NIMBIOS climate-proxies working group November 3-5, 2015

Translate the concentration and relative abundance of pollen spectra from insect-pollinated (entomophilous) taxa found in tropical lake sediments and soil samples into useful, independently validated, reliable paleo-temperature and paleo-precipitation parameters.

Attendees: Mark Bush, Alex Correa, Brian Enquist, Nick Gotelli, Grace Hwang, Crystal McMichael, Crystal, Arzu Unal, Rob van Woesik, Joe Wright

Day 1 (Tuesday Nov 3)

-Welcome. Mark.

Nick Gotelli introduce himself, and 3-4 slides on background

Primary goal

1). Core model (Mark and Rob). Mark introduce. Rob report.

Secondary goals

2). Maxent models for geolocation (Crystal & Brian): Crystal report on MaxEnt

3). System of differential equations (Arzu & Alex). Alex and Arzu report.

4). Simulation experiment (Grace & Rob). Grace report.

5). Pollen representation along environmental gradients: modern + paleo (Mark and Alex). Alex ideas on representation.

Tea/Coffee

6) Modern and fossil-pollen space (Alex and Brian). Alex report.

7) Environmental gradients along elevation gradients: pollen & tree data: Multimodal (Miles & Alex). Alex report.

8). Hyperdominance and pollen representation (Crystal and Miles): Mark and Crystal report on *Iriartea. Submitted to J Ecology*

9) Building priors (Miles, Arzu, Rob). Rob report

Lunch

10) Global Climate Models (GCM) hindcasting and verifying using paleodata (Brian and Mark). Brian.

Tertiary goals

11) Evaluating the assumptions of transfer functions (Alex).Alex report.

12) Sunlight driven responses: corals and tropical rain forests (Joe & Rob). Rob report.

Tea/Coffee

- Consider 12 above; strengths and weaknesses and new directions

Outline immediate objectives for this week.

-Reception for this group would be at 5:00pm on Tuesday Nov. 3.

Dinner: Those who want to meet for dinner (where?)

Day 2

Ideas session: brain-storming on directions, short-term and long-term goals (white board)

Tea/Coffee

- Short-term goals – group work

Team photograph- 11:30

Lunch

- Short-term goals – group work

Tea/Coffee

Short-term goals – group work

Dinner: Those who want to meet for dinner (where?)

Day 3

- Reporting on short-term goals

Tea/Coffee

- Reporting on short-term goals

Lunch

-Round table: allocate tasks for next meeting and long-term goals.

Tea/Coffee

Refined objectives of primary tasks
1). Core model
2). Maxent models for geolocation
3). System of differential equations
4). Simulation experiment
5). Pollen representation along environmental gradients: modern + paleo
6) Modern and fossil-pollen space.
7) Environmental gradients along elevation gradients: pollen & tree data: Multimodal
8). Hyperdominance and pollen representation
9) Building priors.
10) Global Climate Models (GCM) hindcasting and verifying using paleodata.
11) Evaluating the assumptions of transfer functions.
12) Sunlight driven responses: corals and tropical rain forests.
Others

Dinner: Those who want to meet for dinner (where?)

End of workshop

Overview of the first NIMBIOS meeting of the Neotropical pollen-climate transfer working group

(April 20-24, 2015)

Attending: Mark Bush, Alex Correa, Brian Enquist, Grace Hwang, Crystal McMichael, Miles Silman, Arzu Ünal, Robert van Woesik, Joe Wright

We are striving toward constructing a universal 'core' model that will serve to provide transfer functions for statistics on past temperature and precipitation from fossil pollen data. While mean annual temperature or precipitation may be the most useful for the climatic community, the model will probably be built around more directly important biological variables, such as maximum and minimum temperature, length of dry season, interannual climatic variability, and vapor pressure deficits. If we are successful in generating transfer functions for these parameters the mean annual statistics should be resolvable.

