Gaussian process forecasting of sparse ecological time series

Parul V. Patil, *[1], Robert B. Gramacy[1], and Leah R. Johnson[1]

[1]Department of Statistics, Virginia Tech, Blacksburg, Virginia 24061, USA

Abstract

Ecological time series are often unevenly sampled in time. That is, because the sampling processes used are resource intensive, data may be collected infrequently, or with adaptive frequencies triggered by presence of a target variable. When the data are irregularly spaced, standard time series methods may not be directly applicable. Instead, approaches that take inspiration from linear regression (LR) may be appropriate. In this paper, we explore flexible, nonparametric Gaussian process (GP) models as tools for producing forecasts of unevenly sampled observations. Our example is data on abundances of nymphal *Amblyomma americanum* from nine locations spread across the eastern United States, collected by NEON. The data exhibit highly variable sampling regimes and abundance levels across locations and time. We implement two versions of GPs to forecast tick abundance and benchmark models against LR approaches. Both GPs are able to capture population patterns without the need for forecasting additional drivers, such as temperature, or specifying a specific relationship between the response and predictors. We find that GP models provide an effective method to forecast irregularly sampled populations at short to intermediate time scales, outperforming LR and other comparators.

Keywords: *Amblyomma americanum*; Ecological forecasting; heteroskedastic Gaussian process; Lone-star ticks; NEON forecasting challenge; Tick populations.

Open Research statement: This paper uses data that are already published and publicly available together with novel code. Both are provided together, to allow for reproduction of results, at a public git repository linked here: `https://bitbucket.org/parulpatil22/tickgp-code/`.

*Corresponding author: `parulvijay@vt.edu`

1 Introduction

Accurate forecasts of ecological time series are important for a variety of applications. For example, forecasts of algal blooms can ensure proper resource management as well as drinking water quality (Carey et al., 2022) and pest forecasting can allow farmers to spray crops with pesticides in a timely manner improving efficacy while reducing the risk of adverse health effects (Liu et al., 2022; Ponomarev et al., 2019). Forecasting of species dynamics over medium to long-term horizons can also be used to inform policy, for example, forecasting the effects of climate change on *Aedes*-borne viruses to inform public health measures (Ryan et al., 2019) or predicting the growth of chytrid fungus in frogs to help prevent rapid decline of amphibians (Gajewski et al., 2021). Although forecasting populations of species is useful in many domains, it can be challenging as monitoring and sampling populations can be difficult. For instance, observing endangered frogs affected by chytrid in the Sierra Nevada mountains of California requires researchers to haul sampling equipment on foot for hours to collect a single data point. As a result, populations may be irregularly and infrequently sampled which presents challenges for forecasting efforts.

One example of irregularly sampled populations for which short to medium term forecasts would be useful for making decisions are ticks that spread infectious diseases. In the United States and around the world, ticks adversely affect humans and animals, primarily by spreading tick-borne pathogens (Mitchell et al., 2020; Levin et al., 2017; Paddock et al., 2016; Piesman and Eisen, 2008). In the United States, diseases caused by tick-born pathogens include Lyme disease and Babesiosis, both transmitted by the blacklegged tick, *Ixodes scapularis*; and human Ehrlichiosis (Gaines et al., 2014; Raghavan et al., 2014; Childs and Paddock, 2003), Heartland virus (Savage et al., 2013), and Tularemia (Goddard and Varela-Stokes, 2009) (among others) transmitted by the *Amblyomma americanum* (commonly referred to as the lone-star tick). Apart from spreading pathogens, bites from ticks can also cause long term issues in humans such as alpha-gal syndrome which can result in a life long meat allergy (Steinke et al., 2015; van Nunen, 2015; Commins et al., 2011). Predicting tick density, as a proxy for bite risk, can help assess likelihood of pathogen transmission (Léger et al., 2013; Pfäffle et al., 2013). This can be beneficial for resource management, public safety, and preventative health measures (Estrada-Peña, 2001a; Randolph and Rogers, 1997)

To forecast tick density we rely on samples of observed tick activity. Ticks can be sampled in a variety of ways such as tick drags, CO_2 traps, and baited traps. The trap methods are easier as they require less time and effort, but are not as effective as tick drags (Falco and Fish, 1992). The drag method involves an experimenter dragging a cloth across a study area to collect ticks. These are then identified to species or genus level and counts are reported. This method is time and effort intensive, and so adaptive sampling schedules are often used to minimize the chance that a drag is conducted when ticks are not present and active. This can lead to uneven sampling intervals and result in a small dataset with irregular gaps between sampling times. For example, our motivating dataset of tick observations from the National Ecological Observatory Network (NEON) has 385 observations across nine different locations over the past decade. Ideally, the number of observations (assuming weekly resolution) would be 53 weeks $\times$ 10 years $\times$ 9 locations $\approx$ 4,770 observations – an order of magnitude larger than the number of observations we have. This sparsity can make modeling challenging.

Classical time series (TS) models are a popular choice for forecasting (Brugger et al., 2018; Meltzer and Norval, 1992). TS models decompose observations into trend, seasonal, autoregressive (AR), and noise components. The AR term captures dependence between observations separated

by a fixed lag which implicitly assumes regular sampling. In situations with sparse data, successive observations might be separated by variable time intervals. TS models for irregularly spaced data are an option for fitting irregular time seris (Robinson, 1977), however the estimation of an AR term requires extensive data pre-processing including imputation and aggregation over large time intervals. This can dampen the signal and lead to biased estimates of AR coefficients (Little and Rubin, 2019). Identifying seasonality parameters can also be challenging as the phase is not always identifiable due to large gaps. In such situations, classical time series models that rely on evenly spaced data may result in misleading forecasts for unevenly sampled data (Johnson and Munch, 2022).

State space models (SSMs) are also commonly used for modeling irregularly sampled time series (Newman et al., 2023; Auger-Méthé et al., 2021; Jones, 1986). These models are hierarchical and explicitly separate the unobserved latent process (which models the underlying dynamics) from the observed process that is subject to measurement error (Cressie et al., 2009). SSMs depend on a resolution of observations relative to the resolution at which we want to infer the true process (Auger-Méthé et al., 2021). When there are large gaps in the data, the observations may not provide enough information to infer latent states. For example, changes in observations could be due to several factors such as fast-changing dynamics, high process noise, or measurement errors which the model cannot distinguish amongst without prior information (Auger-Méthé et al., 2016). Since parameter estimation is challenging, inference relies on predicting latent states forward in time rather than updating it through the data. Errors in estimation propagate through the model and can produce unreliable estimates that depend strongly on the prior.

We instead propose a Gaussian process (GP) model, which is a non-parametric regression model that leverages relative distances in the input space to inform predictions (Gramacy, 2020; Cressie, 1991). This approach incorporates a noise parameter to separate the mean process (here tick abundance) from the noise (i.e., sampling error, labeling error, etc). The standard formulation of GPs is homoskedastic, that is, the noise process is assumed to be the same in all parts of the input space. However, for our application, we expect that one estimate of noise across all locations would be insufficient and we instead prefer a heteroskedastic GP (hetGP) model. This variant allows both the mean and noise process to vary across the input space. Our contribution is not to simply say use a GP/hetGP and make predictions for tick applications. Instead we propose a framework that checks the following boxes: (1) It relaxes the assumption that the time series must be evenly sampled and is applicable to very sparse datasets such as ours. (2) It does not use environmental/weather predictors (e.g., temperature, humidity) that themselves need to be forecast, nor assume specific functional forms between the response and predictors. (3) It trains one hierarchical model across all locations to allow information sharing rather than training and forecasting individually at each location, as some locations are very sparse compared to others.

The main factor which determines performance of our GP models are the predictors that we choose. For a GP to make meaningful forecasts, we need well defined "closeness" in our input space. Models which determine presence/absence of ticks can provide informative features/predictors for our use. For example, classification models such as generalized linear models (GLM), maximum entropy models (MaxEnt; Raghavan et al., 2016), and classification and regression trees (CART – which are usually used in the field of machine learning (Zannou et al., 2021)) often incorporate weather and climate data such as temperature, moisture, and precipitation (Williams et al., 2015; Springer et al., 2015; Estrada-Peña, 2001b; Haile and Mount, 1987). Additionally, they can incorporate location-based features such as elevation and forest cover (Slatculescu et al., 2020; Hahn et al.,

2016) and host density (Wimberly et al., 2008) which may influence tick abundance (Paddock and Yabsley, 2007). We determine some of our predictors based on features used in these other models. Note that position of features in the input space itself does not directly affect GP performance. Rather, the distance between them is what maters. An absolute value of temperature (or forecasts) are not required, only general patterns. We detail our predictors later in Sec. 3.2.

To benchmark our method, we train additional models as comparators to our GP models. In order to avoid the need to aggregate or interpolate, we do not build TS models and instead focus on two types of comparators. First are a set of linear regression (LR) models with simple predictors. Second are Bayesian adaptive spline surfaces (BASS) which is a common flexible and non-parametric regression framework designed for nonlinear relationships (Francom and Sansó, 2020; DiMatteo et al., 2001). BASS models directly attempt to infer the underlying signal given the data without assuming a time-series structure and adjust the smoothing based on the data (Serra and Krivobokova, 2017; Wood et al., 2002). They can thus model sparse data without requiring imputation. We do not describe BASS in detail (see cites for more) and simply use it as a comparator to our GP methods. To level the playing field, GPs and BASS use the same predictors.

Below, we begin by introducing the problem setup with details of the data source followed by visualization and pre-processing steps (Section 2). Section 3 covers various modeling approaches starting with a linear regression (LR) comparator, which representing a classic approach to this problem (Section 3.1). We then introduce the GP model along with implementation details (Section 3.2) and an extension, namely HetGPs which is our proposed approach for this specific forecasting problem (Section 3.3). Section 4 consists of comparisons of all the modeling approaches across multiple metrics. Lastly, we conclude with a brief discussion of the methods and the future scope (Section 5).

2 Data Description and Preparation

2.1 Data Overview

2.1.1 *Amblyomma americanum*

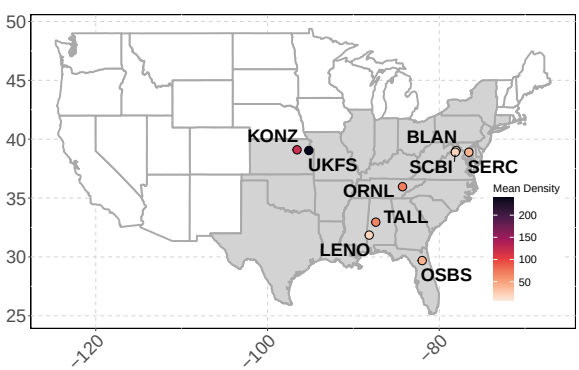
In this study, we focus on modeling the seasonal density of *Amblyomma americanum*. *Amblyomma* ticks follow a three-host life cycle (larval, nymphal, and adult stages – each on a different animal host) that spans approximately 8 months to 2 years, depending on environmental conditions (Troughton and Levin, 2007). At each stage, the tick requires a blood meal from a host, typically mammals, birds, rodents, or humans, to develop further (Kollars Jr. et al., 2000). Feeding usually lasts one to three weeks per stage (Troughton and Levin, 2007). *A. americanum* ticks specifically inhabit wooded areas and primarily feed on white-tailed deer (Allan et al., 2010; Paddock and Yabsley, 2007), though wild turkeys are common hosts for immature stages, especially in the Midwest (Mock et al., 2001). They also aggressively quest for hosts based on carbon-dioxide and movement patterns (Childs and Paddock, 2003).

A. americanum ticks are native to the eastern and southeastern United States and parts of Mexico (Molaei et al., 2019; Monzón et al., 2016; Springer et al., 2014). However, their range has been expanding due to climate change (Rochlin et al., 2022; Raghavan et al., 2019; Cortinas and Spomer, 2013; Masters et al., 2008). Notably, there have been reports of these ticks in northern regions of the U.S. that were historically too cold for their survival (Paddock and Yabsley, 2007).

This expansion is concerning, as *A. americanum* are responsible for a significant proportion of tick-borne diseases in North America (Goddard and Varela-Stokes, 2009; Mixson et al., 2006; Goddard, 2002; Felz et al., 1996).

2.1.2 NEON Forecasting Challenge

Our forecasting effort was based on a subset of tick data obtained from National Ecological and Observatory Network (NEON) aggregated for a tick forecasting challenge set up by Ecological Forecasting Initiative Research Coordination Network (EFI-RCN; Thomas et al., 2023). The target data are counts of nymphal *A. americanum* ticks and also includes the date of collection, the location, the target species, and the iso-week of collection. As part of the NEON challenge, we also have a subsidiary dataset (site dataset; Site Metadata, 2025), with information about the nine locations such as the coordinates, altitude (mean, minimum and maximum), and ecological features such as precipitation, temperature, foliage, etc. In Fig. 1, we show the nine NEON locations along with mean tick density observed at each location in the past decade and total number of observations at each location.



Site	n	n_{test}	Miss_{wk}	$\%\text{Miss}_{wk}$	$\%\text{Miss}_{6wk}$
BLAN	86	18	466	84.4	6.52
KONZ	70	13	482	87.3	23.9
LENO	27	2	525	95.1	70.7
ORNL	50	7	502	90.9	45.7
OSBS	28	7	524	94.9	69.6
SCBI	38	6	514	93.1	58.7
SERC	30	7	522	94.6	67.4
TALL	31	5	521	94.4	66.3
UKFS	25	7	527	95.5	72.8
Total	**385**	**72**	**4583**	**92.3**	**53.5**

Figure 1: *(left)* NEON locations with average tick density at different locations. *(right)* Details such as number of samples collected, used for testing, along with number and % of missing observations.

Ticks are sampled at NEON locations using the "tick drag" method. Specifically a $1m^2$ cloth is manually dragged across the plots in transects of 160m (National Ecological Observatory Network (NEON), 2024). This occurs once every few weeks at a given location during the sampling season which is determined by the foliage at the location. The collected ticks are then taken to the lab for taxonomic identification where the life stage and the taxonomic rank of the species is identified. In Fig. 1 *(right)*, we have provided observed and missing samples for each location assuming a weekly sampling scheme. Notice that almost 90% of data is missing. Assuming a highly adaptive sampling scheme every six weeks, we would still be missing about 50 % data at all locations. This can present as a challenge especially considering that forecasts are required at weekly resolution.

