Eva Deegan, Nathan Kroeze, Hayden Gallo, Rachel Badzioch

Dr. Jason McLachlan

Bayesian Statistics and Biological Forecasting

December 16, 2022

Final Project Report

Introduction

Our forecasting group included four members with various interests and levels of experience coding. We were all interested in learning how to forecast to be able to apply it to our own research. For the NEON forecasting challenge, we decided to create a forecast for NEE, as many of the group members were interested in terrestrial ecosystems. We hoped to create a model that would forecast NEE at least 35 days into the future with multiple drivers.

Site Selection

Like most of the class, we began our analysis with the University of Notre Dame's Environmental Research Center, UNDERC. This decision was largely based out of familiarity with the site and cohesion across the class. However, we found in our exploration and analysis of other sites that UNDERC had qualities that would make it suitable for the forecasting challenge. As a group, we were most interested in a NEON site composed of mostly deciduous forest as this was the most familiar and overlapped well with our research interests. One of the downsides that we encountered was the timing of the challenge and the site location we chose. As fall faded and winter approached, it became apparent that carbon flux would not vary much and likely stay around zero. One of the considerations we made when designing our analysis was flexibility in site selection and we were able to create code that runs for all the NEON sites we tested. As we explored other sites, we encountered the same seasonal challenges in other, more southern, deciduous-dominant forests (Great Smoky Mountains) and a greater number of gaps within the available historical data. Due to these challenges from other sites, we determined that UNDERC would be a suitable site for our analysis.

Historical Data

To begin our exploration of the data and understand the possible trends in the data, we plotted the historical data. This involved graphing NEE over time to understand past trends in NEE over the course of seasons and years. Based on Figure 2, it is obvious that there is a cyclical nature to NEE levels. The NEE levels are higher in summer and peaks in the spring and fall, then lowest in the winter where there is almost no net exchange of carbon. The very low carbon exchange rates during the winter are highly evident after examining Figure 1 which shows UNDERC on 12/6/22 where almost all vegetation is dead. These trends are particularly important for later predictions of NEE. Visualizing these cyclic trends were helpful for future analysis and model fitting.

Random Walk Model

After initially plotting NEE over time, a random walk model was fit to the historical data as our first foray into prediction of NEE. Figure 2 depicts this random walk with the historical data shown in black, the random walk in red, and the 95% credible intervals in blue. This random walk was fit over historical data and then extended 35 days into the future. Due to the fact that there was no net exchange during the winter at UNDERC, we decided to perform this same procedure of fitting a random walk at the Great Smoky Mountains. When we fit the random walk to the historical data at the Great Smoky Mountains, we ran into the same problem of very little to no net carbon exchange during the winter at this site as well. Due to the fact that most sites which fit our specific parameters would face the same challenge of little to no exchange during the winter because of lack of vegetation during the winter months, we decided to keep UNDERC as our site of interest. After fitting the random walk to the historical data and predicting 35 days into the future, our next step was to investigate drivers to add to our model.

Driver Data Exploration

We began to explore the NOAA data to look at various possible drivers for the model after brainstorming which variables we thought would be connected with NEE. Some of the variables we discussed were temperature, shortwave radiation, precipitation, and water temperature. We plotted these variables in time series against NEE to see which ones showed correlation, and would be useful as drivers in our model. Only shortwave radiation showed a linear relationship with NEE, so we knew that we wanted to include this variable in our model. Based on our scientific reasoning, we knew that NEE would be connected with phenology and we wanted to include LAI in our model. However, NOAA didn't have LAI predictions into the future like they did with radiation.

Driver Identification and Selection

We chose to use shortwave radiation as our first driver in the model. We used data predicted up to 35 days in the future. For the second driver in our model, we wanted to include a variable representative of phenology. We ended up using green chromatic coordinate (GCC) to represent phenology, though we would have preferred to use LAI. We hope to collaborate in the future with the Phenofreaks group on forecasting LAI to include in our model.