In this first meeting we discussed the major biological problems at hand, explored the data sources that are available to tackle these problems, and outlined the analytical means to solve them, including a preliminary derivation of the potential components of our 'core' model. It is not immediately clear which of several possible paths will be the most effective in reaching our goal, and so we will take several approaches in parallel, being mindful that we could merge these approaches in near future.

Near the end of the first week we realized that our tasks can be organized in three tiers, with the first tier focusing on the long-term goal of constructing the core model, the secondary tier focusing on components that are necessary for the core model, and the tertiary goals focusing on less essential elements of the core model that eventuated from group discussions. The following outlines some ideas that emerged from the first meeting.

Primary goals

1). Core model

Rationale: Direct empirical measurements of the climate on Earth are limited to approximately the last one hundred years. For this reason, long-term proxies will be required to assess earlier climate fluctuations. Ancient pollen preserved in lake sediments provides one such long-term proxy and has been widely used to infer past climates in temperate and boreal systems. Transfer functions link long-term proxy measures of climate with real populations. To date, the development of transfer functions has focused on temperate systems. Pollen dispersal mechanisms and the sheer diversity of tropical systems offer different challenges and opportunities to that of temperate systems. Furthermore, the biotically important climate parameters that shape tropical ecology may be different to those of temperate regions.

Objectives: Take advantage of improved pollen identification, newly-available-large spatial datasets, and hierarchical Bayesian framework modeling to produce the next generation transfer functions that will translate Neotropical pollen records into reliable paleo-temperature and paleo-precipitation parameters.

Description: Examine functional relationships between pollen assemblages and environmental parameters in the Neotropics. Use pollen records from 12 (potentially more) lake cores across the Neotropics extending across the last 18,000 years. Each core will be considered at time slices of a duration that balances dating resolution and climate homogeneity (e.g. nine sections of 2,000 year-time slices). The central idea is to link the contemporary and paleo- pollen signatures as a system of transfer functions to estimate paleoclimates. Hierarchical Bayesian models will be explored. The spatial cores can be considered as a random variable and the time slices can considered as a fixed variable in the model. For example, $\theta_{i,t} = \alpha + f(X_{i,t}) + \ldots f(N_{i,t}) + \nu_i + u_t + \theta_{i,t-1}$, where θ is the pollen signal, X to N are the predictor variables for core i, ν is the random spatial effect, *u* is the fixed time effect, and $\theta_{i,t-1}$ is the temporal autocorrelation term of the time slices. We will also explore other options, for example with the spatial and temporal error terms fitted conditionally for each core and each time slice using values of its neighbors. The input parameters for the core model are under development in the appendix, at the end of this document. In addition to this approach we will explore machine learning approaches to transfer functions.

Secondary goals

2). Maxent models for geolocation (Crystal & Brian):

Rationale: If pollen is to be used to determine climate, a quick test is to determine if locations with a given set of climates can be predicted based on pollen assemblages. This will be an exercise in predicting the co-occurrence of taxa found in the modern pollen dataset.

Objectives: To map the co-occurrence of modern pollen assemblages from three locations across the Neotropics.

Description: The modern pollen assemblages from 3 lake sites were modeled at the genus level, and joint probabilities were calculated to determine the likely geographic origins of those samples. These preliminary analyses will be re-analyzed with a geographic spread that spans the Neotropics. Maxent model outputs using uncleaned GBIF data will be compared with models ran using the cleaned and scrubbed BIEN dataset. We will also explore comparing Maxent outputs with convex hull models, to assign weightings to our confidences in Maxent outputs.

3). System of differential equations (Arzu & Alex)

Rationale: Complex relationships between pollen abundance and environmental variables have been explored as… yet there is no universal equation that can be utilized....

Objectives: Determine whether we can identify a useful universal function that allows us to fit taxa specific occurrences across environmental gradients.

