Spatio-Temporal Trend Analysis of Historic Bird Arrival Data

Ali Arab*

Abstract
There is increasing interest in spatio-temporal analysis of environmental and ecological response to changes in the climate due to the recent concerns about climate change. In this work we propose a spatio-temporal modeling framework that is suitable for analyses of environmental and ecological data while accounting for spatial and temporal structure, as well as climate effects. As an example, we consider data on bird migration in the United States and analyze the spring arrival dates of Purple Martins between historic data (1905-1940) from the North American Bird Phenology Program and recent data (2001-2010) from the Purple Martin Conservation Association. The proposed approach allows researchers to compare mean arrival dates while accounting for spatial and temporal variability. Our results for Purple Martins showed statistically significant late arrivals in parts of the United States (South, East, Midwest). However, no statistically significant change in mean arrival dates were detected in the Northern U.S. (including Great Lakes area). The proposed approach provides a useful tool for statistical analysis of spatio-temporal data related to climate change studies.

Key Words: Bird migration, climate change, spatio-temporal models, Bayesian

1. Introduction

The study of environmental and ecological response to climate change in the recent years has provided ample evidence of the ecological impacts of recent climate change (e.g., Walther et al. 2002). In particular, bird migration is known to be sensitive to changes in the climate and thus, recently there is increasing interest in analyzing potential changes in the migration patterns of migratory bird that may provide insight on environmental and ecological response to climate change.

The history of bird migration studies dates back to Aristotle who compiled notes on more than 140 species of birds and formalized ornithology as a science (Alerstam 1990; Berthold 2001). Historically, ecologists and ornithologists have studied patterns of bird migration to learn about individual or groups of bird species, as well as to understand the ecological impact of long- and short-term migration on local and global ecosystems. Recently, statistical analysis of bird migration and phenological changes has become increasingly popular in the context of more general problems such as climate change (e.g., Møller et al. 2004; Cox 2010) and epidemiology of infectious diseases that are linked to bird migration such as avian influenza outbreaks (e.g., Liu et al. 2005; Feare 2007; Bourouiba et al. 2010). Often, these analyses require spatial or spatio-temporal models due to the nature of migration data. There are several recent examples of such efforts in ornithology (e.g., Tøttrup et al. 2006; Hüppop and Winkel 2006) and epidemiology literature (e.g., Munster et al. 2007; Onozuka and Hagihara 2008; LaDeau et al. 2008; Si et al. 2009; LaDeau et al. 2010).

In this paper, we focus on the analysis of migratory birds data in order to detect shifts in spatio-temporal patterns of spring arrival dates in the United States (specifically, east of the Rocky Mountains). Notwithstanding the spatial and spatio-temporal nature of the spring arrival process, the literature on analysis of spring arrival dates using spatial and spatio-temporal models is sparse (e.g., Gordo 2007; and Both and et Marvelde 2007, use spatial

*Department of Mathematics and Statistics, Georgetown University, 37th and O streets, Washington, DC 20057 (Email: aa577@georgetown.edu)

models; Fink et al. 2010, Hulbert and Liang 2012, use spatio-temporal models). In this paper, our goal is to develop a straightforward spatio-temporal approach for analysis of spring arrival data. The proposed framework allows us to include weather, climate and other types of predictor variables in the model. The main focus is on developing an exploratory data analysis tool for inferential purposes. However, the flexibility of the proposed framework allows for using this approach for predictive purposes too. As a case study for spatio-temporal analysis of spring arrival dates, we discuss the analysis of historic and recent data on Purple Martins. Section 2 discusses the data and introduces the methodology. Results are given in Sections 3, followed by discussion and conclusions in Section 4.

2. Materials and Methods

2.1 Spring Arrival Data

The Purple Martin (*Progne subis*) is the largest member of the swallow family in North America and are of special interest to birders, in large part, because of the close proximity of their nesting sites to human settlements. Purple Martins spend their non-breeding season in Brazil and migrate to North America to nest. Adult Purple Martins commonly return to the same nesting sites where they were successful in previous years (see e.g., www.purplemartin.org for more information).