Linear Relationships

The linear relationship between shortwave radiation and NEE made sense based on our scientific understanding. More light coming into the ecosystem would mean that there would be more photosynthesis and carbon stored, which would explain why radiation and NEE were correlated. The GCC and NEE correlation also made sense to us because there are more green leaves based on the time of year; green leaves increase in the summer and decrease in the winter. The summer is when more carbon would be stored which would lead to an increase in NEE .

Forecast Iterations and Improvement

We added our first driver, downwelling shortwave radiation, to a dynamic linear model using the "fit_dlm" function from the "ecoforecastR" package. The output of this version of the model had reduced uncertainty in comparison to the random walk model and showed the forecasted NEE values slightly increasing (carbon being released) for 35 days from the beginning of December to the beginning of January (Figure 6). This made sense because our covariate (shortwave radiation) was forecasted by the NOAA to be low during this time period (winter, low sunlight) and the linear relationship between shortwave radiation and NEE caused the NEE forecasts to increase when radiation was low and decrease when radiation was high. Next we added climatology-predicted GCC as a covariate in the same way with the fit_dlm function. The model that was fit through the ecoforecastR dynamic linear model function is shown in Equation 1, where X is NEE, p_t is the phenology (GCC), and s_t is the shortwave radiation at timestep t. β_0 is the intercept, β_p is the slope of the covariate effect of phenology, β_s is the slope of the covariate effect of shortwave radiation, β_X is the slope of the initial conditions effect, and τ_{add} is the additive process error standard deviation. Our model included observation error in the data model and process error in the process model.

Eq 1: $X_{t+1} \sim N(X_t + \beta_0 + \beta_p p_t + \beta_s s_t + \beta_X X_t, \tau_{add})$

Challenges

The most difficult challenges we faced while working on this project arose because of difficulties applying concepts from the homework exercises to our own model. For example, we struggled to quantify the sources of uncertainty in our model following the “Uncertainty Analysis” exercise. In the exercise, the JAGS model was recreated as an R function so different model inputs (parameters, Monte Carlo ensembles, etc) could be turned “on” and “off” to observe how the output changes, however we were unable to follow along with how the JAGS model (which is very different from our own) was reformatted as an R function and how changing the R function inputs would reflect adding or removing uncertainty in the JAGS model. Similarly, we struggled to apply the two data assimilation exercises to our model since the assignments were very complex. Additionally, the data assimilation models again used R function models that were more complicated than our model, and we were unable to format our model in the same way the assignments did or understand which parts of our model we should be changing to reflect the assignment steps since the models were so different.

Final Forecast

Our forecast using our two-driver model for a 35 day period December to January period is shown in Figure 7, where the credible intervals are narrower than in the random walk and radiation-only models. Since we were forecasting NEE into the winter when daily NEE is steady at just above 0, we could not tell how our model was performing and responding to covariates. Therefore, we plotted some additional forecasts to further test our model. First, we forecasted one year into the future to see NEE trends during non-winter seasons (Figure 8). For this forecast, we used climatology predictions for both of our drivers since we do not have NOAA shortwave radiation forecasts after 35 days. This output, though it did not use updating forecasts of our drivers, showed us that the model did respond to seasonal changes in the covariates and captured the cyclical nature of NEE. We also created a hindcast from 7/1/22 to test the model during the summer (when NEE is the most negative and most variable) by cutting off all observed NEE values after July 1 and letting only the covariates drive the model until 1/1/23 (Figure 9). For the hindcast, we did have NOAA-forecasted shortwave radiation drivers but the GCC driver was still climatology-predicted. When compared to the observed NEE values for this period, our hindcast performed fairly well and predicted the negative NEE throughout the summer which increased during the fall to a near-zero NEE during the winter.

Conclusion

This semester-long endeavor of creating a reproducible and iterative forecast of NEE at one of the NEON sites was both exciting and frustrating, rewarding and challenging. There were many gives and takes with the project that resulted in countless scrapped ideas but many skills, concepts, and abilities learned along the way. In the end, we were able to go from very basic data analysis and visualization of NEE against time and other predictor variables to creating a random walk fitted to the historical data to finally creating a forecast with two drivers that predicts future NEE with relative accuracy. We were able to successfully create a forecast that could predict the cyclical nature of NEE at UNDERC. Overall, this project was a success not only in the sense that we were able to create a reproducible forecast but also that we were able to gain many new skills along the way while troubleshooting many pitfalls and challenges in the process.

