Japanese Journal of Statistics and Data Science (2023) 6:847–860
https://doi.org/10.1007/s42081-023-00212-3

ORIGINAL PAPER

Modern Methods and Applications for Biodiversity Studies

Statistical modeling of Peromyscus maniculatus (deer mouse) amounts per trap with spatiotemporal data

Hsin-Hsiung Huang[1] · Qing He[1]

Received: 30 January 2023 / Revised: 27 May 2023 / Accepted: 1 July 2023 /
Published online: 27 July 2023

Abstract

The North American deer mice (Peromyscus maniculatus) have been used as an environmental change indicator in North America. Since precipitation and temperature changes affect plant productivity and deer mouse habitats, they are substantial factors of deer mouse population radical variations. Therefore, modeling their association is important for monitoring dynamic changes of the deer mouse amounts per trap and relationships among weather variables such as precipitation, maximum and minimum temperatures. We acquired the National Ecological Observatory Network (NEON) data of deer mouse monthly amounts in traps for 2013 through 2022 in the contiguous United States from long-term study sites maintained for monitoring spatial differences and temporal changes in populations. We categorize the contiguous United States into six regions associated with climates. The proposed method identifies important factors of temperature and precipitation seasonal patterns with the month and year temporal effect interacting with the proposed climate-related regions.

Keywords Biodiversity · Climate · Deer mouse · Gaussian process · Habitat · Precipitation · Spatial and temporal correlation · Temperature

1 Introduction

The North American deer mice (Peromyscus maniculatus) have been used as an indicator for environmental changes and transformation of the landscape Bedford and Hoekstra (2015) and a sign of reproduction varying by region McLean and Guralnick (2021). Their populations play an important role in understanding environmental

✉ Hsin-Hsiung Huang
hsin-hsiung.huang@ucf.edu

Qing He
carsonqing@knights.ucf.edu

1 Department of Statistics and Data Science, University of Central Florida, Orlando, FL 32816, USA

factors that influence survival and building a prediction model of risks for humans (Loehman et al., 2012; Parsons et al., 2022). The deer mouse, which is abundant across North America, has been commonly used as a biological indicator to monitor environmental factors which impact small mammals' reproduction and biodiversity dynamically since the deer mouse is a major prey species such as foxes, coyotes, weasels, and hawks and also a carrier for infectious diseases including hantaviruses, plague, and Lyme disease (Botten et al., 2000). Therefore, it is important to understand the population dynamics of this species.

We used the National Ecological Observatory Network (NEON) 2022 capture-recapture data [6] sampled by box traps to evaluate the use of indices for small mammal population and community monitoring using multiple trapping grids per site per year over the contiguous United States (see Fig. 1). The NEON data for the deer mice have been collected over 46 sites from 2013 to 2022 in the United States. Each site contains 3 to 8 with a mean of 6 replicate trapping arrays of 100 traps set in grids with 10-m spacing. The traps were assigned monthly for 6 months during the growing season at a subset of sites and 3–4 months for the rest of the sites. At each site, half of the trap arrays were run for an average of 3 nights and the other half were run for a single night Parsons et al. (2022). The NEON dataset has been used to show that climate and urbanization affect the body size of Peromyscus maniculatus Guralnick et al. (2020). Hence, it is important to develop a statistical model that characterizes the dynamic amounts of the captured small mammals changing over time and locations and also provides statistical inferences to identify important spatial and temporal effects corresponding to habitats and climates. We propose a Bayesian model between the spatiotemporal deer mouse population and weather variables such as precipitation (prcp) in tenths of millimeters, maximum temperature (tmax) and minimum temperature (tmin) which are in tenths of degrees Celsius in the United States obtained from the National Oceanic and Atmospheric Administration (NOAA) (Chamberlain et al., 2022) weather data.

Gaussian process (GP) regression has been applied in various fields for modeling nonlinearly spatiotemporal correlated data (Quinonero-Candela & Rasmussen, 2005; He et al., 2023). Using Bayesian Gaussian process regression, we characterize the nonlinear temporal effects at various locations. Based on the modeling results, we identify factors affecting the deer mouse population changes. We found that their population varied by location and time due to the temperature and precipitation changes.

2 Gaussian process regression models

We divide the conterminous United States into six regions by using the absolute values of the latitudes and longitudes as follows: 1. northeast (NE): latitude $> 37°$N and longitude $\leq 88°$W, 2. southeast (SE): latitude $\leq 37°$N and longitude $> 88°$W, 3. northcentral (NC): latitude $> 37°$N and $88°$W $\leq$ longitude $\leq 102°$W, 4. southcentral (SC): latitude $\leq 37°$N and $88°$W $\leq$ longitude $\leq 102°$W, 5. northwest (NW): latitude $> 37°$N and longitude $> 102°$W, and 6. southwest (SW): latitude $\leq 37°$N and longitude $> 102°$W (Schwartz et al., 2013) (see their locations in Fig. 1). The six regions which are corresponding to the Köppen-Geiger climate classification (Geiger, 1954) basically reflect the weather and geographic characteristics of the deer mouse habitats

