Meeting notes for the Across Species paper focused on using SDM methods
Co-lead biweekly meeting → Zoom link; Meeting ID: 956 4808 0311; Passcode: 684940
Upcoming subgroup meetings: meeting slides (all meeting dates)
Manuscript materials; Outline; Workflow; Authorship doc

07/01/2026

Leads finalizing manuscript format, cover letter

Notes:

- In paragraph after questions (intro), change species to individuals to avoid hyperfixation on the definition of high/low latitude species → comment by SR
  - See if it's noticed & if not ask if using species would be a problem
-

06/17/2026 - Subgroup meeting

Goals:

- Get manuscript in general ready for whole paper revisions & send to group by (07/02)
- Aim to submit in August!

To Dos:

- ☑ ~~**Email to group → Liz**~~
  - ☑ ~~**Heads up that we are trying to get manuscript ready for full paper revisions by July 2**~~
  - ☐ **Email to Susan, Chuck, Isaac, and Aaron**
  - ☑ ~~**Ask for recommendations on potential reviewers (need name and institutional email addresses, 5 people)**~~
  - ☑ ~~**Please make sure author info is up to date**~~
- ☐ **Manuscript**
  - ☐ **Rohit & Liz update in text figure references & formatting**
  - ☐ **Include data citations in the reference (phenology & range data)**
    - ☐ **Data availability**
  - ☐ **Abstract → Liz**
  - ☑ ~~**Acknowledgements/data availability/funding/conflicts of interest sections**~~
- ☐ **Other submission materials**
  - ☐ **Cover letter with Qs answered (checklist doc) → Rohit**
- ☐ **Figures**
  - ☑ ~~**Fig 1 suggestions → Hanshi**~~
    - ☑ ~~**Small text → Liz**~~
    - ☑ ~~**Change color combos for raster reference → Liz**~~

    - ☐ **Regression models → Rohit**
        - ☐ **Add data points in the back & let reviewers decide if it's ok**
    - ☑ ~~**Coeff plots → Liz & Rohit**~~
        - ☑ ~~**Give more intuitive names (funct_typeMwoody → Functional Type: Woody)**~~
    - ☐ **Update figure number references → Liz**
- ☐ **Appendix → Saima, Liz, anyone else**
    - ☐ **Update plots in the google folders → Liz**
    - ☐ **Update figure headings and manuscript references → Saima**
    - ☐ **Tables of outputs → Saima**
    - ☐ **Average of results (needs to be averaged) → Liz**
    - ☐ **Average of the RMSE (needs to be averaged) → Liz**
- ☐ **Github**
    - ☐ **Add comment on data doi & citation**
- ☐ **Send manuscript to herbarium collaborators after next meeting → Liz**

Updates:

- Talk about goal date to get tasks finished by
    - Timeframe for leads to get manuscript ready for final revisions
- Checklist doc for submission
- Webinar updates — Sydne
    - Make sure we are clear that we are looking at latitude
    - Cover letter novelty piece: talk about including uncertainty in model (not many studies do that and at this scale and data)
- FIgure 1
    - SR: keep color combos in mind; text is small and make same size as the section headings
- Not including data in but citing source data
- Daijiang traveling the next few months

06/03/2026 – Subgroup meeting

To Dos (*feel free to add name to a task you plan to work on or worked on)*

- ☐ **Manuscript → Everyone!**
    - ☐ **Intro → Everyone**
    - ☐ **Results → Kai, Tong, Sydne**
    - ☐ **Discussion → Everyone**
- ☐ **Other submission materials**
    - ☐ **Cover letter with Qs answered (checklist doc)**
- ☐ **Appendix → Saima, Liz, anyone else**
    - ☐ **Update figure headings and manuscript references → Saima**
    - ☐ **Tables of outputs → Saima**

    - ☐ **Average of results (needs to be averaged) → Liz**
    - ☐ **Average of the RMSE (needs to be averaged) → Liz**
- ☐ **Figures**
    - ☑ ~~**Fig 1 update → Liz**~~
    - ☑ ~~**Update figure text font sizes → Rohit**~~
- ☐ **Scripts**
    - ☐ **Bayes analysis into github/update github → Sydne**
    - ☑ ~~***SDM scripts → Liz**~~
    - ☐ **EML file for data upload → Sydne**
    - ☐ **DOI with Zenodo → Sydne**
    - ☐ **Archiving scripts & Data → Eric**
- ☐ **Zotero → Saima**
    - ☐ **Access to all, but include a comment with the doi**
    - ☐ **Person to update the reference list mark the suggested edit/comment as complete – so we know it's been added**
- ☑ ~~**Outline**~~
    - ☑ ~~**Liz journal paper checklist**~~

Agenda:

- Target Journal: GLobal Change Biology (keep this in mind)
- Methods & Results are updated (with figures)
- Reminder of materials
    - Figures Folder
    - Manuscript folder
    - Appendix
- Some new tasks to be done
    - update appendix materials (with the info that is available
    - average run results for the average run table
    - updating Yi's figure 1 to include the two equations and the hypothesis graphs
- When reading through the manuscript, please keep the

05/20/2026 – Subgroup meeting

Agenda:

To Dos - Next Meeting

- ☐ **Sydne can get bayes analysis into github**
- ☑ ~~**Liz & Rohit Meths/results**~~
- ☐ **Liz SDM script ready**
- ☐ **Sydne & Liz let Eric know about scripts – archiving**
- ☑ ~~**Check-in with Yi on figure 1 reconstruction**~~

- Walk through manuscript

  - Adjustment to …
  - Plots — are there any that should be included or do the present one highlight the story well?
  - Appendix
    - Are there plots we don’t need?
    - To Dos
      - Tables of results from the 9 runs
      - Table of run averages
- EDI → a separate meeting [Eric can help with this]
  - Focus on reproducibility
  - What are the code chunks that are important to this
    - ~peak windows code (low priority)
    - SDM framework
    - MC correlation
    - Regression framework
- GCB
  - Hanshi has published through here
  - Return time was quick – so we will hear back soon
- Figures (slides)
  - Tables
    - run results – include for reviewers to check → Sydne can help
    - Flextable package → Saima
      - Helps format tables
  - Plots
    - Include points underneath the line plots hexagonal density plot) → Sydne
    - Contour plots for the interaction points → Hanshi
    - DL → how are we going to show the slope
  -
- Global Change Biology (checklist)
  - ☐ (file 1) Submission Questions (*In lieu of a cover letter, authors must answer each question below using 50 words or less*)
    - ☐ What scientific question is addressed in this manuscript?
    - ☐ What is/are the key finding(s) that answer this question?
    - ☐ Why is this work important and timely?
    - ☐ Describe how your paper fits within the scope of GCB.
    - ☐ What biological AND global change aspects does it address?
    - ☐ What are the three most recently published papers that are relevant to this manuscript?
    - ☐ Provide the name and institutional email addresses of 5 potential reviewers.
  - ☐ (file 2) Research Article Manuscript
    - ☐ Title
    - ☐ Running Title

        - ☐ List of authors with full names
        - ☐ Institutional affiliations of all authors
        - ☐ Contact Information for corresponding author(s)
        - ☐ Abstract - 300 word limit (no abstract for front material)
        - ☐ Intro
        - ☐ Methods
        - ☐ Results
        - ☐ Discussion
        - ☐ Acknowledgements
        - ☐ See elements for individual paper types, below
        - ☐ Lines numbered – this will help the referees provide feedback
        - ☐ Double spaced
    - ☐ (file 3) Supplemental Materials (Appendix)
    - ☐ Publish data on repository → Eric
        - ☐ EDI?

