed a radar cross-section per bird of 11 cm^2 . The radar cross-section is a measure of reflected energy; this value is typical of medium-sized songbirds and representative of migratory species (12).

Our migration forecast model explained 78.9% of variation in migration intensity over the US (Figs. 2 and 3A). Performance was consistent across years (mean yearly $R^2 = 0.781 \pm 0.010$ SD). We quantified the importance of each predictor by calculating gain, a measure of how much predictions improve by adding a given variable. Air temperature was most important, with an average gain over three times that of the second-ranked predictor, date (Fig. S2). High temperatures coincided with large migration pulses (Figs. 4, S3, and S4). As a predictor of bird migration, temperature likely plays a dual role as an index of spring phenology and a short-term signal for movement, as favorable southerly winds usually accompany warmer air masses. Other important predictors included altitude, longitude, surface pressure, latitude, and wind (Fig. S2).

The model provides informative predictions several days in advance. We evaluated its utility as a true forecast system with archived weather forecasts from the North American Mesoscale Forecast System (NAM) and Global Forecast System (GFS). NAM has higher spatial resolution but is a shorter-range forecast (12-km grid, 3-d range) than GFS (0.5° grid, 7+ d range). We made predictions up to 3 d in advance with NAM and up to 7 d with GFS, expecting performance to degrade with time due to the decreasing accuracy of longer-range weather forecasts. Predictions based on 24-h NAM forecasts explained 75% of variation in migration intensity; 3-d NAM forecasts explained 71%; and 7-d GFS forecasts explained 62% (Fig. S5).

The model captures patterns of bird migration across the United States with high spatial accuracy, particularly in the Central and Eastern regions (Fig. S6). We evaluated spatial accuracy over areas without radar coverage by iteratively removing the data from each radar station, retraining the model on the remaining data, and testing performance on the withheld station. Median R^2 for withheld stations was 0.72, and R^2 was 0.60 or higher for 75% of stations (Fig. S7). Spatial variation in performance likely stems from local influences on migratory behavior (e.g. topography), which our model did not explicitly incorporate.

Previous research suggests that migration behavior and weather conditions in the days immediately preceding a migration event can predict its intensity (e.g. 10). We found that including atmospheric data from the preceding night and 24-h changes in conditions did improve performance, but not dramatically. A model that included atmospheric conditions 24 h before an event explained 80.1% of variation in migration intensity, and further including observed migration intensity from the previous night increased R^2 to 81.3%.

Finally, we used model predictions to estimate the total number of birds actively migrating each night across the United States. Summing predictions countrywide, we infer that

nightly movements frequently exceed 200 million birds (Fig. 3B). Peak passage occurred in the first half of May, when the median predicted movement size was 422 million birds per night. While our model tended to under-predict the largest observed movements (Fig. 3A), a conservative forecast system decreases the risk of taking unneeded mitigation action. More accurately predicting the largest migration events may require explicit modeling of migrant flow across the continent, including responses to topographical features (20).

Migration forecasts will further ecological research while aiding monitoring and mortality mitigation efforts. Accurate predictions can inform decisions to temporarily shut down lights and wind turbines, halt gas flares, choose airplane flight paths, and take other actions to prevent human and avian mortality (10, 21). Global health workers monitoring avian borne diseases can use migration forecasts to anticipate bird movements. Further integration of large citizen science datasets with radar observations will provide the means to study species-specific patterns of behavior at a large scale (22), and studying local variation in migratory behavior will lead to more accurate models of atmospheric bird distributions (23). Migration forecast systems have great potential to aid environmental monitoring and conservation efforts; fully realizing this potential will require the cooperation not just of scientists, but also of governments and agencies that produce and disseminate radar products (21).

References and Notes:

1. S. R. Loss, T. Will, P. P. Marra, Direct mortality of birds from anthropogenic causes. *Annu. Rev. Ecol. Evol. Syst.* **46**, 99–120 (2015).
2. J. D. McLaren, J. J. Buler, T. Schreckengost, J. A. Smolinsky, M. Boone, E. Emiel van Loon, D. K. Dawson, E. L. Walters, Artificial light at night confounds broad-scale habitat use by migrating birds. *Ecol. Lett.* **21**, 356–364.
3. A. Farnsworth, B. M. Van Doren, W. M. Hochachka, D. Sheldon, K. Winner, J. Irvine, J. Geevarghese, S. Kelling, A characterization of autumn nocturnal migration detected by weather surveillance radars in the northeastern USA. *Ecol. Appl.* **26**, 752–770 (2016).
4. B. Erni, F. Liechti, L. Underhill, B. Bruderer, Wind and rain govern the intensity of nocturnal bird migration in central Europe - A log-linear regression analysis. *Ardea*. **90**, 155–166 (2002).
5. B. M. Van Doren, K. G. Horton, A. M. Dokter, H. Klinck, S. B. Elbin, A. Farnsworth, High-intensity urban light installation dramatically alters nocturnal bird migration. *Proc. Natl. Acad. Sci.*, 201708574 (2017).
6. W. J. Richardson, Timing and amount of bird migration in relation to weather: a review. *Oikos*. **30**, 224–272 (1978).
7. F. Liechti, Birds: blowin' by the wind? *J. Ornithol.* **147**, 202–211 (2006).
8. M. U. Kemp, J. Shamoun-Baranes, A. M. Dokter, E. van Loon, W. Bouten, The influence of weather on the flight altitude of nocturnal migrants in mid-latitudes. *Ibis*. **155**, 734–749 (2013).

9. S. A. Gauthreaux, *Direct visual and radar methods for the detection, quantification, and prediction of bird migration*. (Department of Zoology, Clemson University, Clemson, South Carolina, 1980).
- 5 10. J. Van Belle, J. Shamoun-Baranes, E. Van Loon, W. Bouten, An operational model predicting autumn bird migration intensities for flight safety. *J. Appl. Ecol.* **44**, 864–874 (2007).
- 10 11. S. A. Gauthreaux, C. G. Belser, D. van Blaricom, "Using a Network of WSR-88D Weather Surveillance Radars to Define Patterns of Bird Migration at Large Spatial Scales" in *Avian Migration*, P. Berthold, E. Gwinner, E. Sonnenschein, Eds. (Springer Berlin Heidelberg, 2003), pp. 335–346.
12. A. M. Dokter, F. Liechti, H. Stark, L. Delobbe, P. Tabary, I. Holleman, Bird migration flight altitudes studied by a network of operational weather radars. *J. R. Soc. Interface*, rsif20100116 (2010).
- 15 13. G. Hu, K. S. Lim, N. Horvitz, S. J. Clark, D. R. Reynolds, N. Sapir, J. W. Chapman, Mass seasonal bioflows of high-flying insect migrants. *Science*. **354**, 1584–1587 (2016).
14. T. Alerstam, G. A. Gudmundsson, M. Green, A. Hedenström, Migration along orthodromic sun compass routes by arctic birds. *Science*. **291**, 300–303 (2001).
15. J. F. Kelly, K. G. Horton, M. Sykes, Toward a predictive macrosystems framework for migration ecology. *Glob. Ecol. Biogeogr.* **25**, 1159–1165 (2016).
- 20 16. T. D. Crum, R. L. Alberty, The WSR-88D and the WSR-88D operational support facility. *Bull. Am. Meteorol. Soc.* **74**, 1669–1687 (1993).
- 25 17. D. R. Sheldon, A. Farnsworth, J. Irvine, B. M. Van Doren, K. F. Webb, T. G. Dietterich, S. Kelling, "Approximate Bayesian Inference for Reconstructing Velocities of Migrating Birds from Weather Radar" in *Proceedings of the Twenty-Seventh AAAI Conference on Artificial Intelligence*, M. desJardins, M. L. Littman, Eds. (2013), pp. 1334–1340.
18. T. Chen, C. Guestrin, "XGBoost: A Scalable Tree Boosting System" in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (ACM, New York, NY, USA, 2016), *KDD '16*, pp. 785–794.
- 30 19. F. Mesinger, G. DiMego, E. Kalnay, K. Mitchell, P. C. Shafran, W. Ebisuzaki, D. Jović, J. Woollen, E. Rogers, E. H. Berbery, M. B. Ek, Y. Fan, R. Grumbine, W. Higgins, H. Li, Y. Lin, G. Manikin, D. Parrish, W. Shi, North American Regional Reanalysis. *Bull. Am. Meteorol. Soc.* **87**, 343–360 (2006).
- 35 20. A. Aurbach, B. Schmid, F. Liechti, N. Chokani, R. Abhari, Complex behaviour in complex terrain - Modelling bird migration in a high resolution wind field across mountainous terrain to simulate observed patterns. *J. Theor. Biol.* **454**, 126–138 (2018).