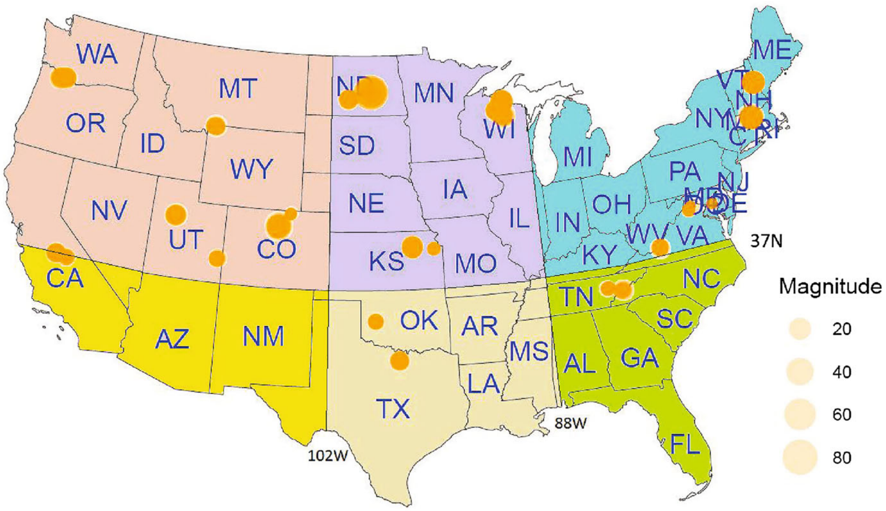

Fig. 1 The map of the traps with the number of the deer mice per trap as the magnitude in the six regions

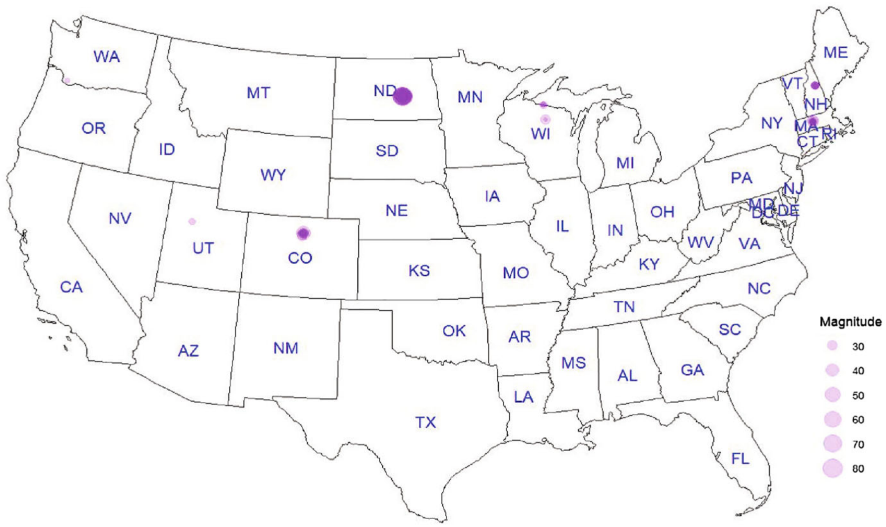

Fig. 2 The map of the traps with the number of the deer mice per trap greater than 20, which are all in the north regions

in the contiguous USA NAVARRO-CASTILLA and Barja (2019). 37° N is called the 37th parallel north, which is approximately the northern boundary of the subtropics (Bannister et al., 2012). The northwest region mostly covers the Rocky Mountains, and the northcentral region relates to the Great Plains. Recent studies have shown that the high-elevation (highland) deer mice have many physiological traits different from the lowland deer mice (Schweizer et al., 2021).

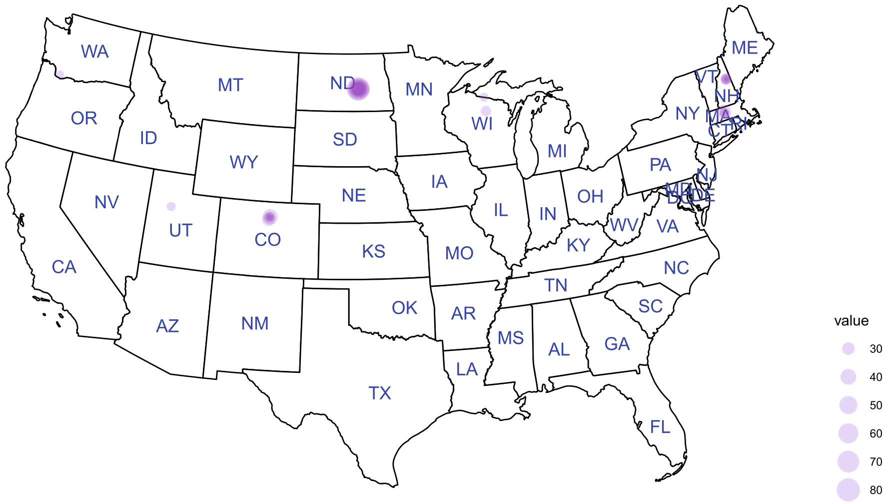

Fig. 3 The map showing the posterior median of the number of the deer mice per trap for those traps of the high-value traps

As the amount of deer mice varies by trap and the distribution is sparse, we also investigate traps that have relatively large amounts of deer mice (named "high-value" traps). These traps contain more than 20 mice, which is the top 1.4% (the right tail of the distribution in the top-left panel in Fig. 4). There are in total 24 high-value traps corresponding to 38 observations from multiple years of record. The high-value amounts per trap are highly correlated with the precipitation, maximum temperature, and minimum temperature (see Fig. 5). Therefore, we propose regression models to the whole data and the high-value data individually. We examined these variables and had the pairwise plots (Figs. 4 and 5) for the whole data and the high-value data (see their locations in Fig. 2) using the R package 'GGally' (Schloerke et al., 2021).