2.1.3 Data Visualization

In this section, we highlight key exploratory observations across locations that inform our modeling decisions. We illustrate the location-level variation in Fig. 2 by showing tick density recorded at three representative sites: Lenoir Landing (LENO), Smithsonian Conservation Biology Institute

(SCBI), and Talladega National Forest (TALL). Tick samples vary across locations due to difference in sampling schemes. For example, samples for some locations, e.g., TALL/SCBI date back to 2014, while LENO only has data 2016 onwards. The sampling season also changes depending on geographic location. Some regions experience lower temperatures and more snowfall which hinders longer tick dragging schemes in winter. However, warmer regions (SCBI/BLAN) do have occasional drags in winter as ticks can remain active year round.

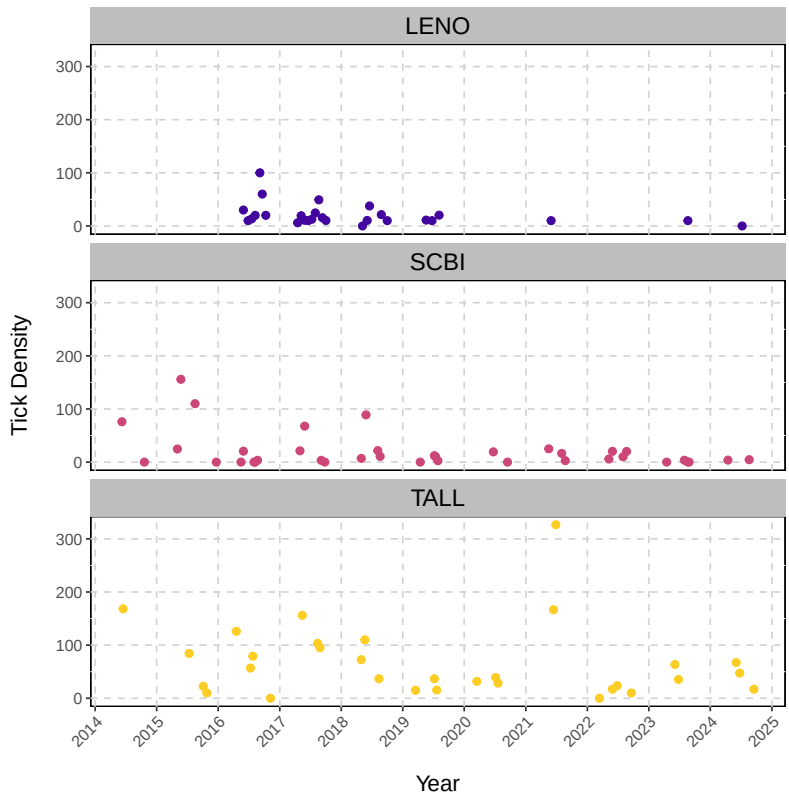


Figure 2: Observed Tick Density across three different NEON locations in the past decade

We also observe a difference in the amplitude of tick density across locations. For instance, peak densities at TALL reach around 300 ticks per plot, while LENO peaks at approximately 100 ticks per plot. Other sites observe over 600 ticks per plot during peak season. We can also identify some seasonal patterns in Fig. 2. Observed density is lower in colder months and higher during warmer months at each location. However, the magnitude of seasonal peaks can vary significantly from year to year, even at the same site.

2.2 Preprocessing Steps

2.2.1 Data Transformation

To forecast tick density, we explore modeling methods that rely on the assumption that the response must be normally distributed around the mean. Since tick density is strictly positive and includes true zeros, we use a transform to ensure this constraint is satisfied. We use a combination of the

square-root and logarithm transform given by:

$$Y' = \begin{cases} \sqrt{Y+1}, & \text{if } (Y+1) > 1. \\ \log(Y+1), & \text{otherwise;} \end{cases} \tag{1}$$

where, Y' is the transformed response and Y is on the original scale, i.e., the observed tick density.

Notice that taking a log of zero is avoided by shifting tick density (add one to original density). Using a square-root transform for larger values prevents over expansion on the back transform (which would be a consequence of using only a log transform). Similarly, using a log transform for smaller values prevents insensible UQ (upper bound < lower bound) on the back transform.

2.2.2 Train-Test Split

To evaluate our methods, we split the data into two sets, training and holdout (testing). For time series data, it is standard to use a cutoff date before which all the data is historic training data. Data after the cutoff is used as the holdout set to evaluate predictions. We selected a cutoff date of 2022-12-31 which provided a sufficient number of points to assess model performance. Because predictions made only at sparse time points do not capture underlying periodic trends well, we constructed a uniform weekly grid from 2014 to the latest date in the dataset. Each of the trained models was used to predict tick density at every week in this grid. Predictions corresponding to all the holdout dates were picked out to evaluate out-of-sample performance.

We benchmarked our method using coverage, root mean square error (RMSE; Appendix S1: Section A), and continuous rank probability score (CRPS; Gneiting and Raftery, 2007). Coverage is the percentage of data points within the 90% predictive intervals (PIs), which ideally should match the nominal level. RMSE measures prediction accuracy ignoring uncertainty, while CRPS accounts for both accuracy and uncertainty and can be computed in closed form for Gaussian distributions. We calculated RMSE and CRPS on the transformed scale for all models. Since we are working with a sparse dataset, RMSE and CRPS are based off very small sample sizes and may not sufficiently capture the model performance. For example, we expect ticks to be inactive in winter and peak in the summer, but some models might predict constant densities with large uncertainty through out the year and still score well on metrics at sparse testing locations. Thus, we also visually assess all the models. In the results below, we preferred to present selected model fits that best illustrate specific characteristics of each modeling approach (see Appendix S1: Section C for all model fits at all locations). Additionally, we induced some dummy reference points to better quantify accuracy. Specifically, we added a zero tick density observation in the middle of the non-sampling season which was typically the last week of January (for all years) at all locations. The addition of the reference points penalizes models which do not predict a zero density in winter and allows us to better quantify model fits.

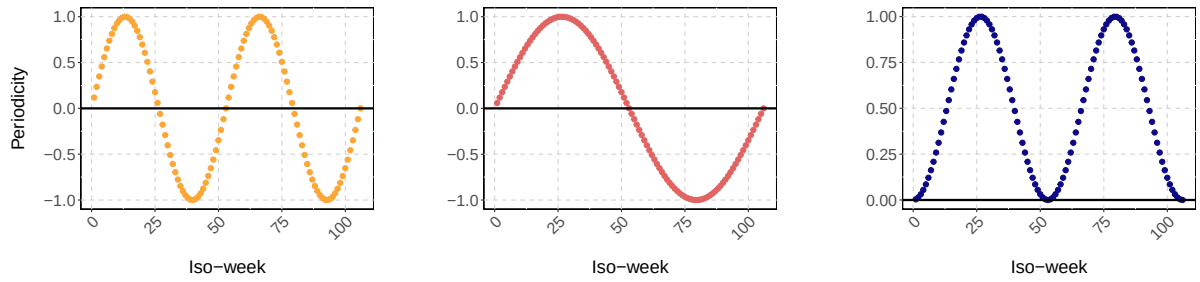
3 Modeling

3.1 Linear regression modeling

We start with using linear regression (LR) since it offers a simple alternative to time series modeling when data are irregularly sampled. The models can be expressed as $Y' = X\beta + \epsilon$ where $\epsilon_i \overset{\text{i.i.d.}}{\sim}$

$\mathcal{N}(0, \sigma^2)$. Here, Y' is the vector of responses (transformed tick density) of size n which corresponds to number of samples. The matrix X has dimensions $n \times (p+1)$, with p predictors appended with one column (of ones) corresponding to the model intercept. The parameter vector $\beta = (\beta_0, \beta_1, ...\beta_p)$ includes the model intercept and p slopes. Random errors ϵ_i are assumed to be independent and identically distributed (i.i.d.), following a normal distribution with mean zero and constant variance σ^2.

We discuss two variants of LR models distinguished by the primary predictor used in each case. The first model is the *LR-Time(L)* model which is our vanilla comparator. We use (L) to represent that the model was trained using a subset of data i.e., data which was collected at that specific **L**ocation rather than data collected at **A**ll locations (denoted by (A) in later sections). This model uses iso-week as a predictor with additional periodicity through sine and cosine predictors – each having a period of 53, representing the number of iso-weeks per year i.e., season in the context of our problem. Observe in Fig. 3a that the sine wave completes two cycles but we suspect that it is not in phase with tick density. Using both sine and cosine predictors allows the model to adjust the phasing. For our second variant, *LR-Temp(L)*, we used minimum temperature as a predictor rather than iso-week for reasons we explain shortly. All linear models are fit using the `stats` package in R (R Core Team, 2022).



(a) sine term, 53-day period. **(b)** sine term, 106-day period. **(c)** sq. sine term, 106-day period.

Figure 3: Periodicity for sine predictor with different phases and formulations explained above. We use (a) for linear models and (c) for Gaussian processes.

The predictors in the *LR-Time(L)* model repeat across years so we visualize the mean predictions as a function of iso-week. At both locations, i.e., UKFS and OSBS in Fig. 4, there is a discontinuity in the predictions across years and overly conservative upper bounds (Appendix S1: Section C). At UKFS (18 training points), the model captures the expected pattern of increase in ticks in summer and decrease in winter. However, at OSBS (21 training points), the model fails to capture this pattern, which we suspect is because the mode of tick-density changes from one season to another, making it challenging to identify the real trend across seasons using only iso-week and periodicity as predictors.

We now move onto our next variant which we call *LR-Temp(L)* model. We consider minimum temperature (*tmin*) as our predictor in place of iso-week for this model. In terms of modeling choices, we had a few options available to us. The first major choice was between weather data provided by NOAA Global Ensemble Forecasting System (Hamill et al., 2022)[1] or historic monthly climatology data obtained from CRU-TS 4.09 (Harris et al., 2020) downscaled with WorldClim

[1]See `https://www.ncei.noaa.gov/products/weather-climate-models/global-ensemble-forecast`.

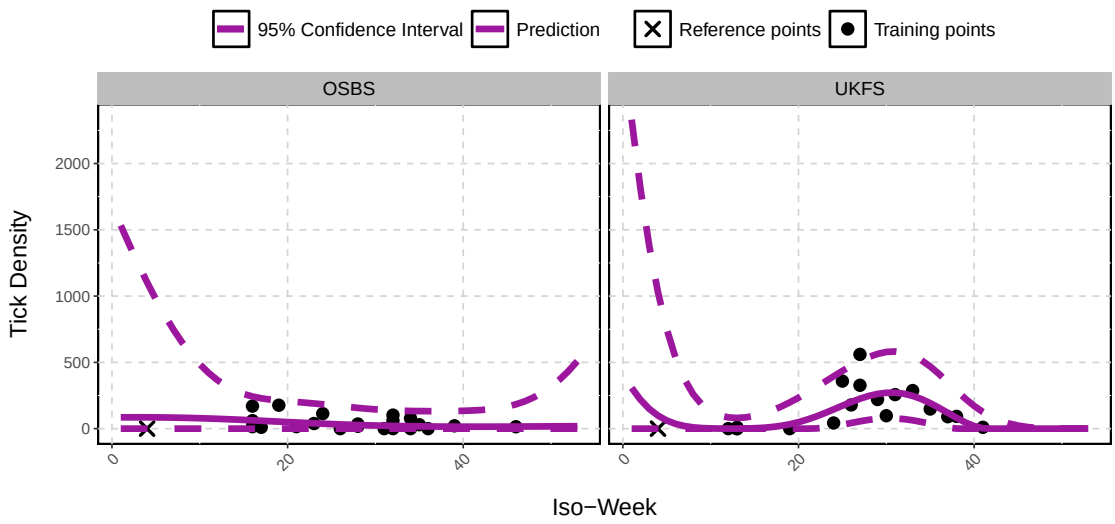


Figure 4: *LR-Time(L)* fit across iso-weeks 1–53 at (a) UKFS and (b) OSBS (solid = mean, dashed = 90% prediction intervals) on the original (untransformed) scale. Black dots show training data.

2.1 (Fick and Hijmans, 2017). The main difference is that weather is often unstable over short periods and using weather forecasts for tick density estimation can lead to severely conservative uncertainty. Instead, we chose to work with climatology, averaging over short-term atmospheric conditions, providing less noisy, more stable forecasts.

The next choice was between different meteorological variables. We had data available for temperature, humidity, and precipitation, all of which are known to affect tick density (Clarke-Crespo et al., 2020; Sagurova et al., 2019; Raghavan et al., 2016; Vail and Smith, 1998). Although humidity and precipitation are important (Clarke-Crespo et al., 2020; Raghavan et al., 2016), they are difficult to measure and are highly correlated with temperature. Thus, we decided to use the minimum temperature as the predictor. We used climatology data from the years 2014 - 2024 at 5 mins resolution to obtain *tmin* (in Celsius) recorded at monthly frequency. To match the weekly frequency of our dataset, we repeat the same temperature for all weeks in the same month. In addition to this predictor, we also use the sine and cosine predictors as in the *LR-Time(L)* model.

We show an example fit of the *LR-Temp(L)* model obtained at location ORNL in Fig. 5. This model performs well on the training data where it predicts low tick density during colder months and captures uncertainty adequately. However, for true forecasts into the future, temperature is unknown and must be forecasted. Forecasting climate variables such as monthly average temperatures even one year in advance is inherently challenging due to short-term variability.

3.2 Gaussian processes

In the previous sections, the LR models have served as a baseline for our analysis and provided insights into potential improvements. These include eliminating environmental factors, training one model for all locations, and obtaining better uncertainty quantification (UQ). In this section, we explore Gaussian processes (GPs; Gramacy, 2020) which are highly flexible and non-parametric models that can be used to learn patterns within data. GP regression models rely primarily on the distance between input values rather than their position, which has made them widely used in

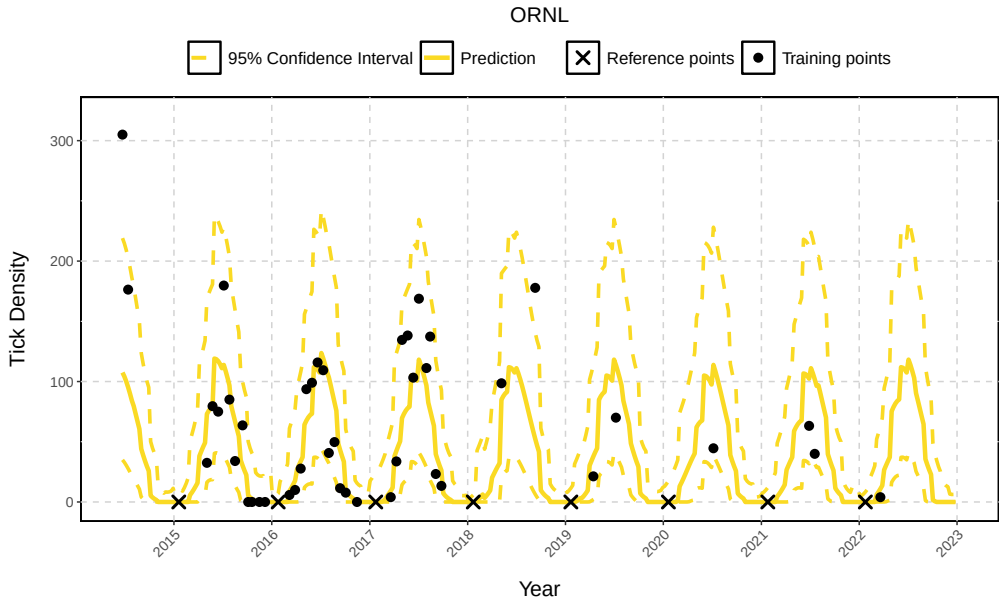


Figure 5: *LR-Temp(L)* fit at ORNL (solid = mean, dashed = 90% prediction intervals) on the original (untrasformed) scale. Black dots show training data.

spatial statistics (where they are known as kriging; Cressie, 1991). GPs have also gained popularity in machine learning due to their ability to provide fast predictions at new locations along with reliable UQ (Rasmussen and Williams, 2005). We apply GPs in our ecological context because they do not require equally spaced time series, are flexible, and are particularly effective at distinguishing signal from noise (Johnson et al., 2018).

