What can pollen tell us about palæo-climate?

Jonathan Rougier

Department of Mathematics
University of Bristol, UK

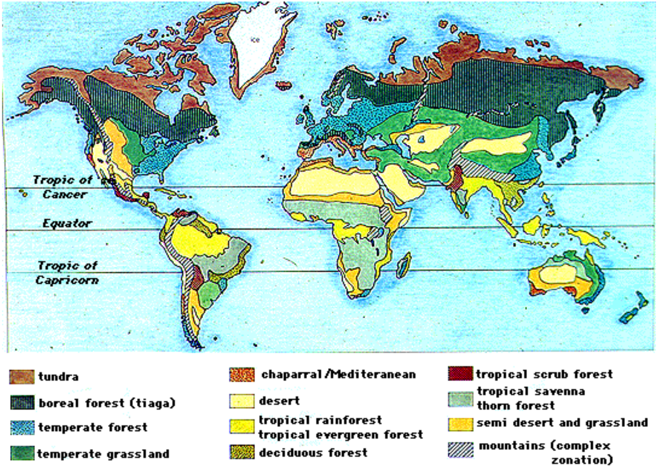
ICAS, Leed, Oct 2010

Obligatory picture of biomes

Source: `http://www.ucmp.berkeley.edu/exhibits/biomes/index.php`

Why reconstruct biomes?

Why reconstruct biomes?



Source: `http://www.life.illinois.edu/bio100/lectures/s97lects/04Ecosystems/BiomeMap.gif`

Why reconstruct biomes?

1. Biomes provide a discrete but rich summary of climate, because plants tend to occupy different subregions of (MTCO, GDD5, precip)-space.

Why reconstruct biomes?

1. Biomes provide a discrete but rich summary of climate, because plants tend to occupy different subregions of (MTCO, GDD5, precip)-space.

2. Reconstructed biomes (including measures of uncertainty) allow us to critique the performance of climate simulators ‘out of sample’, especially the new generation which will be run with dynamical vegetation models.

3. They may also allow us to calibrate climate simulators and vegetation models (i.e. to constrain the values of the uncertain parameters).

Why reconstruct biomes?

1. Biomes provide a discrete but rich summary of climate, because plants tend to occupy different subregions of (MTCO, GDD5, precip)-space.

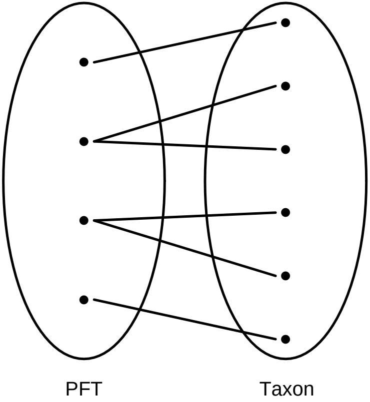
2. Reconstructed biomes (including measures of uncertainty) allow us to critique the performance of climate simulators 'out of sample', especially the new generation which will be run with dynamical vegetation models.

3. They may also allow us to calibrate climate simulators and vegetation models (i.e. to constrain the values of the uncertain parameters).

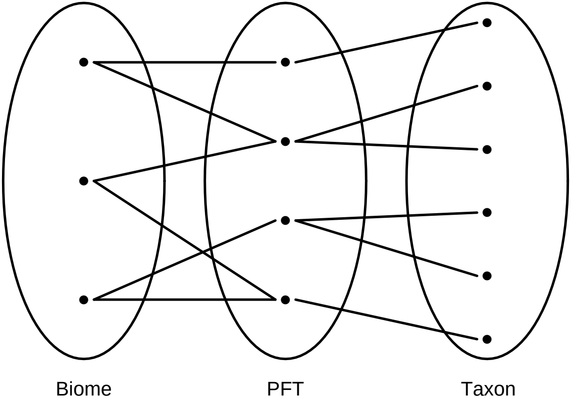
But mainly because . . .