21. S. Bauer, J. W. Chapman, D. R. Reynolds, J. A. Alves, A. M. Dokter, M. M. H. Menz, N. Sapir, M. Ciach, L. B. Pettersson, J. F. Kelly, H. Leijnse, J. Shamoun-Baranes, From agricultural benefits to aviation safety: realizing the potential of continent-wide radar networks. *BioScience*. **67**, 912–918 (2017).
- 5 22. K. G. Horton, B. M. Van Doren, F. A. La Sorte, D. Fink, D. Sheldon, A. Farnsworth, J. F. Kelly, Navigating north: how body mass and winds shape avian flight behaviours across a North American migratory flyway. *Ecol. Lett.* **21**, 1055–1064 (2018).
23. J. Shamoun-Baranes, H. van Gasteren, V. Ross-Smith, "Sharing the Aerosphere: Conflicts and Potential Solutions" in *Aeroecology* (Springer, Cham, 2017), pp. 465–497.
- 10 24. B. M. Van Doren, K. G. Horton, Dataset for "A continental system for forecasting bird migration," figshare (2018); <https://doi.org/10.6084/m9.figshare.6962810>.
25. K. G. Horton, W. G. Shriver, J. J. Buler, A comparison of traffic estimates of nocturnal flying animals using radar, thermal imaging, and acoustic recording. *Ecol. Appl.* **25**, 390–401 (2015).
- 15 26. S. A. Gauthreaux, A Radar and Direct Visual Study of Passerine Spring Migration in Southern Louisiana. *The Auk*. **88**, 343–365 (1971).
27. D. Sheldon, *WSRLIB: MATLAB Toolbox for Weather Surveillance Radar* (2015).
28. K. G. Horton, B. M. Van Doren, P. M. Stepanian, W. M. Hochachka, A. Farnsworth, J. F. Kelly, Nocturnally migrating songbirds drift when they can and compensate when they must. *Sci. Rep.* **6**, srep21249 (2016).
- 20 29. A. Liaw, M. Wiener, others, Classification and regression by randomForest. *R News*. **2**, 18–22 (2002).
30. S. A. Gauthreaux, C. G. Belser, Displays of bird movements on the WSR-88D: patterns and quantification. *Weather Forecast.* **13**, 453–464 (1998).
- 25 31. S. A. Cabrera-Cruz, T. J. Mabee, R. V. Patraca, Using theoretical flight speeds to discriminate birds from insects in radar studies. *The Condor*. **115**, 263–272 (2013).
32. K. G. Horton, B. M. Van Doren, P. M. Stepanian, A. Farnsworth, J. F. Kelly, Seasonal differences in landbird migration strategies. *The Auk*. **133**, 761–769 (2016).
- 30 33. R. P. Larkin, Flight speeds observed with radar, a correction: slow "birds" are insects. *Behav. Ecol. Sociobiol.* **29**, 221–224 (1991).
34. T. Chen, T. He, M. Benesty, V. Khotilovich, Y. Tang, *xgboost: Extreme Gradient Boosting* (2017).
35. B. Bischl, M. Lang, L. Kotthoff, J. Schiffner, J. Richter, E. Studerus, G. Casalicchio, Z. M. Jones, mlr: Machine Learning in R. *J. Mach. Learn. Res.* **17**, 1–5 (2016).

36. S. N. Wood, *Generalized Additive Models: An Introduction with R* (Chapman and Hall/CRC, ed. 2, 2017).
37. V. M. Melnikov, R. R. Lee, N. J. Langlieb, Resonance Effects Within S-Band in Echoes From Birds. *IEEE Geosci. Remote Sens. Lett.* **9**, 413–416 (2012).

5

Acknowledgments:

We thank Andrew Farnsworth, Dan Sheldon, Ben Sheldon, Wesley Hochachka, Valery Melnikov, Giles Hooker, Jacob Calvert, and three anonymous reviewers; **Funding:** Marshall Aid Commemoration Commission (BVD); Edward W. Rose Postdoctoral Fellowship, Leon Levy Foundation, and National Science Foundation grants DBI-1661329, DBI-1661259, and IIS-1633206 (KGH); **Author contributions:** BVD conceived of the study, performed statistical analyses, and wrote the paper; KGH performed radar analyses, shaped the study, and contributed writing; **Competing interests:** Authors declare no competing interests; and **Data and materials availability:** Data and code available from figshare (24).

10

15

Supplementary Materials:

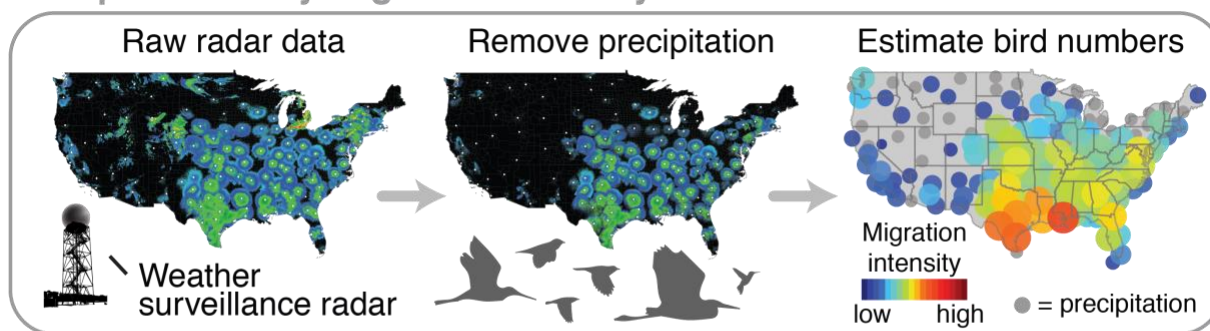
Materials and Methods

Figures S1-S10

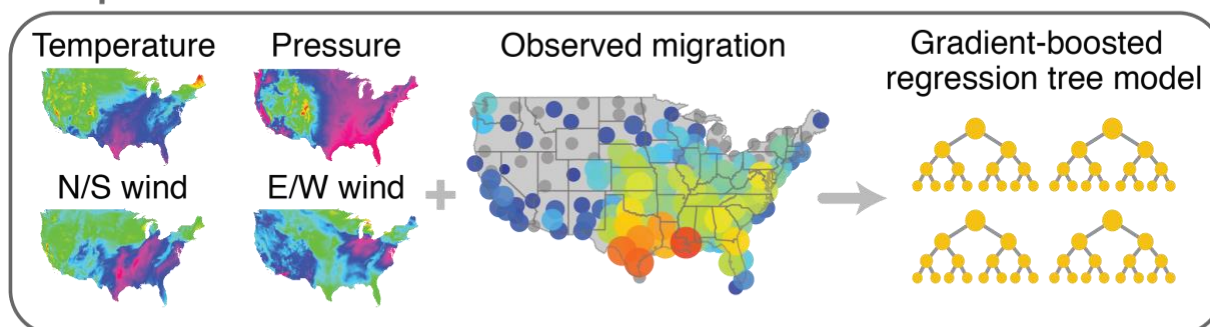
References (25-37)

20

Step 1. Quantify migration intensity at 143 weather radar stations



Step 2. Learn associations with weather conditions



Step 3. Make predictions using weather forecasts

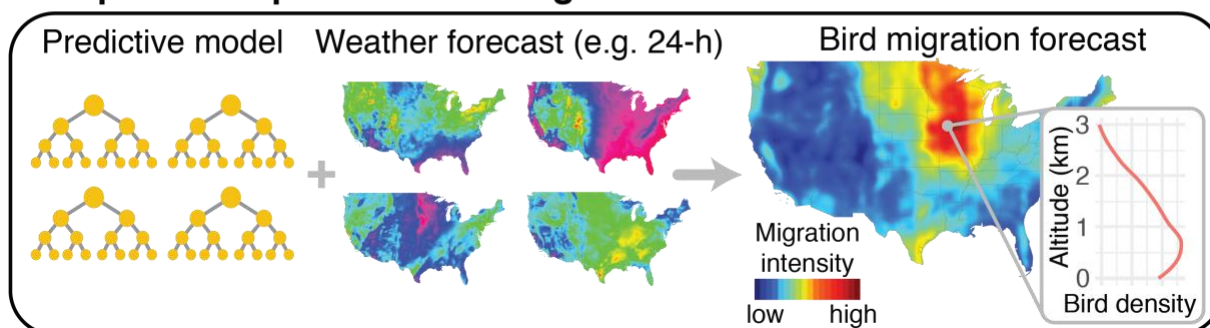


Fig. 1. Methodology for generating migration forecasts. We 1) used weather surveillance radars to quantify 23 years of spring bird migration, 2) modeled migration intensity as a function of observed atmospheric conditions, and 3) used this model to forecast future migration events under predicted weather conditions.

