Latitudinal Shifts in Fern Species: Multilevel Regression Analysis Using iNaturalist Data

Vincent Weng

Introduction

This memo investigates whether different fern species are experiencing shifts in their observed latitudinal ranges over time, and whether specific trends at the species level can be detected using multilevel regression and large-scale inference. This analysis draws on data from the iNaturalist platform, a global citizen science initiative where users contribute geotagged and timestamped observations of living organisms. From this large collection, we focus on a taxonomic subset known as Polypodiopsida (a class of ferns)—comprising nearly 965,000 observations. Each entry includes species identification, geographic coordinates, and the date of observation, offering a rich temporal and spatial dataset for analysis. Understanding how the ranges of plant species are shifting is essential for studying the impacts of climate change, as well as for informing conservation strategies in the face of ongoing environmental disruptions.

Methods

We began by loading the Polypodiopsida subset of the iNaturalist dataset, containing nearly one million fern observations with species name, geographic coordinates, and observation date. The elevation field was entirely missing and was dropped. We then filtered the data to include only observations from 2015 onward and created a new time variable, day, representing the number of days since January 1, 2015, scaled in units of 1,000. This transformation facilitated modeling temporal trends in latitude. As an initial step, we conducted large-scale inference analysis by fitting a separate linear regression for each fern species to estimate the relationship between observation latitude and time. Since testing species simultaneously increases the risk of false positives, we applied a false discovery rate (FDR) correction using the Benjamini-Hochberg procedure. In addition, to assess geographic patterns in the strength of these trends, we plotted the Z-scores of the regression slopes against each species' mean latitude. This allowed us to visually examine whether strong latitudinal shifts were concentrated in specific regions or broadly distributed across latitudes. While this approach enables species-level inference, it treats species independently and ignores the hierarchical structure of the data and uneven sample sizes. To address these limitations, we used a multilevel regression with species-level random intercepts and slopes—allowing joint modeling of population and group-level effects with partial pooling.

Since multiple observations are recorded for each species, and those observations are likely to be correlated, a multilevel model allows us to incorporate both fixed effects for time and random effects for species. Specifically, we included random intercepts and slopes for each species to capture species-specific baseline locations and rates of latitudinal change. This approach provides partial pooling, which improves estimation by borrowing strength across species, especially for those with fewer observations. Compared to separate regressions, the multilevel model more effectively handles within-species variation and enables inference at both population and species levels.

Results

As mentioned earlier, to assess species-level evidence for latitudinal shifts over time, we performed a large-scale inference analysis and applied local false discovery rate (FDR) correction. Figure 1 shows that a notable subset of species exhibit very low FDR values, indicating strong evidence of directional change. We further examined the spatial distribution of these effects by plotting the Z-scores of day slopes against each species' mean latitude (Figure 2). This revealed substantial variation in both directions across all latitudes, with no single region dominating the pattern of movement, suggesting that latitudinal shifts are widespread across fern species.

To complement the large-scale inference approach, we fit a multilevel regression model with random intercepts and slopes for each species, modeling latitude as a function of time. The population-level slope for time was significantly positive (coef = 0.327, $p < 0.001$), indicating an overall trend toward higher latitudes over time across species (Figure 3). However, substantial variation existed at the species level. A histogram of the species-specific slopes revealed a roughly symmetric distribution centered around zero, with many species exhibiting either strong positive or strong negative shifts in latitude (Figure 4). This highlights considerable heterogeneity in responses to temporal change. When plotting these species-specific slopes against mean latitude, we observed no dominant regional trend, suggesting that the magnitude and direction of latitudinal shifts vary broadly across the geographic range of ferns (Figure 5). These findings highlight the importance of using multilevel models to uncover both population-level trends and species-level deviations, especially when data availability differs across groups.