4. It's interesting and challenging!

Biomes, PFTs, and pollen taxa



Biomes, PFTs, and pollen taxa



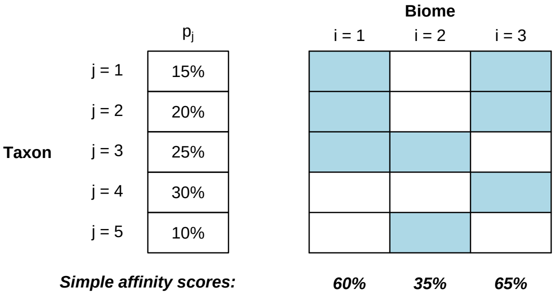
Defines a relationship between biomes and taxa, such that $\mathcal{J}_i$ is the set of taxa that can be reached from biome i.

Current approach: Affinity score

Biomisation is estimating the biome from a pollen assemblage $\{x_j : j = 1, \ldots, n\}$. The dominant method is to *maximise the affinity score*:

$$\mathrm{Aff}(i) := \sum_{j \in \mathcal{J}_i} \hat{p}_j \qquad \hat{p}_j := \{\max(0, p_j - \theta)\}^{\gamma}$$

where p_j is the proportion of taxon j in the assemblage, and typically $\theta = 0.5\%$ and $\gamma = 1/2$.



Current approach: Affinity score

Some observations

1. The affinity score for biome j will be relatively high if and only if this biome contains the well-represented taxa.
2. The choice of $\gamma = 0.5$ down-weights taxa with large proportions, which is probably a crude adjustment for differential rates of productivity and dispersal.
3. But $\gamma = 0.5$ over-weights the contribution from the large number of taxa with small proportions, and so $\theta = 0.5\%$ is need to squash these.

Current approach: Affinity score

Some observations

1. The affinity score for biome *j* will be relatively high if and only if this biome contains the well-represented taxa.

2. The choice of $\gamma = 0.5$ down-weights taxa with large proportions, which is probably a crude adjustment for differential rates of productivity and dispersal.

3. But $\gamma = 0.5$ over-weights the contribution from the large number of taxa with small proportions, and so $\theta = 0.5\%$ is need to squash these.

Statistical concerns

Is $\max_i \mathrm{Aff}(i)$ a good estimator? For example, is there a reasonable underlying statistical model from which it follows that the affinity score is the likelihood function?

- *If so*, we can quantify uncertainty and do hypothesis tests.
- *If not*, what confidence do we have in our reconstructions?

The affinity score is *not* a likelihood function

This follows by considering two biomes, i and i', for which $\mathcal{J}_i \supset \mathcal{J}_{i'}$.

1. Consider the set of pollen taxa that are in $\mathcal{J}_i$ but not in $\mathcal{J}_{i'}$, $\mathcal{I} := \mathcal{J}_i \setminus \mathcal{J}_{i'}$. Suppose that the counts are zero for all the taxa in $\mathcal{I}$. In this case $\mathrm{Aff}(i) = \mathrm{Aff}(i')$.

2. But one would think that getting zero counts in $\mathcal{I}$ was improbable if the biome was i, but probable if the biome was i'. Therefore, statistically, we would want $L(i) < L(i')$, where L is the likelihood function.

3. Extending this argument, if there were small numbers of counts in $\mathcal{I}$ then $\mathrm{Aff}(i) > \mathrm{Aff}(i')$, but we would still want $L(i) < L(i')$ if these could be attributed to a background process.

Statistical model (sketch)

- Aleatory model
  1. The pollen grains on the microscope slide follow independent Poisson processes with rates θ_j that depend on the biome;
  2. A background process with rates λ_j accounts for contamination and other errors such as misidentification and misrecording.

Statistical model (sketch)

- Aleatory model

  1. The pollen grains on the microscope slide follow independent Poisson processes with rates θ_j that depend on the biome;
  2. A background process with rates λ_j accounts for contamination and other errors such as misidentification and misrecording.

