MODELING ANIMAL MIGRATION ROUTES WITH MACHINE LEARNING

Christine Cummings
Penn State University
University Park, Pennsylvania, USA

Justin Rist
Penn State University
University Park, Pennsylvania, USA

Junyan Tian
Penn State University
University Park, Pennsylvania, USA

KEYWORDS
Machine learning, animal migration, prediction, climate change

ABSTRACT

Many species are sensitive to climate change and alter their migration pathways by long distances or change their migration timing in response to changing temperatures. Machine learning algorithms are easy-to-implement tools for creating accurate and informative migration models with geolocator data. The goal of this project is to use an automated machine learning process to predict future migration routes of any animal with historic tracking data. We analyzed two species of waterfowl: (1) barnacle geese near the Barents Sea across four years; (2) tundra swans near Alaska across five years, both datasets were obtained from the Movebank database repository. Using a 75-25 test-train data split, our constructed neural network prediction model predicted the barnacle geese's and tundra swans' location the following day using the current day's location and day of the year with a loss of 0.7-2.0 (corresponding to an average error of 30-100 miles). Training the model to include temperature as a predictor, the model of barnacle geese improved but the model of tundra swans decreased, which suggests that temperature only has an impact on barnacle geese migration location. Future research should use other species data to test this algorithm of predicting migration location.

INTRODUCTION

Many animal species are dependent on migration to reproduce, avoid predators, and find reliable sources of food. Often, the timing or route of an animal migration is determined by changes in weather conditions [1]. In recent years, the phenological weather cycles have rapidly shifted due to accelerated climate change. With changing weather patterns, many animals are migrating at different times, with different pathways, and to different locations. Changes in animal migration patterns can prove to be harmful to both the animal species and to the ecosystems. Reproduction periods can be shortened or lengthened, delicate predator/prey relationships in an ecosystem can be thrown out of balance, and new migration paths can place animals at an increased risk of human interaction and harm [2].

Among all the migratory animals, we are interested in waterflow near the Arctic area because: (1) they have less interactions with human beings; (2) the bird migration in the Arctic area is one of the most important events in the natural world [3]; (3) the climate changing speed in the Arctic area is the fastest on Earth [4]; and (4) previous research indicated that climate change in Arctic area is affecting animals' migration [5]. Barnacle geese are one of the waterflow we included in this study, which predominately live in the Arctic area and migrate every year. A previous simulation study has found that increasing temperatures correlate to and can predict increases in the barnacle goose population [6]. Various indicators of climate change, such as mean daily air temperature, wind, low-altitude cloud cover and timing of spring, have also been associated with barnacle geese migration routes [7, 8]. Additionally, we also included tundra swans (cygnus columbianus) near Alaska as our second waterflow species. Even though the migration speed and route of tundra swans have been found consistent across years, their year-to-year variation was also correlated with environmental conditions [9].

Understanding how climate change affects the location and timing of future migration cycles allow us to make more effective infrastructure planning and policy change. New migration pathways could be made safe from motorways and changes in location-specific ecology could be made more predictable. However, current migration pattern prediction relies on animal-specific expert opinion or specialized mathematical models only applicable to a subset of species [10, 11]. A non-animal-specific tool that accurately and efficiently forecasts future migration locations does not presently exist. The historic data and machine learning tools to build this, however, do exist. Therefore, the present study used the combination of these resources to prove invaluable for maintaining healthy ecosystems and protecting animals from endangerment and extinction.

DATA

This project utilized migratory data of barnacle geese [7, 12] and tundra swans [13] from the Movebank Data Repository, a free archive of migratory data for various animal species, submitted by researchers for public use. Each dataset contains labeled data, from which we extracted the date and the latitude/longitude coordinates to map the movement of each animal over time. The datasets we utilized to develop and test our model contain 6,000 - 20,000 geolocation data points and span four to eight consecutive years.