Interpretation

The combined results from large-scale inference and multilevel regression suggest that many fern species may be experiencing shifts in their latitudinal distributions over time. The large-scale inference approach identified a subset of species with strong signals of directional movement, and the widespread of significant Z-scores across latitudes hints that these shifts are not confined to any specific region. However, because this method treats species independently and does not account for differences in sample size or shared structure, it offers a limited view of overall trends. The multilevel regression model provides a more nuanced picture by simultaneously estimating population-level and species-specific effects through partial pooling. While the population-level slope was significantly positive—indicating an average poleward trend consistent with ecological expectations under climate change—it remains difficult to determine how consistently this trend applies across species.

The broad distribution of species-level slopes in the multilevel model highlights considerable heterogeneity in fern responses. While some species appear to be shifting northward, others remain stable or even shift southward, raising questions about the underlying drivers of this variation. These differences may reflect ecological traits, environmental pressures, or even noise introduced by uneven observation effort. Moreover, the absence of a clear spatial gradient in the slope estimates suggests that latitude alone may not explain the observed dynamics. While this analysis does not fully resolve the complexity of species-level range shifts, it provides a structured framework for identifying general trends and uncovering where species deviate from them. These results could potentially offer a starting point for further investigation and underscore the value of combining broad inference tools with flexible modeling approaches when addressing large-scale ecological questions.