To incorporate with the spatial and temporal correlations, we model y_i the number of deer mice for trap i with covariates x_i including month, year, precipitation (prcp) in tenths of millimeters, maximum temperature (tmax) and minimum temperature (tmin), regions, their interaction terms, spatial random effects and temporal random effects:

$$\begin{aligned}\log(y_i) =& \beta_{\text{region}} x_{i,\text{region}} + \beta_{\text{year}} x_{i,\text{year}} + \beta_{\text{month}} x_{i,\text{month}} + \beta_{\text{prcp}} x_{i,\text{prcp}} \\ &+ \beta_{\text{tmax}} x_{i,\text{tmax}} + \beta_{\text{tmin}} x_{i,\text{tmin}} + \beta_{\text{prcp,region}} \times x_{i,\text{pcrp}} \times x_{i,\text{region}} \\ &+ \beta_{\text{tmax,region}} \times x_{i,\text{tmax}} \times x_{i,\text{region}} + \beta_{\text{year,region}} \times x_{i,\text{year}} \times x_{i,\text{region}} \\ &+ \beta_{\text{month,region}} \times x_{i,\text{month}} \times x_{i,\text{region}} + \epsilon_i, \end{aligned} \tag{1}$$

where $1 \leq i \leq n$, n is the number of traps, and $\epsilon_i \sim N(0, \sigma^2)$.

In the Bayesian regression model (1), year and month are two different sources for the temporal effects where a Gaussian process prior is imposed on the month effects. The Gaussian kernel is based on the temporal distance to ensure that nearby months have similar effects. The month effects repeat every year as a result of the seasonal

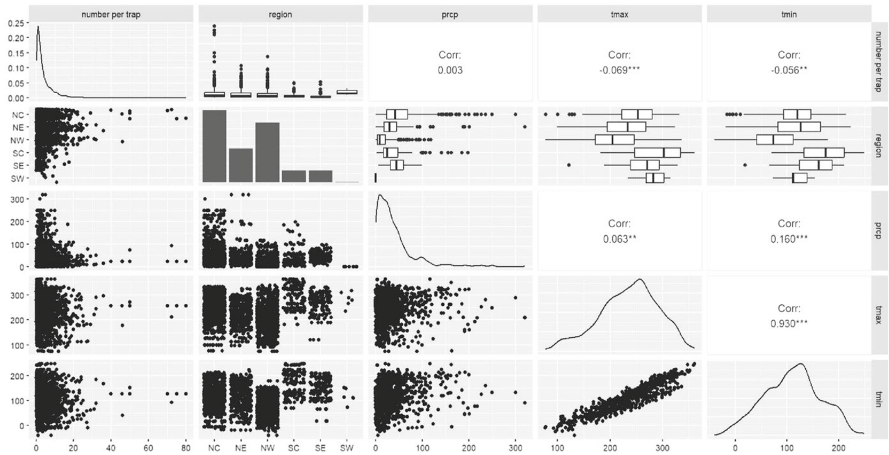

Fig. 4 The pairwise plots for the number of the deer mice per trap, the maximum average temperature (tmax), and the minimum average temperature (tmin) for the whole NEON data

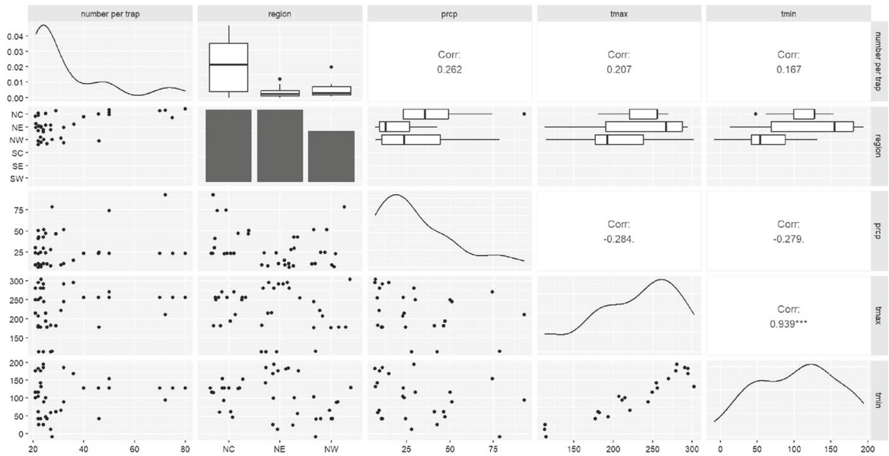

Fig. 5 The pairwise plots for the number of the deer mice per trap (value), the maximum average temperature (tmax), and the minimum average temperature (tmin) for the high-value amounts per trap data

assumption of the population data. The time series plot of the monthly level average counts is in Fig. 7. The plot shows that the abundance of deer mouse is not stable. It is relatively high in years 2014, 2015, 2016 and 2020, and there are two spikes occurring in 2014 and 2020.