GP models are formulated as $Y \sim \mathcal{N}(\mu_n, \Sigma_n)$. LR is technically a specific case of a GP where $\mu_n = (1, X_n)\beta$ and $\Sigma_n = \sigma^2\mathbb{I}$. However, in case of a GP, we focus specifically on the covariance structure Σ_n which is specified through a kernel. This kernel is built so that if training points x_i and x_j are closer together in the input space, the GP assumes responses Y_i and Y_j are more highly correlated. There are many kernels that fit this description, but in our case we use the squared exponential kernel $C_\theta(X_n)$ stated as

$$\Sigma_n \;= \tau^2\left(C_\theta(X_n) + g\mathbb{I}_n\right) \quad \text{where,} \quad C_\theta(X_n)^{ij} = C(x_i, x_j) = \exp\left(-\sum_{k=1}^{d}\frac{(x_{ik}-x_{jk})^2}{\theta_k}\right). \tag{2}$$

Through the kernel, the covariance structure is a function of three hyper-parameters: scale, τ^2 which captures the amplitude of the response; the length-scales, $\boldsymbol{\theta} = \{\theta_1, \theta_2, \ldots, \theta_d\}$ which captures the rate of decay of correlation for each dimension; and the nugget, g, which captures the noise observed in the response. Note that Eq. 2 represents an anisotropic specification where a unique length-scale is estimated for each dimension (as opposed to isotropic kernel where only one common length-scale is inferred for all dimensions.) We specify $\mu_n = 0$ i.e., zero mean (for simplicity) and infer hyper-parameters $\{\tau^2, \boldsymbol{\theta}, g\}$ through Maximum Likelihood Estimation (MLE).

We can then obtain the predictive distribution of $\mathcal{Y}$ at n_p new locations $\mathcal{X}$ by calculating the conditional distribution of $\mathcal{Y}$, given (X_n, Y_n). To do so, first stack the distribution of $\mathcal{Y}$ and Y_n and

jointly write it as,

$$\begin{bmatrix} \mathcal{Y} \\ Y_n \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \Sigma(\mathcal{X},\mathcal{X}) & \Sigma(\mathcal{X},X_n) \\ \Sigma(X_n,\mathcal{X}) & \Sigma_n \end{bmatrix} \right) \tag{3}$$

where $\Sigma(\mathcal{X},\mathcal{X}) = \tau^2(C_\theta(\mathcal{X},\mathcal{X}) + g\mathbb{I}_{n_p})$ is the variance matrix for $\mathcal{X}$ and $\Sigma(X_n,\mathcal{X})$ is the covariance matrix between $\mathcal{X}$ and X_n. Then, using MVN conditioning properties, we have $\mathcal{Y}|Y_n, X_n \sim N(\mu(\mathcal{X}), \sigma^2(\mathcal{X}))$, and

$$\mu(\mathcal{X}) = \Sigma(\mathcal{X},X_n)\Sigma_n^{-1}Y_n \quad \text{and} \quad \sigma^2(\mathcal{X}) = \Sigma(\mathcal{X},\mathcal{X}) - \Sigma(\mathcal{X},X_n)\Sigma_n^{-1}\Sigma(X_n,\mathcal{X}) \tag{4}$$

where $\mu(\mathcal{X})$ is the mean and $\sigma^2(\mathcal{X})$ is the variance at the predictive locations.

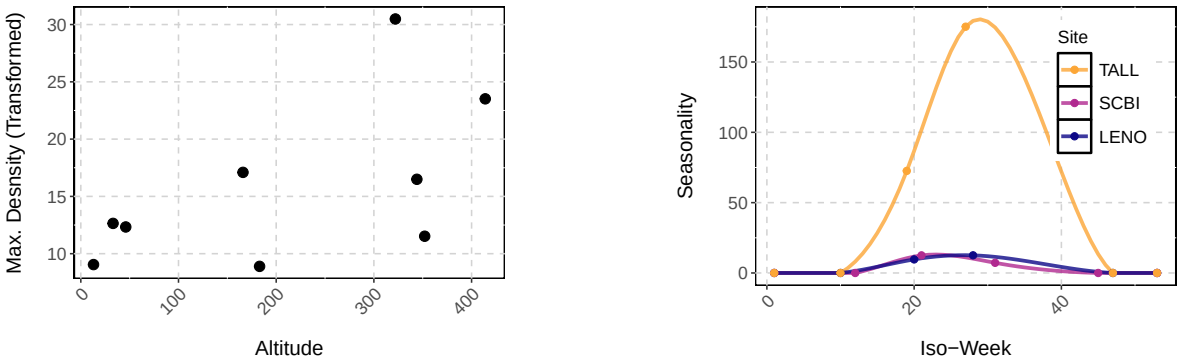
The bulk of our work is in building informative predictors X_n that capture "closeness" in our system. To build predictors, we have some criteria learnt from our previous sections, such as building predictors which do not require forecasting. Further, predictors should be able to capture temporal patterns along with geographical patterns, so we can train the GP model on a unified dataset, i.e., using data from all locations. We demonstrate shortly why this is crucial to obtain good fits in the context of sparse data.

Our first *common predictor* (the same for all locations) includes week number, (X_1^{GP}) which is a grid of time points $\{1, 2, \ldots, n, n+1, \ldots, n+n_p\}$ from the start date (week 1) up to the end date (week $n+n_p$) where week numbers n and n_p correspond to the last week in the training and testing set respectively. By using week number as a predictor, we curate a relationship between the previous week and the current week since the distance between consecutive weeks will be the small. Next, we incorporate information that weeks across years should be similar through periodicity, (X_2^{GP}). This predictor is based on the LR sine predictor, X_2^{LR}, with some modifications i.e., increasing the frequency from 53 to 106 weeks (Fig. 3b) and squaring the sinusoid so we have $X_2^{\text{GP}} = \sin^2(2\pi X_1/106)$. With these adjustment, the curve completes half a cycle in one year and flattens out towards the end of the year allowing for smoother transition between years.

In contrast to our common predictors above, *location specific predictors* are designed to distinguish one location from another. We obtained one location-specific predictor directly from the information provided by NEON namely, mean elevation (X_3^{GP}) of location which is correlated with maximum tick density (Fig. 6a). The intuition behind this is that if observations Y and Y' from our unified dataset are collected at the different geographical locations but having similar elevation, the Euclidean distances between them would be small resulting in higher correlation through Σ_n.

Lastly, we build a seasonality metric (X_4^{GP}), which uses foliage data obtained at each location (Site Metadata, 2025). We use the days-of-year (DOYs) when greenness increases, peaks, decreases, and reaches its minimum. These DOYs are converted to iso-weeks, X_0^{GP}, and are paired with corresponding tick densities Y. Two more data pairs for week 1 and 53, with density set to zero are added for continuity across years. This results in six pairs, $(X_{0_i}^{\text{GP}}, Y_i)_{i=1}^6$ for each location. Then, we fit a cubic spline through these points as shown in Fig. 6b for the subset of locations shown previously in Fig. 2 (see Appendix S1: Section B for one detailed location). Lastly, we use the spline to make predictions Y_i for all $X_{0_i}, i = \{1, 2, \ldots, 53\}$ and map it the corresponding week at that location, creating X_4^{GP}.

This predictor follows the same trend as foliage, i.e., it increases in the summer and decreases in the winter. Each kernel (spline fit) of the seasonality metric is periodic but has a different height which allows us to incorporate some information about the tick density of one location relative to another. For example, when the magnitudes of the seasonality metric are closer together (i.e.,



(a) Max. tick density (transformed) by elevation. **(b)** Seasonality metric for three locations.

Figure 6: (a) and (b) represent location-specific predictors used in the GP model.

iso-weeks 1–10, 45 – 53), the correlation in the responses between these locations is higher. When the magnitudes vary (at iso-week 20–30), the responses will be less similar. Another speciality of this predictor (unlike the sinusoid) is that the peak and the rate of change of the seasonality kernel across iso-weeks is allowed to vary from one location to another.

Using these predictors, we fit two GP models, (i) using data only obtained at one location namely *GP(L)* model through time dependent predictors (X_1^{GP} and X_2^{GP}) and (ii) using data obtained at all locations, *GP(A)* using both time and location predictors (X_1^{GP} – X_4^{GP}). In order to fit the GP, the input space is scaled so all predictors lie within the interval [0,1]. We used an anisotropic GP (Eqs. 2–4) and fit the models using `hetGP` (Binois and Gramacy, 2021) package in `R` (CRAN).

In Fig. 7, we show fits and predictions of both *GP(L)* and *GP(A)* fits at BLAN from 2015 to 2022. Observe that *GP(L)*, as trained only on data obtained at BLAN, cannot adequately learn a trend and regresses towards the mean, possibly due to inadequate data. By contrast, *GP (A)* model "borrows" information from other locations and learns the seasonal pattern, i.e., higher tick abundance during summer and little to no activity in winter. Note that the upper bound of the *GP(L)* model is much smaller than *GP(A)* model. This is because the *GP(A)* is the average of the noise across all locations resulting in overestimation of noise at some locations such as BLAN.

3.3 Heteroskedastic Gaussian processes

In the previous sections, we have observed that the response varies across different locations (see Fig. 2). All the models discussed so far assume constant variance, which can lead to overinflation of noise at some locations with smaller densities (see Fig. 7) and vice-versa. To avoid compromising on UQ, we relax the assumption of constant noise across the input space and allow some flexibility in noise level estimation via a heteroskedastic Gaussian processes (HetGP; Binois et al., 2018). HetGPs model the mean process similar to a GP but additionally infer a latent noise process. Specifically, the model assumes that the noise changes across the input space. Mathematically, this corresponds to replacing $g\mathbb{I} \to \Lambda_n$ in Eq. 2 so $\Sigma_n = \tau^2(C_{\theta_Y}(X_n) + \Lambda_n$ where, $\Lambda_n = \text{Diag}(\lambda_1, \ldots, \lambda_n)$ and

$$\log \boldsymbol{\lambda}_n \sim \mathcal{N}\left(0, \tau_\lambda^2 \Sigma_n\right) \quad \text{where} \quad \Sigma_n = C_{\theta_\lambda}(X_n) + g_\lambda \mathbb{I}. \tag{5}$$

That is, the log-noise process follows a homoskedastic GP with the same kernel structure as in Eq. (2) but a different set of hyper-parameters, θ_λ, τ_λ^2, and g_λ. The predictions follow Eqs. (3–4),

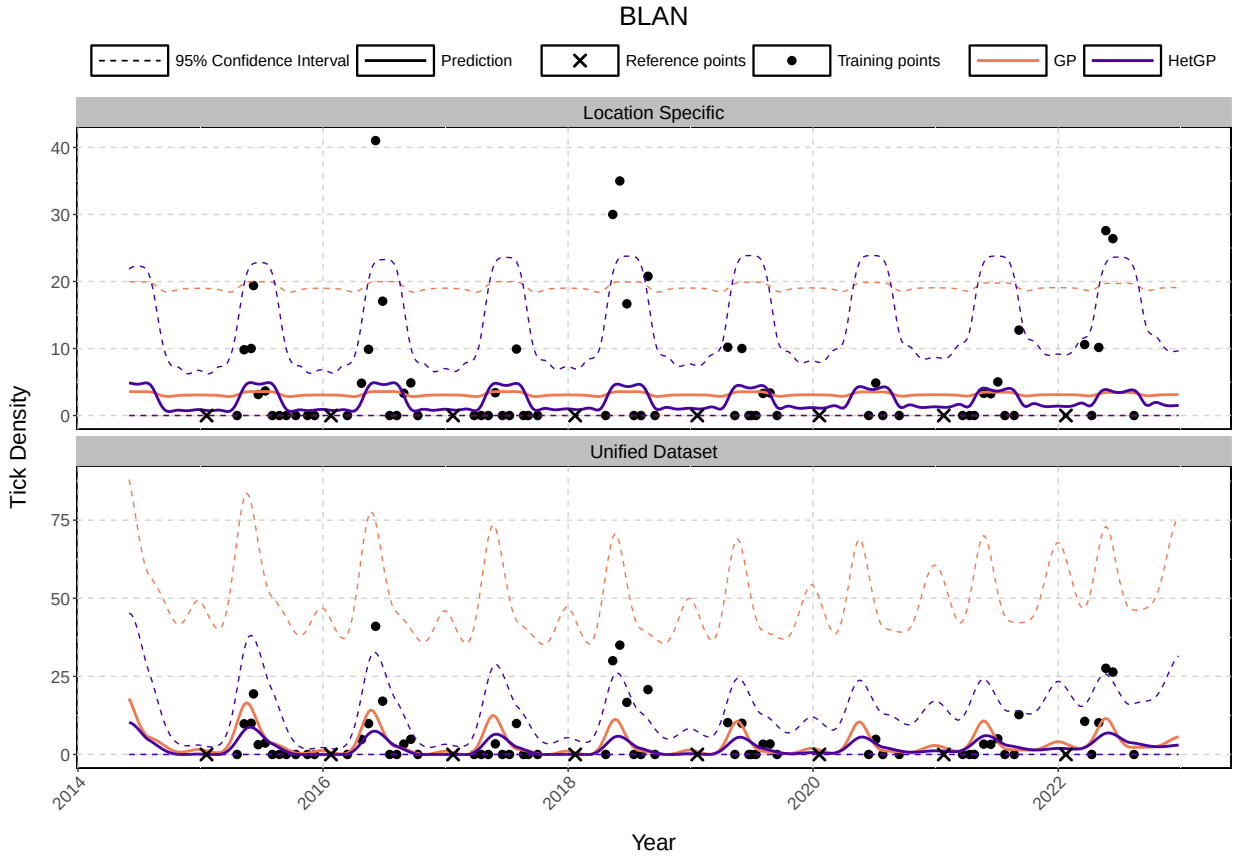


Figure 7: GP (orange) and HetGP (blue) fits at BLAN (solid = mean, dashed = 90% prediction intervals) on the original (untrasformed) scale. Black dots show training data; (*top*) models fit using location specific data only – *GP(L)* and *HetGP(L)*; (*bottom*) *GP(A)* and *HetGP(A)* trained on data from all locations.

with the new Σ_n matrix. To fit the hetGP model, we use the same predictors as the GP scaled to unit-cube and use the `hetGP` library on CRAN. We refer to this model as *HetGP(A)*.