Description: The relative changes in vegetation types will be examined as the proportional contribution of different modern pollen in lake deposits along a temperature and precipitation gradient in the Yucatan peninsula. These relationships will be examined to determine whether they could be used to hindcast vegetational changes. The relationships were set up as a system of differential equations: $dN/dT = a.cos(w.T) + b + cNP$, and $dP/dT = \ldots ?$. For example the temperature and precipitation effects on the proportional representation of genera *Acacia, Begonia,* Amaranthacea, *Sapium* and *Quercus* were considered as a system of differential equations. The state variables of the model are N(T) and P(T) representing the percentage of pollen (N) and precipitation (P) at temperature (T). The (possible) system of differential equations, or alternatively a system of stochastic differential equations, will be investigated as universal functions to predict the rate of change of pollen and precipitation with respect to Temperature. The coefficients of the functions may provide information to elicit informative prior distributions in the core model.

4). Simulation experiment (Grace & Rob):

Rationale: Several species distribution models (SDM) exist that are being used to predict the distribution of species in a rapidly changing climate, yet, we don't know how these models respond to different scales of data resolution. Understanding the strengths of the models is necessary as we consider hindcasting and predicting past climates and other applications such as geographic provenance.

Objectives: We will examine different models, using a dataset defined at different resolutions, to determine which models are applicable under different 'grain sizes'. We then compared the results of our range models to in situ field data, and computed probability of detection (pD) and the probability of false alarms (pFA) ranging from [], and [], respectively.

Description: We explore the applicability of existing SDM to modern pollen assemblages of neotropical flora. We found the area under the curve (AUC) for respective models to be []. We conclude that SDM could be applied to law enforcement applications with appropriate priors (?).

5). Pollen representation along environmental gradients: modern + paleo (Mark and Alex, Mevin)

Rationale: Modern pollen data derived from different habitats should reflect habitat, gross geographic, and climatic differences. This would be the first broad scale analysis of >500 modern pollen samples with the capacity to detect regional and environmental trends in pollen taxa.

Objectives: To provide a spatial analysis of modern pollen data derived from the Neotropics that will highlight climatic patterns within the data.

Description: Modern pollen data will be ordinated to reveal spatial patterning. techniques to be explored include partial redundancy analysis that would remove geographic trends prior to determining relationships with temperature, precipitation and vapor pressure deficits.

6) Modern and fossil-pollen space (Alex and Brian)

Rationale:

Objectives:

Description:Fossil pollen and tree plot data in taxonomic space/trait space, presence/absence, and % and collapsed tree plot data analysis. We will use 510 vegetation plots distributed along the Americas to analyze ecological and taxonomic space reflected by them. Ordinations using presence-absence and abundance data will be explored to identify the relationships of these plots with modern pollen assemblages. Once the ordinations have been validated, fossil samples from different time periods will be projected onto the ordination to explore the degree of analogy of modern and fossil vegetation assemblages. Fossil samples will be from Marine Isotopic Stage 5e, Last Glacial Maximum, deglaciation and Holocene thermal maximum. Three basic questions will be answered: i) what is the relationship between vegetation plots and modern pollen spectra in terms of the ecological and taxonomic space they represent?; ii) Did extreme climates of the past produce no-modern analog vegetation assemblages; and iii) Do rates of ecological change through the late Quaternary match rates of climate change?

7) Environmental gradients along elevation gradients: pollen & tree data: Multimodal (Miles & Alex)

Rationale:

Objectives: To look at the correspondence between vegetation data and pollen data in well characterized vegetation communities along elevation gradients.

Description:

8). Hyperdominance and pollen representation (Crystal and Miles):

Rationale: Reconstructing past climates will depend on the ability to detect changes in the pollen assemblages through time, which were driven by changes in the abundance of plants surrounding the site of reconstruction.

Objectives: To correlate the abundance of genera found in the ATDN tree plot dataset (covering Amazonia) with the percentages of pollen from those same genera found in the modern pollen dataset. We will also determine if the genus-level abundance curves from the plot data follow a gradient of pollination mechanisms.

