Youngjun Woo
UMID: yjwoo
February 24th, 2023

Association between the passage of time and changes in plant habitats

We aim to analyze the association between the passage of time and changes in plant habitats using Plantnet data from Europe. Plantnet data consists of 50 variables related to plants. Among 50 variables, we selected only the following 5 variables, related to species, observation year, and habitats, and proceeded with the analysis.

scientificName	This plant's species name
elevation	Elevation of the place where this plant was observed
decimalLatitude	The latitude of the place where this plant was observed
decimalLongitude	The longitude of the place where this plant was observed
year	The year this plant was observed

Since certain species had too few observations to analyze, only the top 200 species with the most observations were analyzed. Among the top 200 species, the smallest sample size for a species is 2067 observations and the largest 10581 observations. Excluding missing values, the total number of observations for these top 200 species is 720,386.

We used a regression analysis in order to identify habitat changes of each species over time. However, since the habitat information for each species is similar, our data are dependent data. Therefore, when the general regression is used, correlation occurs between residuals of the same species, so the standard deviation of the estimated parameters becomes large, resulting in overconfidence of the predicted slopes. It is possible to think of a method of predicting different regression lines for each species. However, in this case, the regression line of each species has limitations because it cannot refer to any information from other species. On the other hand, multilevel regression can expect a pooling effect because it shares information with other species. Therefore, we proceeded with multilevel regression analysis including species-related random effects.

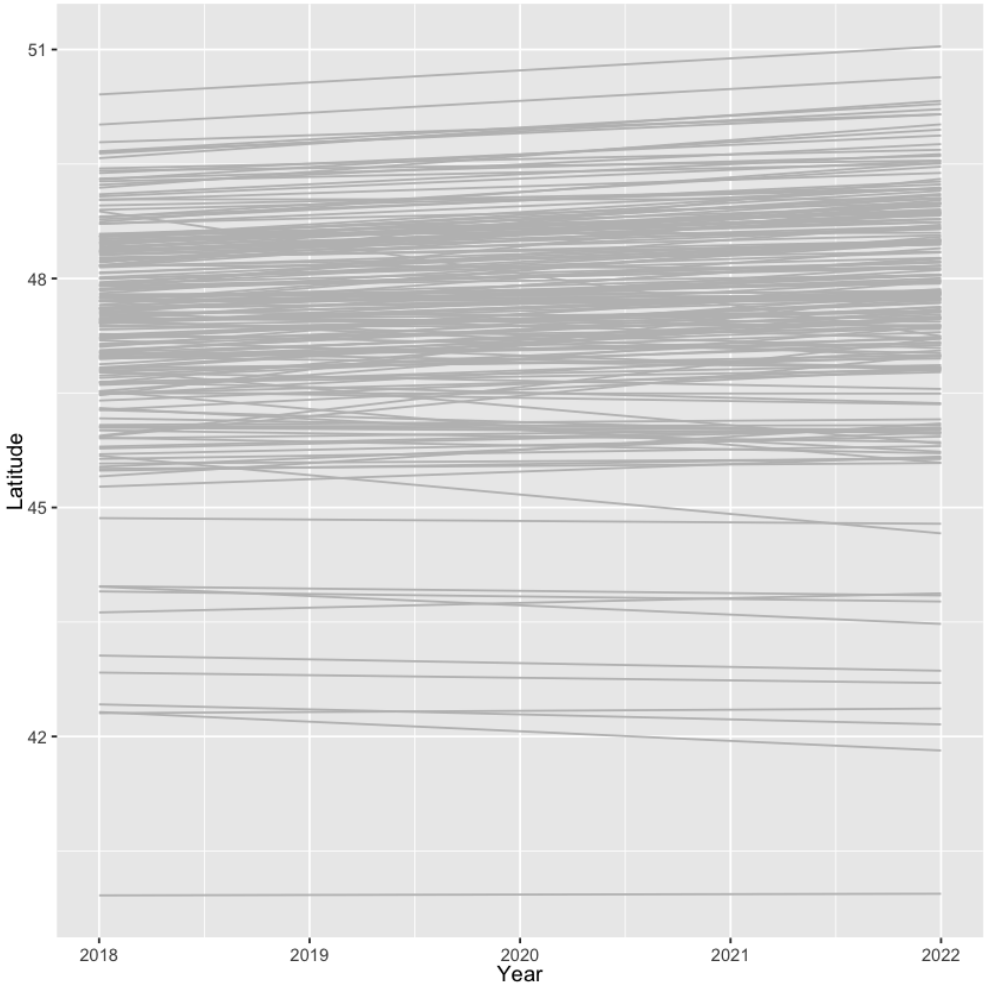
Since we analyze the yearly change of latitude, longitude, and elevation, we set latitude, longitude, and elevation as dependent variables, respectively. In each case, there are 3 combinations that include or exclude random slope and random intercept in the species. Therefore, a total of 3 x 3 = 9 models were analyzed.