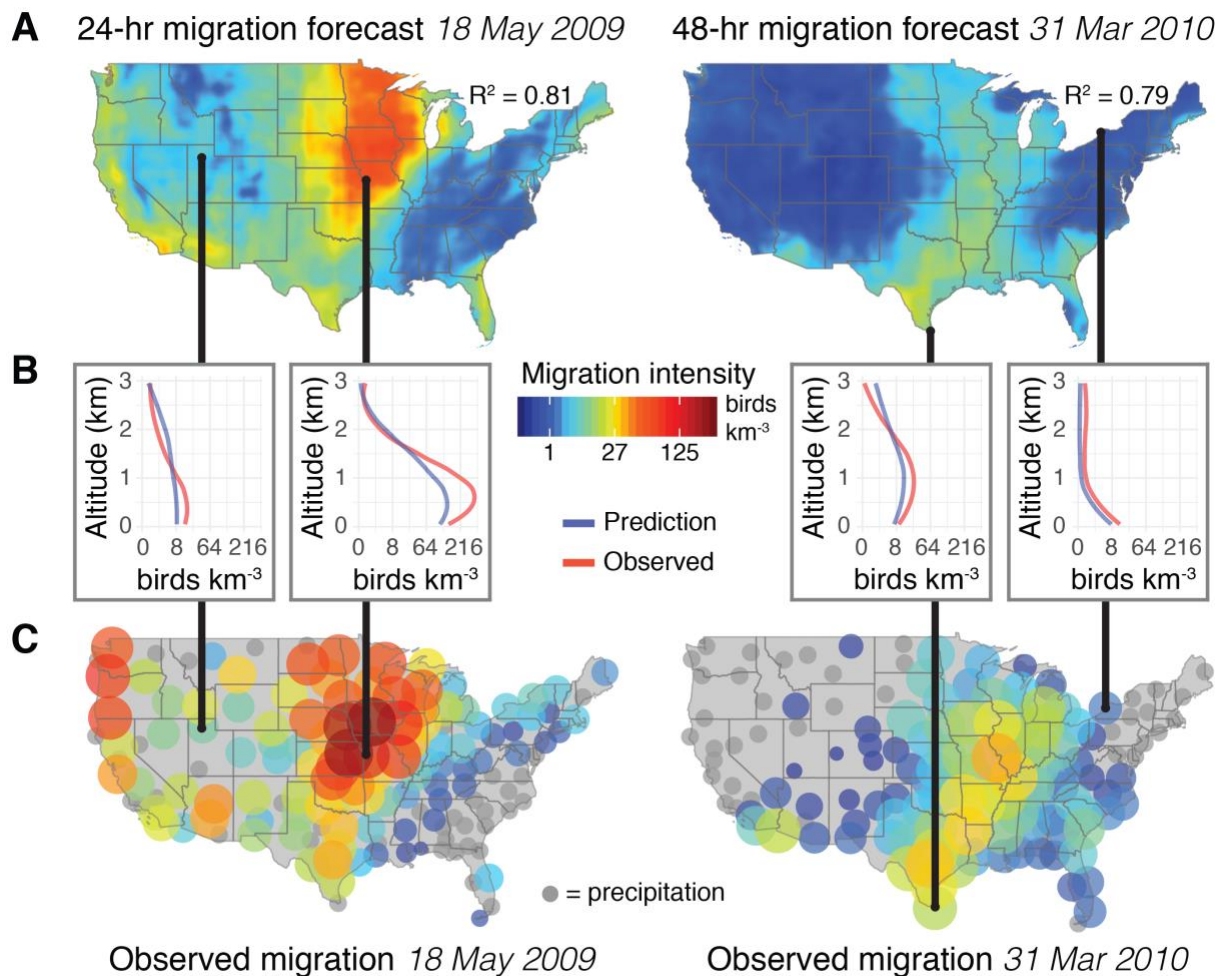


Fig. 2. Migration forecasts and corresponding observed migration. (A) Countrywide migration forecast surfaces showing predicted mean migration intensity across altitudes. (B) Altitudinal profiles at four stations, showing predicted and observed intensity values. (C) Mean migration intensity observed at all radar stations. Gray circles indicate stations where migration intensity could not be measured due to precipitation.

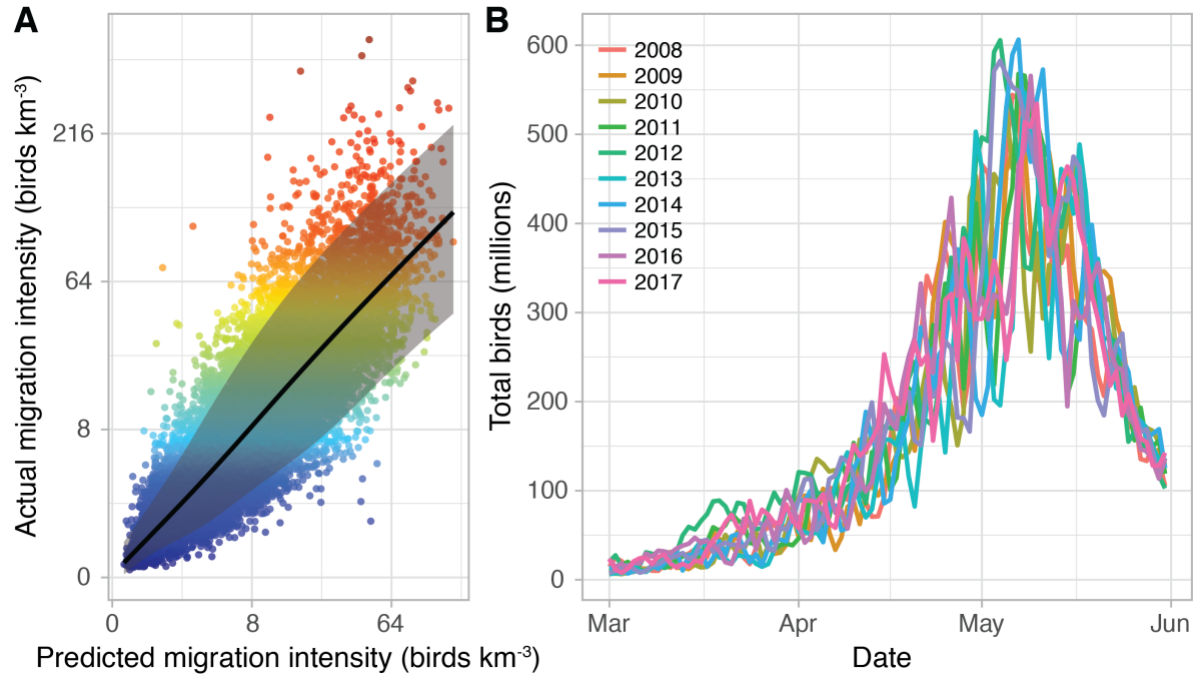


Fig. 3. Accuracy of forecasts and nightly continental predictions. (A) Mean predicted and observed migration intensities for test data, with points colored by observed migration intensity (y-axis). The scatterplot shows values after averaging across altitudes. Shading shows empirical 90% prediction intervals, which covered 90.5% of observed values. (B) Nightly peak migration magnitude estimated across the continental United States for 2008-2017. The size of migratory movements varied dramatically from night to night during the peak of the migration season.

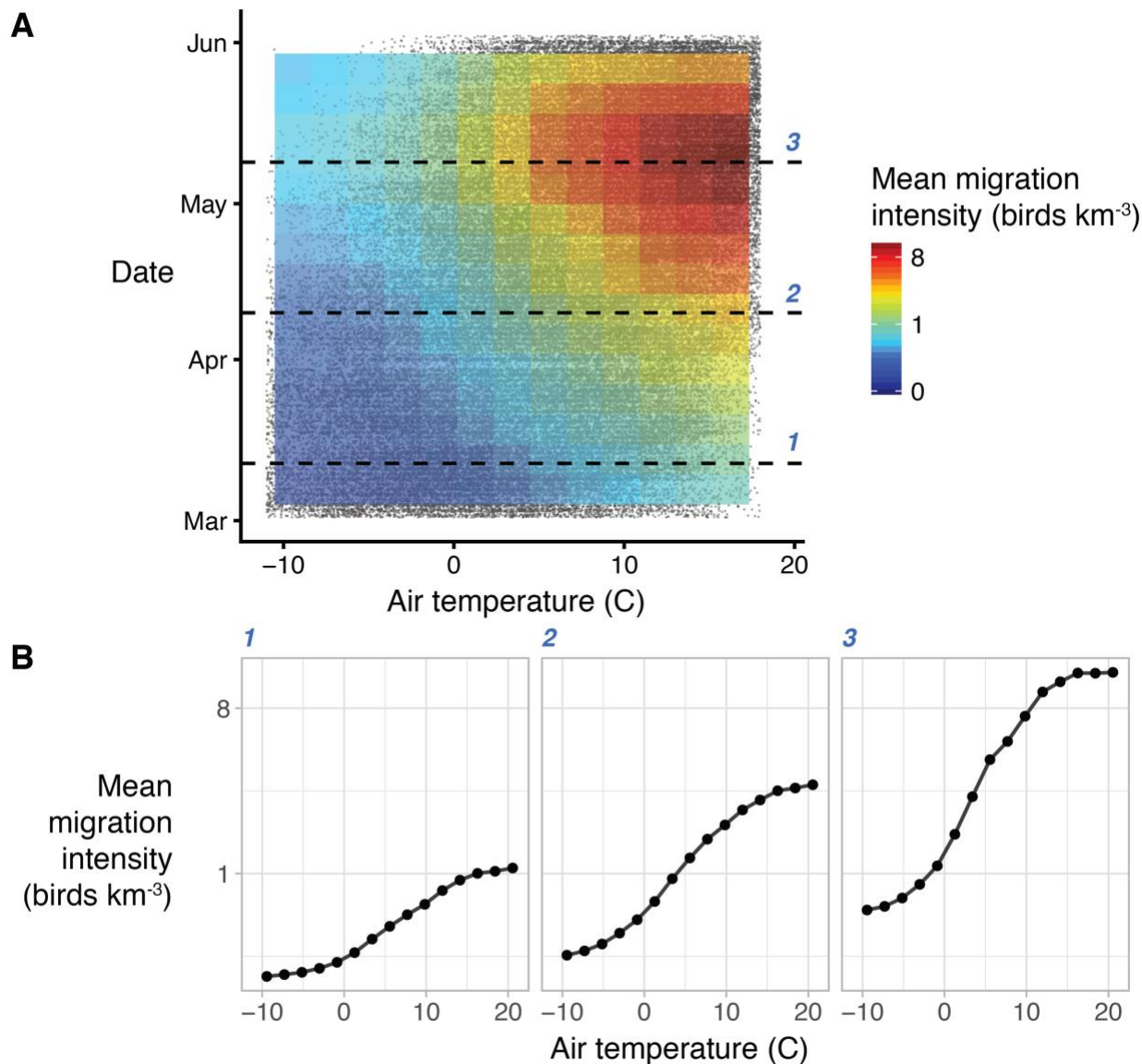


Fig. 4. Migration intensity predictions by air temperature and date. (A) Heatmap colors show migration intensity predictions for date and air temperature values. Each data point on the scatterplot behind the heatmap represents one night from one radar. Only well-supported predictions and corresponding data points are shown (the outer 10% of temperature and date values are excluded). Temperature values correspond to air temperatures at altitudes up to 3000 m. (B) Cross-sections of model predictions at three spring dates. For a given date, the model predicts migration intensity to vary closely with temperature. Note that fewer observations correspond to cold temperatures later in the season.