The North American Bird Phenology Program (NABPP), housed in the United States Geological-Patuxent Wildlife Research Center (USGS-PWRC), was revitalized in 2008 (Zelt et al. 2012). The NABPP houses millions of data index cards on more than 200 bird species collected over a 90 year span between 1881-1970. The NABPP data collection is the product of records collected by a network of over 3,000 volunteers on bird migration, breeding, wintering, and distribution. As of date, over a million handwritten records have been scanned, in an effort to digitize the data and are going through a thorough data validation process. Once validated, the records will be accessible online by biologists, managers, and members of the general public.

Due to low sampling efforts during the early decades as well as the last decades of the existence of NABPP, we chose data during 1905-1940 on arrival dates of Purple Martins. We label these historic records as “old” data in our analysis. We also use data from Courter (2012) on arrival dates of Purple Martins between 2001-2010 collected by the Purple Martin Conservation Association (www.purplemartin.org). In our analysis, we label the recent data as “new” data. Unfortunately, there are no comprehensive and reliable sources (or no straightforward method) to compile data on arrival records of Purple Martins for the period between the 1960s and the late 1990s based on acceptable spatial coverage and sampling effort that is of interest in this study.

Here, we convert the arrival dates to Julian Date (or Day-of-Year) calendar, which is based on the number of days in a calendar year starting January 1st for each year. For example, an arrival date of February 1st, translates to 32 under the Julian Date Calendar, as it is the 32nd day of the year (of course, one has to account for leap years accordingly). We consider a spatial grid with ten irregular sized cells (Figure 1). The spatial grid and cell sizes were decided based on a data criteria which requires each grid cell and for each year to include at least five data points to achieve reasonable variability in arrival data within each cell.

Since we are interested in understanding the relationship between migration patterns and climate, we include climate effects as predictor variables in the model. As an example, we consider data on Winter North Atlantic Oscillation (Winter NAO or WNAO; http://climatedataguide.ucar.edu/). Similarly, other climate indices and weather variables

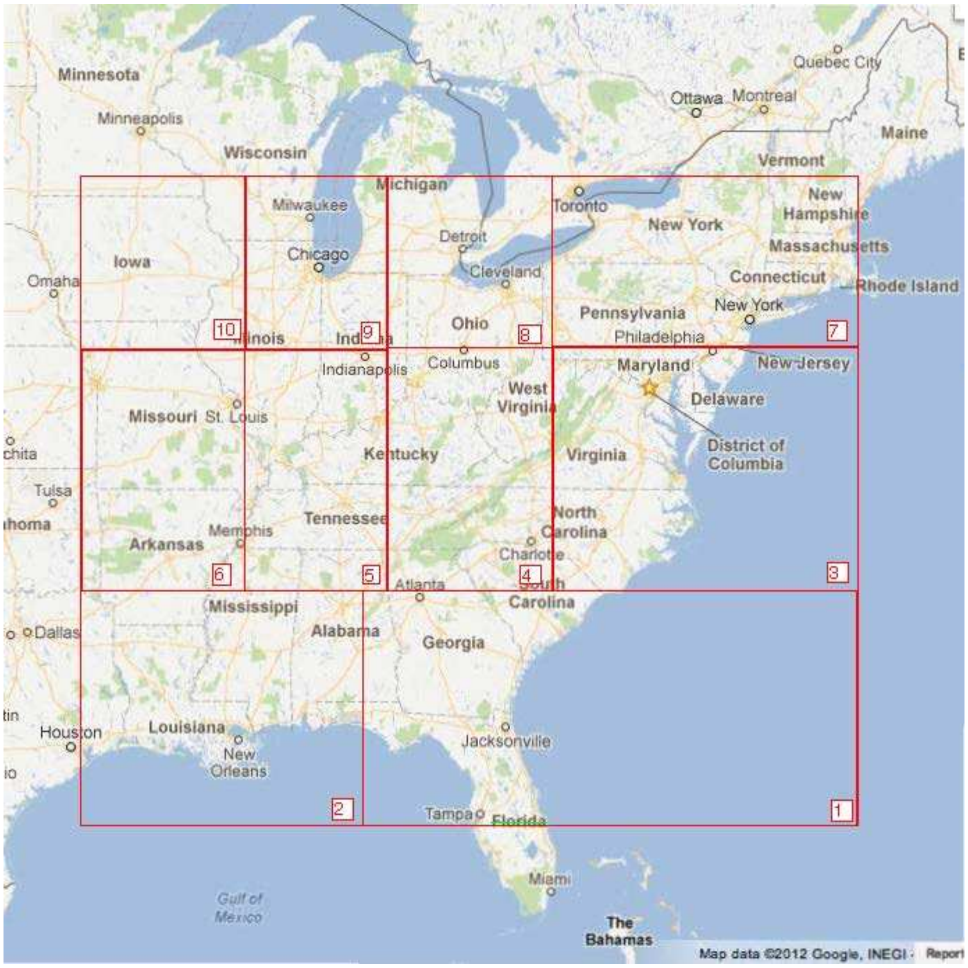

Figure 1: Map with grid cells.

can be easily included in the model.