03/11/2026

Meeting with Rohit, Sydne, Daijiang

11/24/2025 – co-lead meeting

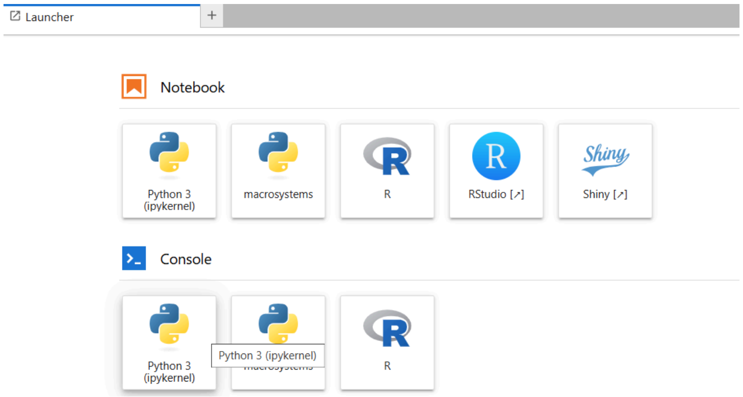
- Submit in February
    - Assign(ed) tasks by December 5th
    - Remember to fill out the authorship agreement doc.
    - Folks finish their tasks by Feb. 6th
        - *Liz send a reminder email in mid-January*
    - Rohit & Liz format manuscript for target Journal (Feb. 9-13th)
    - Last author review period → Feb. 16-27th (2 weeks)
    - Submission → Feb. 27th-March 6th

~~10/01/2025 meeting~~

- Goals:
    - Liz shares some results (hopefully 🫠)
    - Discuss any new trends observed – delegation to rewrite these trends??
    - Choose a target journal
        - Review manuscript format & word count

        - Send to subgroup for edits
-
- Task delegation
    - Conceptual paper → Yi & Hanshi
    - Plot with raw data points
    - Update on the discussion section – based on updated model runs
    - Update the methods section – two model run (nativity & w/o); mention the uncertainty handling; mention the independence between phenology and range
    - Update the intro – if there are big changes to results
    - Write abstract
    - Adjust manuscript format for target journal (citations, cover letter, summary/abstract)
    - Adjust manuscript word count for target journal

09/22/2025

- When using ‘phe-cli’ – mamba install -c conda-forge r-base
- When accessing the `JupyterLabs ESIIL` docker app
    - Run the app & launch
        - 

- Select RStudio – this will take you to a page that breaks; we will bypass it

  a264ef2b6/rstudio/auth-sign-in?appUri=%2F

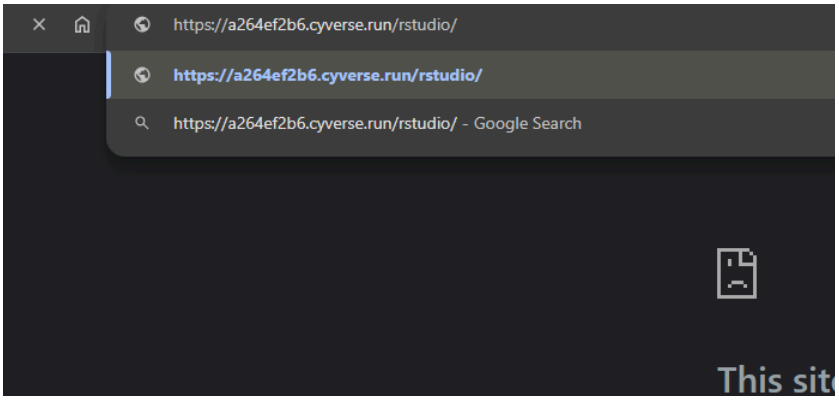
- On the same page, replace the last end of the URL with ‘**.cyverse.run/rstudio/auth-sign-in**’, so the full URL is ‘https://<id>.cyverse.run/rstudio/auth-sign-in’, and click/hit search

  - 

    a264ef2b6.cyverse.run/rstudio/auth-sign-in

- The page will load and stop on a blank screen

- On the same page, replace the end of the URL (again) with ‘**.cyverse.run/rstudio/**’ so the full URL is ‘https://<id>.cyverse.run/rstudio/’, and hit search

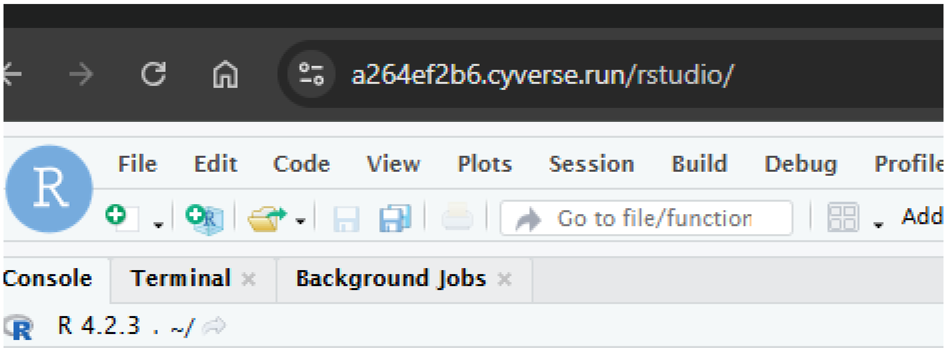
  - 

    https://a264ef2b6.cyverse.run/rstudio/

    https://a264ef2b6.cyverse.run/rstudio/

    https://a264ef2b6.cyverse.run/rstudio/ - Google Search

- Voilà, the RStudio GUI should now appear!



```
 version 4.2.3 (2023-03-15) -- "Shortstop Beagle"
```

09/08/2025 In-person meeting

Liz presentation:
144 species with 50 or more observations for each time period T1 and T2

Daijiang: Even if the specimen says no flower, it's tough to say it's a true absence.
Katie: We circled around that if in a herbarium specimen, if they are collected as non-flowers, then it is hard to assume that there would be flowers.