Supplementary Materials for

A continental system for forecasting bird migration

Benjamin M. Van Doren, Kyle G. Horton

Correspondence to: benjamin.vandoren@zoo.ox.ac.uk

This PDF file includes:

Materials and Methods
Figs. S1 to S10

Materials and Methods

Doppler radar

We used the NEXRAD network operated by the National Oceanic and Atmospheric Administration to characterize spring migratory movements (March 1st to May 31st) from 1995 to 2017. These radars scan 360° at multiple elevation angles (e.g. 0.5°, 1.5°... 4.5°), fully sampling the airspace every 5 to 10 minutes. We downloaded radar scans from Amazon Web Services (<https://s3.amazonaws.com/noaa-nexrad-level2/index.html>), selecting those in a 30-minute window centered on three hours after local sunset. We chose this time window because it approximates the time of peak nocturnal migration across the United States (e.g., 3, 25). However, there is spatial variation in the time of peak nocturnal migration (e.g., stations along the Gulf of Mexico experience peak migration earlier in the night, 26), so in some areas our predicted totals will be conservative. We processed scans using the WSRLIB weather surveillance radar package for MATLAB (27). To characterize migration intensity, we used radar reflectivity factor, a measure of reflectance to the radar. To sample the airspace from 0-3 km above ground level, we extracted radar reflectivity factor values 5-37.5km from each radar (28) and cast them into vertical profiles with 100-m altitudinal resolution. We converted radar reflectivity factor (dBZ) to radar reflectivity (dB η) using the equation $\eta[\text{dB}] = Z[\text{dBZ}] + \beta$, where $\beta = 10\log_{10}(10^3\pi^5|K_m|^2/\lambda^4)$. We set the radar wavelength (λ) to 10.71 cm, the average for NEXRAD radars (16) and set the refractive index ($|K_m|^2$) to 0.93 for liquid water. This yielded $\beta = 13.35$. We converted dB η to η using the equation $\eta = 10^{\text{dB}\eta/10}$, yielding units of $\text{cm}^2\text{km}^{-3}$. To estimate numbers of birds from η , we divided η by a radar cross-section of 11 cm^2 . This resulted in units of birds km^{-3} .

To mitigate the influence of time-invariant clutter (e.g., buildings, terrain, wind turbines), we applied binary clutter masks to each low elevation scan prior to the construction of the vertical profile of migration intensity. Masks were generated by summing a minimum of 100 low elevation scans (0.5° elevation), starting on January 1st (16:00 UTC to 18:00 UTC) and continuing to January 15th. This time window falls well outside typical migration periods. If 100 samples were not tallied by January 15th, the window of selection was expanded until the threshold was met. We classified any pixel above the 85th percentile of the summed reflectivity as clutter and masked it from our calculation of migration intensity.

To discriminate radar scans contaminated with precipitation from those containing only clear air or bird-dominated signal (hereafter termed “clear”), we created a random forest classifier using the package “randomForest” (29). We trained the classifier on 157,279 manually classified nocturnal scans (generated every 5-10 minutes) selected from a 3-hour period on March 15th, April 15th, and May 15th for every radar and every year in the training set (fig. S8). We designed this sampling to capture any geographic, seasonal, or longitudinal patterns in the data. We extracted derived predictor variables from profiles of radar reflectivity, groundspeed, and summaries of the number of density values above 35 dBZ (extreme densities characteristic of intense precipitation). We generated 1,000 trees and set the minimum terminal node size to 50. Overall, the model showed a 5.6% classification error. We used the algorithm to classify a total of 979,326 scans. As an additional step to reduce the inclusion of precipitation incorrectly classified as clear, we used only scans with a probability of being clear >70% (rather than a majority rule, i.e. >50%).

Identifying and removing flying insects from weather surveillance radar data has been a long-standing challenge for ornithologists. The standard method of ameliorating insect

contamination is to filter data by airspeed values, because samples with large average airspeeds must be dominated by strong flyers (30, 31). One downside to this approach is that it may also remove samples containing slow-flying birds or a mix of birds and insects. Here, we filter vertical profiles by removing altitude bins with mean airspeeds of 5 m s^{-1} or less (3, 22, 32), a cutoff value chosen to remove the majority of insects (30, 33). Slow-flying samples represented 16% of total summed reflectivity. We examined the sensitivity of our analysis to this step by comparing model predictions with and without insect filtering (fig. S9).

Weather reanalysis

The North American Regional Reanalysis, or NARR (19), compiles data from numerous observational data sources to produce a best estimate of weather conditions that occurred in North America. The reanalysis is published in 3-hour intervals across a 32-km grid. We downloaded NARR data for 1995-2017 and extracted the following parameters: air temperature ($^{\circ}\text{C}$), geopotential height (m), zonal and meridional wind components (m s^{-1}), surface pressure (Pa), relative humidity (%), visibility (m), vertical velocity (Pa s^{-1}), mean sea level pressure (Pa), total cloud cover (%), albedo (%), total precipitation (kg m^{-2}), convective available potential energy (J kg^{-1}), and snow cover (%). For variables available at multiple pressure levels, we extracted data from the surface to 300 mb. To match weather data to radar stations, we averaged data within 37.5 km of each radar station. We then calculated altitude above ground level by subtracting surface geopotential height from the geopotential height at each pressure level, and we used linear interpolation to calculate vertical profiles of weather data at 100-m altitude bins from 0-3000 m. Finally, we matched radar profiles to weather profiles by taking the observation closest in time for each radar station. Pairwise correlations among predictor variables were generally low (fig. S1).

Weather forecasts

The North American Mesoscale Forecast System, or NAM (<https://www.ncdc.noaa.gov/data-access/model-data/model-datasets/north-american-mesoscale-forecast-system-nam>), generates weather forecasts out to 84 hours; forecasts are hourly from 1-36 hours and subsequently every 3 hours until hour 84. Forecast models are run every 6 hours. NAM predictions are made on a 12-km grid. We downloaded 0Z NAM forecast data for 2008-2017, extracted the same parameters as for NARR, and matched NAM data to radar stations in the same manner as for NARR.

In addition to NAM, we used the Global Forecast System, or GFS (<https://www.ncdc.noaa.gov/data-access/model-data/model-datasets/global-forecast-system-gfs>) to generate longer-range migration forecasts. GFS forecasts with 0.5° spatial resolution currently extend out to 384 hours at 3-hour increments, but this range has improved with time; in 2010, the range was 180 hours. We downloaded 0Z GFS forecast data for 2010-2017 and extracted weather predictions up to 7 days (168 h) in advance.

Supervised learning

We used gradient boosted trees to predict bird migration from weather data (Fig. 1). We used the R implementation of XGBoost (18, 34), a highly efficient and scalable gradient boosting framework. The algorithm automatically detects nonlinear effects and complex interactions among predictors, and it is not hindered by predictor collinearity. We trained an XGBoost model on NARR weather data, with the cube root of bird density as our response variable.

We divided our dataset into three groups: a training set, for learning; a validation set, for hyperparameter tuning; and a test set, to evaluate performance. We split the dataset by whole days instead of individual data points to prevent any spatial autocorrelation from inflating performance metrics. From 2,115 total days (comprising 3,434,703 altitude bins across 143 radar stations after filtering steps), we randomly selected 75% of days for training, 10% for validation, and 15% for testing.

We tuned model hyperparameters with grid searches across hyperparameter space (fig. S10). For our first search, we set the learning rate *eta* to 0.05 while varying maximum tree depth *max_depth* between 8-16. Trees of these depths are complex, but predicting bird migration across the entire United States from March to May at 30 different altitude bins is a complex problem. We used the *early_stopping_rounds* argument to stop the algorithm after 10 boosting iterations in which performance on the validation set failed to improve. Larger trees perform better on training data, but trees that are too large lower performance due to overfitting. We therefore used the validation dataset to select the best-performing value of maximum tree depth. We then tested the following modifications to additional parameters that can prevent overfitting: decreasing *subsample* from 1.0 to 0.70, increasing *min_child_weight* from 1 to 5, and increasing *gamma* from 0 to 1 or 10. We tried all 12 combinations of these modifications. The best combination of parameters was the following: *max_depth* = 12, *min_child_weight* = 5, *gamma* = 1, *colsample_bytree* = 1, and *subsample* = 0.7. Using the best combination of hyperparameters, we further lowered the learning rate to 0.01 and set *early_stopping_rounds* to 100 to determine the optimal number of boosting iterations for that learning rate. Lower learning rates decrease the contribution of each tree to the model, making the boosting algorithm more conservative and further preventing overfitting, but lower learning rates require more iterations. With this information, we fit a final model with learning rate = 0.01 on the combined training and validation sets. We then evaluated its performance on the test dataset (15% of data), which had been withheld from all training and validation. To assess performance, we calculated two metrics: root mean square error and the coefficient of determination (or R^2). We calculated R^2 by dividing the sum of squared errors by the total sum of squares, and then subtracting this value from 1. An R^2 value of 0 indicates that the model does not explain the data any better than a simple null model that predicts the mean for each observation, while a negative R^2 value indicates that the model explains the data worse than this null model.