2.2 Spatio-Temporal Model

Let $\mathbf{Y}_t = (Y_{1,t}, \ldots, Y_{n,t})'$ denote the vector of mean arrival days for the grid cells ($n = 1, \ldots, 10$) over the total number of years in the study ($t = 1, \ldots, 46$; 36 years in the old data for 1905-1940, and 10 years in the new data for 2001-2010) where $Y_{i,t}$ represent the mean arrival days for the i-th grid cell in the t-th year. Using a hierarchical modeling framework (Berliner 1996) which relies on three stages of data, process, and parameter models, we define the following *Data Model*

$$\mathbf{Y_t} \sim \mathbf{N}(\mathbf{m_t}, \sigma^{\mathbf{2}}\mathbf{I}) \tag{1}$$

where the observed arrival days in (1) are assumed to be conditionally independent (conditioned on a process model that accounts for spatial and temporal dependence).

The *Process Model* is defined based on a time series threshold modeling approach (Tong 1983; Geweke and Terui 1993)

$$\mathbf{m}_t = \boldsymbol{\mu}_0 + \begin{cases} b_{0,1} + \boldsymbol{\mu}_{1,sp} + b_{1,1}\mathbf{X_t} + \mathbf{e_{1,t}} & \text{if } 1 \leq t \leq 36 \text{ (years 1905-1940)} \\ \\ b_{0,2} + \boldsymbol{\mu}_{2,sp} + b_{1,2}\mathbf{X_t} + \mathbf{e_{2,t}} & \text{if } 37 \leq t \leq 46 \text{ (years 2001-2010)} \end{cases}$$

where $\boldsymbol{\mu}_0 = (\mu_{0,1}, \ldots, \mu_{0,n})'$ denotes the spatially-varying common mean for the old and the new data, $\boldsymbol{\mu}_1 = (\mu_{1,1}, \ldots, \mu_{1,n})'$ denotes the spatially-varying mean specific to the old data, and $\boldsymbol{\mu}_2 = (\mu_{2,1}, \ldots, \mu_{2,n})'$ denotes the spatially-varying mean specific to the new data. The predictor data on Winter NAO is given in the variable $\mathbf{X_t}$ with different coefficients for old ($b_{1,1}$) and new data ($b_{1,2}$). Parameters $b_{0,1}$ and $b_{0,2}$ represent the constant means for the old and new data, respectively. Also, we consider different autoregressive error processes $\mathbf{e}_{1,t}$ and $\mathbf{e}_{2,t}$, for the old and new data, respectively.

The autoregressive error processes are based on the following AR(1) models (e.g., Cressie and Wikle 2011):

$$\mathbf{e}_{1,t} = \nu_1 \mathbf{e}_{1,t-1} + \eta_{1,t}, \quad \eta_{1,t} \sim N(0, \sigma^2_{\eta 1}) \tag{2}$$

$$\mathbf{e}_{2,t} = \nu_2 \mathbf{e}_{2,t-1} + \eta_{2,t}, \quad \eta_{2,t} \sim N(0, \sigma^2_{\eta 2}) \tag{3}$$

and the spatial structure for the spatially-varying parameters $\boldsymbol{\mu}_p$, for $p = 0, 1, 2$, are based on a Conditional Autoregressive (CAR) model (see e.g., Cressie 1993; Banerjee et al. 2004; Arab et al. 2008)

$$\mu_{p,l}|\mu_{p,s}, \tau^2_{p,l} \sim N(\bar{\mu}_{p,l} + \sum_{s \in N_l} c_{p,ls}(\mu_{p,s} - \bar{\mu}_{p,s}), \tau^2_{p,l}). \tag{4}$$

where $l, s = 1, \ldots, n$, and $c_{p,ls}$'s are weights defined such that $c_{p,ls} = 1$ for $l \neq s$, $c_{p,qq} = 0$ for $q = 1, \ldots, n$, and $c_{p,ls}\tau^2_{p,l} = c_{p,sl}\tau^2_{p,s}$.