2.1 The posterior inference

Denote the complete design matrix for (1) by $\mathbf{X}$, and the design matrix that does not have month information (i.e., month main effects and their interaction effects with region) by D. Denote the $n \times 12$ design matrix that has the month information by L,

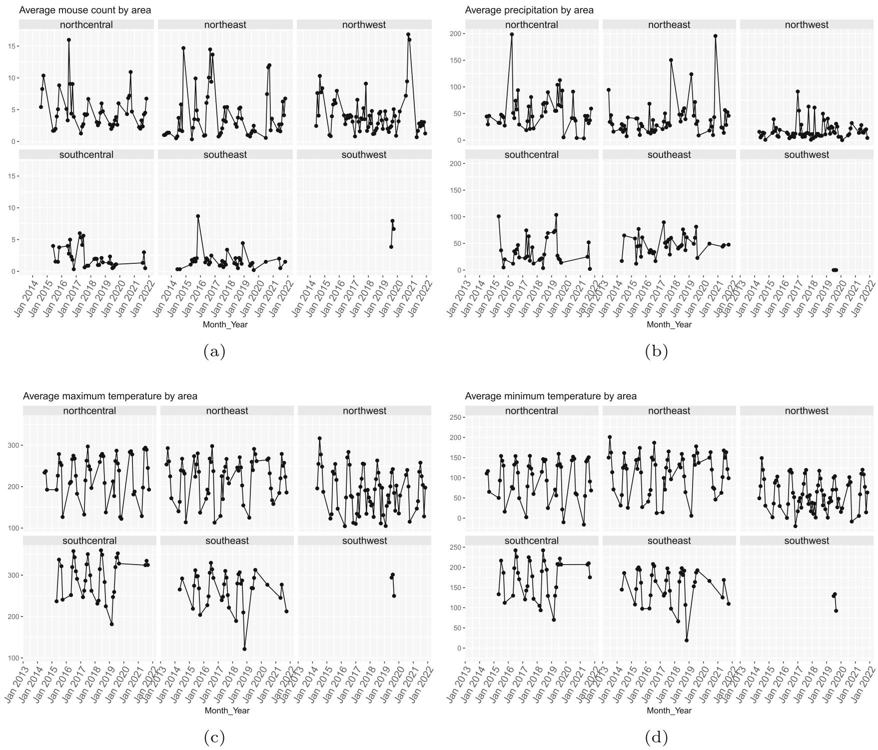

Fig. 6 **a** The sample size per trap of the deer mice of the north central region has high peaks in 2016 and 2022. **b** Average precipitation for the six regions. **c** Average maximum temperature for the six regions. **d** Average minimum temperature for the six regions

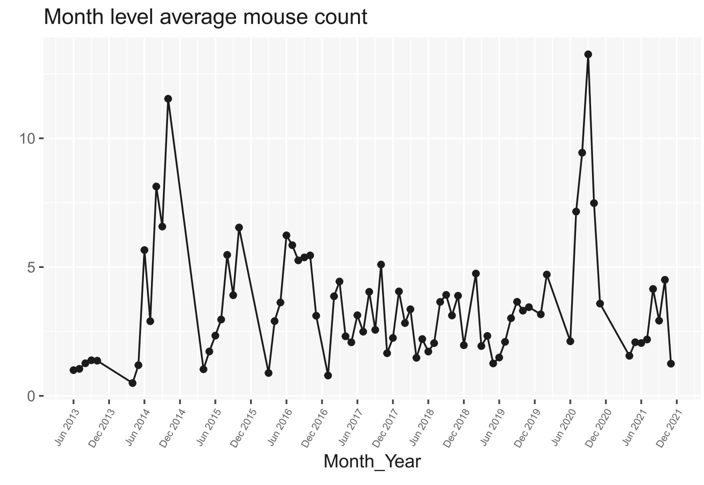

Fig. 7 The time series plot of the monthly level average mouse counts

where n is the number of traps, $\ell_{i,j} = 1$ if i-th observed trap is collected from month j and 0 otherwise, for $i = 1, \ldots, n$, $j = 1, \ldots, 12$. Similarly, denote the design matrix for the interaction between month and region by LR. Then, the design matrix $\mathbf{X}$ can be written as $\left[D\ L\ LR\right]$. For the parameters of interest, they can be written as $\gamma^T = [\vec{\beta}_D, \vec{\beta}_{\text{month}}, \vec{\beta}_{\text{month,region}}]^T$. The regression model can be simplified as follows:

$$y \mid \gamma, \sigma^2 \sim \mathrm{N}(y \mid \mathbf{X}, \gamma, \sigma^2).$$

We assume that the month effects,

$$\vec{\beta}_{\text{month}} := (\beta_{\text{month}=1}, \ldots, \beta_{\text{month}=12})^T \sim \mathrm{N}(\vec{0}_{12\times 1}, K),$$

where K is a positive definite covariance matrix with element $k_{jj'} = \sigma^2 \exp(-\phi(j - j')^2)$, $j, j' = 1, 2, \ldots, 12$, where $(j - j')^2$ is the Euclidean distance between the j-th month and the j'-th month, σ^2 denotes the temporal variability, and $\phi > 0$ determines the speed that the correlation decays with distance. The posterior median and 95% credible interval (C.I.) of the parameters are shown in Table 1. We define a significant variable if the 95% credible intervals (C.I.) of its corresponding regression coefficient does not contain 0 (i.e. the lower bound and upper bound of the C.I. having the same signs). We manually select variables from the full model using all the variables and remove insignificant variables step-by-step (i.e. the backward selection). When the 95% C.I. of the regression coefficient (parameter) contain 0, the corresponding variable is insignificant and we remove it in the model.