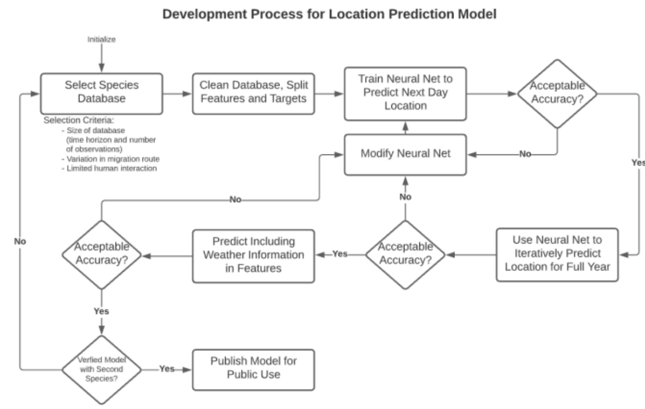
We obtained historic weather data using the Meteostat Python Package. This API allows us to obtain average daily temperature, minimum/maximum daily temperature, total precipitation, snow depth, wind direction, average wind speed, wind peak gust, sea level air pressure, and total sunshine duration from the nearest weather station based on the date and location coordinates. Some weather stations, however weather stations only report some of these data points. We recorded the relevant data from the stations within a 20-mile radius of the geolocation. Geolocations that did not contain the required weather data were eliminated. We noted the distance between the animal geolocation and the weather station.

We predict the animal's future location based on our features which contains the day of the year, the animal's latitude, the animal's longitude, and temperature data. Only the first recorded geolocation from each day was included. We will compare our predictions to our targets containing the animal's latitude and longitude on the following days to determine accuracy.

METHOD

The team proposes a three phased approach to developing a generalizable animal-migratory predictor. An overview of the process is shown in the following figure. Phase I identifies candidate animal species and corresponding weather data to train models and test performance. Phase II utilizes proposed animal location data as an input for a neural network to predict the location of an individual animal on the next day, then expanding to location prediction over longer time-horizons. Phase III expands the inputs of the Phase II model to include weather data to increase accuracy of location prediction.

The team methodology creates a generalized model that will be able to predict migration patterns, only needing inputs of historic tracking data, weather data, and location information. All information that is readily available for many animal species.



Phase I:

The team first defines attributes that would make a candidate species well-suited for testing the effectiveness of the model. The attributes selected in order of importance are as follows:

1. Sufficient availability of tracking data for given animal species, in terms of both time-horizon and total number of data entries
2. Significant variation in animal migration patterns, in terms of both timing and location of migration
3. Limited interaction of migration routes with densely human-populated areas.

Although the proposed model could be extended to animal datasets with limited entries or frequent human interaction, these characteristics would make it difficult to evaluate the effectiveness of the proposed model during preliminary stages. Animal species with highly consistent migration would provide no challenge for the model; similarly, making evaluation of the model difficult. The proposed model can be enhanced to address these challenges in future extensions of the project.

Each location entry in sufficiently large datasets available in the MoveBank data repository are plotted by year using Python. File size of each dataset allow the team to evaluate criteria 1, plots of the dataset allow the team to evaluate criterium 2 and 3. Barnacle geese near the Barents Sea and tundra swans near Alaska datasets are identified to be applied to neural network training and testing. Python generated plots of the mapping information for both species are shown in the appendix in Figures A-1 and A-2.

Phase II:

First, the selected datasets are cleaned and formatted to evaluate the prediction accuracy of the designed neural network on animal movements without consideration for weather. This phase of the process validates method selection and allows the model framework to be refined before the addition of weather data.

Tracking data is filtered to only include a single animal from the species population. Inputs for the simplified model include the date and location of the animal on the previous day with the target of animal location on the given date. A neural network created using the keras package in python is trained using the first (by date) 80% of the dataset.

The model initially outputs the location of the animal one day in advance of the provided day. After the accuracy of +1-day predictions is sufficient, the network then iteratively outputs next day location off the previous day prediction for all days in the testing set. To test the accuracy of these predictions, the actual animal location of all observations in the testing set (final 20% of dataset by date) is visually compared to model predictions.