Description: We will take the most common genera and families found in the Amazon (from ter Steege et al 2013), and compare those data with the relative abundances of pollen types that appear in Amazonian pollen records. We expect that the abundance/percentage of pollen types will be positively correlated with the abundances of the most common genera in Amazonia (from plot data). We also predict that the most common genera found in the modern pollen dataset and tree plot dataset will exhibit anemophilous pollination mechanisms. Rarer taxa are expected to exhibit entomophilous or zoophilous pollination mechanisms.

9) Building priors (Mevin, Kiona …Miles, Arzu, Rob...)

Rationale:

Objectives:

Description:

10) Global Climate Models (GCM) hindcasting and verifying using paleodata (Brian and Mark)

Rationale: The accuracy of GCMs is known to be influenced by assumptions regarding atmospheric and oceanic states, and an inability to capture complex topography. A common consequence is for GCMs to underpredict cooling and overpredict drying relative to indicator-species based pollen-climate reconstructions, which have their own problems of over and under-prediction. Multiproxy paleoecological data, including data derived from pollen-based climate transfer functions and isotopic data from cave calcite may provide a better comparator for GCM outputs, that can then help to inform the next generation of GCMs, ultimately leading to more credible climate reconstructions for South America.

Objectives: To provide improved temperature and precipitation estimates of past conditions in South America through informed hindcasting.

Description: We will assess the coherence of different ways to assess paleoclimate in South America. Taking selected time slices and locations, the GCMs will be downsampled to provide climatic estimates. Those data will be contrasted with data derived from fossil pollen and cave calcite data.

Tertiary goals

11) Evaluating the assumptions of transfer functions (Alex):

Rationale: The use of transfer functions allows predictions of … . These functions however rely on a suite of assumptions, several of which are problematic and need revisiting.

Objectives: Reconsider or repent

Description: Transfer functions rely on four basic assumptions that are independent of the used method: i) biological uniformitarianism holds, ii) relationships between pollen and the environment are consistent and systematic through time and environmental settings; iii) the environmental variable that is being reconstructed is the main driver of pollen assemblages through time; and iv) methodological assumptions of the chosen methods are met. We will address each one of those assumptions using modern and fossil pollen from the Yucatan Peninsula.

12) Sunlight driven responses: corals and tropical rain forests (Joe & Rob)

Rationale: Climate variability provides the proximate cues that initiate reproduction and alters resource levels that limit reproduction in many species. For both reasons, climate change is likely to alter reproductive schedules and effort. These outcomes are well documented at higher latitudes, but remain unexplored for tropical forests and coral reefs.

*Objectives:*Review the proximate and ultimate controls of the timing reproduction for trees in tropical forests and corals in tropical reefs and explore the potential consequences of climate change.

Description: The numbers of propagules produced, the distances they disperse and the timing of their arrival all have important implications for intra- and interspecific interactions, species composition and potentially biodiversity. The flowering of trees in the humid tropics and the spawning of reef corals are both highly influenced by solar insolation cycles. Although past climates has forced convergence of these important drivers, contemporary changes in climate are likely to cause these reproductive processes to diverge on the land and in the sea.

APPENDIX. Parameters for the core model and general to do list.

Contemporary parameters

Q1: Climatology. What are we going to use for modern climatologies?

How accurate/valuable are NASA products? In general, what are the climatologies we are going to use, what are important indicators of performance (how do we assess that? Do we take what other people have done based on more limited climatological data, or do we do some kinds of new assessments ourselves? For example, a more nuanced view of variability, or adding things like VPD, etc.?). What is the mechanism for making the determination? I can talk with the NASA folks, but beyond that, how do we choose?

Q2: What are we going to use for presence data?

1. BIEN 3
2. Additional Data from plots (see below)

Q3: What are we going to use for plot data?

1. Go through BIEN and see what we have.
2. Find other plots/networks to fill in gaps. (ATDN, RAINFOR, ABERG, “AnTDN”, Madre de Dios plots).

Q4: What are we going to use for paleoclimatologies?

1. Model type and collaborators
2. Time slices
3. How to deal with discrepancies among models (e.g. model averaging, ensembles, etc.)