Inference is conducted in a Bayesian framework using a Markov Chain Monte Carlo (MCMC). The Bayesian framework requires that we define prior distributions for unknown parameters (also called the *Parameter Models* in the hierarchical framework). We define, the following relatively non-informative prior distributions for the unknown parameters

$$\begin{aligned}
b_{j,k} &\sim N(\mu = 0, \sigma^2 = 100), \quad j = 0, 1, \ k = 1, 2, \\
\sigma^2 &\sim InvGamma(\text{mean} = 1, \text{var} = 100) \\
\nu_1 &\sim Uniform(-1, 1) \\
\nu_2 &\sim Uniform(-1, 1) \\
\sigma^2_{\eta_1} &\sim Uniform(0, 100) \\
\sigma^2_{\eta_2} &\sim Uniform(0, 100)
\end{aligned}$$

We also define the following prior distribution for the variance components of the CAR Priors (i.e., hyperparameters for the CAR priors)

$$\tau^2_p \sim InvGamma(\text{mean} = 1, \text{var} = 100) \ \ p = 0, 1, 2. \tag{5}$$

Also, for the AR(1) models in (2) and (3), we need to define initial states, $\mathbf{e}_{1,0}$ and $\mathbf{e}_{2,0}$. We assign the following prior distributions for these initial states

$$\begin{aligned}
\mathbf{e}_{1,0} &\sim N(0, \sigma^2_{\eta_1}) \ \ \text{(Old data; years } 1905, \ldots, 1940) \\
\mathbf{e}_{2,0} &\sim N(0, \sigma^2_{\eta_2}) \ \ \text{(New data; years } 2001, \ldots, 2010)
\end{aligned}$$

Note that we have already defined prior distributions for the variance parameters (i.e., hyperparameters) $\sigma^2_{\eta_1}$ and $\sigma^2_{\eta_2}$.

Table 1: Posterior results for the model parameters.

Parameter	Posterior Mean	Posterior St. Dev.	95% Credible Interval
$b_{0,1}$	16.36	6.73	(3.177, 29.51)
$b_{0,2}$	1.675	6.721	(-11.55, 14.83)
$b_{1,1}$	35.31	4.298	(26.7, 43.73)
$b_{1,2}$	-11.06	9.605	(-30.52, 7.325)

Table 2: Posterior results for the overall difference in spatial means of the old $(b_1 + \boldsymbol{\mu}_{1,sp})$ and new data $(b_2 + \boldsymbol{\mu}_{2,sp})$.

Grid Cell	Posterior Mean	95% Credible Interval
1	18.02	(9.42, 26.41)
2	21.27	(12.63, 29.69)
3	17.50	(8.89, 25.84)
4	22.75	(14.20, 31.06)
5	19.90	(11.36, 28.20)
6	10.49	(1.89, 18.88)
7	14.39	(5.74, 22.77)
8	7.61	(-0.98, 15.95)
9	7.74	(-0.87, 16.10)
10	7.22	(-1.40, 15.63)

3. Results

The MCMC algorithm was implemented in OpenBUGS (http://www.openbugs.info/) for 100,000 iterations. We discarded the first 10,000 iteration for "burn-in" and based our inference on the remaining 90,000 iterations. The MCMC algorithm achieved convergence rapidly within the first few thousand iterations. Convergence was assessed using visual inspection, as well as autocorrelation of the MCMC chains.

Our results show significant changes in arrival dates of Purple Martins in recent years. Table 1 shows the posterior inference for the regression parameters. Table 2 shows the inference for the overall difference in total means for the new and old data (combined mean effect of the constant and spatially-varying means, $b_{0,k} + \boldsymbol{\mu}_k$ for $k = 1, 2$).

Figure 2 shows boxplots of the posterior distributions of the common spatially-varying mean for the two periods. Figures 3 and 4 show boxplots of the posterior distributions of the spatially-varying specific to the old and new data, respectively. Figure 5 shows boxplots of the posterior distributions of the difference of the spatially-varying means for the old and new data. Critically, the spatial structure (strong latitudinal effect and mild to weak longitudinal effect) pronounced in Figures 2-4 provides strong justification for the need for spatially-varying mean parameters in the model.