Referring back to Fig. 7, observe that *HetGP(L)* captures a mean trend, but is unable to capture varying noise across the input space. At other locations, we also see *HetGP(L)* predictions regressing towards the mean (Appendix S1: Section C). By contrast, *HetGP(A)* trained on all available data provides better UQ. The bounds are tighter in winter where the model confidently predicts no tick activity but the UQ increases during summer months when there is higher noise. We provide a full comparison of percentage coverage along with interval widths in Section 4 showing that the *HetGP(A)* outperforms all models with appropriate bounds.

Since the *GP(L)* and *HetGP(L)* models are hit-or-miss, we focus on *GP(A)* v/s *HetGP(A)* to explain noise process estimation. The predictive variance is a combination of the process uncertainty (variability in the response) and noise (unexplained variation), which we suspect varies from one location to another and across seasons. GPs average across the noise in the input space using one noise level, g. This often results in over/under estimation (e.g., BLAN). HetGPs provide different noise level estimates across every week at every location. This results in adaptive prediction intervals that effectively capture input dependent noise and improve uncertainty estimation.

We use Fig. 8 to demonstrate the capabilities of *HetGP(A)*. We have two locations where the

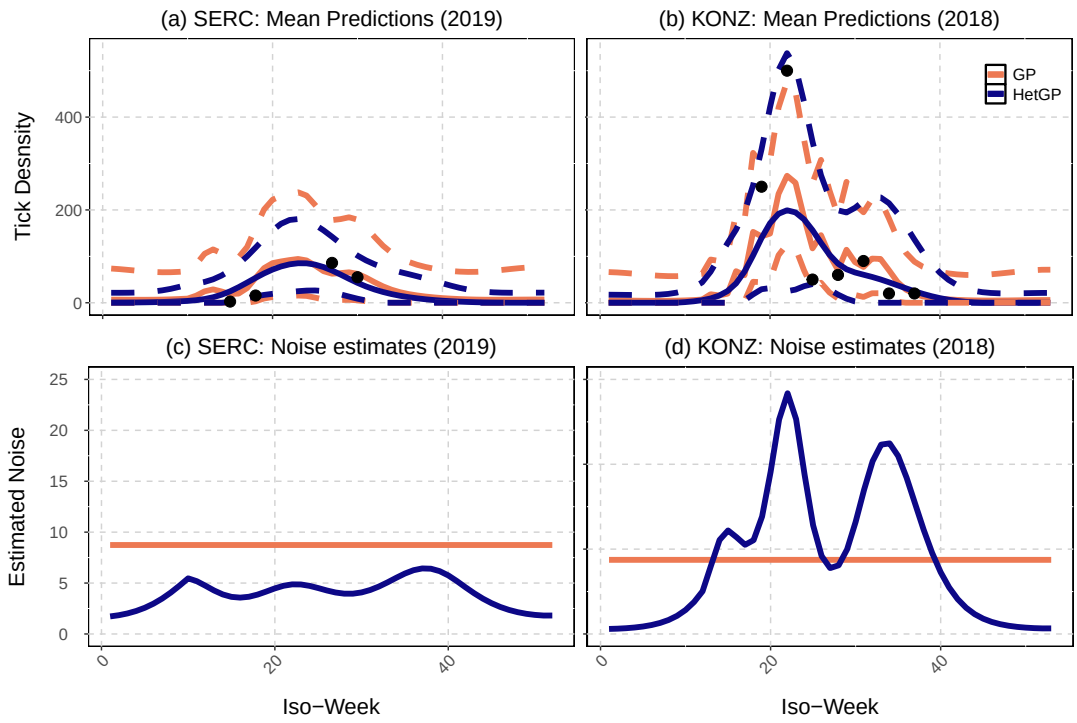


Figure 8: Predicted tick densities (solid = mean, dashed = intervals) and estimated noise levels for two locations obtained from regular GP (orange) and HetGP (blue) model.

GP(A) and *HetGP(A)* provide the same coverage. Notice in Fig. 8a at SERC that noise estimated by the *GP(A)* model is consistently higher for the same coverage. At KONZ (Fig. 8c), we observe that *HetGP(A)* captures both seasonal and site-specific variation. For example, at KONZ during high-variability summer weeks ($n' \in \{18, \ldots, 37\}$), the model estimates higher noise ($\boldsymbol{\lambda} n' > g$), leading to wider intervals whereas, in winter, noise is lower ($\boldsymbol{\lambda}'_n < g$ for $n' \notin \{14, \ldots, 26, 29, \ldots, 39\}$), resulting in tighter bounds (Fig. 8d). This pattern is different than what is seen at SERC even though the locations are fit together with the same model. This adaptability enables *HetGP(A)* to provide better coverage and more reliable forecasts.

4 Comparative Prediction Performance

In this section, we evaluate the performance of several models using metrics discussed in 2.2.2. We consider models such as LR-Temp, BASS, GP, and HetGP. We use two versions of all models: (1) Models trained using data at specific location (notated by L) using time dependent predictors such as week and periodicity and (2) models trained using all the available data which also use location specific predictors. For the *LR-Temp(A)* model, we use a categorical input to distinguish between locations. Lastly, we include *LR-Time(L)* model as our baseline.

We evaluate within-sample fits using data before 2023 by analyzing Q-Q plots of standardized residuals (Fig. 9). Residuals are computed using training data and model predictions, then standardized by dividing by the predicted standard error. The Q-Q plot compares predicted vs. observed tick densities, and closeness to the Q-Q line indicates better fit and UQ. GP models predict

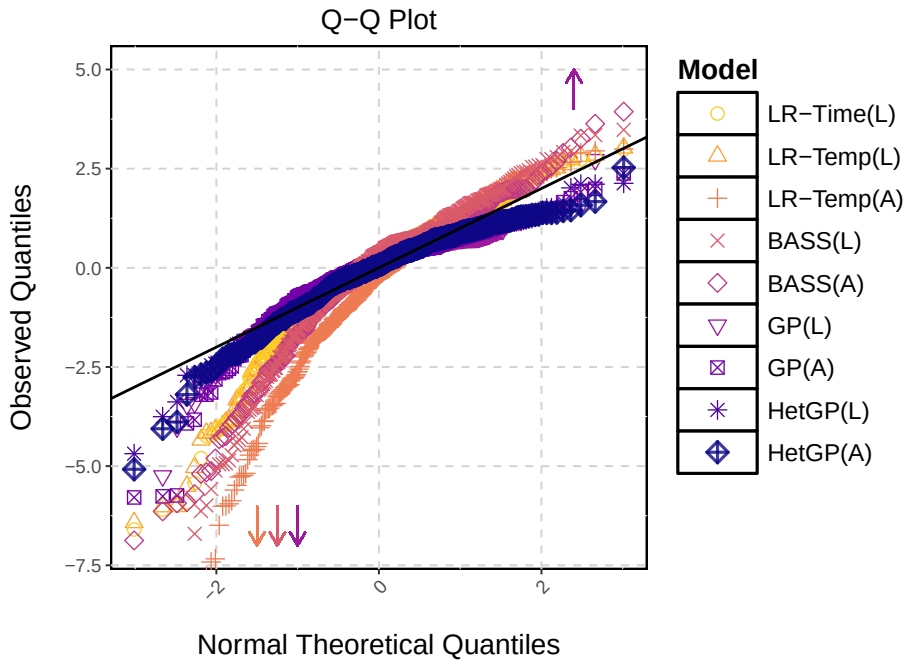


Figure 9: Q-Q Plot showing the standardized residuals of all models for all points in the training data

well at lower densities and stray away from the line at higher densities, slightly under-estimating tick abundance. All other models under predict at lower tick densities, which is arguably worse as detecting an absence of ticks when there are really ticks could lead to lack of preventive measures and increase in tick-borne diseases.

Model	In-sample			Out-of-sample		
	Coverage (%)	Median Width	Mean Width	Coverage (%)	Median Width	Mean Width
LR-Time(L)	95.49	10.90	15.18	87.90	13.15	17.22
LR-Temp(L)	95.49	11.21	11.77	87.90	11.49	11.92
LR-Temp(A)	91.93	14.02	14.09	89.14	14.06	14.11
BASS(L)	91.47	11.43	12.17	84.81	11.95	12.75
BASS(A)	92.17	10.95	10.95	87.90	10.97	10.98
GP(L)	95.04	10.21	10.49	92.59	12.66	12.79
GP(A)	94.24	10.33	10.49	89.07	11.37	11.56
HetGP(L)	91.55	9.76	9.76	86.05	13.09	13.25
HetGP(A)	94.15	8.04	8.34	89.44	10.27	10.76

Table 1: Coverage (%) and interval widths for both in-sample and out-of-sample data for 90% prediction intervals.

We assess sample coverage and interval widths to evaluate whether prediction intervals provide adequate UQ. These metrics were calculated on the transformed predictive bounds to capture the full width of the intervals (back transforming results in lower PIs bounded at zero). The coverage of all models along with the mean and median width of the interval across all locations, for all observations is shown in Tab. 1. The in-sample coverage is typically between $90 - 95\%$ for a 90%

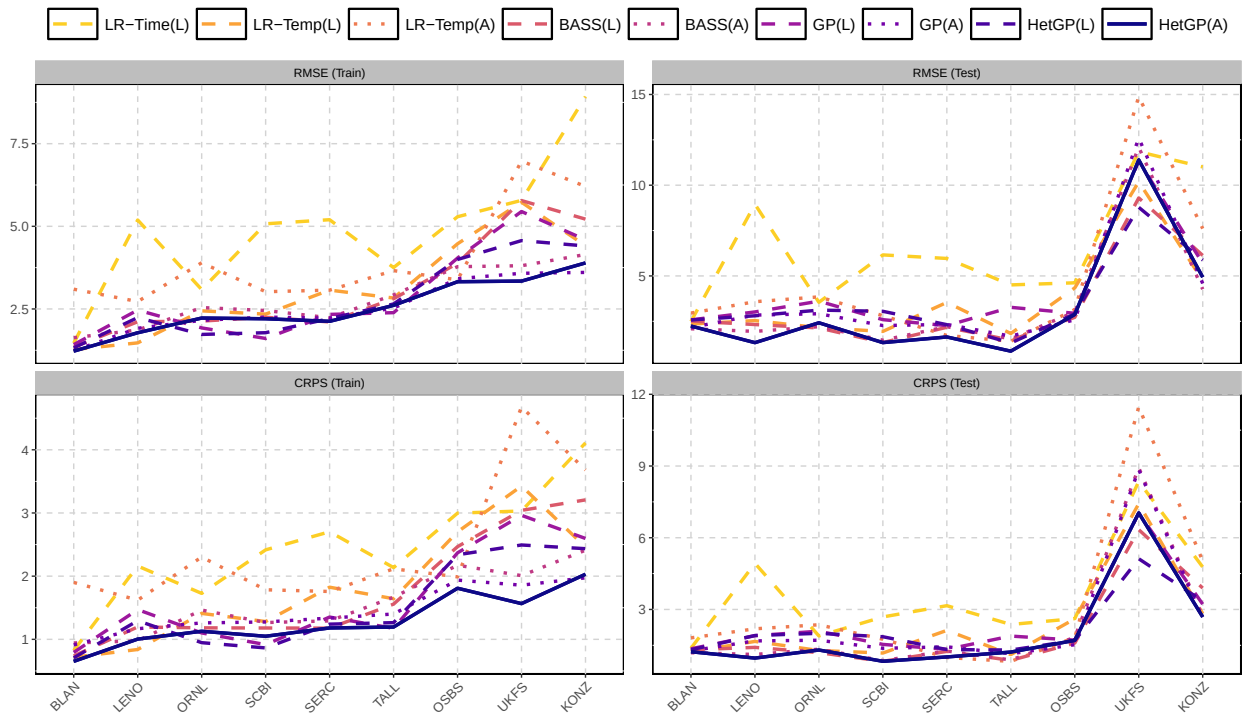


Figure 10: Train and test CRPS and RMSE. Models trained only using location-wise data are in *dashed* and on all locations in *dotted*. Our approach: *HetGP* model is shown in *solid* purple line.

prediction interval for all models whereas, the out-of-sample coverage is between $85 - 90\%$ with only the *GP(A)* providing $> 90\%$ coverage. Notice that *HetGP(A)* is closest to the nominal level for out-of-sample predictions and has smaller widths than all other models for in- and out-of sample data.

Lastly, we also evaluate in- and out-of-sample RMSE and CRPS. The RMSE and CRPS for the *HetGP(A)* model is generally lower at all locations compared to all other models. One situation where the *HetGP(A)* does not convincingly outperform other models is at UKFS in the testing (out-of-sample) set. Here, the magnitude for the testing data responses were much larger than previously recorded at the location. Additionally, the site is very sparsely sampled with only 25 total observations (Fig. 1). The *HetGP(A)* model simply learns patterns in historical data and forecasts these into the future with an assumption of stationarity. The results can sometimes be poor when there is a drastic change in the range of the response not present in historical training data. Also, recall Fig. 7 where *GP(L)* model clearly does not learn a pattern and simply provides a linear mean prediction. We notice similar behavior in the *BASS(L)* model at other locations. Judging purely on basis of metrics, we see all GP, HetGP, and BASS seem to perform similarly at BLAN, even despite the addition of reference points. Yet only *HetGP (A)* captures both the mean and noise process as evident in the Fig. 7.

5 Discussion

We proposed using Gaussian processes to forecast tick populations for one future season (here one year) and have shown their capability in settings with sparse data. Our framework leveraged a

core property of GPs: reliance on relative distances between predictor/input values rather than their position in the input space. Additionally, we show that modeling noise separately via HetGPs provides better UQ and coverage. Ultimately we observed that the HetGP models outperformed all the other competitors.