Appendix

Figure 1. Photo of UNDERC phenology camera on 12/6/2022. The camera shows winter conditions and indicates extremely low levels of carbon exchange (NEE).

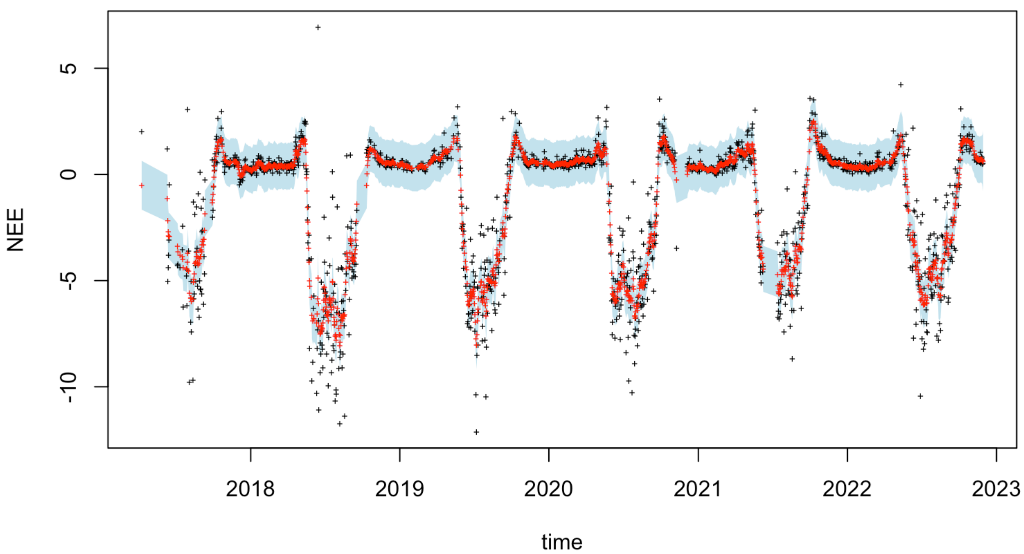

Figure 2. Historical data for NEE at UNDERC NEON site. Trends show a consistent cyclical pattern with low NEE in the winter and high NEE in the summer. Observed NEE is shown in black, random walk modeled NEE is shown in red, and 95% credible intervals are shown in blue.

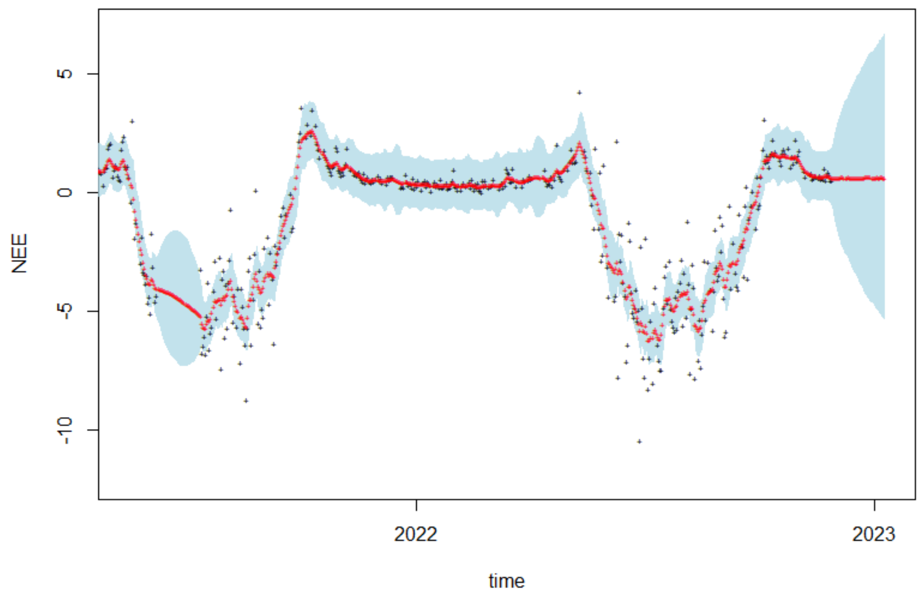

Figure 3. Random walk-modeled NEE forecasted 35 days into the future at UNDERC. Observed NEE is shown in black, modeled NEE is shown in red, and 95% credible intervals are shown in blue. The forecast showed relatively constant NEE in the previous winter which indicated it would be difficult to test the accuracy of our model this winter.