Tong: Different locations have different springs, how do we define a spring?
Yi: Same months for all locations were considered for the spring months.
Daijiang: As our purpose is to compare between range and phenology, and not to see the effect of change, it hardly matters what month we consider for spring.

Daijiang: What does the value indicate?
Liz: change in prob. Of peak flowering within the time window of three months (March, April and May)

Daijiang: Why not use DOY?
Liz: We had to module the resolution of the time as we did not have exact date of flwoering for most observation.

Tong: Do we have uncertainty into it and is it Bayes?
Sydne: This is random forest (Not Bayes) and we need to propagate uncertainty into it

Liz:’One raster for phenology and one for range change for each species

Sydne: The number of data points behind each lat-band (also a comment from committee)

Tong: What negative range is:
Liz: the change is across time. So negative change value indicate higher probability of range change in historic time period.

Daijiang: We would need a band of latitudes.(plotting histogram?) show the raw data

Sydne: Is it just physics or biology? (question from Liz's defense)
Daijiang:

Tong: Discussion about the negative range change latitudes needed.

Katie: Changing the term “status” to nativity.

Daijiang: invasive vs introduced

Yi: If species can change their phenology more, they can also expand more (depending upon dispersal ability - Sydne)
Second conclusion seems to describe the species' ability to disperse

Tong: We can put dispersal syndrome (the mode of dispersal in species).

Betsy: As we chose invasives so in a sense we are saying they have more dispersability.

Saima: Successful dispersal change depends on the time of arrival of seeds?
Liz: We do not include germination success, but we can find a way to put something about it in the discussion.

07/31/2025 - Co-lead meeting

- To Dos
    - Have Sydne & Rohit review methods by next Weds then send to the group
    - Delegation email → Rohit
    - Results → Liz
    - Discussion first pass
        - Describe results → liz
        - Find support for results → Rohit

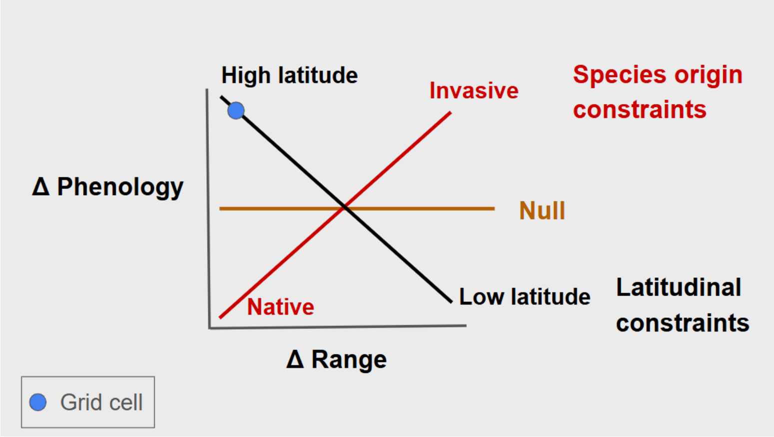
- In-person meeting
    - Think about some stuff that can be delegated to the group

  - Review methods
  - Merge the intros (with the caveat that some details may change depending on the results) → Saima
  - figures/table for the manuscript/appendix (some time allocated during the in-person meeting)
    - Or provide ideas
    - Make a refined figure of the hypothesis → Rohit
    - Please make captions for the table/figure you are working on, so the group can review
  - Results → Liz
  - Discussion → work on during in-person meeting; but will send updates as we work on the first draft
  - Please keep track of your contributions! (Author contribution)
- Think about an agenda for an in-person meeting
  - Monday & Tuesday:
    - 9:30am-10am verbal discussion of results and what they mean; have a target journal
    - Break
    - 10:15am-11:15am: hour of writing (independent or as a group)
  - Weds:
    - 2 hrs – Figures, manuscript revisions etc.,
    - If more days are needed, we will fill up gaps where available

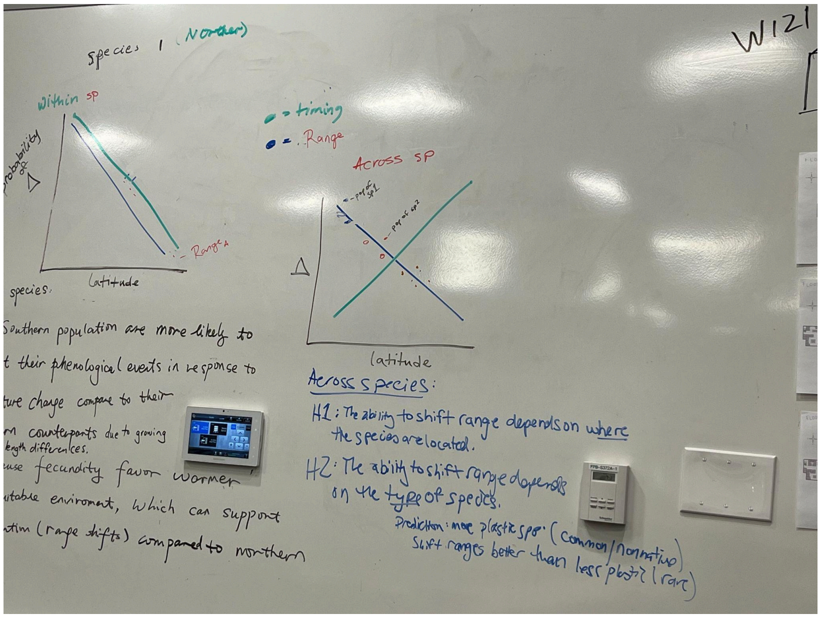
- Liz is running the second model with spatial effects & random species effect
  - Maybe don't need to include simple regression model
  - Put sdm stuff in appendix – see what we should report for SDM

07/17/2025 - Sub-group meeting

- Slides
- Analyses
  - Working conceptual fig



- ○ Some previous drafts



species.

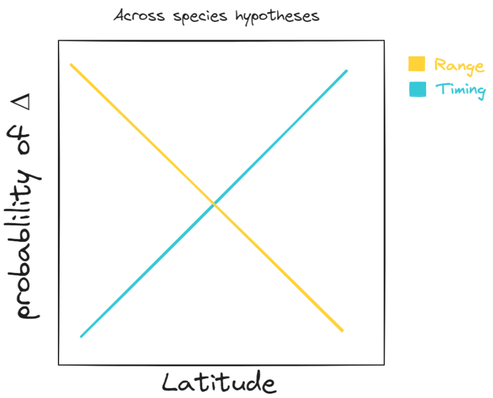
Southern population are more likely to
t their phenological events in response to
ture change compare to their
n counterparts due to growing
ength differences.
use fecundity favor warmer
itable environment, which can support
tion (range shifts) compared to northern

Across species:

H1: The ability to shift range depends on where the species are located.