After testing with both animal datasets, the deep neural network proposed consists of four layers of relu activation, dense connections between layers, and the number of nodes being reduced by half at every layer. Prediction results as compared to actual results are included in Appendix B. The neural network generated movement predictions along a reasonable path but often failed to predict change in direction of migration paths. The inclusion of weather data in the model, Phase III of the study, provides more dimensions to inputted data that the team hypothesize would allow for the network to predict the change in trajectory of animal migration.

Phase III:

The final stage extends the simplified model to include information about weather conditions in the location of the animal. The weather conditions inputted include daily low, high and average temperature.

The inputs for the model are the same as Phase II with the addition of the weather conditions to the location of the animal on the previous day. The target and network configuration for the model remain the same.

Training and testing of the model are conducted in the same manner as in Phase II (individual animal location sets for each species) and the same model validation process is utilized. Prediction results as compared to actual results are included Appendix B. The neural network generated movement predictions along a reasonable path and predicted the correct changes in direction for both swan and geese; however, the trajectory of movement appears a bit un-natural as compared to actual animal movement.

RESULTS

The models proposed in Phase II and Phase III were applied to four years of barnacle geese data and five years of tundra swan data. In all cases, the model was trained 10 times and the best model (lowest testing dataset loss) was saved. Table 1 contains a summary of the model prediction accuracy for the two datasets. Training loss represents the weighted error of the model's predictions of next day location using the training data used to train the model. The next day prediction loss is the weighted error of the animal's predicted next day location of testing data set. The migration route continuous prediction loss is the weighted error of the model at predicting the animal's location and migration for 1-1.5 years. This route prediction used the initial testing set datapoint to predict the next day location. Each successive day was dependent on the previous day's prediction, forming a migration path. Our predictions are shown compared to the actual values on a map in Appendix B-1 and B-2.

Table 1: Loss Comparison for Barnacle Geese and Tundra Swan Predictions

	Phase	Training set prediction loss	Next day prediction loss	Migration route prediction loss
Barnacle Geese	Phase II	0.70	0.72	12.16
	Phase III (incl. temp.)	0.67	0.67	8.2
Tundra Swan	Phase II	1.9	2.0	11.6
	Phase III (incl. temp.)	2.4	2.4	15.2

For both the barnacle geese and tundra swan predictions, the next-day predictions were accurate, however predicting the animal's full migration route was less accurate. For the barnacle geese prediction, the inclusion of temperature data increased the model prediction accuracy (reducing loss by 7% for next day prediction and by 33% for migration route prediction). This suggests barnacle geese migration is impacted by temperature. The model can then be used with simulated temperatures to forecast how barnacle gees migration could be impacted by global increases in temperature. Simulating a global temperature increase of 2 degrees Celsius, the model predicts that the geese migrate sooner and further south. These results suggest that global temperature rise will alter the barnacle geese migrations. The tundra swan prediction, however, is worsened by the inclusion of temperature in the model. Including temperature increased loss by 26% for next day prediction and by 31% for migration route prediction. This implies that tundra swan migration is largely independent of temperature.

When applying the model to other animal species datasets, it is recommended to do some research in understanding how temperature impacts the migration patterns of the species of interest. If the temperature has significant impact, we recommend that weather should be included as a dimension for input into the model. If research suggests temperature does not impact the migration patterns, we recommend excluding this dimension from model input to increase the accuracy of results.

CONCLUSION

This study used machine learning algorithm to predict animals' migration routes near Arctic area. We included barnacle geese data and tundra swans data from the Movebank Data Repository. Our result indicated that we could better predict the next-day locations of barnacle geese and tundra swans compared with the year-long migration prediction. After including temperature as a predictor, only the model of predicting barnacle geese's next-day migration improved. Therefore, temperature had an impact on barnacle geese's migration routes, but not tundra swans' migration. Previous research has also suggested that the migration speed and route of tundra swans were consistent across years [9].