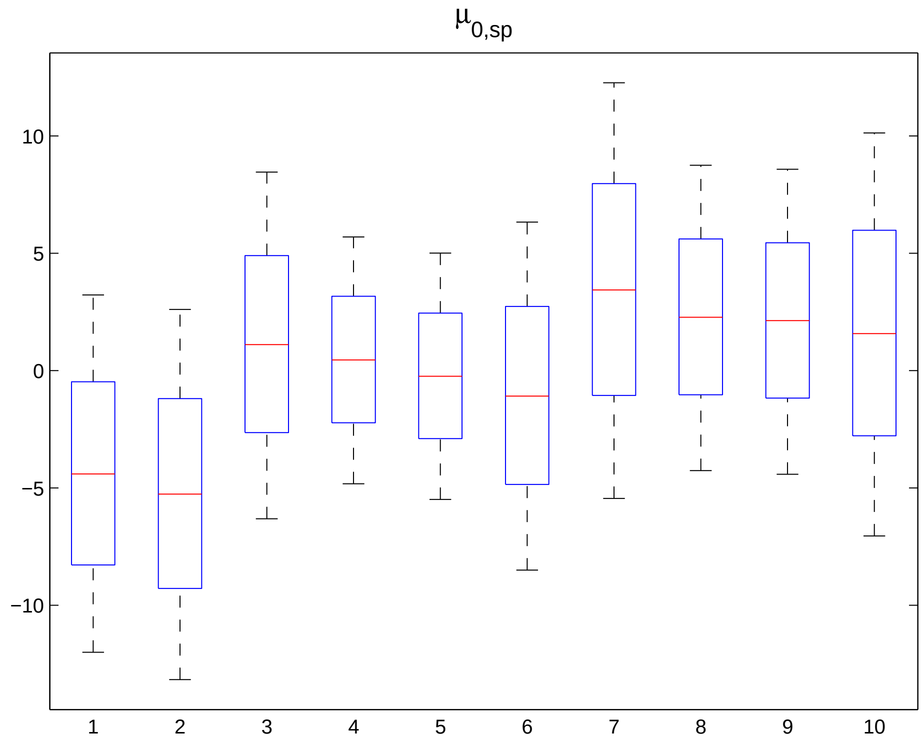

Figure 2: Common spatially-varying mean for the two periods.

4. Discussion and Conclusions

Our model results show significant shifts in the mean arrival days of Purple Martins in the South, East and part of North West of the United States (See Table 2) with significantly earlier arrivals for the recent data compared to the old data. This may be an indication of the linkage between the recent changes in the climate (i.e., global warming) and shifts in the Purple Martin migration patterns.

Also, we have detected a significantly positive association between the Winter NAO index and the mean arrival days for the old data (1905-1940). No significant effect of Winter NAO was detected for the new data (See Table 1). We suspect that this may be mainly due to low variability in the Winter NAO data for the recent data since the 2000s Winter NAO values are mostly negative with low variability (e.g., the standard deviation of Winter NAO values for the old period is more than 2.5 times the standard deviation of the values for the recent period.).

As mentioned in the previous section, the inherent spatial latitudinal and longitudinal structure verifies the importance of considering spatially-varying mean parameters. Critically, our results how that shifts in arrival patterns of Purple Martins are not constant over space.

Potential future directions include analysis of multivariate arrival dates for closely related bird species, and characterize the potential association between the changes in the arrival dates and climate change. In this work, as an example, we used a climate index (Winter NAO) as a predictor variable in the model. However, for a thorough investigation of the link between changes in the climate and shifts in migration patterns, one should consider other related weather variables (e.g., temperature, precipitation) and climate indices