Table 1 below shows the AIC of three random effect combinations when latitude is set as a dependent variable. AIC appears the lowest when random intercept and random slope are

included for each species. In other words, since the model with the lowest AIC includes a random slope according to the year, there is a significant difference in latitude change by year for each species. The Plot 1 below shows latitude by year in the predicted model for each species. Most species show a trend of increasing latitude for 5 years.

<Latitude dependent variable>	AIC
Random intercept	3726034
Random slope with year	3857802
Random intercept + random slope with year	3725424

Table 1: AIC results for the models that use the latitude as dependent variables

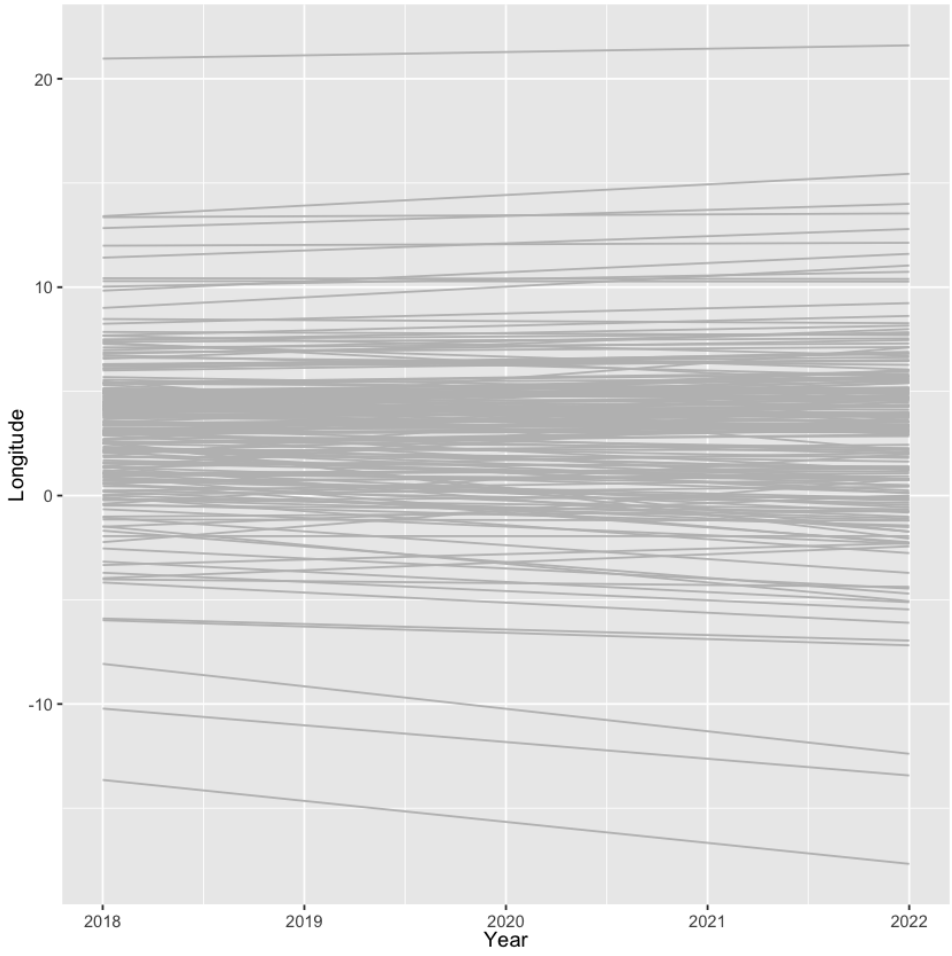


Plot 1: Predicted relationship between latitude and year for each species

Table 2 below shows the AIC of three random effect combinations when longitude is set as a dependent variable. AIC appears the lowest when random intercept and random slope are included for each species. In other words, since the model with the lowest AIC includes a random slope according to the year, there is a significant difference in longitude change by year for each species. The Plot 2 below shows longitude by year in the predicted model for each species. In the case of longitude, there does not appear to be a greater increase or decrease with year than latitude, except for a few species.