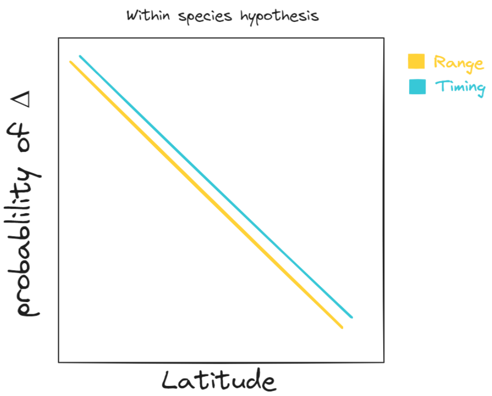
H2: The ability to shift range depends on the type of species.

Prediction: more plastic sp: (common/nonnative) Shift ranges better than less plastic (rare)



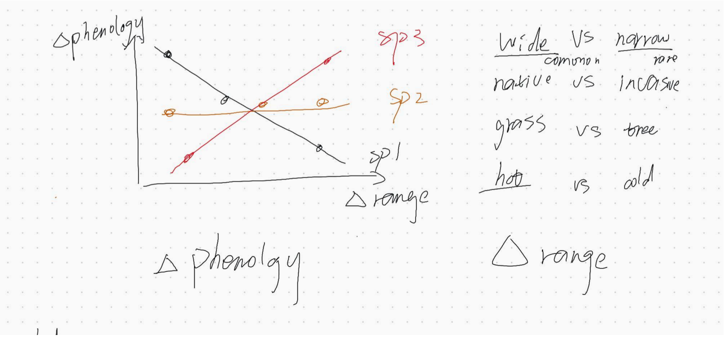
H1: The ability species to shift range depends on where the species are located. Species at low latitude are more likely to shift their range compare to species at higher latitudes. Species at high latitudes are more likely to shift timing of phenological events than species at lower latitudes.

H2: The ability to shift range depends on the relative plasticity of the species. More plastic species are more likely to be able to shift ranges than less plastic species.



H1: Southern populations are more likely to shift their phenological events in response to temperature change compare to northern populations.

1: Due to different length in growing season
2: fecundity favours warmer and stable environments, which support the migration of southern populations than northern populations.



- ○ Difference values are either -1, 0, 1 – not enough for a regression
- ○ How do we want to proceed/conceptualize regression for the values

    - Can’t do continuous change in probability bc then would get negative values which is out of bounds conceptually
- Drafts
    - Intro
    - Methods (regression not yet finished)

Meeting notes:

Liz presentation:

Using SDM output to test the hypotheses

Updated data and data sources

Discussion for regression:

Katie: do we have data to discuss the range?
Liz: We have. They are tied to latitude range

Katie: With 0 and 1, are we going for logistic regression?
Liz: We could.

Kai: They are binary responses. The sdm output will be probability and you have to binarize them. There should be a direct way to directly use probability value.

Liz: We got -ve value which is outside the range of probability.

Sydne: Can we talk more about how to work with prob. Value? There is a paper in the drive on different way to calculate range change.

Saima: what is negative value indicating for phenology and range?
Liz: for phenology it would be like having less flower at that peak. For range it would be something like less suitable habitat.

Sydne: r we overthinking it? We could somehow make it continuous.

Katie: 1, 0, and -1 say us about the directionality of change, but we lose the magnitude of change.

Liz: my committee suggested range and phenology are two different time scale event. One thing we could look at the magnitude of the slope.

Katie: So our range change is now limited to latitude?

Kai: we can look at the preliminary result first.

Sydne: does that mean having all results in terms of 0, 1, and -1.
Katie: yes, to visualize it.

Sydne: we still need to binarize whether to know whether there is presence or absence.

Kai: binary and then using glm would be tricky. If we can do this without bianrizing, it would save us a lot of effort. There might be more direct way to do it.

Kai: if we reframe the question to suitability, we can use the prob. Scale.

Sydne: prob. Are bounded between 0 and 1. If we do difference, we would have value between 1 and -1. Are you thinking about all grid cells?

Kai: for any of the grid cells.

Kai: x-y (one prob. - another prob.) is no longer a prob. Binarization is complicated and when you look at difference between 2, it is continuous. So i dont see need to do binarization.

Liz: after the difference we will have single raster with the differences. And we will still have negative value. (Sydne: yes)

Liz: each point is going to be grid cell. I can work on it and send the email with how it look like.

Should we include phylogenetic part?

Sydne: with alot of species, its wise to use phylogenetic part to see if the changes you see are simply due to their phylogenetic relationship.

Liz: one of the first thing to look at could be across species, have a separate model for change in phenology and change in range.

Kai: start with no random effect, just fixed effect at the beginning. And then start adding RE stepwise.

Liz: summarising: no binarization, start with fixed effect stepwise modeling.

Sydne: We could still have a sub-group meeting before September.
- Think about what times/days we want the group to works on subgroup tasks
- The entirety of in-person meeting is the entire week // travel stipends for ubering
- Some ecocomDP stuff in the works
- Hanashi modeling and will have a meeting with subgroup
- Liz and Rohit think about some stuff that can be delegated to the group
-

07/01/2025 - co-lead meeting

- **To Dos**
    - **Rohit** → refining intro **// biomod & regression (INLA r, brms, bayes GLM or GLMM) papers //** ~~BIEN data uploads~~ // slides for sub-group meeting
    - **Liz** → ~~finish running 6 pseudo-absence species to determine random forest prevalence // binarization threshold: average predictions // run ALL species for random forest only // create mock data of temporal differences for each species~~
    - ~~Email group for meeting w/ slides of updates results~~

- Occurrence data → Rohit uploaded most of it to Cyverse; getting error codes for some files; getting some more BIEN data points
- Phenology SDM → Liz ran all true-absence species and some pseudo-absence species
    -

06/26/2025 - Co-lead meeting

- **ToDos**
    - ~~Biomod papers~~ → Rohit
    - Liz → ~~show an example of disk w/ lowered max dist;~~ run species from the smaller species list & run a chunk of species w/ little true absences
    - Parallelization on Cyverse for data downloads & model running
- Pseudo-absences → choosing random method
    - Rohit: Random makes more sense biologically. random selection provides a relatively unbiased sample of the *available environment*.
      It assumes that the species has access to the full range of environments in the study area, a reasonable assumption for mobile or widely dispersed species.

Random pseudo-absences help when real absences are unavailable (e.g., opportunistic presence data from citizen science). In such contexts, random background sampling allows a fair comparison of presence vs. availability across environmental gradients.