Despite these advantages, GP models retain some of the challenges typical of more common time-series approaches. For example, GP models are best suited for short-term forecasting (e.g., one season/year or less) as these models typically assume stationarity which may not hold for longer horizons. GPs also do not include mechanistic components that describe the dynamics of the population being modeled. Unlike differential equation based models which study rate of change of population explained by factors such as demographics, temperature, or humidity, GP models do not posit either mechanisms or causality. Thus, they cannot be used to explore possible changes in populations across time due to climactic shifts or to implement control measures.

Understanding the noise process was essential to obtain better predictions. However, the current HetGP implementation lacks flexibility in specifying how noise varies across input dimensions. In ecological datasets, noise levels often fluctuate due to seasonal and geographical factors. In this paper, the HetGP model fits an individual noise level at every point in time across all locations. Although this worked well in our case, it may lead to overfitting in some scenarios. Future work could explore alternative implementations where the noise process only changes in some dimensions of the input space.

The goal of our case study was to develop a short-term forecasting model for a sparse dataset such as ticks. These forecasts provide insights into tick patterns and enable health departments and policy makers to take action and ensure disease prevention (Kessler et al., 2019; Goddard and Varela-Stokes, 2009). We can protect wildlife, livestock and pet infections (Sykes, 2023; Springer et al., 2015). However, we see broader applications using the GP framework. Using a similar recipe to ours, ecologists can use this model on other irregularly sampled datasets with large gaps such as mosquitoes, endangered species, etc.

References

Allan, B. F., Goessling, L. S., Storch, G. A., and Thach, R. E. (2010). "Blood meal analysis to identify reservoir hosts for *Amblyomma americanum* ticks." *Emerging Infectious Diseases*, 16, 3, 433.

Auger-Méthé, M., Field, C., Albertsen, C. M., Derocher, A. E., Lewis, M. A., Jonsen, I. D., and Mills Flemming, J. (2016). "State-space models' dirty little secrets: even simple linear Gaussian models can have estimation problems." *Scientific reports*, 6, 1, 26677.

Auger-Méthé, M., Newman, K., Cole, D., Empacher, F., Gryba, R., King, A. A., Leos-Barajas, V., Mills Flemming, J., Nielsen, A., Petris, G., et al. (2021). "A guide to state–space modeling of ecological time series." *Ecological Monographs*, 91, 4, e01470.

Binois, M. and Gramacy, R. B. (2021). "`hetgp`: Heteroskedastic Gaussian process modeling and sequential design in R." *Journal of Statistical Software*, 98, 1–44.

Binois, M., Gramacy, R. B., and Ludkovski, M. (2018). "Practical heteroscedastic gaussian process

modeling for large simulation experiments." *Journal of Computational and Graphical Statistics*, 27, 4, 808–821.

Brugger, K., Walter, M., Chitimia-Dobler, L., Dobler, G., and Rubel, F. (2018). "Forecasting next season's *Ixodes ricinus* nymphal density: the example of southern Germany 2018." *Experimental and Applied Acarology*, 75, 281–288.

Carey, C. C., Woelmer, W. M., Lofton, M. E., Figueiredo, R. J., Bookout, B. J., Corrigan, R. S., Daneshmand, V., Hounshell, A. G., Howard, D. W., Lewis, A. S., et al. (2022). "Advancing lake and reservoir water quality management with near-term, iterative ecological forecasting." *Inland Waters*, 12, 1, 107–120.

Childs, J. E. and Paddock, C. D. (2003). "The ascendancy of *Amblyomma americanum* as a vector of pathogens affecting humans in the United States." *Annual Review of Entomology*, 48, 1, 307–337.

Clarke-Crespo, E., Moreno-Arzate, C. N., and López-González, C. A. (2020). "Ecological niche models of four hard tick genera (Ixodidae) in Mexico." *Animals*, 10, 4, 649.

Commins, S. P., James, H. R., Kelly, L. A., Pochan, S. L., Workman, L. J., Perzanowski, M. S., Kocan, K. M., Fahy, J. V., Nganga, L. W., Ronmark, E., Cooper, P. J., and Platts-Mills, T. A. (2011). "The relevance of tick bites to the production of IgE antibodies to the mammalian oligosaccharide galactose." *Journal of Allergy and Clinical Immunology*, 127, 5, 1286–1293.e6.

Cortinas, R. and Spomer, S. (2013). "Lone Star Tick (Acari: Ixodidae) Occurrence in Nebraska: Historical and Current Perspectives." *Journal of Medical Entomology*, 50, 2, 244–251.

Cressie, N. (1991). *Statistics for Spatial Data*. A Wiley-Interscience Publication. J. Wiley.

Cressie, N., Calder, C. A., Clark, J. S., Hoef, J. M. V., and Wikle, C. K. (2009). "Accounting for uncertainty in ecological analysis: the strengths and limitations of hierarchical statistical modeling." *Ecological Applications*, 19, 3, 553–570.

DiMatteo, I., Genovese, C. R., and Kass, R. E. (2001). "Bayesian curve-fitting with free-knot splines." *Biometrika*, 88, 4, 1055–1071.

Estrada-Peña, A. (2001a). "Forecasting habitat suitability for ticks and prevention of tick-borne diseases." *Veterinary Parasitology*, 98, 1-3, 111–132.

— (2001b). "Forecasting habitat suitability for ticks and prevention of tick-borne diseases." *Veterinary Parasitology*, 98, 1-3, 111–132.

Falco, R. C. and Fish, D. (1992). "A comparison of methods for sampling the deer tick, *Ixodes dammini*, in a Lyme disease endemic area." *Experimental & Applied Acarology*, 14, 165–173.

Felz, M. W., Durden, L. A., and Oliver Jr, J. H. (1996). "Ticks parasitizing humans in Georgia and South Carolina." *The Journal of Parasitology*, 505–508.

Fick, S. E. and Hijmans, R. J. (2017). "WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas." *International Journal of Climatology*, 37, 12, 4302–4315.

Francom, D. and Sansó, B. (2020). "BASS: An R package for fitting and performing sensitivity analysis of Bayesian adaptive spline surfaces." *Journal of Statistical Software*, 94, 1–36.

Gaines, D. N., Operario, D. J., Stroup, S., Stromdahl, E., Wright, C., Gaff, H., Broyhill, J., Smith, J., Norris, D. E., Henning, T., et al. (2014). "Ehrlichia and spotted fever group Rickettsiae surveillance in *Amblyomma americanum* in Virginia through use of a novel six-plex real-time PCR assay." *Vector-Borne and Zoonotic Diseases*, 14, 5, 307–316.

Gajewski, Z., Stevenson, L. A., Pike, D. A., Roznik, E. A., Alford, R. A., and Johnson, L. R. (2021). "Predicting the growth of the amphibian chytrid fungus in varying temperature environments." *Ecology and Evolution*, 11, 24, 17920–17931.

Gneiting, T. and Raftery, A. E. (2007). "Strictly proper scoring rules, prediction, and estimation." *Journal of the American Statistical Association*, 102, 477, 359–378.

Goddard, J. (2002). "A ten-year study of tick biting in Mississippi: implications for human disease transmission." *Journal of Agromedicine*, 8, 2, 25–32.

Goddard, J. and Varela-Stokes, A. S. (2009). "Role of the lone star tick, *Amblyomma americanum* (L.), in human and animal diseases." *Veterinary Parasitology*, 160, 1, 1–12.

Gramacy, R. B. (2020). *Surrogates: Gaussian process modeling, design, and optimization for the applied sciences*. Chapman and Hall/CRC.

Hahn, M. B., Jarnevich, C. S., Monaghan, A. J., and Eisen, R. J. (2016). "Modeling the geographic distribution of *Ixodes scapularis* and *Ixodes pacificus* (Acari: Ixodidae) in the contiguous United States." *Journal of Medical Entomology*, 53, 5, 1176–1191.

Haile, D. and Mount, G. (1987). "Computer simulation of population dynamics of the lone star tick, *Amblyomma americanum* (Acari: Ixodidae)." *Journal of Medical Entomology*, 24, 3, 356–369.

Hamill, T. M., Whitaker, J. S., Shlyaeva, A., Bates, G., Fredrick, S., Pegion, P., Sinsky, E., Zhu, Y., Tallapragada, V., Guan, H., et al. (2022). "The reanalysis for the global ensemble forecast system, version 12." *Monthly Weather Review*, 150, 1, 59–79.

Harris, I., Osborn, T. J., Jones, P., and Lister, D. (2020). "Version 4 of the CRU TS monthly high-resolution gridded multivariate climate dataset." *Scientific Data*, 7, 1, 109.

Johnson, B. and Munch, S. B. (2022). "An empirical dynamic modeling framework for missing or irregular samples." *Ecological Modelling*, 468, 109948.

Johnson, L. R., Gramacy, R. B., Cohen, J., Mordecai, E., Murdock, C., Rohr, J., Ryan, S. J., Stewart-Ibarra, A. M., and Weikel, D. (2018). "Phenomenological forecasting of disease incidence using heteroskedastic Gaussian processes: A dengue case study." *The Annals of Applied Statistics*, 12, 1, 27–66.

Jones, R. H. (1986). "Time series regression with unequally spaced data." *Journal of Applied Probability*, 23, A, 89–98.

Kessler, W. H., Ganser, C., and Glass, G. E. (2019). "Modeling the distribution of medically important tick species in Florida." *Insects*, 10, 7, 190.

Kollars Jr., T. M., Oliver Jr., J. H., Durden, L. A., and Kollars, P. G. (2000). "Host Associations and Seasonal Activity of *Amblyomma americanum* (Acari: Ixodidae) in Missouri." *Journal of Parasitology*, 86, 5, 1156 – 1159.

Léger, E., Vourc'h, G., Vial, L., Chevillon, C., and McCoy, K. D. (2013). "Changing distributions of ticks: causes and consequences." *Experimental and Applied Acarology*, 59, 1, 219–244.

Levin, M. L., Zemtsova, G. E., Killmaster, L. F., Snellgrove, A., and Schumacher, L. B. (2017). "Vector competence of *Amblyomma americanum* (Acari: Ixodidae) for Rickettsia rickettsii." *Ticks and Tick-borne Diseases*, 8, 4, 615–622.

Little, R. J. and Rubin, D. B. (2019). *Statistical analysis with missing data*. John Wiley & Sons.

Liu, C., Zhai, Z., Zhang, R., Bai, J., and Zhang, M. (2022). "Field pest monitoring and forecasting system for pest control." *Frontiers in Plant Science*, 13, 990965.

Masters, E. J., Grigery, C. N., and Masters, R. W. (2008). "STARI, or Masters Disease: Lone Star Tick–Vectored Lyme-like Illness." *Infectious Disease Clinics of North America*, 22, 2, 361–376. Tick-borne Diseases, Part I: Lyme Disease.

Meltzer, M. and Norval, R. (1992). "The use of time-series analysis to forecast bont tick (*Amblyomma hebraeum*) infestations in Zimbabwe." *Experimental & Applied Acarology*, 13, 4, 261–279.

Mitchell, C. L., Lin, F.-C., Vaughn, M., Apperson, C. S., Meshnick, S. R., and Commins, S. P. (2020). "Association between lone star tick bites and increased alpha-gal sensitization: evidence from a prospective cohort of outdoor workers." *Parasites & vectors*, 13, 1–4.

Mixson, T. R., Campbell, S. R., Gill, J. S., Ginsberg, H. S., Reichard, M. V., Schulze, T. L., and Dasch, G. A. (2006). "Prevalence of Ehrlichia, Borrelia, and Rickettsial agents in *Amblyomma americanum* (Acari: Ixodidae) collected from nine states." *Journal of Medical Entomology*, 43, 6, 1261–1268.

Mock, D. E., Applegate, R. D., and Fox, L. B. (2001). "Preliminary survey of ticks (Acari: Ixodidae) parasitizing wild turkeys (Aves: Phasianidae) in eastern Kansas." *Journal of Medical Entomology*, 38, 1, 118–121.

Molaei, G., Little, E. A., Williams, S. C., and Stafford, K. C. (2019). "Bracing for the worst—range expansion of the lone star tick in the northeastern United States." *New England Journal of Medicine*, 381, 23, 2189–2192.

Monzón, J. D., Atkinson, E. G., Henn, B. M., and Benach, J. L. (2016). "Population and evolutionary genomics of *Amblyomma americanum*, an expanding arthropod disease vector." *Genome Biology and Evolution*, 8, 5, 1351–1360.

National Ecological Observatory Network (NEON) (2024). "Ticks sampled using drag cloths (DP1.10093.001)."

Newman, K., King, R., Elvira, V., de Valpine, P., McCrea, R. S., and Morgan, B. J. (2023). "State-space models for ecological time-series data: Practical model-fitting." *Methods in Ecology and Evolution*, 14, 1, 26–42.

Paddock, C. and Yabsley, M. (2007). "Ecological havoc, the rise of white-tailed deer, and the emergence of *Amblyomma americanum*-associated zoonoses in the United States." *Wildlife and Emerging Zoonotic Diseases: The Biology, Circumstances and Consequences of Cross-Species Transmission*, 289–324.

Paddock, C. D., Lane, R. S., Staples, J. E., and Labruna, M. B. (2016). "Changing paradigms for tick-borne diseases in the Americas." In *Global health impacts of vector-borne diseases: workshop summary*, 221–258. National Academies Press Washington, DC, USA.

Pfäffle, M., Littwin, N., Muders, S. V., and Petney, T. N. (2013). "The ecology of tick-borne diseases." *International Journal for Parasitology*, 43, 12-13, 1059–1077.

Piesman, J. and Eisen, L. (2008). "Prevention of Tick-Borne Diseases*." *Annual review of entomology*, 53, 323–43.

Ponomarev, V., Napalkova, V., and Klobukov, G. (2019). "Forecasting the phenology of gypsy moth development and the time frames for the pest control based on sum of effective temperatures according to latitudinal origination of population." In *AIP Conference Proceedings*, vol. 2063. AIP Publishing.

R Core Team (2022). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Raghavan, R. K., Goodin, D. G., Hanzlicek, G. A., Zolnerowich, G., Dryden, M. W., Anderson, G. A., and Ganta, R. R. (2016). "Maximum entropy-based ecological niche model and bioclimatic determinants of lone star tick (*Amblyomma americanum*) niche." *Vector-Borne and Zoonotic Diseases*, 16, 3, 205–211.

Raghavan, R. K., Neises, D., Goodin, D. G., Andresen, D. A., and Ganta, R. R. (2014). "Bayesian Spatio-Temporal Analysis and Geospatial Risk Factors of Human Monocytic Ehrlichiosis." *PLOS ONE*, 9, 7, 1–8.