In an XGBoost model, correlated or uninformative predictors generally have little negative effect; they will generally not be incorporated during tree construction. However, extraneous predictors increase computational time and data storage requirements, making the forecast system more unwieldy to operationalize. For this reason, we sought to remove uninformative predictors. Using the *xgboost* package, we calculated the gain, a predictor importance metric that quantifies how much a tree improves by adding a split on a given variable. Gain values are scaled to sum to 1. After the first grid search step, we identified and eliminated predictors with gain values less than 0.01 and restarted the tuning procedure. In this manner, we eliminated albedo, vertical velocity, convective available potential energy, total precipitation, and snow cover. This left 12 variables in the final model: ordinal date, height above ground level, latitude, longitude, air temperature, relative humidity, zonal wind, meridional wind, surface pressure, mean sea level pressure, visibility, and total cloud cover.

We trained and tested two further modifications to the final model: one which also included additional conditions variables from the previous night (winds, temperature, and surface pressure) and their 24-hour changes, and another which included these lagged weather variables

plus migration intensity measured during the previous night. Our aim here was to determine how much additional explanatory power we could achieve with a model that took into account recently observed conditions and behavior.

Performance and importance

To assess performance of the final model using weather forecasts instead of reanalysis (i.e. NARR) data, we tested the model using archived NAM forecasts made 1-3 days in advance. We did the same for GFS forecasts made 1-7 days in advance. Because GFS does not contain a visibility variable, we first retrained the model without visibility included in order to conduct this evaluation.

To assess model performance at unobserved spatial locations, we performed a cross-validation where we randomly removed one station (out of 143 total) from the dataset, retrained the model on the remaining data, and tested its performance on the withheld station.

We identified the predictor variables that were most important for model predictions using gain, a measure of the variable's importance in making accurate predictions. We also generated partial dependence plots using the R package *mlr* (35) to explore how these variables influence predictions. Here, we used a learning rate *eta* of 0.05 instead of 0.01 to make computation tractable.

Prediction intervals

We constructed empirical prediction intervals using residuals from XGBoost predictions for the validation dataset. We fitted a generalized additive model (36) on squared XGBoost residuals against the XGBoost-predicted value to account for an error variance that increased with the magnitude of the predicted value. The generalized additive model produced an estimated error variance for each predicted value, which we used to construct 90% prediction intervals using 0.05/0.95 Gaussian quantiles. We constructed separate models for upper and lower limits to allow for asymmetry in the width of the interval, and we used the Gamma distribution family in the generalized additive model to constrain the predicted variances to be non-negative.

Forecast output and estimation of nightly migration magnitude

Using our validated migration forecast model, we made predictions across the entire 12-km NAM grid. For smooth presentation, we averaged predictions across 9×9 cell blocks. We also used our model to estimate the total number of birds migrating over the continental United States each night. For this we used the NARR dataset because it is the best retrospective estimate of occurred conditions. For each 32-km NARR grid cell covering the continental United States, we multiplied the bird density estimate by the area of the cell and summed totals across all grid cells for each night.

NEXRAD radars operate at slightly different carrier frequencies (and hence different wavelengths) to reduce interference from neighboring radars, and this variation may introduce noise into estimates of total bird numbers if radars differ substantially in wavelength (37). However, such noise is likely to be minor because (1) most radars operate at more similar wavelengths than the example presented in (37), (2) variation in carrier frequency is not correlated with geography (i.e. no consistent spatial bias would be introduced), and (3) including wavelength as a model predictor to account for any systematic difference in detected bird densities did not appreciably change model performance.

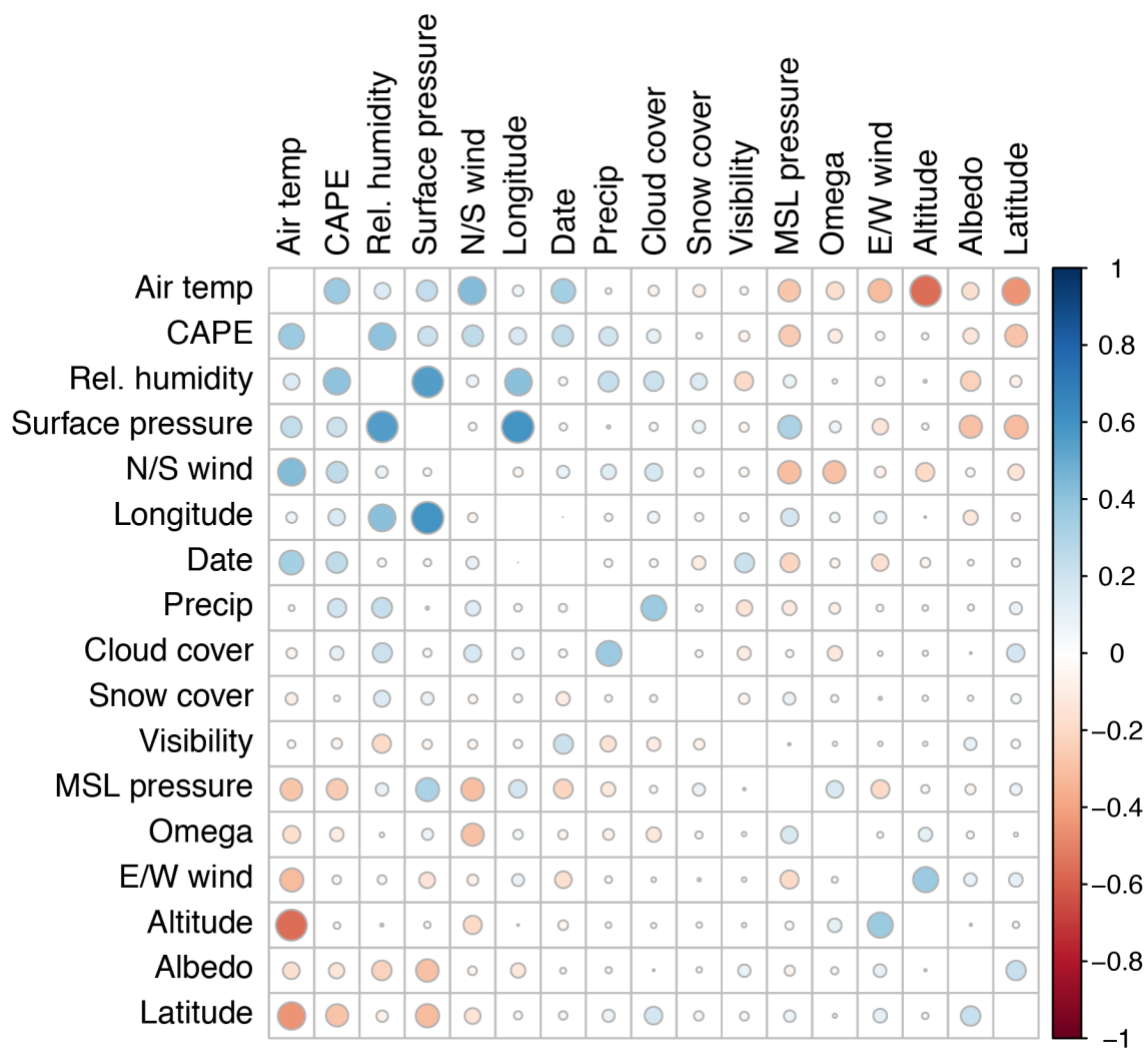


Fig. S1. Spearman rank correlations among all pairs of predictor variables. No pair of predictors had absolute Spearman or Pearson correlation coefficients greater than 0.60.