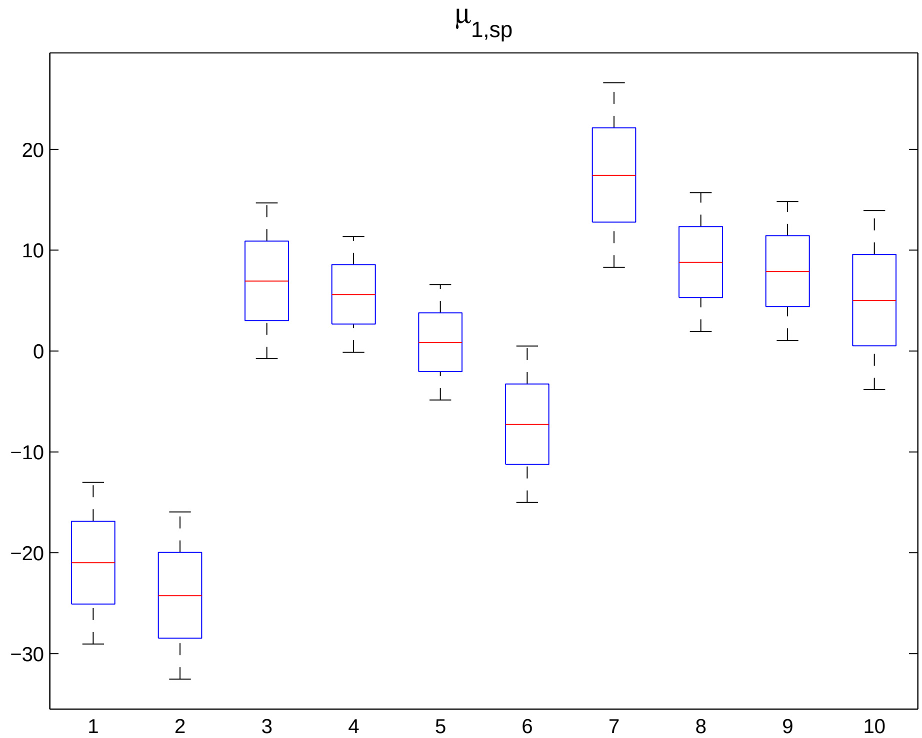

Figure 3: Spatially-varying mean for the 1905-1940 data.

(e.g., North Pacific (NP); Atlantic Multi-decadal Oscillation (AMO); information on El Niño and La Niña seasons).

5. Acknowledgements

We would like to thank Jessica Zelt and Sam Droege from the United States Geological Survey-Patuxent Wildlife Research Center and the North American Bird Phenology Program for providing data. Also, We would like to thank Jason Courter (Taylor University) and Ron Johnson (Clemson University), and the Purple Martin Conservation Association for providing data for the recent period. This work was partially supported by a Georgetown University Junior Faculty Fellowship during Spring 2011 and partial travel support was provided by a Georgetown University Non-Competitive Grant-in-Aid.

REFERENCES

Alerstam, T. (1990), *Bird Migration*, Cambridge, UK: Cambridge University Press.

Arab, A., Hooten, M. B., and Wikle, C. K. (2008), "Hierarchical Spatial Models". In *Encyclopedia of Geographical Information Science*. Editors: Shashi Shekhar and Hui Xiong. pp. 425-431. Springer: New York.

Banerjee, S., Carlin, B.P., and Gelfand, A.E. (2004), *Hierarchical modeling and analysis for spatial data*. Chapman & Hall/CRC, Boca Raton.

Berliner, L.M. (1996), "Hierarchical Bayesian Time Series Models". In K. Hanson and R. Silver (eds), *Maximum Entropy and Bayesian Methods*, Kluwer Academic Publishers, Dordrecht, pp. 15–22.

Berthold, P. (2001), *Bird Migration: A General Study* (2nd ed.), Oxford, UK: Oxford University Press.