Raghavan, R. K., Peterson, A. T., Cobos, M. E., Ganta, R. R., and Foley, D. H. (2019). "Current and Future Distribution of the Lone Star Tick, *Amblyomma americanum* (L.) (Acari: Ixodidae) in North America." *PLoS ONE*, 14.

Randolph, S. and Rogers, e. D. (1997). "A generic population model for the African tick *Rhipicephalus appendiculatus*." *Parasitology*, 115, 3, 265–279.

Rasmussen, C. E. and Williams, C. K. I. (2005). *Gaussian Processes for Machine Learning*. The MIT Press.

Robinson, P. (1977). "Estimation of a time series model from unequally spaced data." *Stochastic Processes and their Applications*, 6, 1, 9–24.

Rochlin, I., Egizi, A., and Lindström, A. (2022). "The Original Scientific Description of the Lone Star Tick (*Amblyomma americanum*, Acari: Ixodidae) and Implications for the Species' Past and Future Geographic Distributions." *Journal of Medical Entomology*, 59, 2, 412–420.

Ryan, S. J., Carlson, C. J., Mordecai, E. A., and Johnson, L. R. (2019). "Global expansion and redistribution of Aedes-borne virus transmission risk with climate change." *PLoS neglected tropical diseases*, 13, 3, e0007213.

Sagurova, I., Ludwig, A., Ogden, N. H., Pelcat, Y., Dueymes, G., and Gachon, P. (2019). "Predicted northward expansion of the geographic range of the tick vector *Amblyomma americanum* in North America under future climate conditions." *Environmental Health Perspectives*, 127, 10, 107014.

Savage, H., Godsey, M., Lambert, A., Panella, N., Burkhalter, K., Harmon, J., Lash, R., Ashley, D., and Nicholson, W. (2013). "First Detection of Heartland Virus (Bunyaviridae: Phlebovirus) from Field Collected Arthropods." *The American Journal of Tropical Medicine and Hygiene*, 89.

Serra, P. and Krivobokova, T. (2017). "Adaptive empirical Bayesian smoothing splines."

Site Metadata (2025). "NEON Sites Metadata: `https://raw.githubusercontent.com/eco4cast/neon4cast-ci/main/catalog/sites/10.1002/fee.2616`."

Slatculescu, A. M., Clow, K. M., McKay, R., Talbot, B., Logan, J. J., Thickstun, C. R., Jardine, C. M., Ogden, N. H., Knudby, A. J., and Kulkarni, M. A. (2020). "Species distribution models for the eastern blacklegged tick, *Ixodes scapularis*, and the Lyme disease pathogen, *Borrelia burgdorferi*, in Ontario, Canada." *PLoS One*, 15, 9, e0238126.

Springer, Y. P., Eisen, L., Beati, L., James, A. M., and Eisen, R. J. (2014). "Spatial distribution of counties in the continental United States with records of occurrence of *Amblyomma americanum* (Ixodida: Ixodidae)." *Journal of Medical Entomology*, 51, 2, 342–351.

Springer, Y. P., Jarnevich, C. S., Barnett, D. T., Monaghan, A. J., and Eisen, R. J. (2015). "Modeling the present and future geographic distribution of the lone star tick, *Amblyomma americanum* (Ixodida: Ixodidae), in the continental United States." *The American Journal of Tropical Medicine and Hygiene*, 93, 4, 875.

Steinke, J. W., Platts-Mills, T. A., and Commins, S. P. (2015). "The alpha-gal story: Lessons learned from connecting the dots." *Journal of Allergy and Clinical Immunology*, 135, 3, 589–596.

Sykes, J. E. (2023). "Tick-borne diseases." *Veterinary Clinics: Small Animal Practice*, 53, 1, 141–154.

Thomas, R. Q., Boettiger, C., Carey, C. C., Dietze, M. C., Johnson, L. R., Kenney, M. A., McLachlan, J. S., Peters, J. A., Sokol, E. R., Weltzin, J. F., Willson, A., Woelmer, W. M., and contributors, C. (2023). "The NEON Ecological Forecasting Challenge." *Frontiers in Ecology and the Environment*, 21, 3, 112–113.

Troughton, D. R. and Levin, M. L. (2007). "Life Cycles of Seven Ixodid Tick Species (Acari: Ixodidae) Under Standardized Laboratory Conditions." *Journal of Medical Entomology*, 44, 5, 732–740.

Vail, S. G. and Smith, G. (1998). "Air temperature and relative humidity effects on behavioral activity of blacklegged tick (Acari: Ixodidae) nymphs in New Jersey." *Journal of Medical Entomology*, 35, 6, 1025–1028.

van Nunen, S. (2015). "Tick-induced allergies: mammalian meat allergy, tick anaphylaxis and their significance." *Asia Pacific Allergy*, 5, 1, 3–16.

Williams, H. W., Cross, D. E., Crump, H. L., Drost, C. J., and Thomas, C. J. (2015). "Climate suitability for European ticks: assessing species distribution models against null models and projection under AR5 climate." *Parasites & vectors*, 8, 1–15.

Wimberly, M. C., Yabsley, M. J., Baer, A. D., Dugan, V. G., and Davidson, W. R. (2008). "Spatial heterogeneity of climate and land-cover constraints on distributions of tick-borne pathogens." *Global Ecology and Biogeography*, 17, 2, 189–202.

Wood, S. A., Jiang, W., and Tanner, M. (2002). "Bayesian mixture of splines for spatially adaptive nonparametric regression." *Biometrika*, 89, 3, 513–528.

Zannou, O. M., Ouedraogo, A. S., Biguezoton, A. S., Abatih, E., Coral-Almeida, M., Farougou, S., Yao, K. P., Lempereur, L., and Saegerman, C. (2021). "Models for studying the distribution of ticks and tick-borne diseases in animals: A systematic review and a meta-Analysis with a Focus on Africa." *Pathogens*, 10, 7, 893.

Appendix

A Performance Metrics

1. **Residual Mean Square Error (RMSE)**

   RMSE provides a measure of prediction error. Let y_i be the response observed at input location x_i and μ_i denote the prediction at that location. Let n be the sample size. Then the RMSE can be given as follows:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \mu_i)^2}{n}} \tag{6}$$

2. **Continuous Rank Probability Score (CRPS)**

   CRPS provides a measure of uncertainty quantification. Let y_i be the response with posterior predicted mean μ_i and standard deviation σ_i. The CRPS is defined as follows:

$$CRPS = \frac{1}{n}\sum_{i=1}^{n}\left[\sigma_i\left(\frac{1}{\sqrt{\pi}} - 2\phi(z_i) - z_i(2\Phi(z_i) - 1)\right)\right] \qquad z_i = \frac{y_i - \mu_i}{\sigma_i} \tag{7}$$

B Spline Fitting for Seasonality Metric

Here, we provide a detailed explanation of how we fit a spline to obtain the seasonality metric for one location. (TALL)

1. Obtain points for one location to fit the spline. This is based on the foliage information provided by NEON.

Foliage Trend	**DOY**	X_0	**Y**	(X_{0_i}, Y_i)
Continuity		1	0	(1, 0)
Starts increasing	75	10	20	(10, 20)
Reaches Maximum	135	19	174.85	(19, 174.85)
Starts Decreasing	195	27	55.08	(27, 55.08)
Reaches Minimum	330	47	0	(47, 0)
Continuity		53	0	(53, 0)

Table 2: Fitting a spline for Seasonality Metric; Section 3.2 (Fig. 6).

2. Use (X_{0_i}, Y_i) pairs and fit a spline using R cubic.spline = spline(X,Y, n=53, method = "natural") Here n = 53 indicates we want 53 points which will be X = 1, 2 ...53 and corresponding Y.

3. Obtain Y for all $X = 1, 2, ...53$ from the fitted spline. If any Y is negative, we bound it at 0. This gives us a positive seasonality metric.

C Location-wise Data Visualization and Model Fitting

In this section, we show the fits obtained from all the models: *LR-Time, LR-Temp, BASS, GP*, and *HetGP*, at each location. We include both the *(L)* and *(A)* methods trained using a subset and all of the data respectively. Observe that at locations which is very irregularly sampled (e.g., BLAN, LENO, etc.), *HetGP(L)* and *BASS(L)* also suffer from mean regression (similar to the *GP(L)* model described in Fig. 7). Additionally, we observe that training *LR-Temp(A)* and *BASS(A)* using all the available data also behaves similar to the *GP(A)* averaging across the noise over all locations resulting in overly conservative bounds at some locations (e.g., BLAN) and very poor coverage at others (e.g., UKFS).

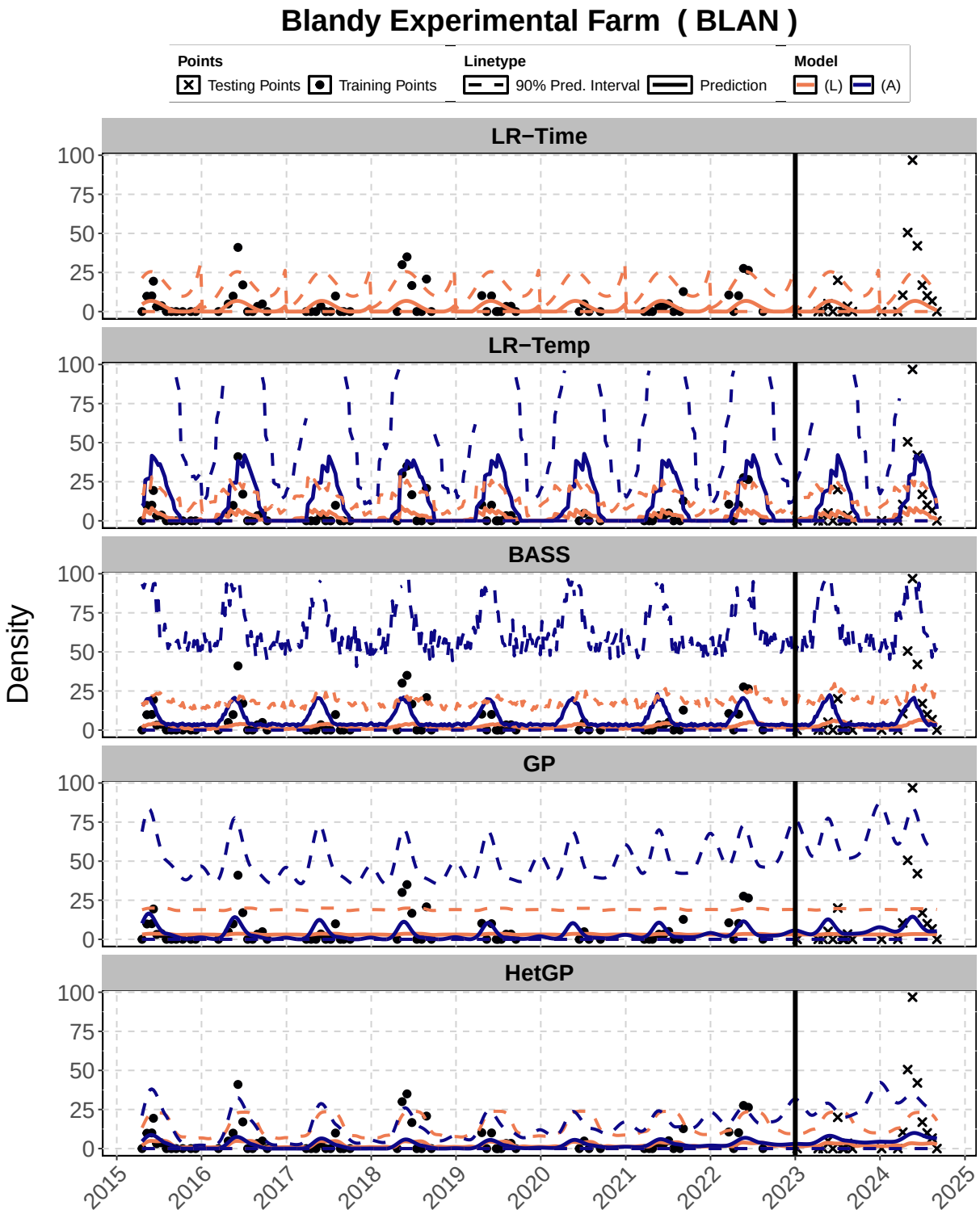


Figure 11: *LR-Time, LR-Temp, BASS, GP*, and *HetGP* – (L) and (A) models at BLAN along with 90% prediction intervals on training and testing dataset

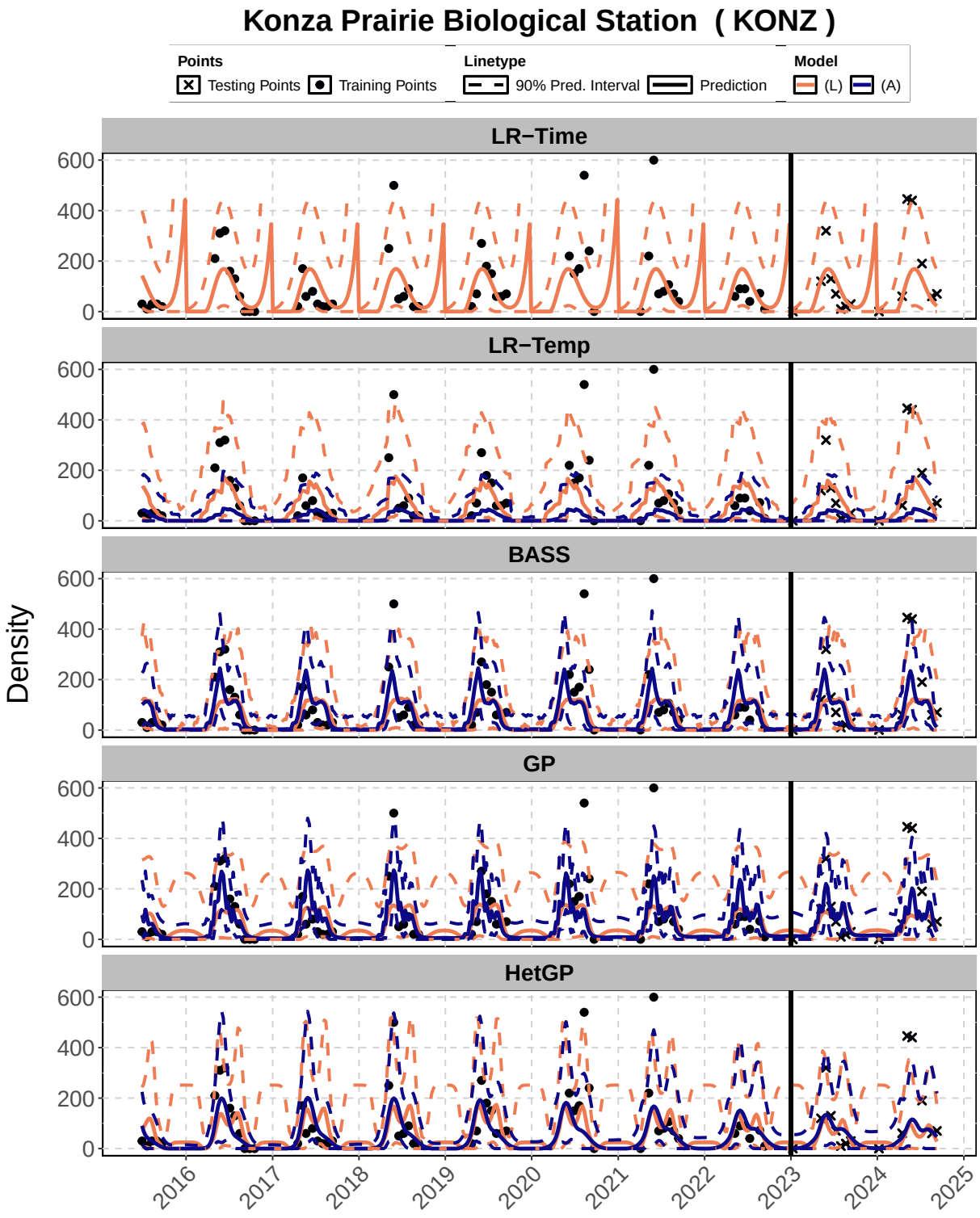


Figure 12: *LR-Time, LR-Temp, BASS, GP*, and *HetGP* – (L) and (A) models at KONZ along with 90% prediction intervals on training and testing dataset