<Longitude dependent variable>	AIC
Random intercept	6097733
Random slope with year	6136332
Random intercept + random slope with year	6097463

Table 2: AIC results for the models that use the longitude as dependent variables

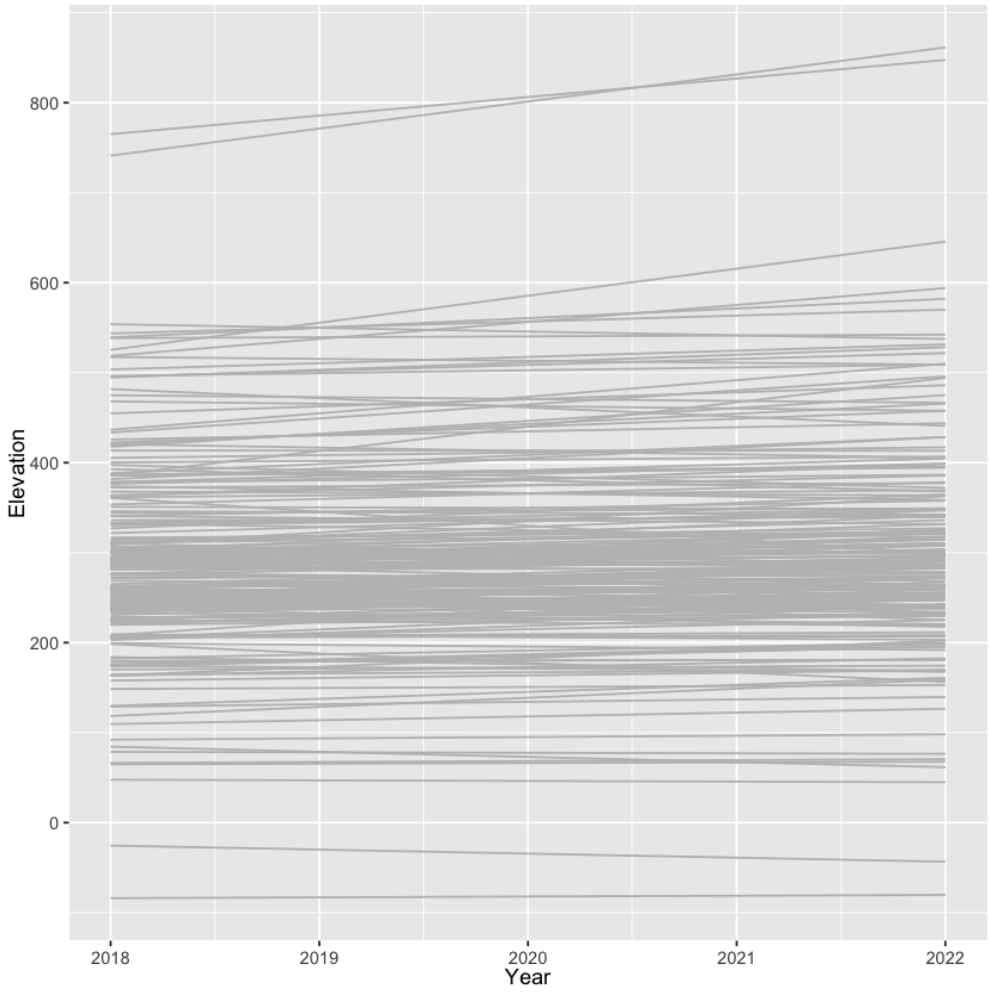


Plot 2: Predicted relationship between longitude and year for each species

Table 3 below shows the AIC of three random effect combinations when elevation is set as a dependent variable. AIC appears the lowest when random intercept and random slope are included for each species. In other words, since the model with the lowest AIC includes a random slope according to the year, there is a significant difference in elevation change by year for each species. The Plot 3 below shows elevation by year in the predicted model for each species. The elevation, unlike the latitude, also does not change significantly over the years.

<Longitude dependent variable>	AIC
Random intercept	10201043
Random slope with year	10289040
Random intercept + random slope with year	10200765

Table 3: AIC results for the models that use the elevation as dependent variables



Plot 3: Predicted relationship between elevation and year for each species