  - Liz: Pseudo-absences can help detect absence in flowering for indv of a species who flowered earlier or will flower later; random clusters observation closer to active observations and disk method makes the assumption of geographic range but heavily focuses on areas that were not samples // random may be better to encompass areas where indv may have already flowered early/late for reasons we were not able to capture in the model (geodiversity//intense climate periods…)

06/19/2025 - Co-lead meeting

To Dos:

- ~~Intro rough draft → Rohit~~
- ~~Pull the occurrence data from BIEN, FIA, and idigbio and put them into one data file with an extra column for data source. Once done, put data files in Data folder → Rohit~~
- ~~Check out maxent in the ensembling for the presence-only → Liz~~
  - Could not get maxent models to run – sticking with pseudo-absences via biomod2
- ~~Email for subgroup meeting & poll (week of July 14th - do not include Wednesday)~~
  - Poll: https://www.when2meet.com/?31043997-MEyIN
  - Goal: talk about SDM results
  - View first draft of intro & methods
  - Regression analyses
  - Target Journals

Objective: Merge intro outlines!

- Fill with papers to include

- [Liz's outline](); [Rohit's outline]()
- [Methods]()
- Liz's role in the subgroup post-summer

06/05/2025 - Co-lead meeting

- To Dos
  - ~~Pull the occurrence data from BIEN, FIA, and idigbio and put them into one data file with an extra column for data source. Once done, put data files in Data folder~~ → Rohit
  - ~~Check out maxent in the ensembling for the presence-only~~ → Liz

    - Writing -
        - Each write an outline & merge → Rohit & Liz (done by next meeting)

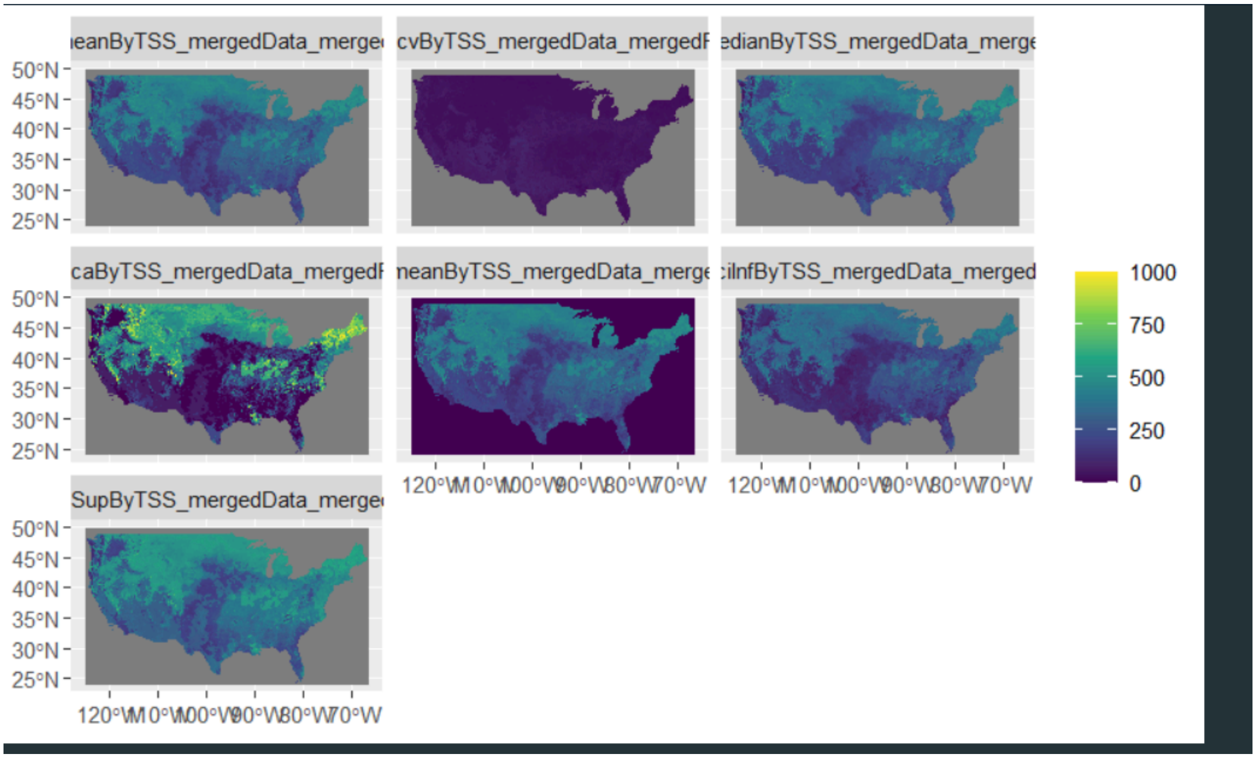
- Liz adjusted species list to include overlap of NPN and Herbarium only (orginially included NEON as a separate entity) → 520 species; 263 when filtering for active observations >= 50
- Katie shared peak flowering code, Liz adjusted to include quantile grouping based on observations → validity in determining varying peaks across species and region, rather than choosing a subjective & static range
- Species list 188 after filtering for peak flower observations & for species with >= 50 active observations; at least 10 native parings
    - Issue: not all species have enough true absence observations;
    - Solution 1: filter for species that have true absence observations >= 10 ; 25 species with not many native pairs (4)
    - Solution 2: presence only method (MaxEnt) on species that are absence-poor
    - Solution 3: use pseudo-absences for the species that are absence poor

04/23/2025 - Co-lead meeting

- To Dos
    - Create biblio google sheet → Rohit
    - Create outline for coding → Liz
- Abstract
- Slides on preliminary results – Liz
- Occurrence – add BIEN (current & historical?) & FIA (current-only, presence only?)
- Goals for virtual meeting
    - Bibliography – lit search for an hour, each person pulls 2-3 papers, put in literature folder, input basic annotations into an excel sheet (title, in-text citation, abstract/summary (can be copy-paste or in your own words), objective(s), main results/takeaways, how it relates to this project (1-2 sentences)
    - Coding workshop
        - Gather & format species occurrence data (BIEN, FIA) with our dataset
        - Spatial-temporal thinning
        - Pseudo-absences

04/16/2025 - Co-lead meeting

- Updates from Liz on prelim models (Acer rubrum) 🎉
    - Was able to run first spatial projection for historical – need to project within species range only
        - Still tweaking & figuring out the labeling



- Abstract writing and planning for virtual meeting

04/09/2025 - Subgroup meeting

- Attendees: Sydne, Eric, Liz, Rohit, kai, Katie, Saima
- **To-Dos:**
    - Coding workflow // break up coding for group to work on
    - rFIA → pull present/absent occurrence points
    - rbien → adapt gbif occurrence range maps to dataset
        - *Can pull species ranges but have yet to find something for species distribution modeling (found this paper but uses MaxEnt)*

- Presentation on progress → Liz y Rohit
- Virtual meeting discussion
- ESIIL had an annual synthesis review – Sydne represented our group

Sydne:
Where are the leftmost points left out for MCP?
For the range map, do we think FIA current range would be best?. This is useful for current data, not necessarily phenology data.
Little tree map

Kai

rFIA package in R will help you access FIA data. Link to rfia

- Use FIA data for current species occurrence (presence absences); will mainly cover native woody species
- Best to get as much data as we can

Sydne - focusing on Yi's project on April 29th meeting and Liz/Rohit will focus on their project on May 5th

Kai: There must be a way gbif give range estimation for species
Sydne: need to consider the historic/current time periods ; but could peek into their approach towards modeling the range

Sydne - rbien package to check out, in reference to Kai's suggestion

Katie: If we can co-develop the workflow before the actual coding.