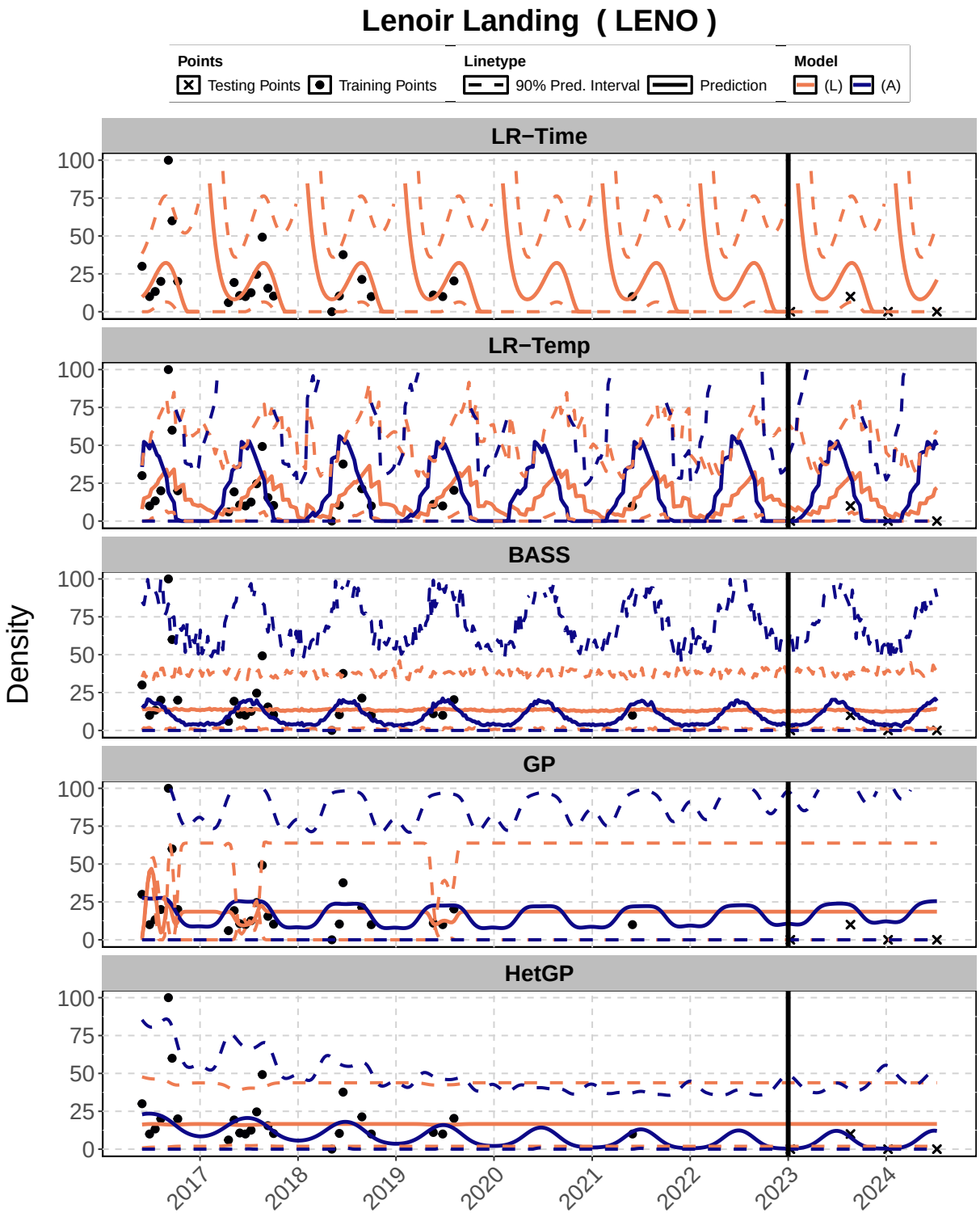


Figure 13: *LR-Time, LR-Temp, BASS, GP*, and *HetGP* – (L) and (A) models at LENO along with 90% prediction intervals on training and testing dataset

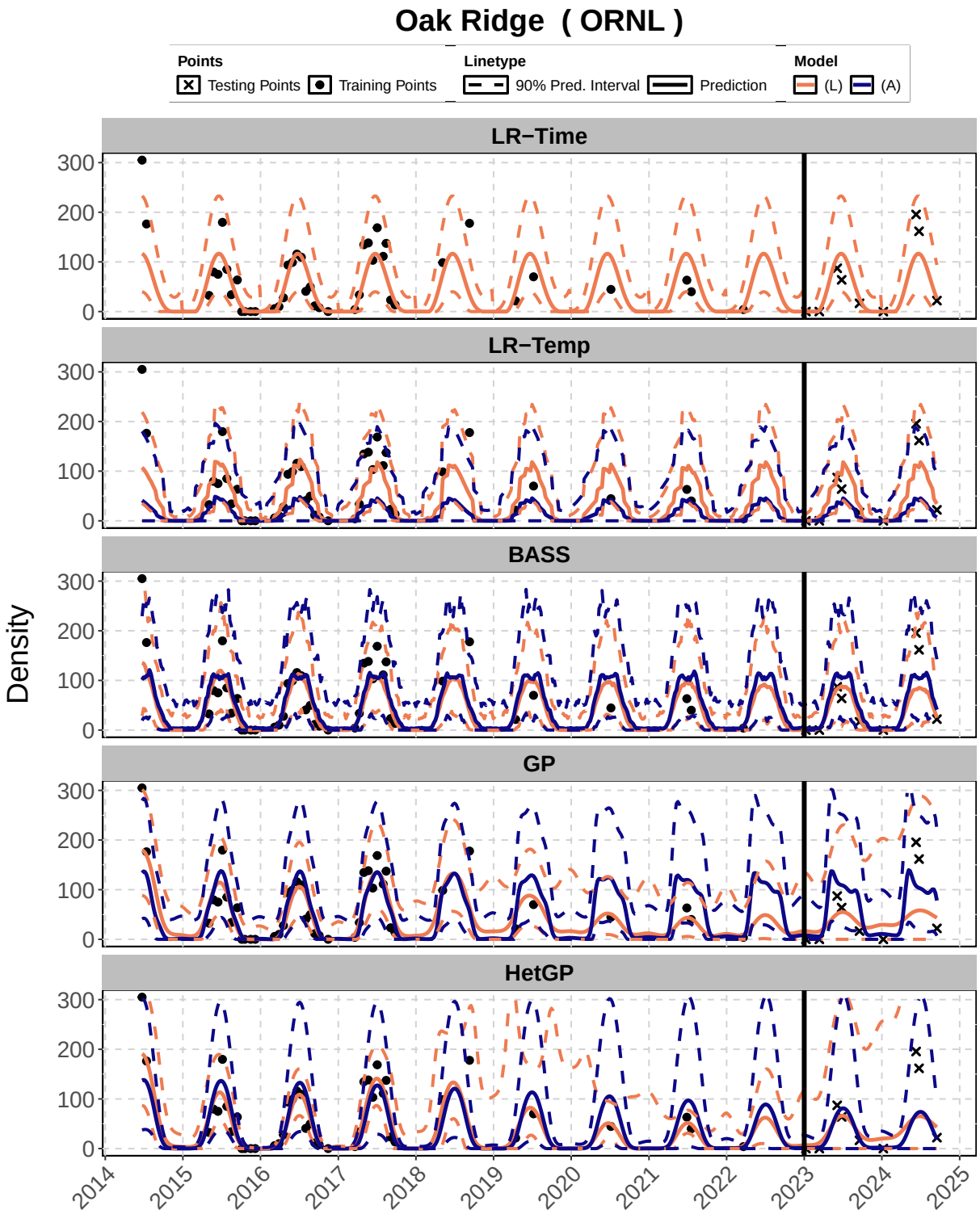


Figure 14: *LR-Time, LR-Temp, BASS, GP*, and *HetGP* – (L) and (A) models at ORNL along with 90% prediction intervals on training and testing dataset

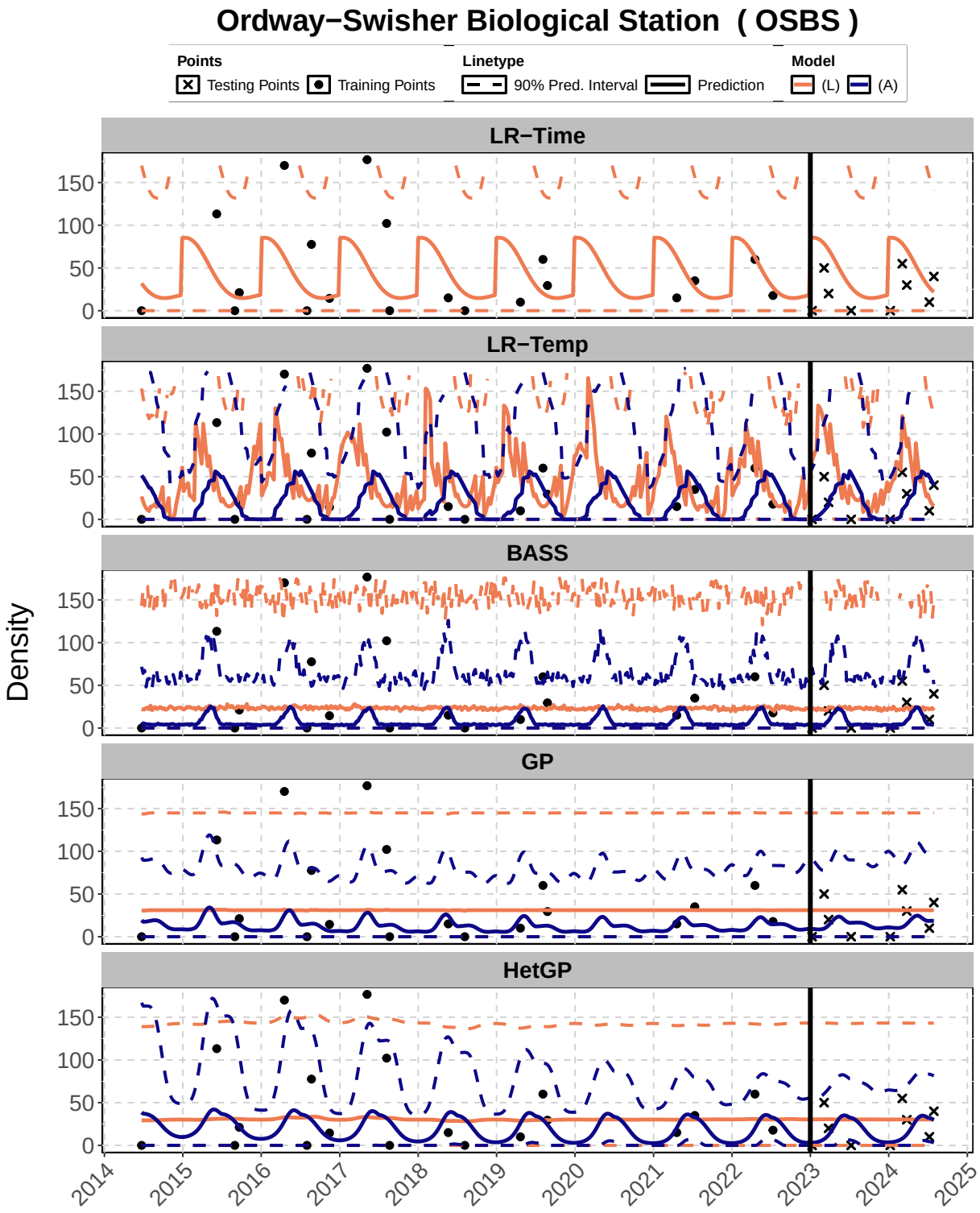


Figure 15: *LR-Time, LR-Temp, BASS, GP*, and *HetGP* – (L) and (A) models at OSBS along with 90% prediction intervals on training and testing dataset