By creating an accessible, generalized, and simplified migration forecasting system, governments and other environmental conservation organizations will be able to use the tool to better prepare for changes in the migration patterns of animals. The improvement in prediction will allow for the resource-limited groups to design policy to promote initiatives that will have a high impact on the well-being of environments and animal species. Improving the well-being of animals and the environment directly impacts the well-being of the human population, as stable and healthy ecosystems are vital for maintaining natural resources that permeate all aspects of human life.

This project impacts fields of engineering by providing an example of how technical tools can be applied in novel ways outside of the traditional scope of engineering to the benefit of society. Engineers are inherently problem-solvers, and the problem of climate change is a threat to society that can be mitigated with the use of the latest technological tools available.

CODE

For more information on the project and to download the code used, visit the project website: https://jsrist0028.github.io/animalmigration/website/

REFERENCES

1. Kölzsch, Andrea, Bauer, Silke, de Boer, Rob, Griffin, Larry, Cabot, David, Exo, Klaus-Michael, van der Jeugd, Henk P. and Nolet, Bart A. "Forecasting spring from afar? Timing of migration and predictability of phenology along different migration routes of an avian herbivore." *Journal of Animal Ecology* Vol. 84 No. 1 (2014): pp. 272-283. DOI:10.1111/1365-2656.12281

2. Seebacher, Frank and Post, Eric. "Climate change impacts on animal migration." *Climate Change Responses* Vol. 2 No. 5 (2015): pp. 1-2. DOI: 10.1186/s40665-015-0013-9

3. Johnson, S. R., & Herter, D. R. "Bird migration in the Arctic: a review". *Bird migration*. (1990): pp. 22-43.

4. Vincent, Warwick F. "Arctic climate change: Local impacts, global consequences, and policy implications." In *The Palgrave handbook of Arctic policy and politics.* Palgrave Macmillan (2020): pp.507-526.

5. Davidson, Sarah C., Bohrer, Gil, Gurarie, Eliezer, LaPoint, Scott, Mahoney, Peter J., Boelman, Natalie T., ... & Hebblewhite, Mark (2020). "Ecological insights from three decades of animal movement tracking across a changing Arctic". *Science* Vol. 370 No. 6517: 712-715. DOI: 10.1126/science.abb7080

6. Trinder, Mark N., Hassell, David and Votier, Stephen. "Reproductive performance in arctic-nesting geese is influenced by environmental conditions during the wintering, breeding and migration seasons." *Oikos* Vol. 118 No. 7 (2009): pp. 1093-1101. DOI: 10.1111/j.1600-0706.2009.17429.x

7. Shariati-Najafabadi, Mitra, Darvishzadeh, Roshanak, Skidmore, Andrew K., Kölzsch, Andrea, Exo, Klaus-Michael, Nolet, Bart A., ... and Toxopeus, Albertus G. "Environmental parameters linked to the last migratory stage of barnacle geese en route to their breeding sites." *Animal behaviour* Vol. 118 (2016): pp. 81-95. DOI: 10.1016/j.anbehav.2016.05.018

8. Hupp, Jerry W., Ward, David H., Soto, David X. and Hobson, Keith A. "Spring temperature, migration chronology, and nutrient allocation to eggs in three species of arctic-nesting geese: Implications for resilience to climate warming." *Global Change Biology* Vol. 24 No. 11 (2018): pp. 5056-5071. DOI: 10.1111/gcb.14418

9. Nuijten, R. J. (2020). *Bewick's Swans in a Changing World: Species Responses and the Need for Dynamic Nature Conservation.* Universiteit van Amsterdam.

10. Moore, Thomas T. "Climate change and animal migration." *Envtl. L.* Vol. 41 (2011): pp. 393.

11. Kristensen, Nadiah P., Johansson, Jacob, Ripa, Jörgen and Jonzén, Niclas. "Phenology of two interdependent traits in migratory birds in response to climate change." *Proceedings of the Royal Society B: Biological Sciences* Vol. 282 No. 1807 (2015). DOI: 10.1098/rspb.2015.0288.