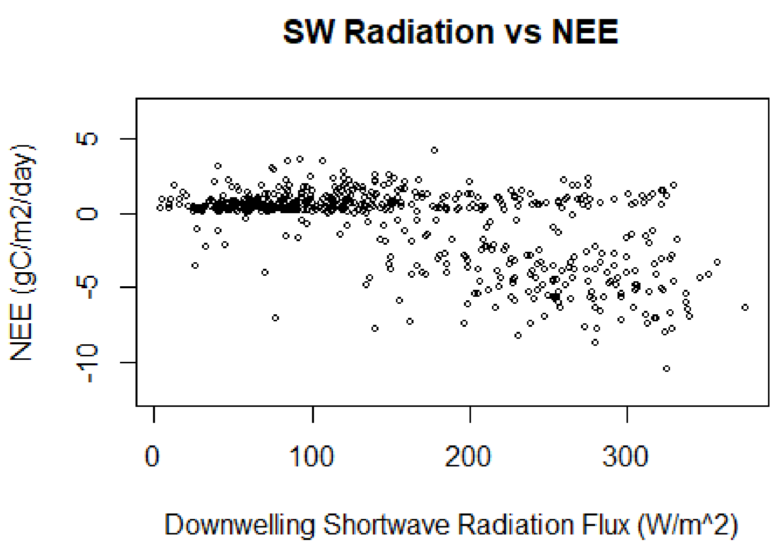

Figure 4. Relationship between NEE and shortwave radiation. Driver exploration showed a negative linear correlation between the two variables.

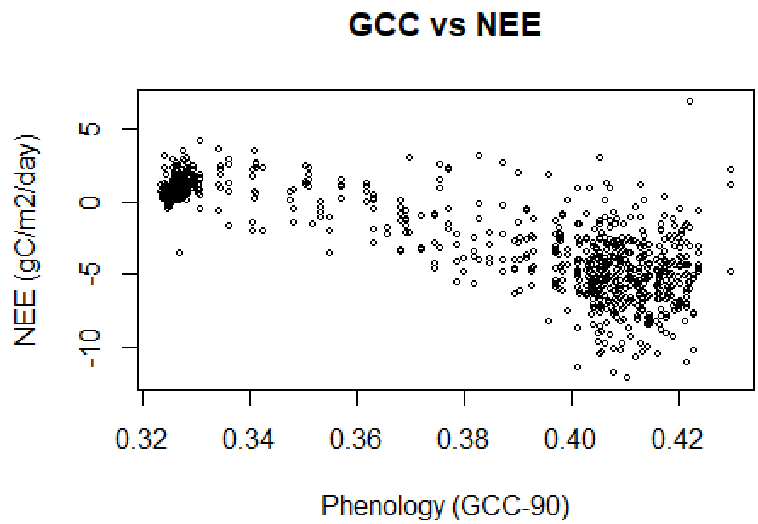

Figure 5. Relationship between Green Chromatic Coordinate (GCC) climatology data and NEE. Driver exploration showed a negative linear correlation between the two variables.