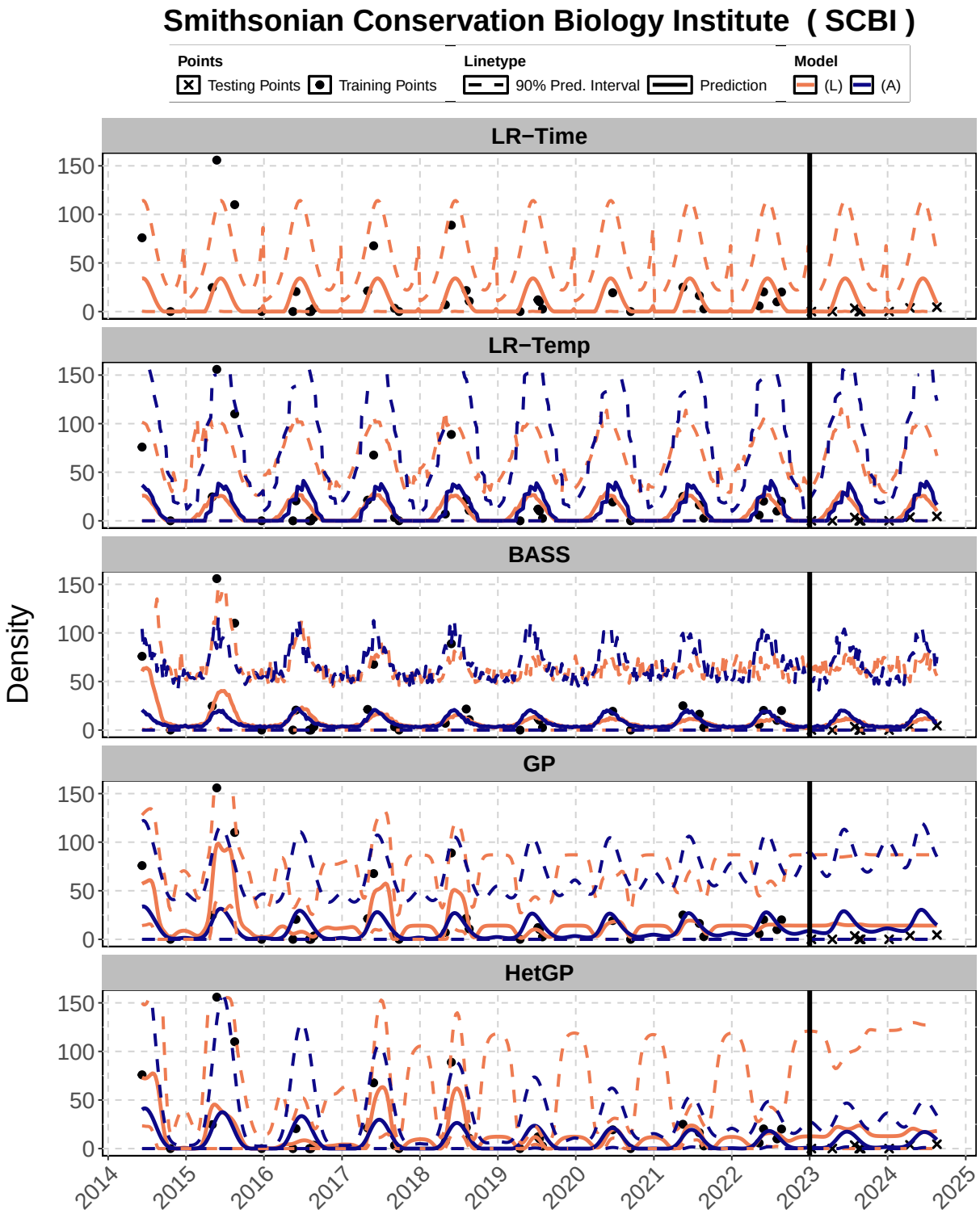


Figure 16: *LR-Time, LR-Temp, BASS, GP*, and *HetGP* – (L) and (A) models at SCBI along with 90% prediction intervals on training and testing dataset

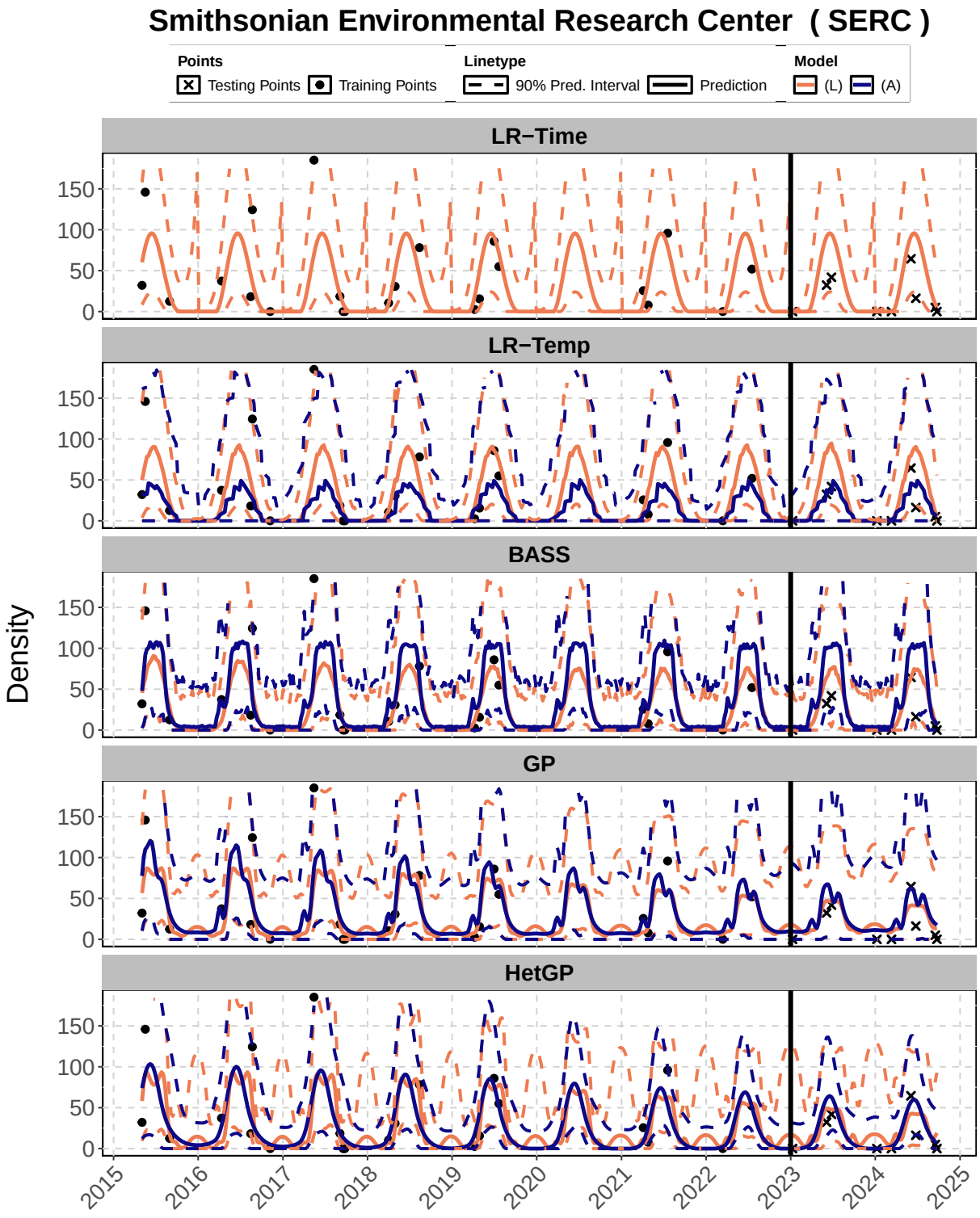


Figure 17: *LR-Time, LR-Temp, BASS, GP*, and *HetGP* – (L) and (A) models at SERC along with 90% prediction intervals on training and testing dataset

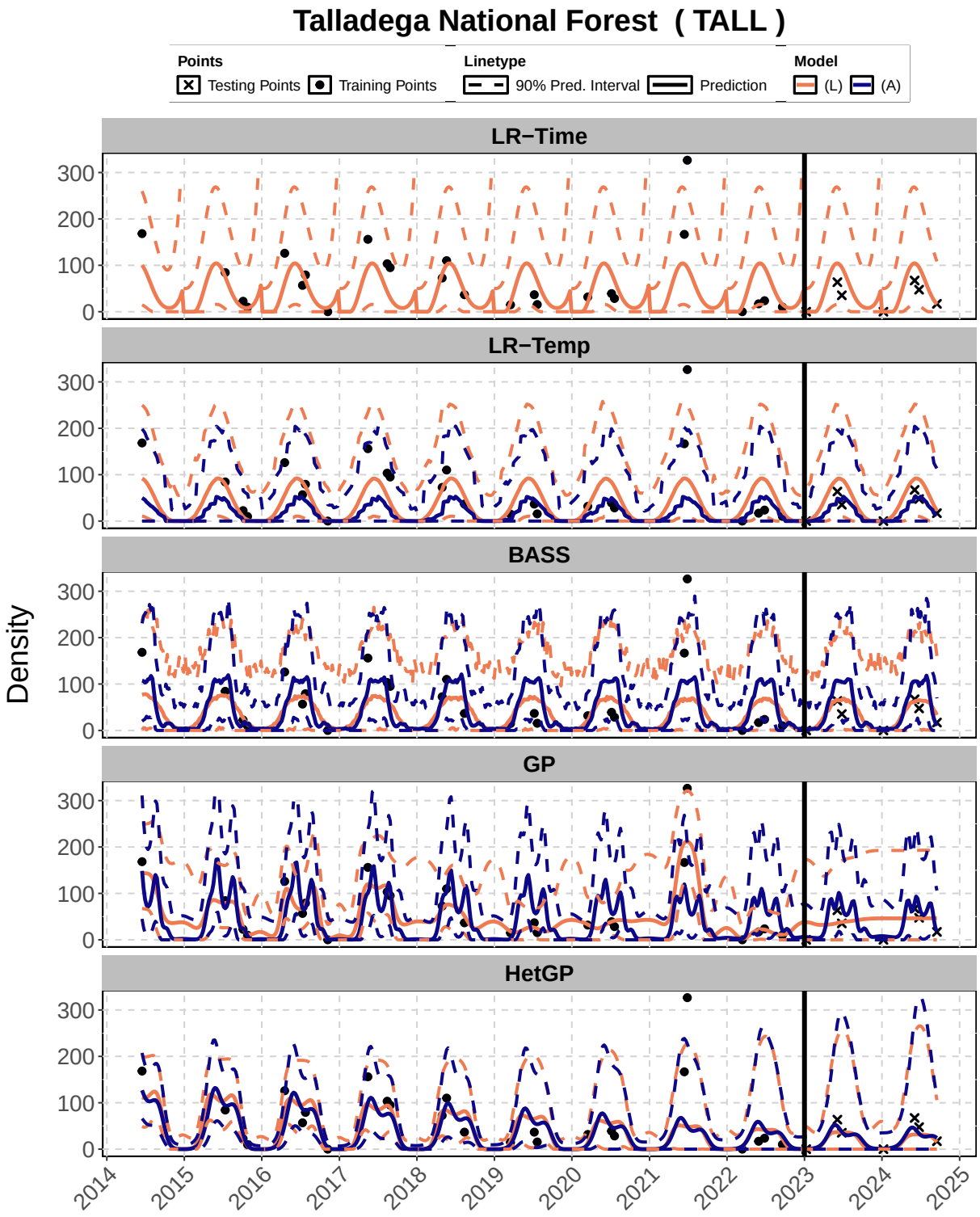


Figure 18: *LR-Time, LR-Temp, BASS, GP*, and *HetGP* – (L) and (A) models at TALL along with 90% prediction intervals on training and testing dataset

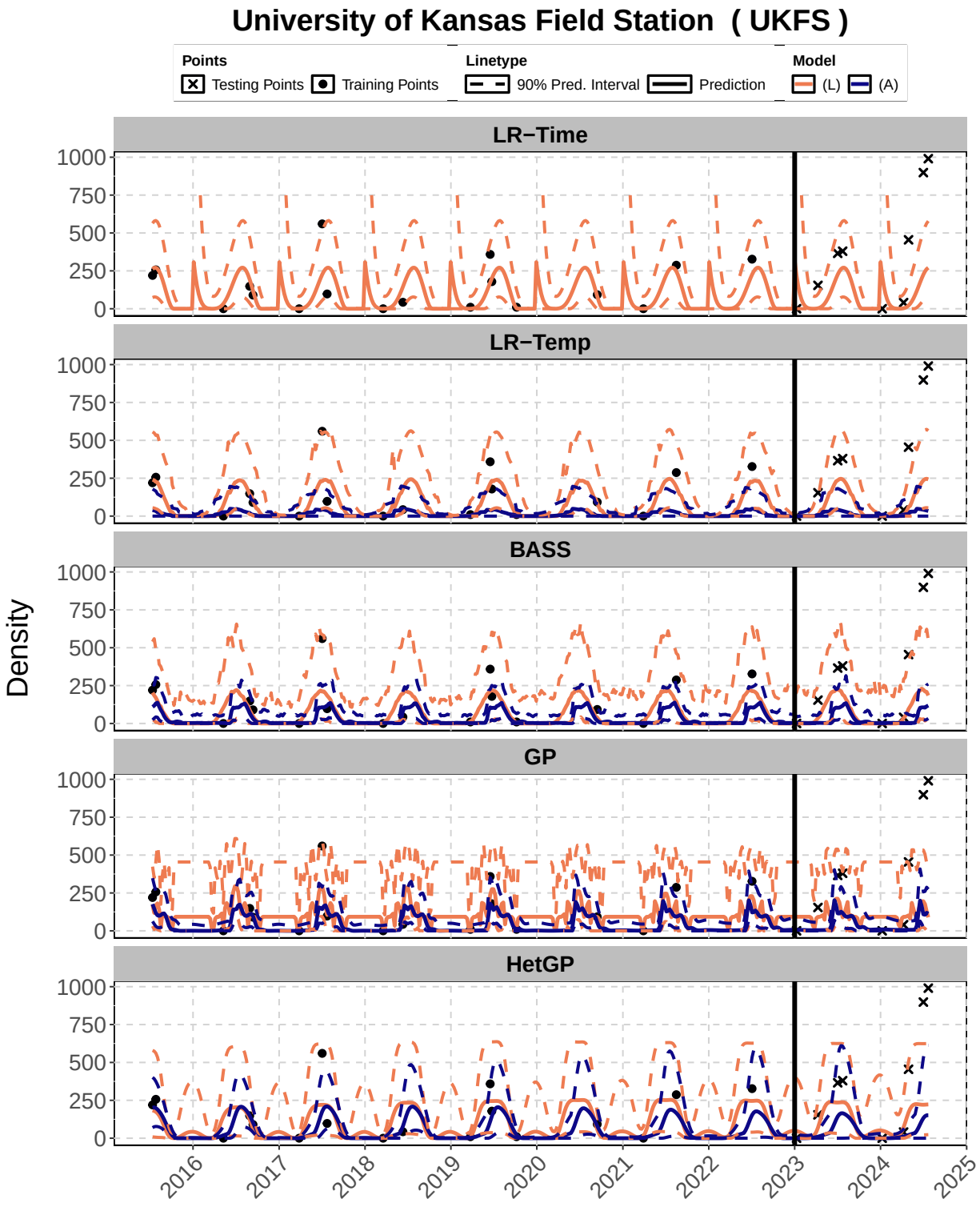


Figure 19: *LR-Time, LR-Temp, BASS, GP*, and *HetGP* – (L) and (A) models at UKFS along with 90% prediction intervals on training and testing dataset