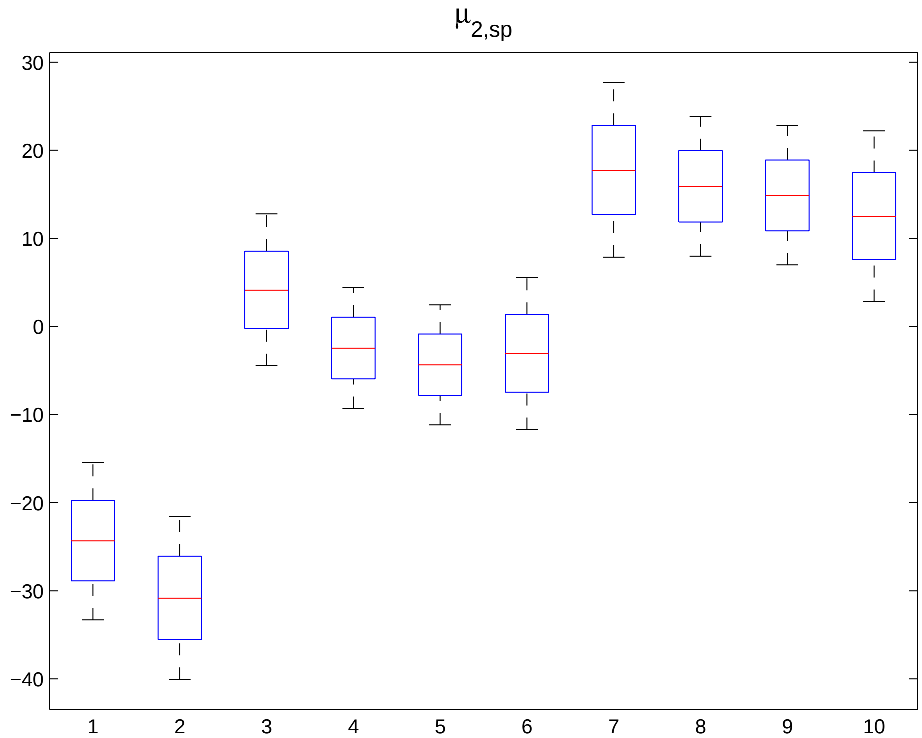

Figure 4: Spatially-varying mean for the 2001-2010 data.

Both, C., and te Marvelde, L. (2007), "Climate Change and Timing of Avian Breeding and Migration Throughout Europe". *Climate Research*, 35, pp. 93–105

Bourouiba, L., Wu, J., Newman, S., Takekawa S., and Natdorj, T., et al. (2010), "Spatial Dynamics of Bar-Headed Geese Migration in the Context of H5N1", *Journal of the Royal Society Interface*, 7, pp. 1627–1639.

Courter, J.R. (2012), "Avian Phenology, Climate, and Land-use Conservation Assessed at Broad Spatial and Temporal Scales Using Interdisciplinary Approaches and Citizen Science". Unpublished PhD Dissertation. Clemson University.

Cox, G.W. (2010). *Bird Migration and Global Change*. Island Press, Washington, DC.

Cressie, N. (1993), "Statistics for Spatial Data". Wiley, New York.

Cressie, N., and Wikle, C. K. (2011), *Statistics for Spatio-Temporal Data*. New York, NY: Wiley.

Feare, C.J. (2007), "The Role of Wild Birds in the Spread of HPAI H5N1", *Avian Diseases*, 51(s1), pp.440-447

Fink, D. , Hochachka, W. M., Winkler, D., Shaby, B., Hooker, G., Zuckerberg, B., Munson, M. A., Sheldon, D., Riedewald, M., and Kelling, S. (2010), "Spatiotemporal Exploratory Models for Largescale Survey Data", *Ecological Applications*, 20, pp. 2131–2147.

Geweke, J., and Terui, N. (1993), "Bayesian Threshold Autoregressive Models for Nonlinear Time Series", *Journal of Time Series Analysis*, 14, pp. 441–454.

Gordo, O., (2007), "Why Are Bird Migration Dates Shifting? A Review of Weather and Climate Effects on Avian Migratory Phenology". *Climate Research*, 35, pp. 37–58

Hüppop,O., and Winkel, W. (2006), "Climate Change and Timing of Spring Migration in the Long-Distance Migrant *Ficedula Hypoleuca* in Central Europe: The Role of Spatially Different Temperature Changes along Migration Routes", *Journal of Ornithology*, 147(2), pp. 344–353.

Hurlbert, A. H., and Liang Z. (2012), "Spatiotemporal Variation in Avian Migration Phenology: Citizen Science Reveals Effects of Climate Change", *PLoS ONE*, 7, p. e31662.

LaDeau, S.L., Marra, P.P, Kilpatrick, A.M., and Calder, C.A. (2008), "West Nile Virus Revisited: Consequences for North American Ecology". *Bioscience*, 58, pp. 937–946.

LaDeau, S.L., Calder, C.A., Doran, P., Marra, P. (2010), "West Nile Virus Impacts in American Crow Populations are Associated with Human Land Use and Climate". *Ecological Research*, 26, pp. 909–916.

