Description of the data file

**Background of the dataset**

The dataset comprises a subset of the USDA’s Forest Inventory and Analysis (FIA) census from 2009. Details regarding the preparation of the dataset can be found in:

Johnson, D.J., Beaulieu, W., Bever, J.D., Clay, K. 2012. Conspecific negative density dependence and forest diversity. Science 336: 904-907.

Eppinga, M.B., Baudena, M., Johnson, D.J., Jiang, J., Mack, K.M.L., Strand, A.E., Bever, J.D. 2018. Frequency-dependent feedback constrains plant community coexistence. Nature Ecology and Evolution 2: 1403-1407.

**Overview of the data folder**

The data has been saved as a single file:

* FIAData.txt

This file contains no headers. The dataset contains information on more than 200,000 forest plots within the contiguous United States, all situated east of the 100th meridian. The plots are divided over ~ 65,000 locations, meaning that at each location, 3-4 plots have been recorded (referred to as subplots). For each plot, all adult trees have been identified within a 24-foot (about 7.32 m) radius. Within each plot, seedlings have been recorded within a subsection with a 6.8 foot radius (about 2.07 m), referred to as a microplot. The unit of observation within the dataset is species, that is, each row indicates the number and size of adults and seedlings of a species within a certain (sub)plot. More specifically, the data contains the following 9 variables:

* *Latitude:* Part of the GPS location of the plot (i.e. the same for all subplots within a location), in decimal degrees.
* *Longitude:* Part of the GPS location of the plot (i.e. same for all subplots within a location), in decimal degrees. Note that values are negative, as all locations are on the western hemisphere.
* *CellID:* ID that divides the dataset up into 2 by 2 degree grid cells. Counting starts (with 0) in the southwest (Texas), moving northwards within the westernmost column of grid cells until the Canadian border is reached. Then, counting continues in the next column, starting with the southernmost cell. The last cell (203) is in the most northeastern region of Maine.

* *PlotID:* A fourteen-digit code that identifies a location, where 1-4 subplots have been measured.
* *Subplot:* The subplot number of the location, subplots are labeled: 1, 2, 3 or 4.
* *Species:* A two to four digit species code (i.e. ranging from 12 to 8513), identifying the tree species measured. Species codes correspond to the FIA’s coding scheme for species, see the fourth and sixth columns of the following table: <https://www.nrs.fs.fed.us/fia/data-collection/field-guides/ver6.0/Appendix%203.pdf>. (Note that ‘0012’ in this Table refers to ’12’ in the dataset, etc.).
* *ad:* The number of adults of a particular species that were observed in the subplot.
* *ba:* The cumulative basal area, expressed in m2, of the adult trees of a particular species observed in the subplot.
* *Seed:* The number of seedlings of a particular species that were observed in the microplot.