Figures

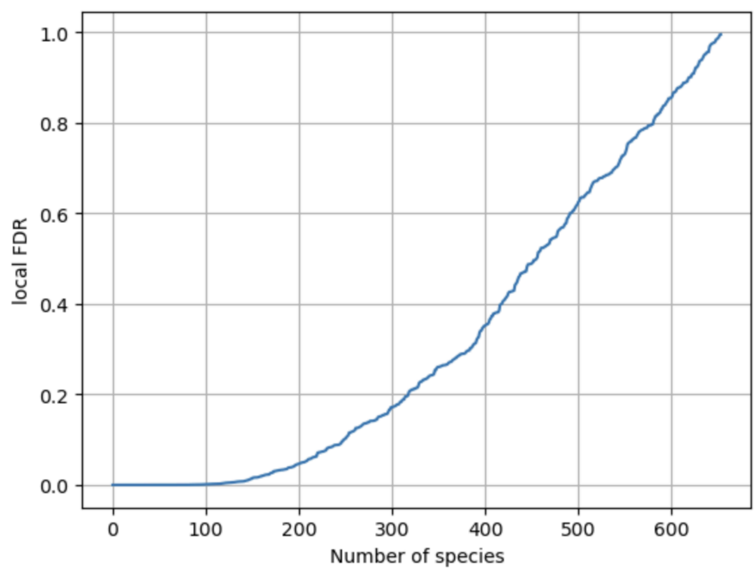

Figure 1: Local False Discovery Rates Across Species

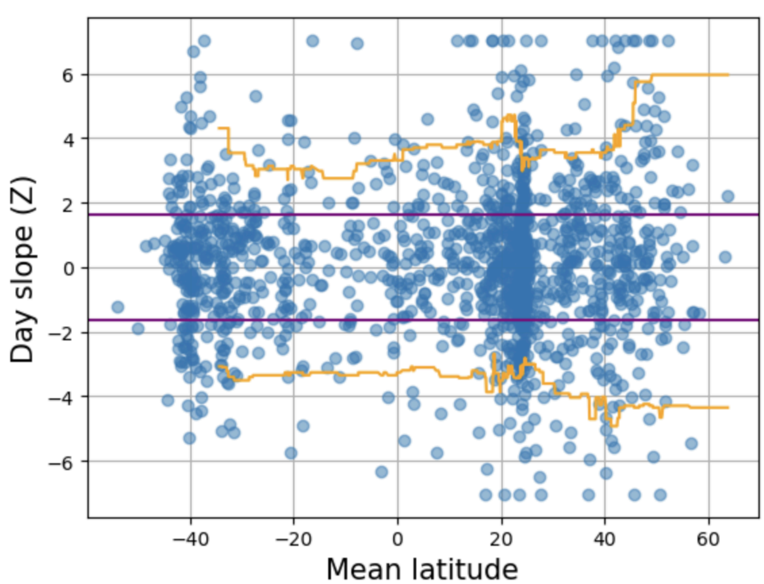

Figure 2: Day Slope Z-scores vs. Mean Latitude Across Species

```
            Mixed Linear Model Regression Results
=============================================================
Model:             MixedLM Dependent Variable: decimalLatitude
No. Observations: 930045  Method:             REML
No. Groups:        3058    Scale:              75.4103
Min. group size:   1       Log-Likelihood:     -3337979.4252
Max. group size:   58798   Converged:          Yes
Mean group size:   304.1
-------------------------------------------------------------
                    Coef.  Std.Err.   z    P>|z| [0.025 0.975]
-------------------------------------------------------------
Intercept           6.694    0.482 13.885 0.000  5.749  7.639
day                 0.327    0.061  5.373 0.000  0.208  0.446
Group Var         646.299    1.998
Group x day Cov    -6.297    0.150
day Var             2.923    0.021
=============================================================
```

Figure 3: Summary of Multilevel Model Fit (Latitude ~ Day)

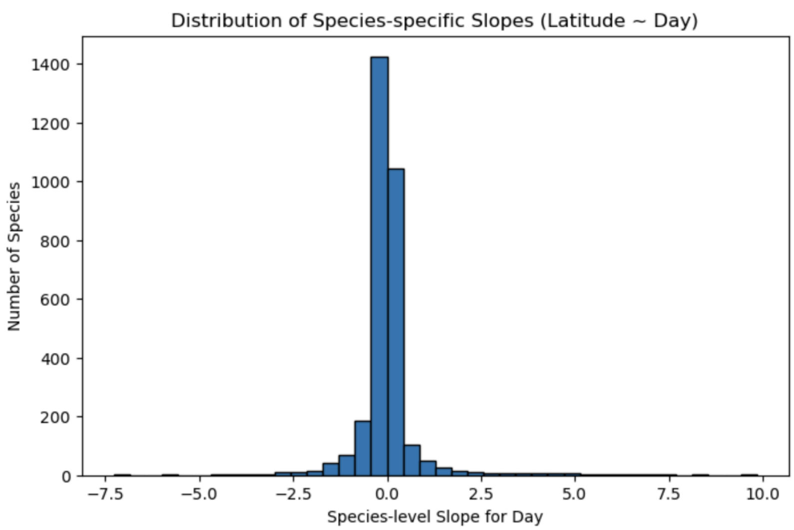

Figure 4: Distribution of Species-Specific Temporal Slopes

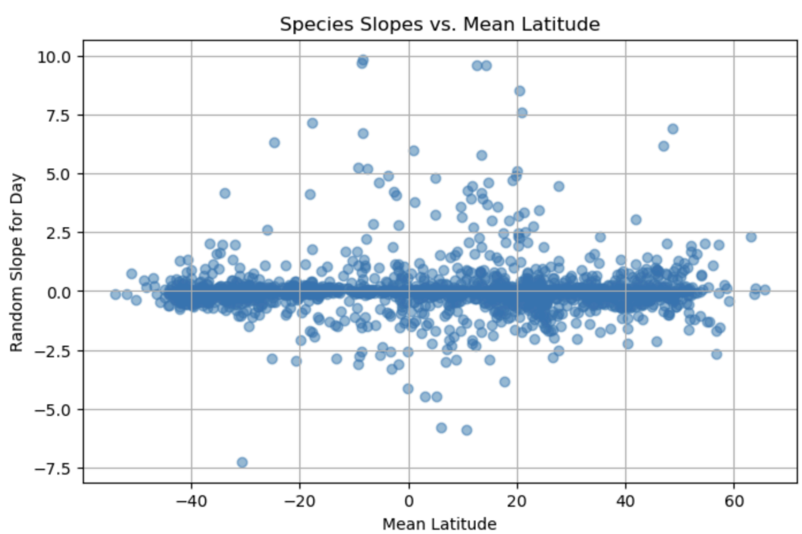

Figure 5: Species-Specific Slopes vs. Mean Latitude

Focus	
Methods	
Writing	
Findings	