Notes about the PLS Composition Data From Chris.

**Midwest composition modeling**

**Context**

The current domain includes the states of Minnesota, Wisconsin, Michigan, Illinois, and Indiana. Data from Minnesota and Wisconsin are essentially complete. Data from Illinois are a dispersed sample as generated by Notre Dame plus focused data from Chicagoland and near St. Louis (provided by the Chicago Arboretum and Dick Brugam, respectively). Data from Indiana are a small sample of convenience from Notre Dame. Data from Michigan are extensive and nearly complete in the upper peninsula and northern lower peninsula and a dispersed sample from the southern lower peninsula.

This product is the estimated composition at the resolution of the PalEON 8 km Albers grid, estimated with a spatial multinomial model, developed by Andy [Thurman](https://paleon.geography.wisc.edu/doku.php/people;andrew_thurman) and Chris[Paciorek](https://paleon.geography.wisc.edu/doku.php/people;chris_paciorek).

**Version 0.2 results, March 2014**

These results are based on the [western composition dataset, version 0.3](https://paleon.geography.wisc.edu/lib/exe/fetch.php/data_and_products;public_data;westerncompv0.3.csv).

At the moment, there are 519,754 trees in the dataset, of which 2,251 are in the PalEON 'Other hardwood' category.

This [settlement vegetation conversion table (version 0.2)](https://paleon.geography.wisc.edu/lib/exe/fetch.php/data_and_products;public_data;level3s_v0.2.csv) shows how we group PalEON Level 3a taxa into Level 3s for use in the statistical model. Here are some more details:

* Modelled taxa are: Ash, Basswood, Beech, Birch, Black gum/sweet gum, Cedar/juniper, Elm, Fir, Hemlock, Hickory, Ironwood, Maple, Oak, Pine, Poplar/tulip poplar, Spruce, Tamarack, Walnut, Other hardwood
* Atlantic white cedar and chestnut are not modeled because, while they exist in the eastern composition analysis, there are zero trees in the western portion of the domain.
* Included in the Other hardwood category are the PalEON taxa: Alder, Buckeye, Cherry, Dogwood, Hackberry, Locust, Mulberry, Other hardwood, Sycamore, Willow
* The PalEON “unknown tree” category is not included in any category, but in the future may be included in “Other hardwood” as most are likely to be hardwoods.

The MCMC was run for 150,000 iterations, with the first 25,000 discarded as burn-in, and the remaining iterations subsampled (to save on storage and computation) to give 500 posterior draws.

* [Raw data plots](https://paleon.geography.wisc.edu/lib/exe/fetch.php/data_and_products;public_data;plscomposition_western_0.2_rawdata.pdf)
* [Fitted proportions](https://paleon.geography.wisc.edu/lib/exe/fetch.php/data_and_products;public_data;plscomposition_western_0.2_fits.pdf) (Bayesian posterior mean)
* [Uncertainty](https://paleon.geography.wisc.edu/lib/exe/fetch.php/data_and_products;public_data;plscomposition_western_0.2_uncertainty.pdf) (Bayesian posterior standard deviation)
* [Version 0.2 data product](https://paleon.geography.wisc.edu/lib/exe/fetch.php?tok=2de280&media=http%3A%2F%2Fwww.stat.berkeley.edu%2Fshare%2Fpaciorek%2Fpaleon%2FPLScomposition_western_0.2-release.nc) (500 MCMC draws) [note that loading this into R using the RNetCDF package may omit three of the taxa so use of ncdf4 is recommended]
* [code on Github](https://paleon.geography.wisc.edu/lib/exe/fetch.php?tok=000224&media=https%3A%2F%2Fgithub.com%2FPalEON-Project%2Fcomposition%2Ftree%2Fversion0.2) (version0.2 branch)

Comment from Steve [Jackson](https://paleon.geography.wisc.edu/doku.php/people;steve_jackson) (3/19/14):

In version 0.2 results in March, Cherry is lumped into Other. It is interesting that taxa like sycamore, cherry, hackberry, etc. are not very important. Their rarity seems significant, especially cherry, which is likely to be ‘fairly’ abundant in FIA data – i.e. on the modern landscape. Cherry is a disturbance indicator, and second-growth forests from New England to the Midwest have a lot of it. I don’t think we need to devote a map to cherry, but the observation that it’s so rare, even compared to walnut and gum, is telling, especially with respect to the controversies over Native American influence on presettlement forest. If Abrams and Nowacki are right, I’d expect to see a LOT more cherry in the 19th Century.