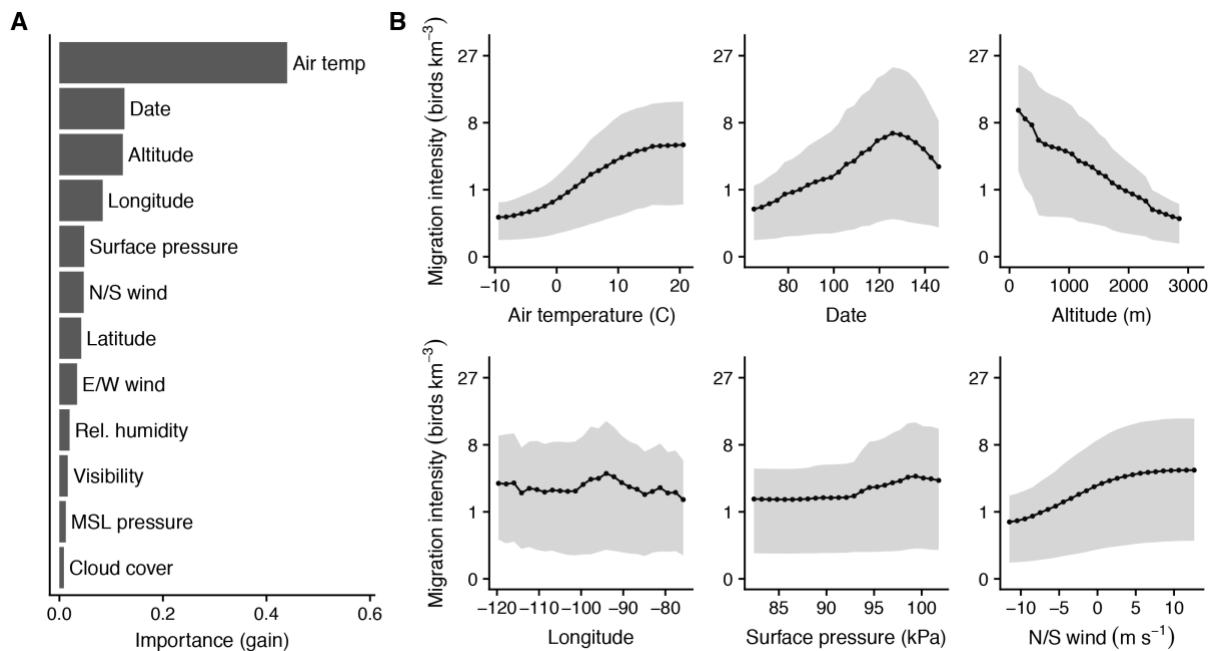


Fig. S2. Predictor importance and partial dependence. (A) Predictor importance measured by gain. Gain is a measure of each variable’s importance in making accurate predictions. (B) One-dimensional partial dependence plots for the six most important predictor variables. Solid lines show the mean and shading shows the middle 50% of predicted y-values. Note that this is not a confidence interval; it shows the marginal distribution of y-values over the values of all other predictors and should be expected to be wider than confidence limits. Narrower shading indicates that the predictor explains a greater proportion of variance in the predicted values.

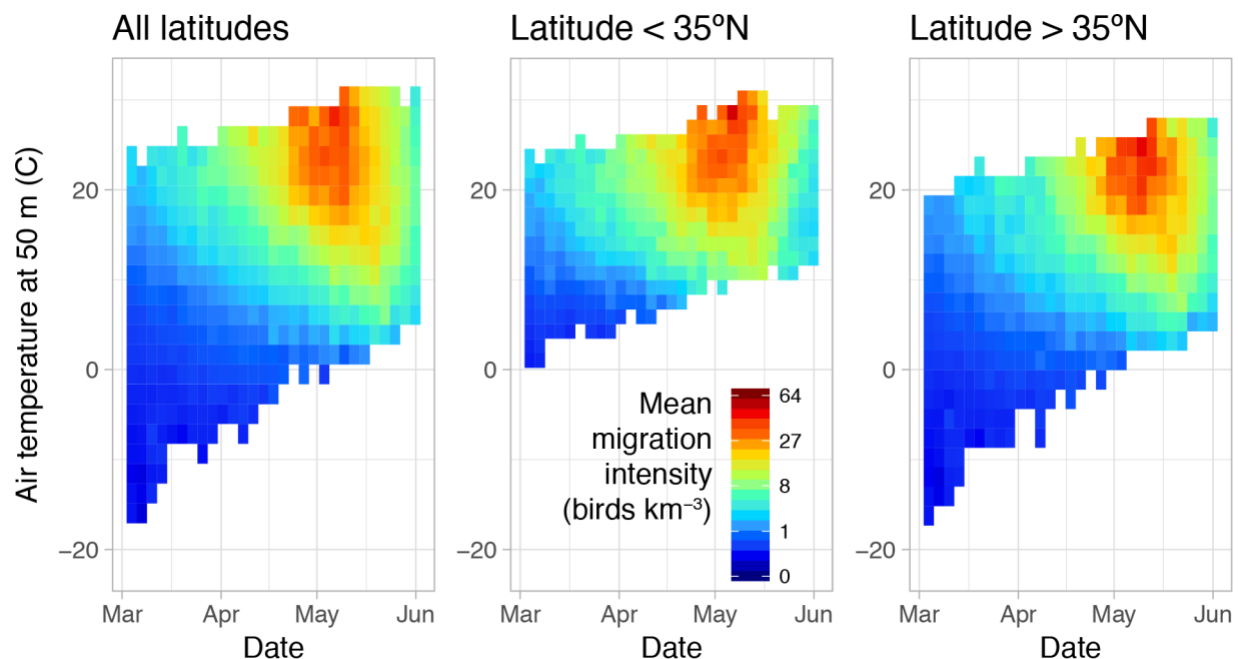


Fig. S3. Mean observed migration intensity by date and temperature close to surface. For a given date, the highest migration intensities occurred where temperatures close to the surface were warmest, especially at higher latitudes. This figure summarizes raw data and does not show model output.

5

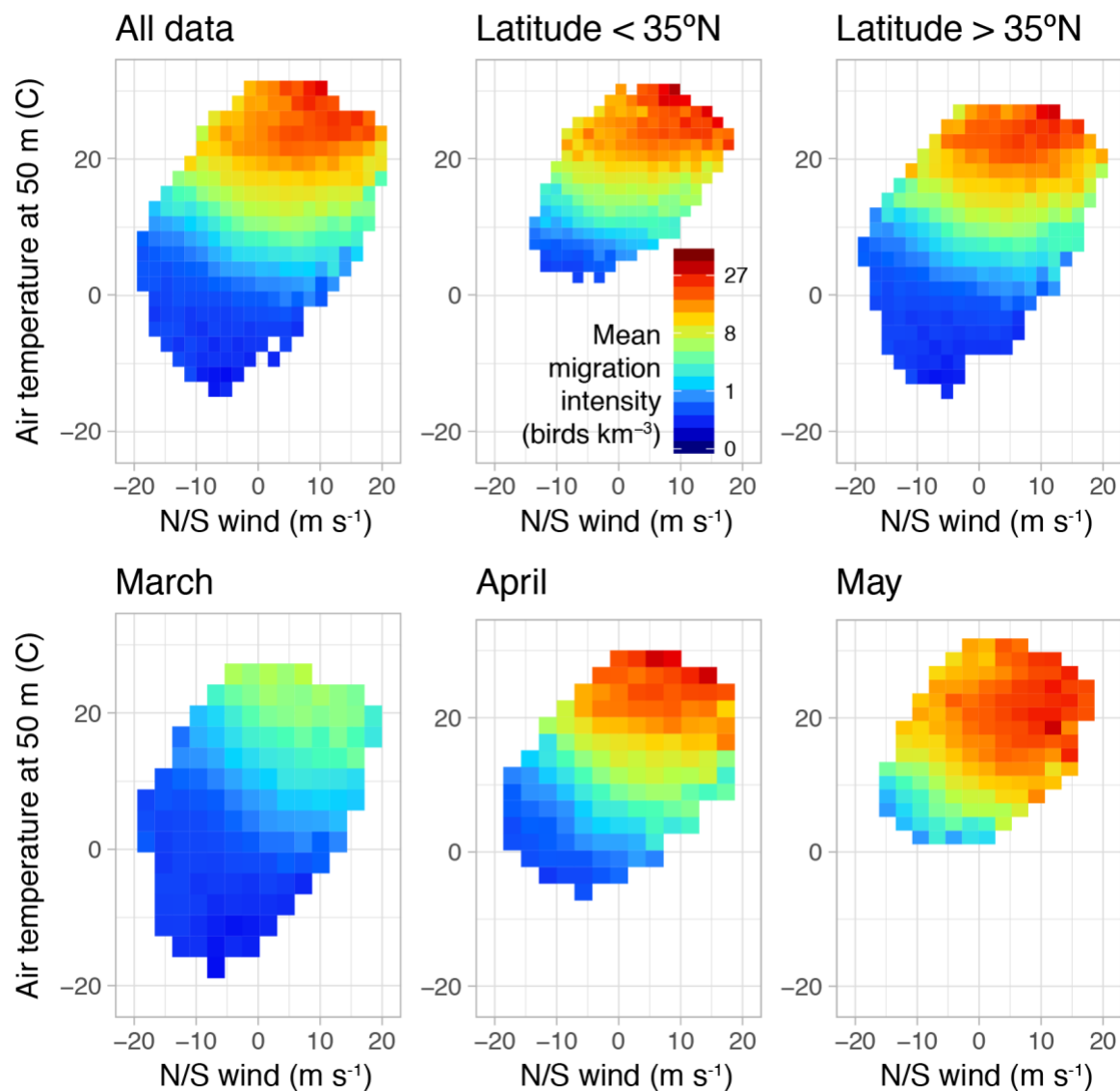


Fig. S4. Mean observed migration intensity by temperature close to the surface and wind direction. For given wind conditions, the highest migration intensities occurred where temperatures close to the surface were warmest, especially at higher latitudes. This figure summarizes raw data and does not show model output.