- Epistemic model

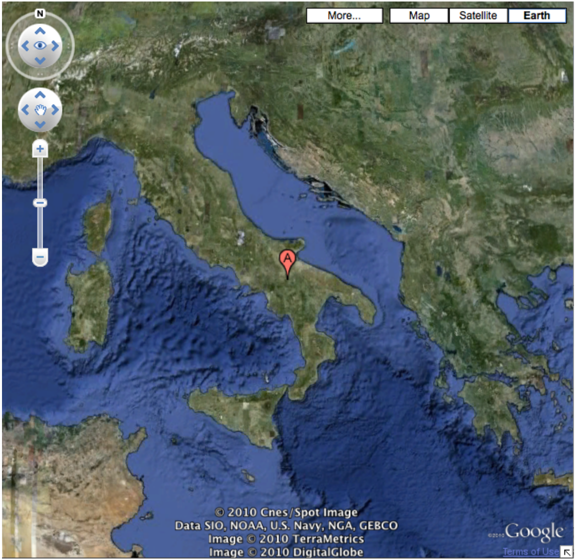
  1. The biome rates θ_j are zero if $j \notin \mathcal{J}_i$, and otherwise $\text{Gamma}(\alpha_{ij}, \beta_{ij})$;
  2. The background rates λ_j are $\text{Gamma}(\alpha_{ij}^{\lambda}, \beta_{ij}^{\lambda})$;
  3. To account for differential productivity and dispersal, set $\beta_{ij} = \beta_i / \hat{\alpha}_j C_j$, likewise for β_{ij}^{λ}.

Statistical model (sketch)

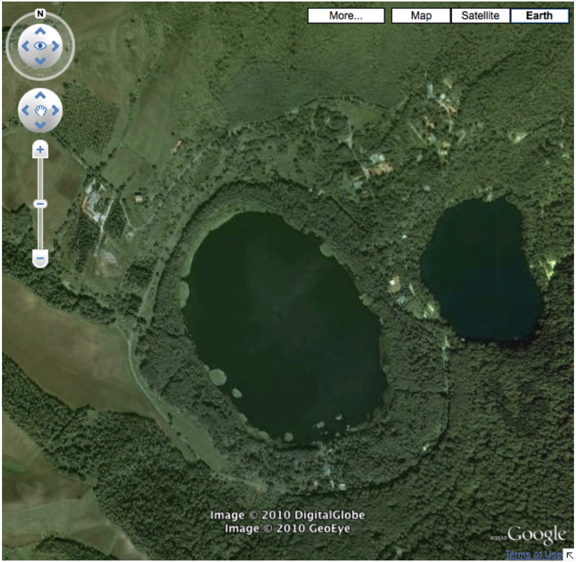
- Aleatory model
  1. The pollen grains on the microscope slide follow independent Poisson processes with rates θ_j that depend on the biome;
  2. A background process with rates λ_j accounts for contamination and other errors such as misidentification and misrecording.
- Epistemic model
  1. The biome rates θ_j are zero if $j \notin \mathcal{J}_i$, and otherwise Gamma$(\alpha_{ij}, \beta_{ij})$;
  2. The background rates λ_j are Gamma$(\alpha_{ij}^{\lambda}, \beta_{ij}^{\lambda})$;
  3. To account for differential productivity and dispersal, set $\beta_{ij} = \beta_i / \hat{\alpha}_j C_j$, likewise for β_{ij}^{λ}.
- Special vague case

  Set $\alpha_{ij} = \alpha_{ij}^{\lambda} = 1$. All of the β's can then be set, in the very simple case, according to a single rate $\kappa \in (0, 1)$ which represents the expected proportion that comes from the background. *One intuitive tuning parameter!*