We rewrite the covariance matrix as $K = \sigma^2 M$ for convenience, where M is a 12 by 12 positive definite matrix corresponding to the exponential part of the Gaussian kernel. We choose one region as the baseline, and assume a diagonal prior covariance for region contrast effects $\vec{\beta}_{\text{region}}$, denoted by $R = \text{diag}\,(r_l)$ for $l = 1, \ldots, k-1$ where k is the number of regions ($k = 6$ for the whole dataset and $k = 3$ for the high-value subset). Since the regions are far from each other, we assume their region contrast effects are mutually independent and thus

$$\vec{\beta}_{\text{region}} := \left(\beta_{\text{region}=1}, \ldots, \beta_{\text{region}=k-1}\right)^T \sim \mathrm{N}(\vec{0}_{(k-1)\times 1}, R).$$

We further define the element of the 12 by 12 positive definite matrix M_l for the covariance between $\beta_{\text{month}=i,\text{region}=l}$ and $\beta_{\text{month}=j,\text{region}=l}$ as $(M_l)_{ij} = r_l M_{ij}$, where $l = 1, \ldots, k-1$. The matrix form of this joint region (spatial effects) and month (temporal effects) covariance structure is the Kronecker product of R and M (i.e., $R \otimes M$), which is a natural model of spatio-temporal effects. It reduces the number of parameters, and leads to more accurate estimators (Werner et al., 2008). Besides, we assume a normal prior with mean $\mu_{\vec{\beta}_D}$ and covariance $\sigma^2 V_{\vec{\beta}_D}$ on $\vec{\beta}_D$. With that, a conjugate prior for inference of γ and σ^2 can be defined conveniently for the Bayesian model. Let $\mu_0^T = [\mu_{\vec{\beta}_D}, \vec{0}_{12\times1}, \ldots, \vec{0}_{12\times1}]^T$ and

$$V_0 = \begin{bmatrix} V_{\vec{\beta}_D} & 0 & 0 & \ldots & 0 \\ 0 & M & 0 & \ldots & 0 \\ 0 & 0 & M_1 & \ldots & 0 \\ 0 & 0 & 0 & \ldots & 0 \\ 0 & 0 & 0 & \ldots & M_{k-1}. \end{bmatrix}$$

The Bayesian regression model has a conjugate prior for parameters γ and σ^2 as follows:

$$p(\gamma, \sigma^2) = p(\gamma \mid \sigma^2)p(\sigma^2),$$

where

$$\gamma \mid \sigma^2 \sim \mathrm{N}(\gamma \mid \mu_0, \sigma^2 V_0),$$
$$\sigma^2 \sim \mathrm{IG}(\sigma^2 \mid a_0, b_0).$$

The prior on γ incorporates a normal prior for β_D and a GP prior with mean 0 and covariance $\sigma^2 M$ (i.e., K) for the monthly main effects, as well as a normal prior on month and region interaction effects. The joint prior of (γ, σ^2) consists of a normal conditional prior of $\gamma \mid \sigma^2$ and an inverse-gamma prior of σ^2. The conjugate normal-inverse-gamma prior $\gamma, \sigma^2 \sim \mathrm{NIG}(\gamma, \sigma^2 \mid \mu_0, V_0, a_0, b_0)$ has a posterior which is also an normal-inverse-gamma distribution as follow:

$$p(\gamma, \sigma^2 \mid \mathbf{X}, y) \propto p(\gamma \mid \mathbf{X}, y, \sigma^2) \times p(\sigma^2 \mid \mathbf{X}, y),$$

where

$$\begin{aligned} \gamma \mid \mathbf{X}, y, \sigma^2 &\sim \mathrm{N}(\gamma \mid \mu_n, \sigma^2 V_n), \\ \sigma^2 \mid \mathbf{X}, y &\sim \mathrm{IG}(\sigma^2 \mid a_n, b_n), \\ V_n &= (V_0^{-1} + \mathbf{X}^T\mathbf{X})^{-1}, \\ \mu_n &= V_n(\mu_0^T V_0^{-1} + \mathbf{X}^T y), \\ a_n &= a_0 + \frac{n}{2}, \\ b_n &= b_0 + \frac{1}{2}(y^T y + \mu_0^T V_0^{-1}\mu_0 - \mu_n^T V_n^{-1}\mu_n). \end{aligned}$$

A chain with 10,000 iterations including 2000 burn-ins is used for the posterior estimation. In the Bayesian posterior sampling, γ, ϕ, and σ^2 are updated using the posterior distribution described above, while ϕ is updated through the Metropolis-Hasting algorithm (Hastings, 1970) conditioned on σ^2. For the inverse-gamma prior for σ^2, we set it as $\mathrm{IG}(a_0 = 2, b_0 = 1)$. In the precision matrix for the region variable, we set $r_i = 1$ for $i = 1, \ldots, k-1$. We use $\mathrm{Gamma}(a_\phi = 5, b_\phi = 1)$ as the target distribution

Table 1 The significant variables of the proposed Bayesian GP regression model using the northeast (NE) region as the baseline to fits for the deer mouse sample size per trap in the southeast region

Parameter	Posterior median (95% C.I.)
prcp	−0.002 (−0.004, 0)
tmax	−0.003 (−0.005, −0.001)
southeast	−1.047 (−1.956, −0.143)
northcentral::prcp	0.002 (0, 0.005)
northwest::prcp	0.007 (0.004, 0.011)
southeast::prcp	0.013 (0.005, 0.022)

and a normal distribution $\mathrm{N}(\phi_t, 1)$ to propose a candidate ϕ^* given a current value at ϕ_t at the t-th iteration. As a result, the proposal distribution is a multivariate normal $\mathrm{MVN}(0_{M\times 1}, K_{\sigma,\phi^*})$, where K_{σ,ϕ^*} is an $M \times M$ covariance matrix whose (i, j)th element is $\sigma^2 \exp\left(-\phi^* \times (D_t)_{i,j}\right)$, and D_t is an $M \times M$ matrix whose (i, j)th element $(D_t)_{i,j}$ measures the difference between months i and j, and M is the number of months used in the model. $M = 12$ for the whole data, and $M = 6$ (May to October) for the high-value subset. The acceptance probability A for ϕ^* is as follows:

$$A = \min\left(1, \frac{\pi(\phi^*)}{\pi(\phi_t)}\right) = \min\left(1, \frac{p(\phi^*)q(\phi_t \mid \phi^*)}{p(\phi_t)q(\phi^* \mid \phi_t)}\right),$$

where p is the target distribution and q is the proposal distribution.