12. van der Jeugd H, Oosterbeek K, Ens BJ, Shamoun-Baranes J, Exo K (2014) Data from: Forecasting spring from afar? Timing of migration and predictability of phenology along different migration routes of an avian herbivore [Barents Sea data]. Movebank Data Repository. doi:10.5441/001/1.ps244r11

13. Ely, C. R. and B. W. Meixell. 2016. Demographic outcomes of diverse migration strategies assessed in a metapopulation of tundra swans. Movement Ecology. doi:10.1186/s40462-016-0075-8

APPENDIX A

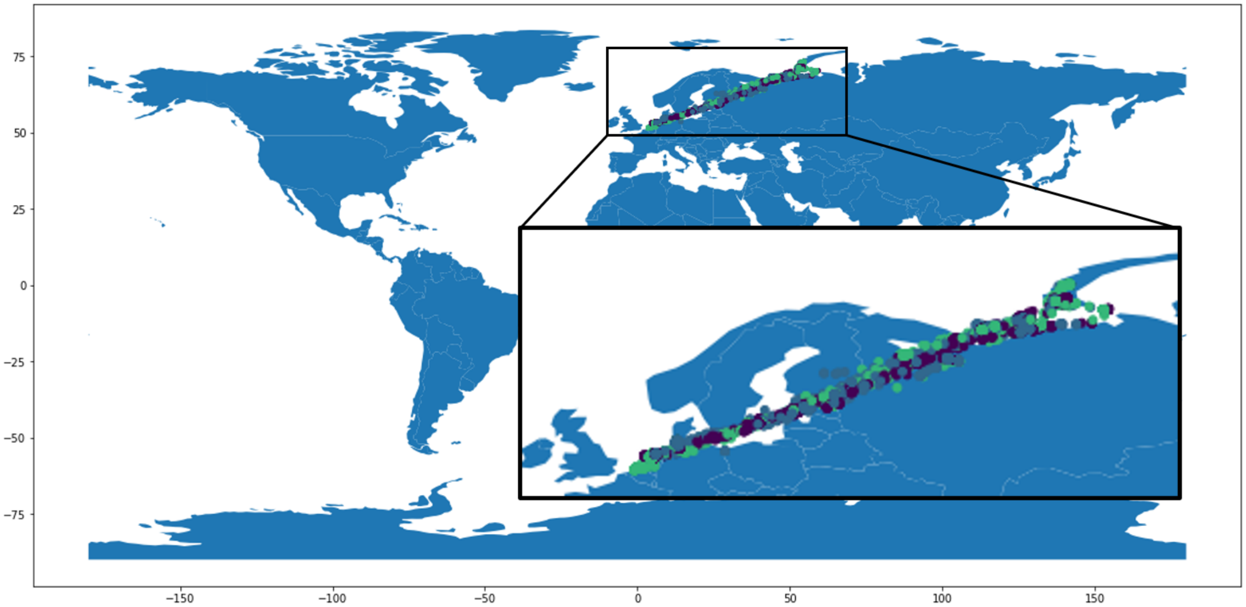


Figure A-1: Tracked movements of 14 barnacle geese over 4 years (January 2008 - April 2011)