So it’s worth bookmarking this observation and noting it somewhere in the text of what comes from this.

**Eastern township composition modeling**

**Context**

The current domain includes the states of New England, New York, Pennsylvania, New Jersey, and Ohio. The data from Ohio are a convenience sample and will be augmented in the future. There is also data from one township in northern Delaware and one in Quebec, augmented by a convenience sample of data from Ohio.

This product is the estimated composition at the resolution of the PalEON 8 km Albers grid, estimated with a spatial multinomial model using the eastern township data. The primary statistical trick is to include latent variables for each tree that indicate the grid cell the tree is in, based on the grid-township overlap, the estimated spatial composition field, and the taxon of the tree.

**Version 0.2 results, March 2014**

These results are based on the [1372.08 version](https://paleon.geography.wisc.edu/lib/exe/fetch.php/data_and_products;public_data;1372centroid_polygonsver08.zip) of the data from most of the domain and the [471.02 version](https://paleon.geography.wisc.edu/lib/exe/fetch.php/data_and_products;public_data;oh471centroid_polygonsver02.zip) of the Ohio data.

This version is essentially unchanged from Version 0.1 (the same input data and taxa are used) except that MCMC mixing has improved due to a change in the MCMC sampler and this is based on a longer MCMC run as well.

At the moment, there are 420,956 trees in the dataset, of which 9,729 are in the PalEON 'Other hardwood' category.

This [settlement vegetation conversion table (version 0.2)](https://paleon.geography.wisc.edu/lib/exe/fetch.php/data_and_products;public_data;level3s_v0.2.csv) shows how we group PalEON Level 3a taxa into Level 3s for use in the statistical model. Here are some more details:

* Modelled taxa are: Ash, Atlantic white cedar, Basswood, Beech, Birch, Black gum/sweet gum, Cedar/juniper, Chestnut, Elm, Fir, Hemlock, Hickory, Ironwood, Maple, Oak, Pine, Poplar/tulip poplar, Spruce, Tamarack, Walnut, Other hardwood
* Included in the Other hardwood category are the PalEON Level 3a taxa: Buckeye, Cherry, Dogwood, Hackberry, Mulberry, Sycamore, Other hardwood

The MCMC was run for 150,000 iterations, with the first 25,000 discarded as burn-in, and the remaining iterations subsampled (to save on storage and computation) to give 500 posterior draws.

* [Raw data plots](https://paleon.geography.wisc.edu/lib/exe/fetch.php/data_and_products;public_data;plscomposition_eastern_0.2_rawdata.pdf) (this is a big file - 15Mb)
* [Fitted proportions](https://paleon.geography.wisc.edu/lib/exe/fetch.php/data_and_products;public_data;plscomposition_eastern_0.2_fits.pdf) (Bayesian posterior mean)
* [Uncertainty](https://paleon.geography.wisc.edu/lib/exe/fetch.php/data_and_products;public_data;plscomposition_eastern_0.2_uncertainty.pdf) (Bayesian posterior standard deviation)
* [Version 0.2 data product](https://paleon.geography.wisc.edu/lib/exe/fetch.php?tok=7f03ba&media=http%3A%2F%2Fwww.stat.berkeley.edu%2Fshare%2Fpaciorek%2Fpaleon%2FPLScomposition_eastern_0.2-release.nc) (500 MCMC draws) [note that loading this into R using the RNetCDF package may omit some of the taxa so use of ncdf4 is recommended]
* [code on Github](https://paleon.geography.wisc.edu/lib/exe/fetch.php?tok=000224&media=https%3A%2F%2Fgithub.com%2FPalEON-Project%2Fcomposition%2Ftree%2Fversion0.2) (version0.2 branch)