3 Results and discussion

The NEON small mammal data set provides the opportunity to explore variation in captured amounts of small mammals over large number of locations and periods. Despite the challenge of having missing data of temperature and precipitation, using a Bayesian hierarchical model with Gaussian processes facilitates estimation of the deer mouse population changes with associated spatial and temporal factors. We aim to estimate substantial factors affecting the amounts of deer mice per trap with replicate trapping arrays. The results of our study support previous findings that Gaussian processes can be used for quantifying the correlated spatial and temporal variations. The model fits are shown in Tables 1 and 2. We notice that the high-value amounts of deer mice are from the Northern United States, which have relatively smaller precipitation and lower temperature than the Southern United States. This may explain why the signs of of the posterior medians for prcp and tmax are both negative in the whole data. The proposed Bayesian model could be extended to allow additional pooling across species, sites, months, and years to increase the power to generate estimates for rare species.

We found that the mean of the number of deer mice captured per trapper in the northcentral region (North Dakota) significantly increased in June 1 and September 1, 2016 and the precipitation amounts on these days are much higher than other days. The amounts are 60.1 on June 1, 2016 and 72.2 on September 1, 2016 and the temperatures (see Fig. 9) are between 9 and 27 degrees of Celsius which are suitable

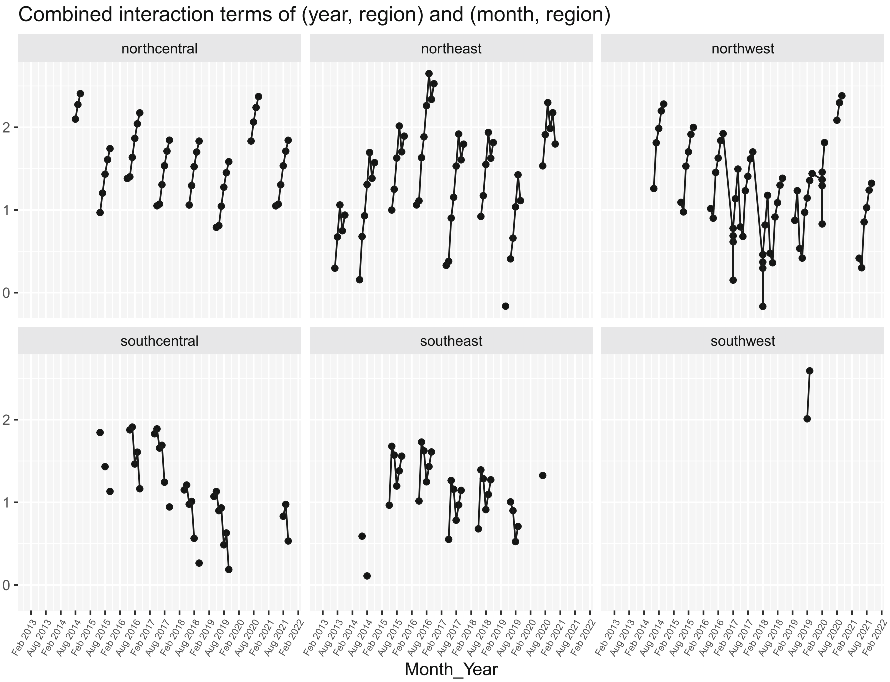

Fig. 8 The combined interaction terms of (month, region) and (year, region) for each region. The temporal effects by using the sum of the interaction terms of year:region and month:region from the GP model. Each region has a monthly pattern and the yearly pattern reveals each region's temporal effects change relatively high and low

Table 2 The significant variables of the proposed Bayesian GP regression model for the high-value data using the northeast (NE) region as the baseline

Parameter	Posterior median (95% C.I.)
prcp	0.008 (0, 0.017)
tmax	0.021 (0.015, 0.027)
tmin	−0.016 (−0.025, −0.007)

All the weather variables are significant. The opposite signs of the maximum temperature and minimum temperature may be caused by their collinearity among them and between precipitation

for deer mouse breeding (Joyner et al., 1998). This is consistent with the previous study Gorosito and Douglass (2017) that rainfall is a crucial factor affecting deer mouse population dynamics. High precipitation which fosters plant growth in a warm weather increases food availability and hence bolsters the deer mouse populations (Stinson & Fisher, 1953; Hansson, 1979; Chappell et al., 2004; Gorosito & Douglass, 2017). The combined interaction terms of (month, region) and (year, region) show the pattern for each region for both the whole and high-value datasets (see Figs. 8 and 10).