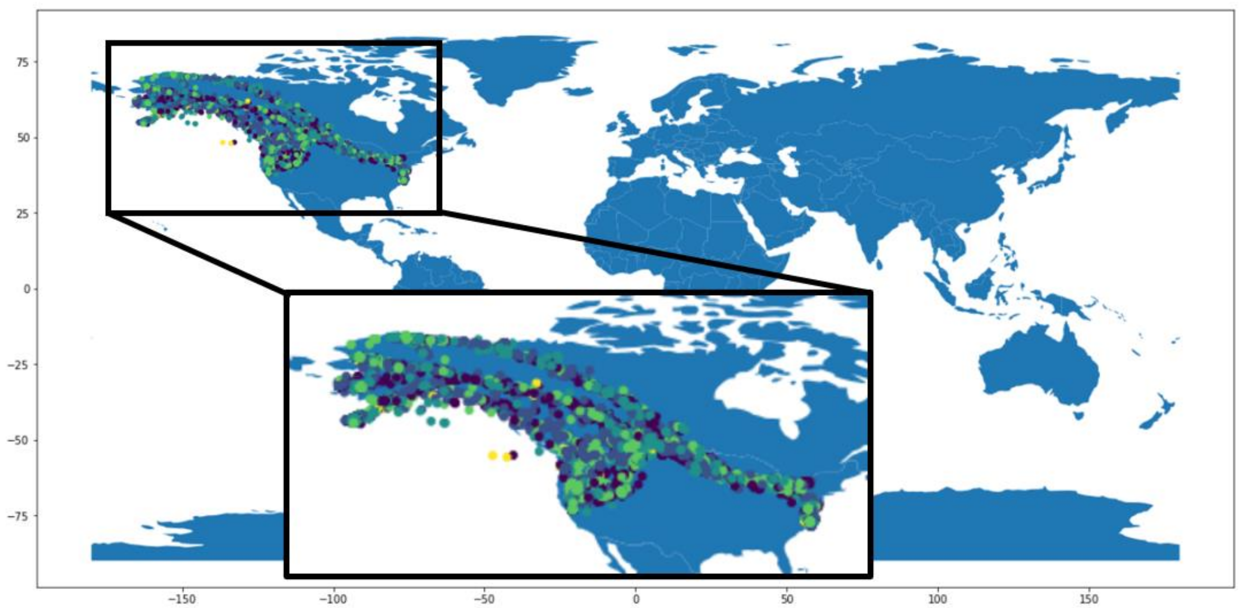


Figure A-2: Tracked movements of 49 swans over 5 years (January 2008- February 2012)

APPENDIX B

Next day and continuous migration route prediction with and without temperature. B-1) For barnacle goose model prediction, the results with temperature are more accurate. Additionally, the next day location prediction is more accurate than the migration route continuous prediction. B-2) For the swan model prediction, the results with temperature are less accurate than the results from the model that was not trained to include temperature. Similar to the goose prediction results, the swan's next day location prediction is more accurate than route prediction.