To Dos: Sydne: Liz & Rohit work on workflow for the project indicating where you are for each part of coding

- Liz break up analysis rmd script so we can divide and conquer

0402/2025 - Co-lead meeting

- **To Dos**
    - **Liz** → compile subgroup meeting materials
    - **Rohit** → ~~note taker~~
    - **Liz** → create range maps for temporal subsets
- Rohit made code for GBIF range estimations // idigbio points (P/A)
- Liz showing progress in maxent modeling so far
    - Compile single raster for each covariate
    - Was using dismo for maxent but not compatible with terra so switching to ENMeval (detailed vignette)
    - Still sticking with the Maxent-Maxlike comparison but learned that you can extract confidence intervals from biomod2 (confusion matrix)
- Look at Rohits range code & subset between time periods
- Virtual meeting agenda for our subgroup
    - If we have results can ask group to compile biobliographies
    - If stuck in analysis can work on that as a group
    - Either way look for literature

3/19/2025 - Co-lead meeting

- To Dos
    - Liz: Will start working on SDM modeling for one species – MaxEnt & MaxLike
        - *Acer rubrum* – to get the workflow set

    - Rohit: Range constraints (historical or present range?) present range should be okay.

- Great sub-group meeting :D
    - Good feedback on species range – use what is commonly done then adjust
    - Adapt Brian’s paper for species occurrence & range – because non-native tree ranges is based on where people decide to plant them and not entirely “naturally” spreading, so informing range based on climatic niche
    - Try both MaxEnt & MaxLike and compare results – if similar then us MaxLike to leverage uncertainty outputs – informs uncertainty in regression modeling for hypothesis
        - NOTE: both software take **presence-only** data, so we would not use the absence information nor need to simulate pseudo-absences
    - Use all species with viable sample sizes then use species pairing as a sub-analysis
- Rohit has quals in mid-April and May — Will be busy around those times
- A new folder for compiling manuscript information
    - Methods doc to start structuring information, fill as we go

03/12/25 - Subgroup meeting

- Slides on progress & future directions
- Attendants: Kai, Eric, Katie, Sydne, Rohit, Liz

Katie: Sample size of two species is concerning.

Rohit: For non-native species, we have substantial data for all four species. We can perform a different set of analyses for non-native species.

Questions to be discussed:

1. Spp occurrence data from same dataset or different source (GBIF) ?

2. How to establish range constraints? (GBIF range estimation, or IUCN)

Sydne:
We were thinking about using Bayes analysis to capture uncertainty.

Kai:
Does that focus more on the prediction part? Machine Learning model far better than traditional statistical method. If it focuses on inference then Bayes can be an approach to consider

We are more interested in the change in range in the first place. That should be first priority. Then focus on delta phenology and then you will be placed well to test your hypotheses.

Sydne
So, using machine learning to estimate the change in range and phenology to fill up the empty grids.
Different ML approaches can be used to estimate this.

Eric:
What is the way to estimate uncertainty? Some R packages do not give a good estimation of uncertainty.

Sydne: Do we need a good estimation of uncertainty? Or we are fine without good uncertainty.

Eric
BIOMOD? We were suggested to use individual modeling approach (because biomod2, at least in R, just uses default tuning parameters for the ML approaches).

Kai: A tutorial about estimation
https://daniel-furman.github.io/Python-species-distribution-modeling/

There should be a way to estimate uncertainty from ML approach

SDM is a tool to augment data from very large sample.

Sydne:
You see a prob of occurrence but not the uncertainty associated with it.

Sydne:
Maxlike? gives a prob. And uncertainty associated with it. That can be an option.

Kai: about occurrence, start with what are the most used approaches for range estimation.

Sydne: Go with maxent and maxlike? and see if maxlike give good estimation of uncertainty.

Liz:
Range estimation and phenology shift estimation

Kai: SDM is tool to augment occurrence data like you have different model for estimate sparsely distributed flowering data.

Katie: historical data for non-native species could be where people decided to put it not where they exactly would be.

Kai: a paper in science about temp niche for trees by Brian McGill

Sydne: Can adapt Brian's range estimations for the range shift portion then implement our phenology portion

Liz: Please fill out the doodle poll for virtual meeting (all time/days that work)

Rohit: package announcement (paper)

03/05/2025 - Co-lead meeting

- ~~ToDos~~
  - ~~Email to Steve (see if he can even help us) → Rohit~~
  - ~~Presentation slides~~
  - ~~Update species spatial/temporal figures → Liz~~

- Merging annotations with occurrence
  - Filter necessary columns
  - Merge by coreid column
- Occurrence pseudo-absences
  - Species range
    - Current species distribution – GBIF ** → ask Steve if he can help with that and range estimation for species
    - BONAP – historical is present
    - IUCN
  - Quick & dirty – use other species locations as pseudo absence that fall into focal species range
  - Using climate data an dan r package
- Next week's agenda
  - Review of processed iDigBio samples (histograms, maps, sample size comparison) // P/A
    - Betsy checking for herbarium overlap
    - Occurrence data from same data or different source (GBIF)
  - Parameter/model selection
  - Discussion of analysis
    - SDM portion (Bayes glm)
    - Regression portion → plyogenetic linear mixed models

02/19/2024 - Co-lead meeting

Action Plans

- ~~Email to group about pivot in phenophase extraction & why, the adjusted temporal extent & species list → Rohit~~
- ~~Pull images & metadata & merge for all species → Rohit~~
- ~~Refine species sample sizes for each time period → Liz~~
    - ~~Filter columns to what we need – keep link to image link!~~
- ~~Process iDigBio in imageant before group meeting → Rohit, Liz~~
    - ~~Get through Acer spp at least – then go as far as we can go (impromptu meeting next week to learn about imageant)~~
- ~~Agenda → Liz~~
    - ~~Present processed iDigBio data and adjusted sample size stats~~
        - ~~Call for help on any remaining species observations~~
    - ~~Conversation of the type of SDM model we should use & P/A-P/O approach~~
    - ~~Conversation about the regression type for the second portion of analysis (e.g. PGLMM)~~

Meeting Notes:

- Rohit will be studying for quals in March!
- SDM thinking
    - Looking into Bayes GLM or MAXLIKE (or something that measures uncertainty for presence-only data) → convo for next subgroup meeting
    - P/A approach: will need to calculate pseudo-absences for herbarium
        - Need to thin herbarium & contemporary samples to match sampling intensities
        - Can use BONAP to provide confined info on where species did not occur
        - Can use a buffer or convex hull around observations
    - P approach: will need a software to quantify uncertainty
        - Need to thin herbarium & contemporary samples to match sampling intensities
- Liz: Checking harmonization code – Yi noticed some weird info for
- Rohit's meeting with Daijiang on phenophase extraction (Phenobase)
    - Low confidence on the low training data – need more data from iDigBio
    - Pivot to leaf machine with Yi and Kai?
        - Make sure we have enough data to give them an estimate of true
    - We can use imageant and go through a semi-manual process
        - Rohit: doable since we have less than 10k samples in total across species
        - Liz: we can focus on Acer spp first to start on a preliminary analysis
- Adjustment of temporal extents and adjusted species list (n<30-50) – [Liz]
    - **1910-1950 & 1985-2025 maximized the most observations per time period**
    - **Focal Genera: *Acer, Rubus, Juglans, Ilex***

- # **Original time period:**
  ## T1: 1910-1950; T2: 2009-2025
  Identify species with n <= 40/50
  T1:
  - Hymenocallis littoralis 8
  - Hymenocallis occidentalis 32
  - Ilex aquifolium 30
  - Juglans regia 27
  T2:
  - Hymenocallis littoralis 2
  - Hymenocallis occidentalis 18
- # **Adjusted time periods:**
  ## T1: 1915-1950; T2: 1990-2025
  Identify species with n <= 40/50
  T1:
  - Hymenocallis littoralis 8
  - Hymenocallis occidentalis 29
  - Ilex aquifolium 30
  - Juglans regia 27
  T2:
  - Hymenocallis littoralis 5
- ## ****T1: 1910-1950 T2: 1985-2025****
  Identify species with n <= 40/50
  T1:
  - Hymenocallis littoralis 8
  - Hymenocallis occidentalis 32
  - Ilex aquifolium 30
  -Juglans regia 27
  T2
  - Hymenocallis littoralis 10

02/12/2025 - Subgroup Meeting

Zoom link, 11am CST/12pm EST, Meeting ID: 662 845 1685

Action plans:

- ☐ Check for overlap in herbarium-iDigBio specimens →
- ☑ ~~Adjust time ranges – 1915-1950, 1990-2025 → Liz~~
  - ☐ Or based on climate data
- ☑ ~~Remove species pairs with n < 40/50 total observations for T_1 & T_2 → Liz~~
- ☑ ~~Yi send over PRISM climate data – for Acer, inquiring about all spp~~
- ☐ Rohit chat with Daijiang about phenophase extraction

Attendees: Daijiang, Sydne, Kai, Betsy, Yi, Katie, Eric, Rohit, Liz

Agenda:

- Short presentation
    - Information on species pairings & climate variables
    - *Review of Goals & Objectives*
- Main asks for the group
    - Help to pull phenological information from iDigBio data (phenophase info in metadata)
    - Source for pulling climate data for temporal anomalies
        - Will need to consider data availability for the early 1910s-1950s
- Spring Goals (workflow)
    - Prepare data for ecocomDP/ESIIL Hackathon?
    - Have a simple SDM ready by April/May (prelim results)
        - Using subset data to start an analysis workflow & prelim results

Meeting notes:

Daijinag: Difference between herbarium and idigbio observations?

Things to do: Check whether there is overlap between observation between herbarium and idigbio

How many observation needed for herbarium: (**lowest sample size: Yi: 40-50 range**

Sydne: small sample size (less than 40ish) species can removed.

Katie: Sticking to the species where we have enough sample size.

Rohit: Increase the time span of historical and current timeline

Betsy: What is the historical period gap is different than from current time gap?

Liz: it happened as we wanted to capture time before major climate change impacts, and the emergence of NPN

Bestsy: broadening the current climate period, and making the time period comparable maybe 20 or 30 years window
If you wanted to standardize the time window to 35 years - you could use 1915-1950 and 1990-2025, for historical and current records
That would be in line with what has been done in the past, in terms of bracketing the period of dramatic climate change

Sydne: If we keep 30 or 40 years window, it will still be good.

Kai: What happens between 1950 and 2009 or are you comparing three time period?

Sydne and Kai: for now we compare two time frame with a gap in between.

Kai: best way to choose time frame is look at climate data.

Liz: Climate covariates

Kai: Chelsa climate data could be good option

Yi: PRISM climate data and I can help with that. – runs further back & still high res

Yi: Leaf Machine2 process: We extract the data on gbif and send them the link.
Paper link: https://bsapubs.onlinelibrary.wiley.com/doi/10.1002/aps3.11548

- Can work with flowers
- Weaver, W. N., & Smith, S. A. (2023). From leaves to labels: Building modular machine learning networks for rapid herbarium specimen analysis with LeafMachine2. Applications in Plant Sciences, 11(5), e11548. https://doi.org/10.1002/aps3.11548

Liz: work on phenophase extraction and getting climate data

Rohit: Will talk with Daijiang if phenobase can help with idigbio observation

Liz: co-leads meet biweekly (link at top of document), let Rohit and Liz know if you'd like to be included in the google calendar event for those; As a subgroup will meet once a month until April (i.e. 2 meetings left) – google calendar invite has been sent out but will send reminder emails beforehand

01/29/2025 - Co-lead meeting

- Zoom link
- Action plans
    - **Rohit:**
        - ~~Send Liz iDigBio species observations~~
        - **Explore rgbif**
    - **Liz**
        - ~~Combine figures into slides~~
        - ~~Add slide for climate covariate discussion~~
- Check-in
- Slides for subgroup meeting
    - Main objectives for the meeting
        - Sharing Species pairings & sample sizes
        - Input/Help with phenophase extraction
        - Input/Help with climate covariate

12/10/2024

- **Action plans**
  - ~~Meet before subgroup meeting to go over presentation slides~~
  - **Liz:**
    - ~~Include plots to Rohit's slides~~
    - ~~Send poll for subgroup meeting~~
    - ~~Make an agenda for subgroup meeting~~
- Rohit's updates:
  - Plots with species records & sample size (native-nonnative pairings)
  - Has idigbio data
  - Stuck: How to pull phenophase information from idigbio specimens
- Liz:
  - Slides on observations across datasets (NEON, NPN, Herbarium)
  - Do we slim the current time period to 2009/2013 (since NEON & NPN start then)
  - Update Rohit's slides with additional plots
- **Agenda/talking points for the next subgroup meeting:**
  - **How to pull phenology information from iDigBio data**
  - **Do we slim the current time period to 2009/2013 (since NEON & NPN start then) – or keep it as is (Sydne recommends we slim)**
    - **Maybe include LTER to fill the temporal gap**
  - **Talk about where to pull climate data for temporal anomalies**
  - **Prep data for ecocomDP/ESIIL Hackathon**
  - **Have a simple SDM ready by May/before in-person meeting? – w/ subset of data**