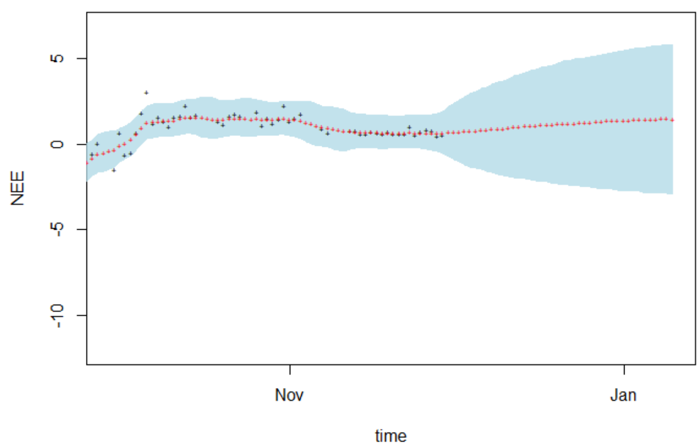

Figure 6. 35 day forecast with shortwave radiation as a driver in a dynamic linear model. Observed NEE is shown in black, modeled NEE is shown in red, and 95% credible intervals are shown in blue.

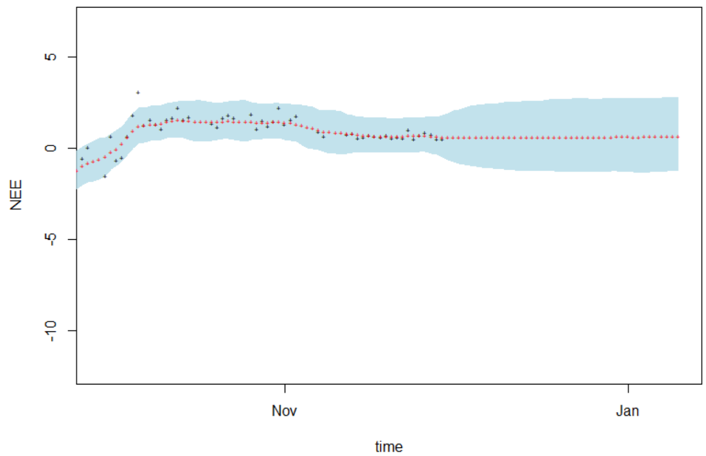

Figure 7. 35 day forecast with shortwave radiation and green chromatic coordinate as drivers in a dynamic linear model. Observed NEE is shown in black, modeled NEE is shown in red, and 95% credible intervals are shown in blue.