Liu, J., Xiao, H., Lei, F., Zhu, Q., Qin, K., Zhang, X.-w., Zhang, X.-I., Zhao, D., Wang, G., Feng, Y., Ma, J.,

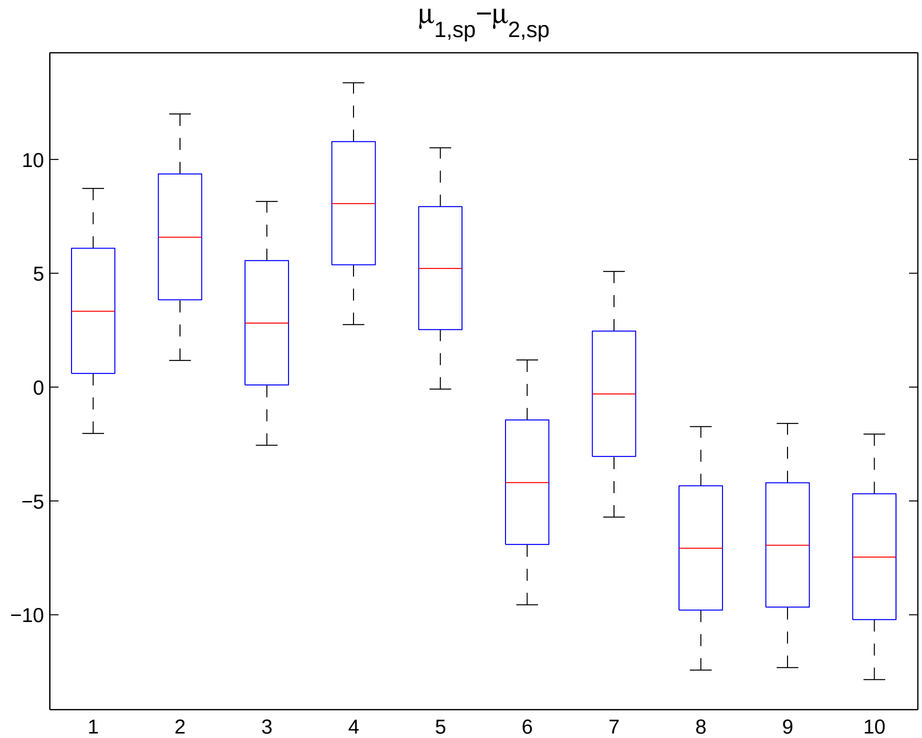

Figure 5: Difference between the spatially-varying means for 1905-1940 and 2001-2010 data.

Liu, W., Wang, J., Gao, G.F. (2005). "Highly Pathogenic H5N1 Influenza Virus Infection in Migratory Birds". *Science*, 309, p. 1206.

Møller, A., Fiedler, W., and Berthold, P. (2004), *Birds and Climate Change*. Advances in Ecological Research Number 35. Elsevier Academic Press, London, UK.

Munster, V.J., Baas, C., Lexmond, P., Waldenström, J., Wallensten, A., et al. (2007), "Spatial, Temporal, and Species Variation in Prevalence of Influenza A Viruses in Wild Migratory Birds". *PLoS Pathogens*, 3(5), pp. 630-638.

Onozuka, D., Hagihara, A. (2008). "Spatial and Temporal Dynamics of Influenza Outbreaks". *Epidemiology*, 19, pp. 824-828.

Sherry, T.W. (2011),"Bird Migration and Global Change; Effects of Climate Change on Birds", *The Auk*, 128(4), pp. 803–805.

Si Y., Skidmore, A.K., Wang, T., de Boer, W.F., Debba, P., Toxopeus, A.G., Li, L., and Prins H.H. (2009), "Spatio-Temporal Dynamics of Global H5N1 Outbreaks Match Bird Migration Patterns", *Geospatial Health*, 4(1), 2009, pp. 65–78.

Tong, H. (1983), *Threshold Models in Non-Linear Time Series Analysis*. New York: Springer Verlag.

Tøttrup, A. P., Thorup, K., and Rahbek, C. (2006). "Patterns of Change in Timing of Spring Migration in North European Songbird Populations". *Journal of Avian Biology*, 37: 84-92.

Walther , G.-R. , Post, E., Convey, P., Menzel, A., Parmesan, C., Beebee, T. J. C., Fromentin, J.-M., Hoegh-Guldberg, O., and Bairlein, F. (2002), "Ecological responses to recent climate change," *Nature*, 416, 389–395.

Zelt, J., Courter, J., Arab, A., and Droege. S. (2012), "Reviving a Legacy Citizen Science Project to Illuminate Shifting Migratory Bird Patterns", *The International Journal of Zoology*, doi:10.1155/2012/710710.