The posterior median for each sampling point is shown in Fig. 3, which shows similar patterns with the actual values in Fig. 2. To measure the goodness of fit of the proposed model, we apply 5-fold cross-validation which randomly divides the high-

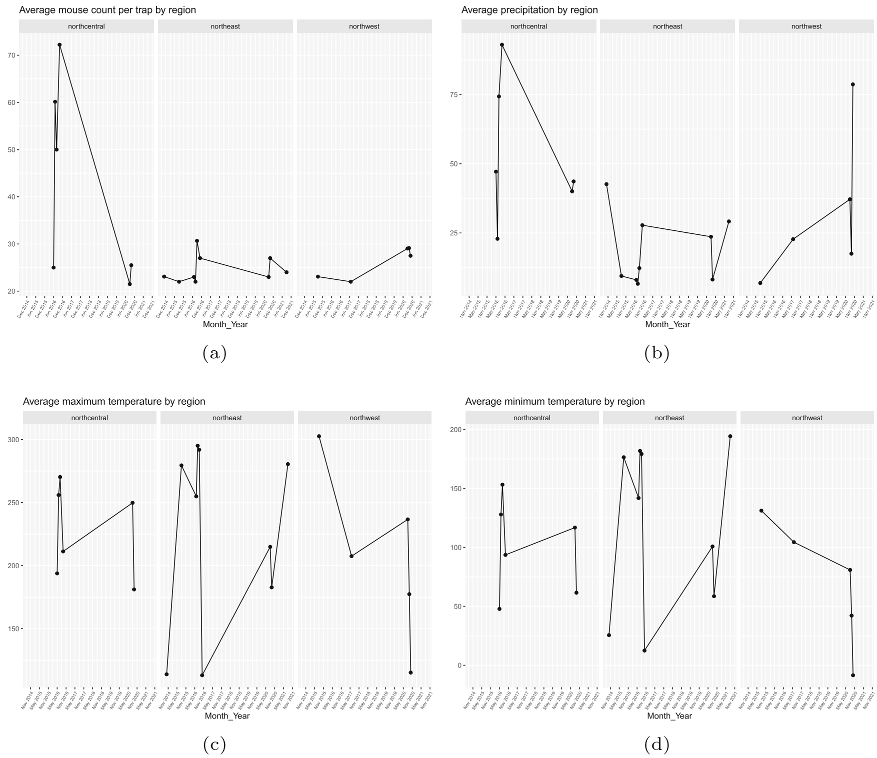

Fig. 9 The average deer mouse counts per trap **a** has a related pattern to that of the precipitation by region (**b**). The maximum temperature by region **c** and minimum temperature by region **d** have similar trends

value data into training (80%) and testing (20%) using stratified sampling based on year and month. Parameter estimates are obtained from the MCMC posterior means and used to predict the mouse counts for the testing dataset. The test MAPE (mean absolute percentage error) has the mean 0.0452 and the standard deviation 0.0328. We also apply linear mixed regression using the month variable as the random effect and features–prcp, tmax, tmin, NC, and NW. The resulting average MAPE is 0.0953 with standard deviation 0.0764. The results indicate that the proposed model fits the data better than the linear mixed model.

4 Conclusion

Statistical modeling and inference for the deer mouse amounts per trap of the NEON data which may across locations and time with periodic trends is challenging. The proposed Bayesian regression models of the deer mouse amounts per trap was used with the proposed regions, weather features, and temporal effects using a Gaussian process to ensure accurate estimation and meaningful interpretation. Our model fitting results show that the deer mouse amounts per trap generally follow the seasonal pattern

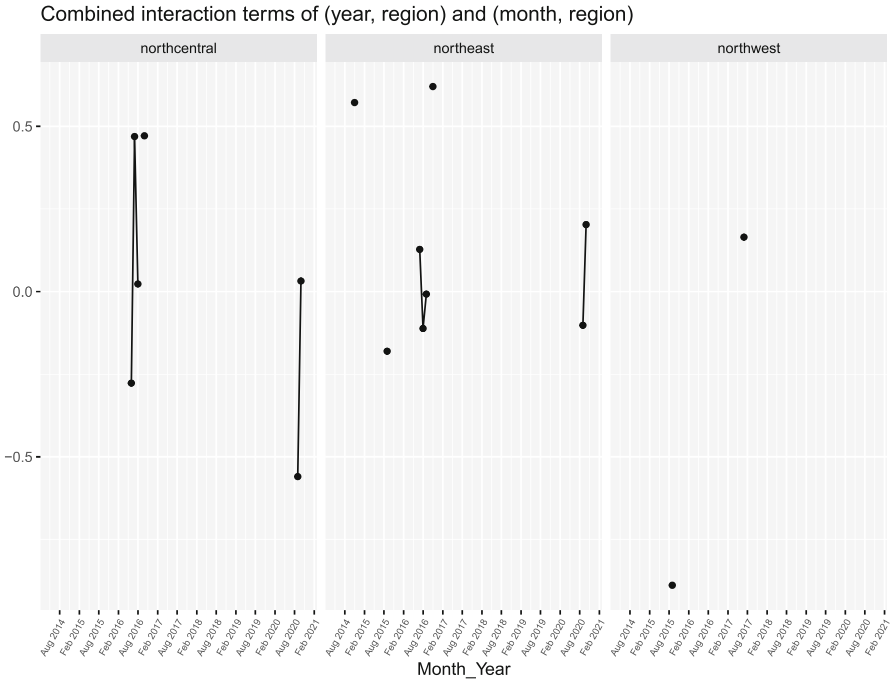

Fig. 10 The combined interaction terms of (month, region) and (year, region) for each region using the high-value data. The temporal effects by using the sum of the interaction terms of year:region and month:region from the GP model. Each region has a monthly pattern and the yearly pattern reveals each region's temporal effects change relatively high and low

and the proposed regions approximately explain the geographic attributes of deer mice. When the rainfall is extraordinarily high and the temperature is not too low, it causes a sudden increase in food availability. Consequently, the deer mouse populations expand, and the amounts per trap change accordingly. Our finding results are consistent with the literature (Hansson, 1979; Loehman et al., 2012; Gorosito & Douglass, 2017; Guralnick et al., 2020). The proposed method broadens the statistical analysis in the biodiversity data and it can be applied to modeling data with spatial and temporal correlated features and identifying important factors.