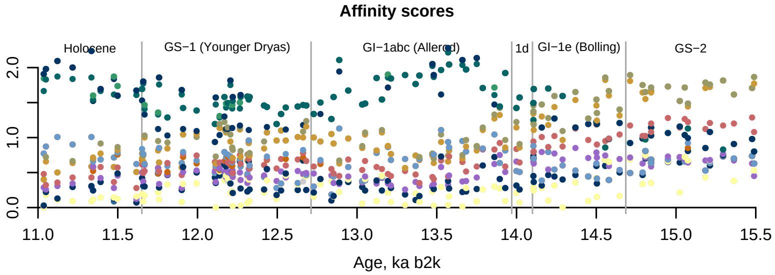
Reconstruction: Monticchio, S. Italy



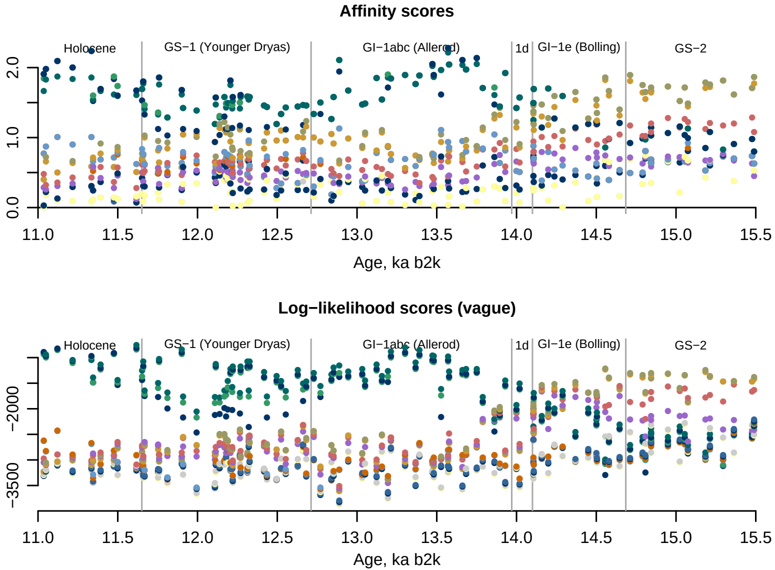
Reconstruction: Monticchio, S. Italy



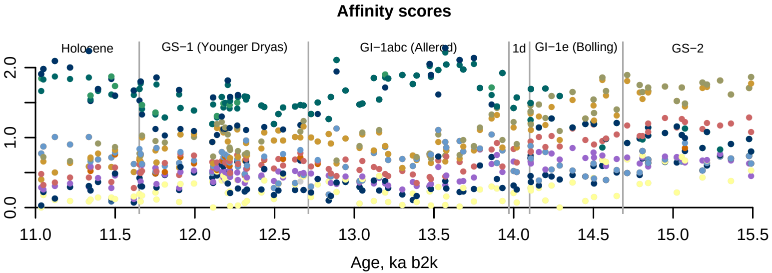
Reconstruction: Monticchio, S. Italy

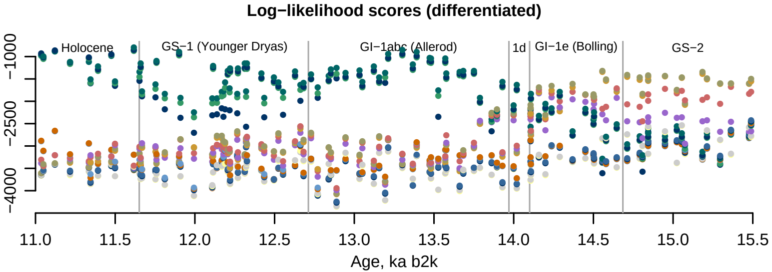


Reconstruction: Monticchio, S. Italy

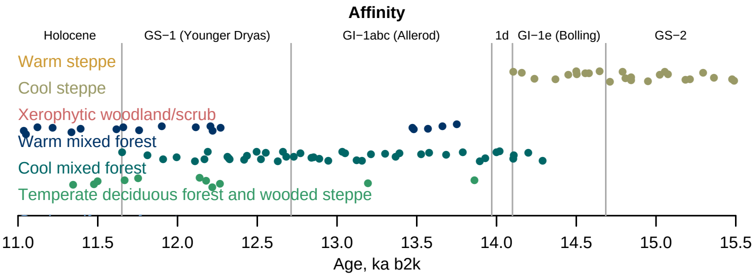


Reconstruction: Monticchio, S. Italy

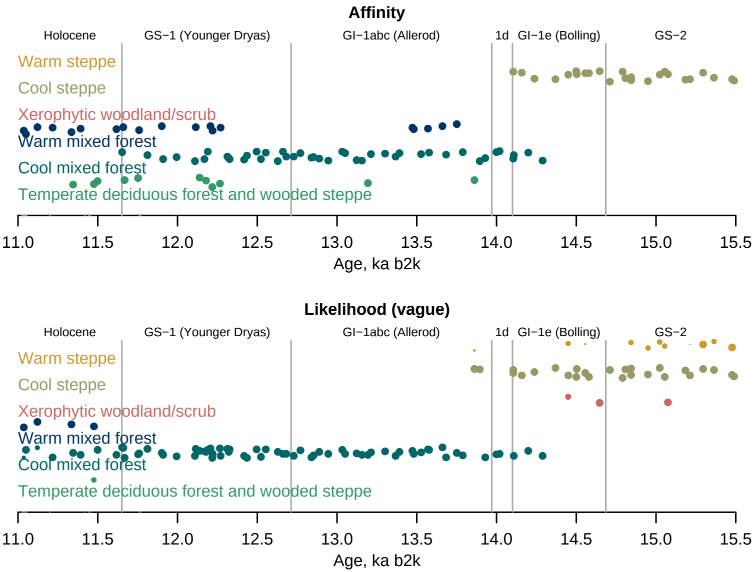




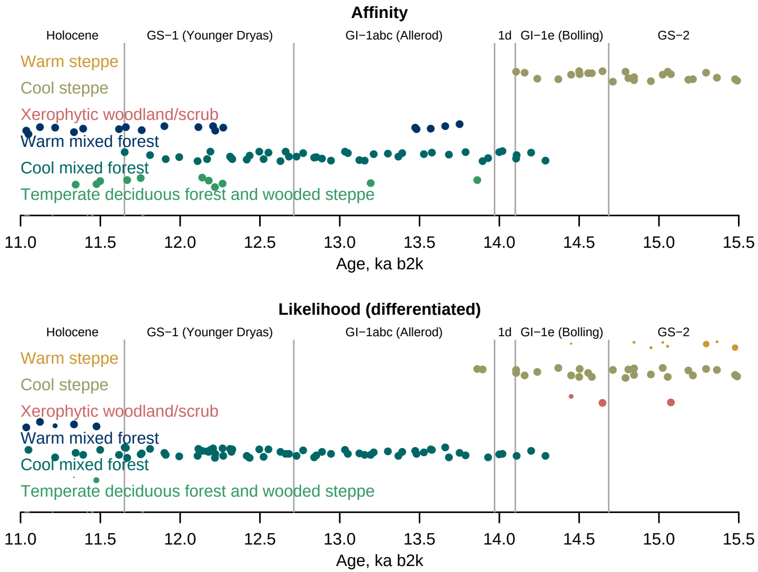
Another visualisation, Monticchio



Another visualisation, Monticchio



Another visualisation, Monticchio



Final observations

- Statistics is about doing sensible things when faced with uncertainty. In particular, how to make good estimates, like what the biome was at site x and time t BP.

- A crucial aspect of the statistical approach is its clarity. One is obliged to make explicit statements of one's judgements, that can be discussed and challenged.

- The benefit is the development, one hopes, of a consensus, and the much richer inference that is possible with more structured judgements.

Many thanks to: Prof Judy Allen, Dr Heather Binney, Dr Philip Brohan, Prof Caitlin Buck and the SUPRAnet project, Prof Mary Edwards, Prof Brian Huntley, Prof Jim Smith.