5

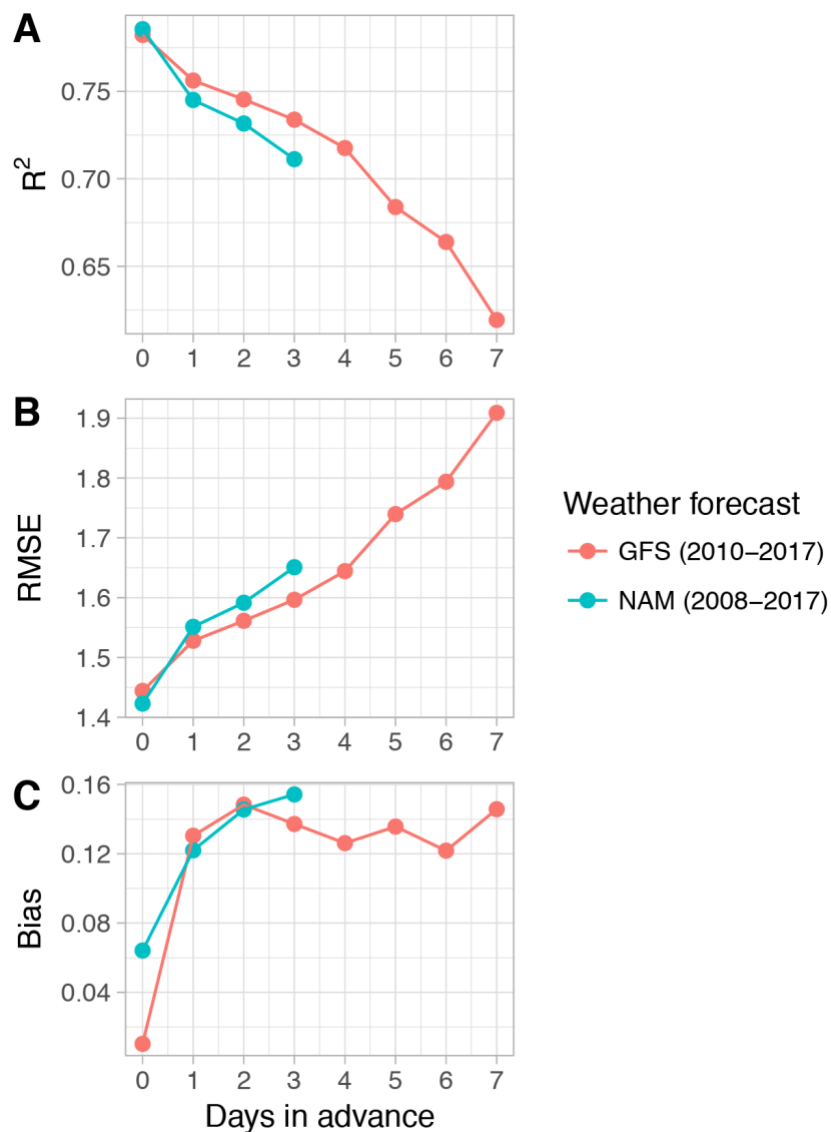


Fig. S5. Model performance using weather forecast data. We evaluated performance on the test dataset using the Global Forecast System (GFS) and North American Mesoscale (NAM) forecast system.

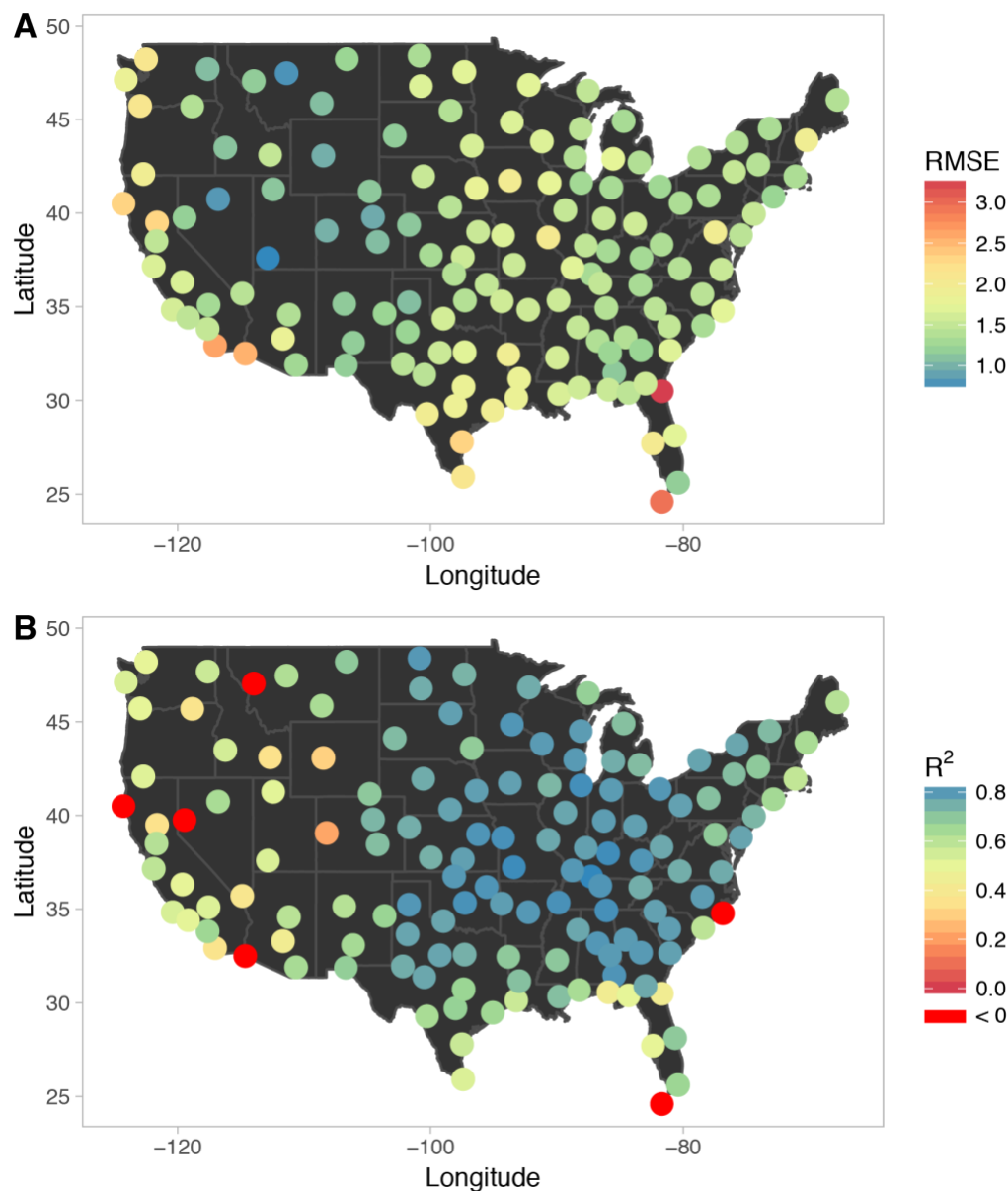


Fig. S6. Relative and absolute performance at radar stations withheld from the training dataset. Performance was best at interior sites, especially in the central and eastern United States. At a small minority (4%) of withheld sites, relative performance was poor ($R^2 < 0$, indicating performance was worse than a naive model that predicts the mean response for each observation), which may be due to local influences such as topography (e.g. see Florida Keys).

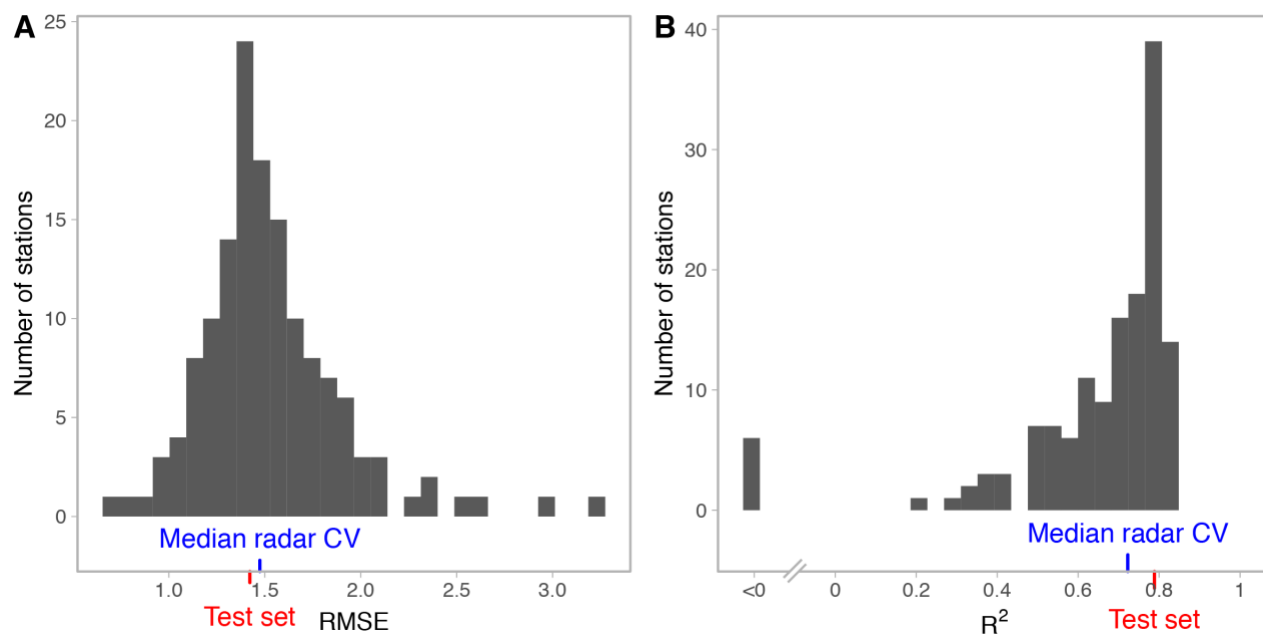


Fig. S7. Model performance at unobserved locations. Histograms show the distribution of (A) absolute and (B) relative performance metrics for radar stations that were withheld from the training dataset. The blue tick marks show the median value across sites, and the red tick marks show the corresponding value for the randomly-selected test set (all locations included).