Acknowledgements This research was supported in part by the National Science Foundation grant (DMS-1924792).

Declarations

Conflict of interest The authors have no competing interests to declare that are relevant to the content of this article.

References

Bannister, J. R., Vidal, O. J., Teneb, E., & Sandoval, V. (2012). Latitudinal patterns and regionalization of plant diversity along a 4270-km gradient in continental chile. *Austral Ecology, 37*(4), 500–509.

Bedford, N. L., & Hoekstra, H. E. (2015). The natural history of model organisms: Peromyscus mice as a model for studying natural variation. elife 4, 06813.

Botten, J., Mirowsky, K., Kusewitt, D., Bharadwaj, M., Yee, J., Ricci, R., Feddersen, R. M., & Hjelle, B. (2000). Experimental infection model for sin nombre hantavirus in the deer mouse (peromyscus maniculatus). *Proceedings of the National Academy of Sciences, 97*(19), 10578–10583.

Chamberlain, S., Hocking, D., & Anderson, B. (2022). Package 'rnoaa'.

Chappell, M. A., Garland, T., Jr., Rezende, E. L., & Gomes, F. R. (2004). Voluntary running in deer mice: speed, distance, energy costs and temperature effects. *Journal of Experimental Biology, 207*(22), 3839–3854.

Geiger, R. (1954). Klassifikation der klimate nach w. köppen (classification of climates after w. köppen). Landolt-Börnstein-Zahlenwerte und Funktionen aus Physik, Chemie, Astronomie, Geophysik und Technik, alte Serie, 603–607.

Gorosito, I. L., & Douglass, R. J. (2017). A damped precipitation-driven, bottom-up model for deer mouse population abundance in the northwestern united states. *Ecology and Evolution, 7*(24), 11113–11123.

Guralnick, R., Hantak, M. M., Li, D., & McLean, B. S. (2020). Body size trends in response to climate and urbanization in the widespread north american deer mouse, peromyscus maniculatus. *Scientific Reports, 10*(1), 1–13.

Hansson, L. (1979). Food as a limiting factor for small rodent numbers. *Oecologia, 37*(3), 297–314.

Hastings, W. K. (1970). Monte carlo sampling methods using markov chains and their applications.

He, Q., Harrison, C. W., Huang, H.-H. (2023). Detection of anomalies in traffic flows with large amounts of missing data. The New England Journal of Statistics in Data Science, 1–11. https://doi.org/10.51387/23-NEJSDS20

Joyner, C. P., Myrick, L. C., Crossland, J. P., & Dawson, W. D. (1998). Deer mice as laboratory animals. *ILAR Journal, 39*(4), 322–330.

Loehman, R. A., Elias, J., Douglass, R. J., Kuenzi, A. J., Mills, J. N., & Wagoner, K. (2012). Prediction of peromyscus maniculatus (deer mouse) population dynamics in montana, usa, using satellite-driven vegetation productivity and weather data. *Journal of Wildlife Diseases, 48*(2), 348–360.

McLean, B. S., & Guralnick, R. P. (2021). Digital biodiversity data sets reveal breeding phenology and its drivers in a widespread north american mammal. *Ecology, 102*(3), 03258.

National Ecological Observatory Network (NEON): Small mammal box trapping (DP1.10072.001). National Ecological Observatory Network (NEON) (2022). https://doi.org/10.48443/H3DK-3A71. https://data.neonscience.org/data-products/DP1.10072.001/RELEASE-2022

NAVARRO-CASTILLA, Á., & Barja, I. (2019). Stressful living in lower-quality habitats? body mass, feeding behavior and physiological stress levels in wild wood mouse populations. Integrative Zoology *14*(1), 114–126.

Parsons, A. W., Clark, J. S., & Kays, R. (2022). Monitoring small mammal abundance using neon data: are calibrated indices useful? Journal of Mammalogy.

Quinonero-Candela, J., & Rasmussen, C. E. (2005). A unifying view of sparse approximate gaussian process regression. *The Journal of Machine Learning Research, 6*, 1939–1959.

Schloerke, B., Cook, D., Larmarange, J., Briatte, F., Marbach, M., Thoen, E., Elberg, A., & Crowley, J. (2021). GGally: Extension to 'ggplot2'. R package version 2.1.2. https://CRAN.R-project.org/package=GGally.

Schwartz, M. D., Ault, T. R., & Betancourt, J. L. (2013). Spring onset variations and trends in the continental united states: past and regional assessment using temperature-based indices. international Journal of Climatology *33*(13), 2917–2922.

Schweizer, R. M., Jones, M. R., Bradburd, G. S., Storz, J. F., Senner, N. R., Wolf, C., & Cheviron, Z. A. (2021). Broad concordance in the spatial distribution of adaptive and neutral genetic variation across an elevational gradient in deer mice. *Molecular Biology and Evolution, 38*(10), 4286–4300.

Stinson, R., & Fisher, K. C. (1953). Temperature selection in deer mice. *Canadian Journal of Zoology, 31*(4), 404–416.

Werner, K., Jansson, M., & Stoica, P. (2008). On estimation of covariance matrices with kronecker product structure. *IEEE Transactions on Signal Processing, 56*(2), 478–491.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.