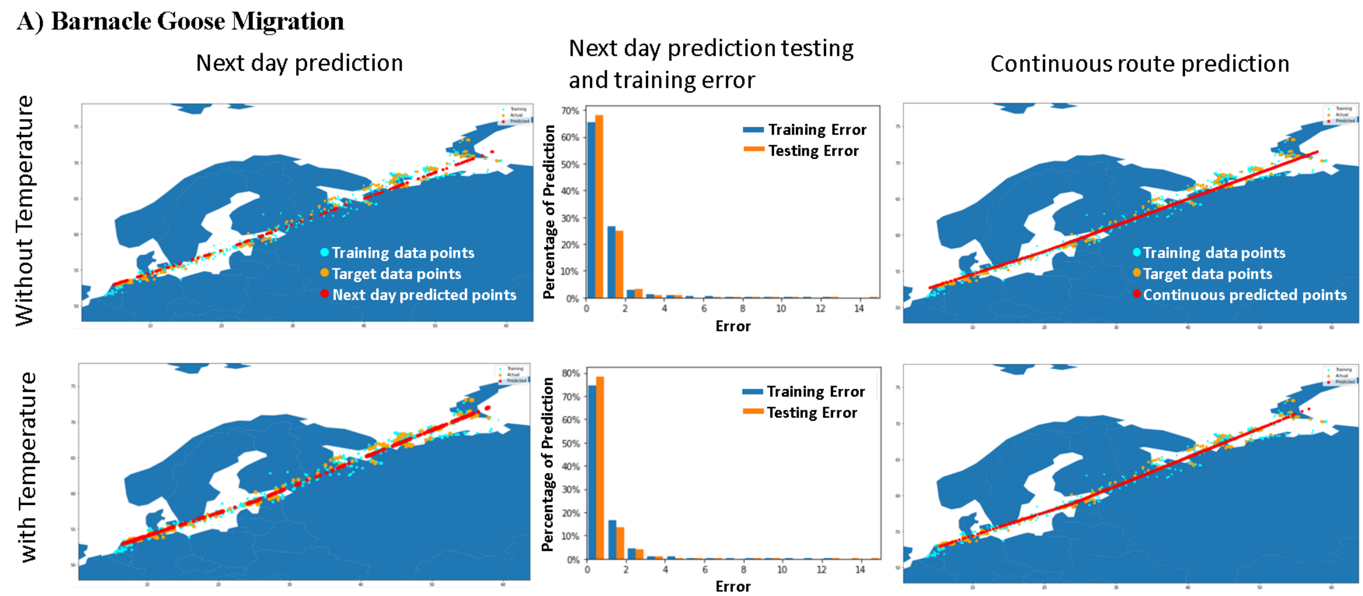


Figure B-1

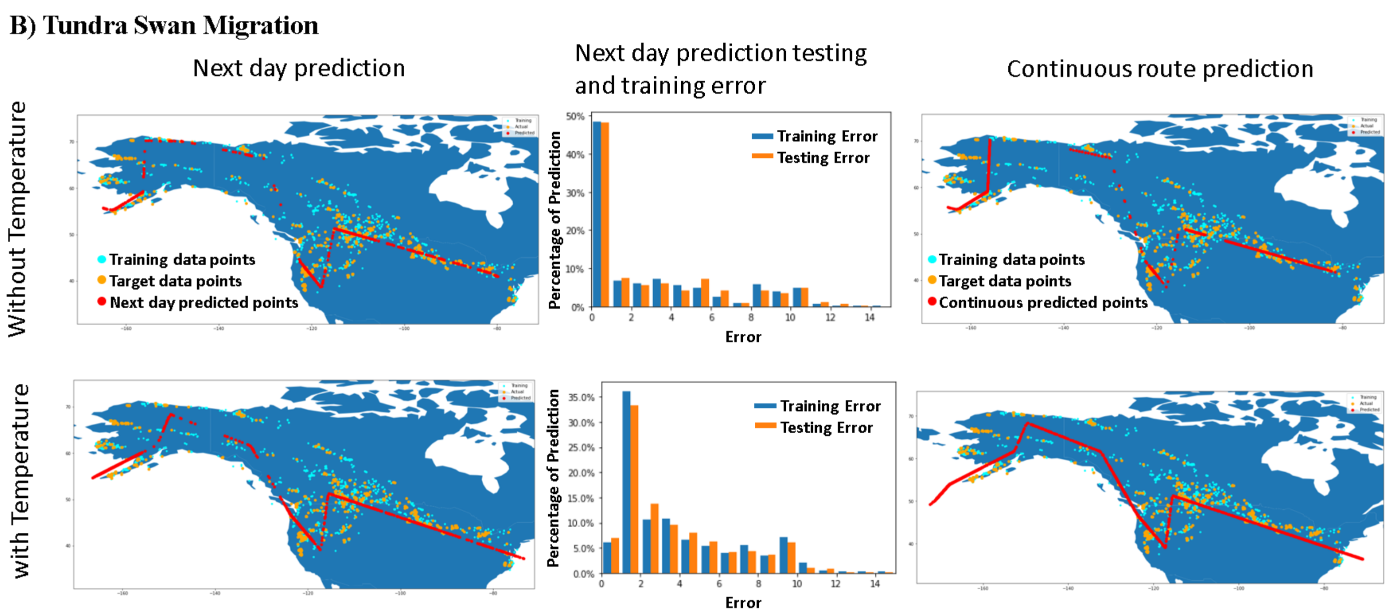


Figure B-2