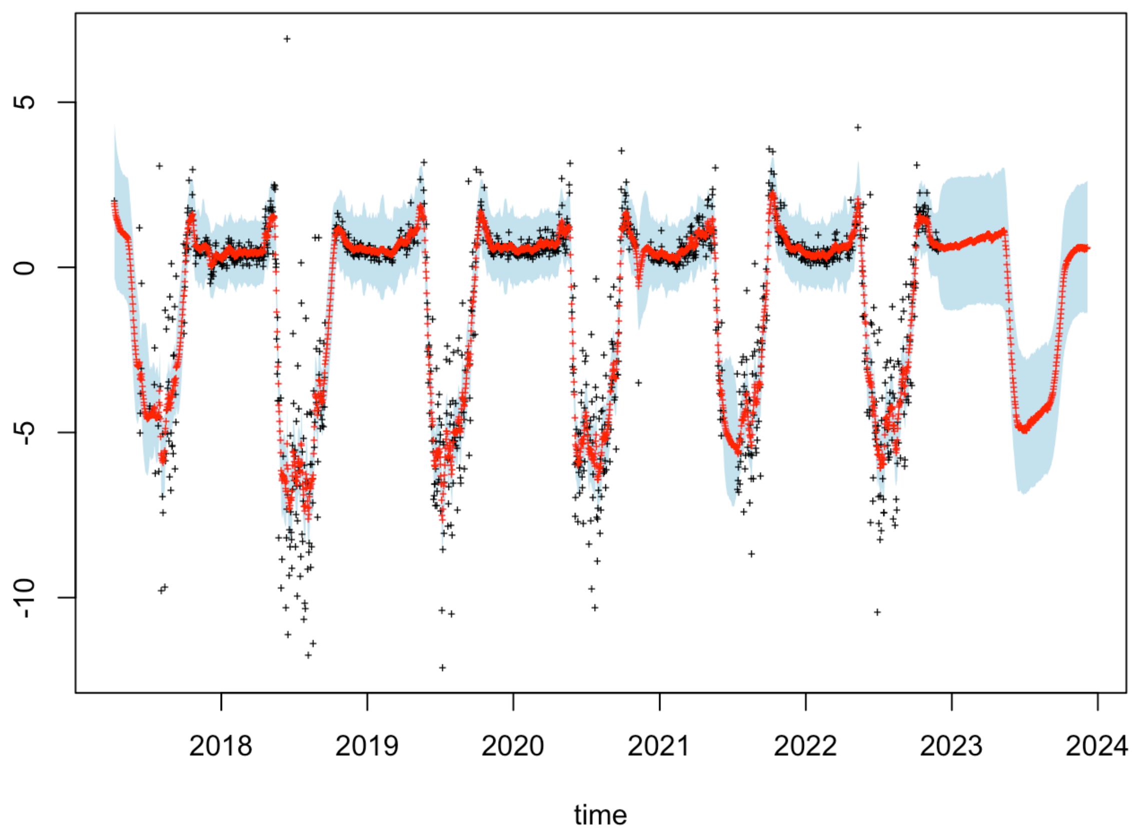

Figure 8. Final forecast 1 year into the future using climatology drivers and data. Observed NEE is shown in black, modeled NEE is shown in red, and 95% credible intervals are shown in blue.

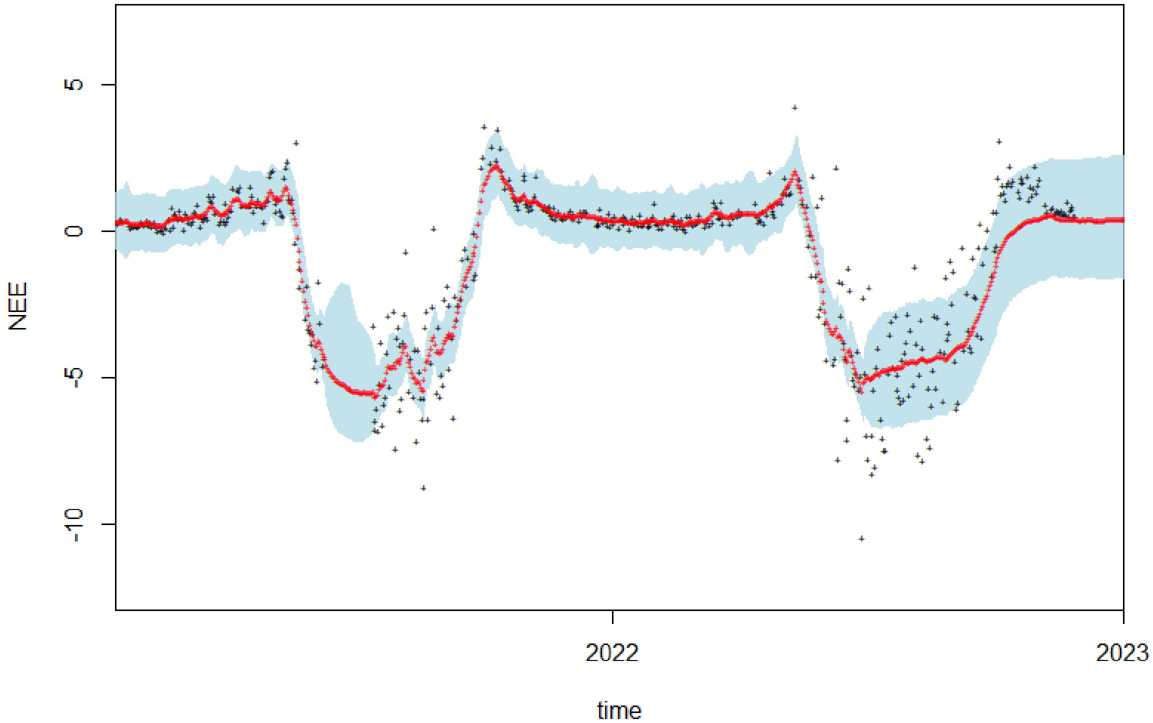

Figure 9. Hindcast of from July 1, 2022 to test the model in the summer. Observed NEE is shown in black, modeled NEE is shown in red, and 95% credible intervals are shown in blue.