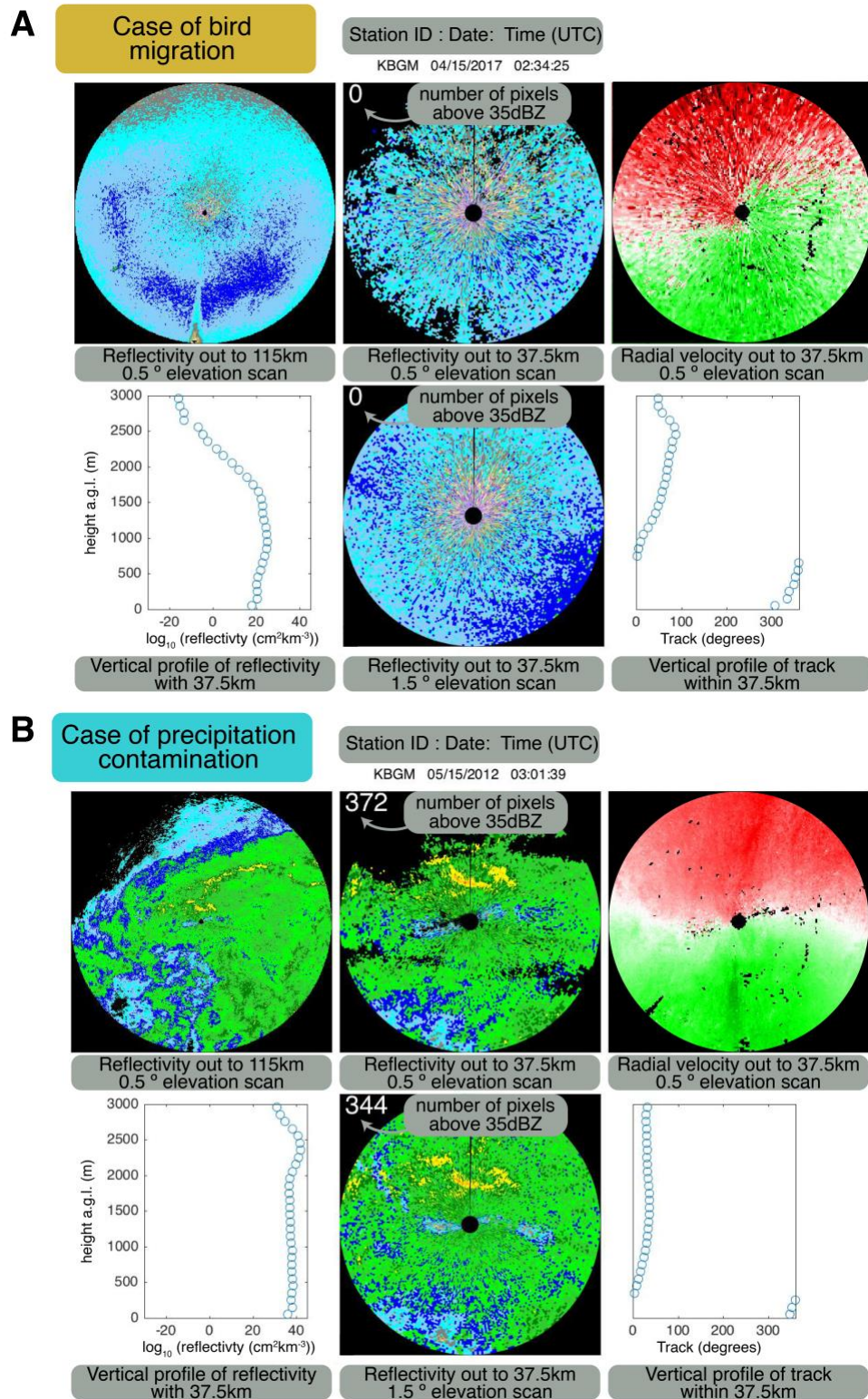


Fig. S8. Composite images used for classifying radar scans containing precipitation. Panels show example cases of (A) bird migration and (B) precipitation contamination.

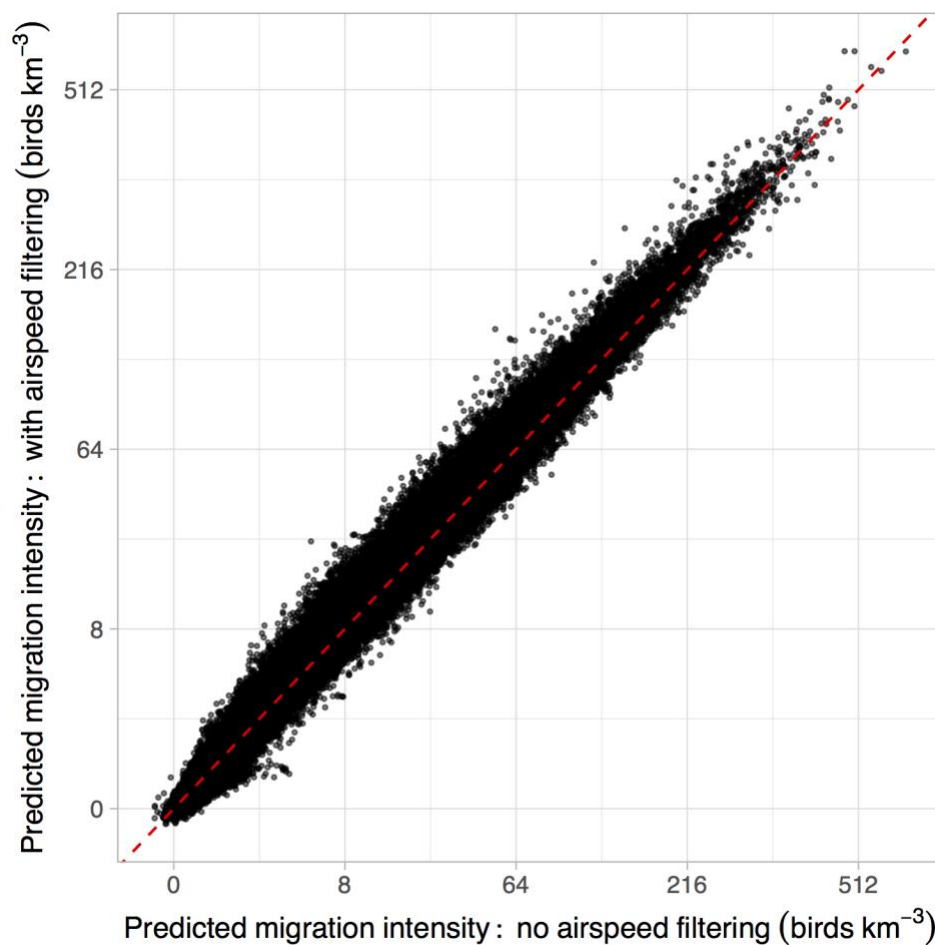


Fig. S9. Model results are robust to possible insect contamination. We removed altitude bins with mean airspeed $\leq 5 \text{ m s}^{-1}$ to limit the inclusion of flying insects in our dataset. Predictions made by a model trained without airspeed filtering corresponded closely to those made by the final model with airspeed filtering (Pearson's $r = 0.995$). Prediction error for these two models was comparable (RMSE = 1.422 with filtering; RMSE = 1.442 without filtering). The dashed red line is the identity line.

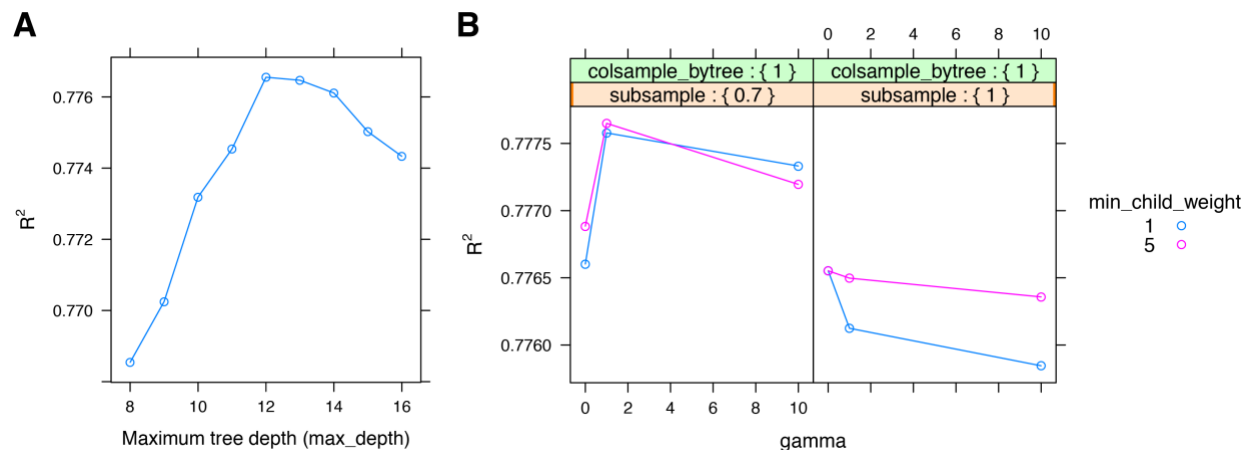


Fig. S10. Hyperparameter tuning using validation dataset. (A) We first varied maximum tree depth, and second (B) varied three parameters that may aid in limiting overfitting. Note that hyperparameter tuning affected performance only to a small degree (~1% in A and ~0.3% in B).

5