11/26/2024 - Co-lead meeting

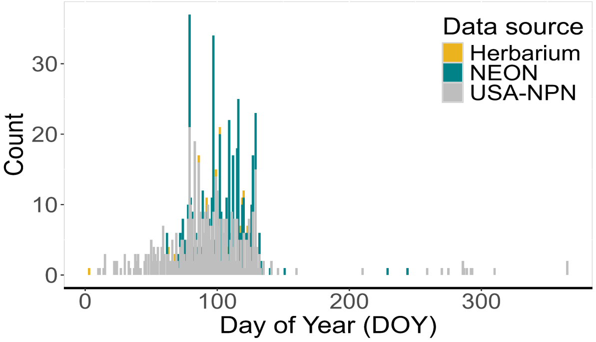
- Action plans
  - Rohit
    - Add our prospective time stamps to the histogram plot for iDigBio data
    - Present plots to next group meeting – Ask Daijiang about image processing
  - Liz
    - ~~Make visuals for NEON/NPN/Herb data (with phenophase info) – send to Rohit~~
    - Work on updated species list – still timing out :(but met with Eric
    - ~~Take notes at next group meeting~~
    - ~~Make a poll for next semester's meetings~~ → https://www.when2meet.com/?27855624-Xgvnj to get an idea of what everyone's month looks like to set biweekly meetings for co-leads and a monthly subgroup meeting
- Updates:

  - Rohit
    - Parsed through the iDigBio data collections based on initial species list
    - Lots of obs in 1950-60s then drops
    - Chunks of observations do not have GPS but they are included in the metadata
    - Still need to pull phenophase information
  - Liz
    - Herbarium dataset citations - (website includes herbarium data platforms & hosts)
      - Park, Isaac et al. (2023). Herbarium-Derived Phenological Data in North America [Dataset]. Dryad. https://doi.org/10.25349/D9WP6S. Accessed February 2023
      - Park, D., A. Williams, E. Law, A. Ellison, and C. Davis. 2023. Assessing Plant Phenological Patterns in the Eastern United States Over the Last 120 Years ver 5. Environmental Data Initiative. https://doi.org/10.6073/pasta/bfb70a1701ef23f686fcc73840e6ae17 (Accessed 2024-09-24).
    - Still working on species list – move through with what we have so far until then
- Pulling GPS & phenophase info from the metadata
  - Can make queries – some info in metadata did not match up with search keys
  - Filtered data that had gps location in metadata with some image
  - Next steps are to pull phenology information from filtered data
  - Ask Daijiang if he can process images
- Move forward with current species list & NEON/NPN data to make simple models

10/12/2024 – Co-lead meeting

- Updates
  - Liz: Still running code for updated species list; working to push another manuscript in submission and will be a bit slower; presented ideas to a group for feedback and Susan Mazer suggested changing tradeoffs to conastraints
  - Rohit: Also working to push publication; pulling phenology data from iDigzbio may be difficult (after a talk with Daijiang); think about how to pull phenology data from iDigBIO
- Bety’s notes
- **Actions plans**
  - **Rohit:**
    - ~~looking to pull phenology data from iDigBio & some exploratory ana~~lysis
    - For the next meeting talk about extracting info from phenology data
  - **Liz**
    - **Update species list**
    - ~~List on herbarium~~

10/29/2024 – Co-lead meeting

- Updates on the Outline
    - Details on how to define the range and how to calc range shift
    - Details on what we are considering in phenology and what phenological events
        - Herbarium the limiting factor – only indicates peak flowering, especially for historical periods
        - Phenological event: peak period for each species
        - How to define a season?
        - [season] *Acer rubrum* active flowering - 2013-2021
        - 

        - [season] Can visually inspect the peaks for different species & identify a general range – give a buffer to account for movement in peak informed by literature
        - Feedback from Liz's presentation during lab group
            - Consider the different rates of warming between latitudes
            - Nest data within ecoregion to account for different physical barriers to range movement and can help with computations
            - Additional grouping (if relevant & enough data) – gradient in origin status: native, non-native, pest, invasive
- Analysis workflow
- **Actions plans**
    - **Rohit** → **i**dentify obs for current species (flowering phenophases/angiosperms no grasses) list from iDigbio
        - Take notes during next week's meeting
    - **Liz** → keep working on the updated species list (all phenophases)
        - ~~Compile conversation topics for next week's group meeting (facilitate updates)~~

09/19/2024 – Sub-group meeting

Zoom, Password: 974342, Meeting ID: 870 2010 4747, 11am EST

1. Review of hypothesis & methods – Liz
2. Assigning tasks/open discussion

3. Authorship guidelines
4. Planning for future meetings

- **Action plans**
    - Liz:
        - Look through species counts in herbarium
        - ~~Make sure Chuck’s data is open (Susan’s/Isaac’s is published) -~~ citations in workflow
    - Rohit:
        - ~~Look through native status~~
        -

- Rohit
    - Shifts in phenology independent of range shift
    - Zooming into the range rather than looking at the range
    - Changes in centroid rather than entire geographic range
- Eric
    - Connect predictions between phenological responses with range shifts (include it as a covariate)
    - Differences between values in pixels
- General interest on the SDM approach
    - Looking at some phenological response in SDM framework
- Eric
    - What will the temporal range look like
    - Mismatch in the temporal aspects — longer time in range & shorter phenophase occurrence
    - Rasterize the output
- Yi
    - Trying to find covariance btw phenology change and climate change — improving response – in Shijia’s paper phenology is a fixed trait of species
- Sydne
    - Referenced screenshot in meeting notes (Kai’s change in latitude to range)
    - How are we defining/quantifying change?
    - Latitudinal bands a bit fuzzy
- Kai
    - SDM can help smooth out noise in the data – value in running models
- Eric
    - Use each pixel as an observation to look at the change in prob change
    - Rather than latitudinal bands, we can do grid cells
    - Latitude as one predictor, or elevation/climate
- Kai
    - If the range defined at the margins then lat bands makes sense
    - Look at SDM literature at how to quantify change over time

    - Biotime?
    - Lots of SDM papers we can pull from
- Sydne
  - Ok with including the entire time period for herbarium and then adding on “earlier” datasets
    - Binning time
- Yi
  - Comparisons between data’s time sensitivity or location has shown to be consistent
- Kai
  - T_1 & T_2 are the outputs of the model rather than the raw data making the datasets comparable & mergeable – already accounting for differences in sampling in the model
- Sydne
  - What species are we using?
  - Annuals
- Kai
  - **Invasive & native species – shows dramatic changes in range**
    - E.g. Cheatgrass – Bethany Bradley
- Sydne:
  - are there comparable native species, use remote sensing or herbarium
- Yi
  - Have a common place for data, for when there is overlap between groups
- Sydne
  - Congregate all data in cyverse
-