**State-space Canopy Conductance (StaCC) Modeling Framework Manual**

David M. Bell

December 31, 2014

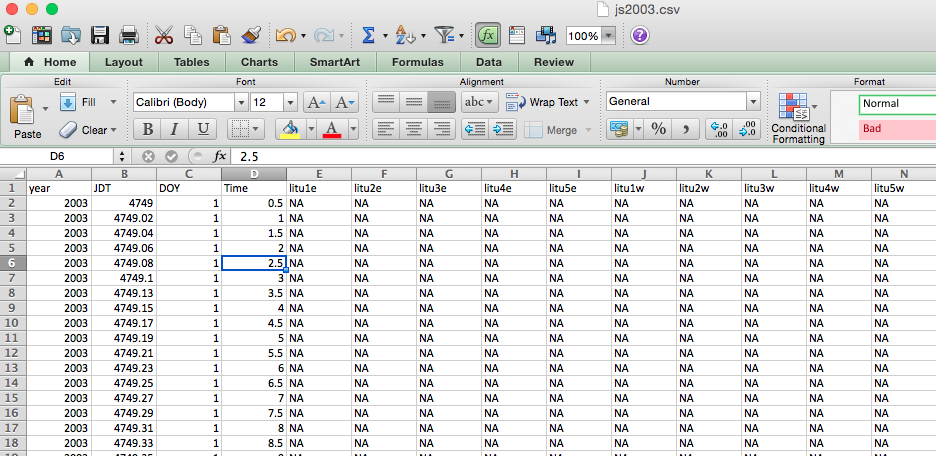
**1. Introduction**

Uncertainties in ecophysiological responses to environment limit our ability to estimate key inputs for ecosystem models. Advanced statistical frameworks provide coherent methodologies for relating observed data, such as stem sap flux density, to unobserved processes, such as canopy conductance and transpiration (Clark et al. 2011). Because terrestrial ecosystem water and carbon dynamics are intimately linked, improved prediction of canopy processes based on sap flux density data is a high priority for forest ecosystem modelers.

To address this need, we developed a hierarchical Bayesian State-Space Canopy Conductance (StaCC) model linking canopy conductance and transpiration to tree sap flux density from a four-year experiment in the North Carolina Piedmont, USA (Bell et al. *in review*). The following document describes model setup, implementation, and analysis of results. We avoid in depth discussion of the statistical details of the modeling framework here, focusing instead on practical concerns, such as data formatting and prior specification. It is strongly recommended that users familiarize themselves with our previous publications before attempting to use or modify the statistical framework (Ward et al. 2013, Bell et al *in review*).

**2. Data Formatting**

*Sap flux density data*: Sap flux density data are stored in a .csv file (e.g., “js2003.csv”; Fig. 1). The file contains at least five columns with each row representing an individual time interval. We use 30-minute intervals in our previous work, so each row is a single 30-minute period of time. The column “year” contains the year of the time interval. The column “JDT” contains the Julian date of the time interval with units in days. The column “DOY” contains the day of year of the time interval. The column “Time” contains the time of day of the time interval with units in hours. All remaining columns are sap flux data (g m-2 s-1), with each column representing a single sap flux probe. Missing data are represented by “NA”. Each column is named based on the species and the sensor. For the example data (Fig. 1), all of the sap flux probes are measuring *Liriodendron tulipifera* located in either the eastern or western portion of a study area. Thus, we have “litu1e”, “litu2e”, “litu3e”, “litu4e”, “litu5e”, “litu”1w”, “litu2w”, “litu3w”, “litu4w”, and “litu5w”.



**Figure 1. Example of the sap flux density file.**

*Leaf area data*: Leaf area data are stored in a .csv file (e.g., “LAI2003.csv”). The file contains at least five columns with each row representing an individual time interval. The first four columns (“year”, “JDT”, “DOY”, and “Time”) are the same as for the sap flux density data file. We use 1-day intervals in our previous work, so each row represents a single day. All remaining columns are leaf area index data (m2 m-2), with each column representing a single species. Each column is named based on the species.

*Sapwood area data*: Sapwood area data are stored in a .csv file (e.g., “SA2003.csv”). The file contains at least five columns with each row representing an individual time interval. The first four columns (“year”, “JDT”, “DOY”, and “Time”) are the same as for the sap flux density data file. We use 1-day intervals in our previous work, so each row represents a single day. All remaining columns are sapwood area index data (m2), with each column representing a single species. Each column is named based on the species.

*Photosynthetically Active Radiation data*: Photosynthetically active radiationdata are stored in a .csv file (e.g., “PAR2003.csv”). The file contains five columns with each row representing an individual time interval. The first four columns (“year”, “JDT”, “DOY”, and “Time”) are the same as for the sap flux density data file. We use 30-minute intervals in our previous work, so each row is a single 30-minute period of time. The last column is photosynthetically active radiationdata (??) for the study site

*Soil moisture data*: Soil moisture data are stored in a .csv file (e.g., “sm2003.csv”). The file contains five columns with each row representing an individual time interval. The first four columns (“year”, “JDT”, “DOY”, and “Time”) are the same as for the sap flux density data file. We use 30-minute intervals in our previous work, so each row is a single 30-minute period of time. The last column is soil moisture data (??) for the study site

*Air temperature data*: Air Temperature data are stored in a .csv file (e.g., “Ta2003.csv”). The file contains five columns with each row representing an individual time interval. The first four columns (“year”, “JDT”, “DOY”, and “Time”) are the same as for the sap flux density data file. We use 30-minute intervals in our previous work, so each row is a single 30-minute period of time. The last column is air temperature data (°C) for the study site

*Vapor pressure deficit data*: Vapor pressure deficit data are stored in a .csv file (e.g., “VPD2003.csv”). The file contains five columns with each row representing an individual time interval. The first four columns (“year”, “JDT”, “DOY”, and “Time”) are the same as for the sap flux density data file. We use 30-minute intervals in our previous work, so each row is a single 30-minute period of time. The last column is vapor pressure deficit data (m2) for the study site.

*Probe data*: Individual probe metadata is stored in a .csv file (e.g., “HW\_JSid.csv”). Each row of the file is data for individual sap flux probes including the name of the site (“Site”), sub-site (“Ring”), the tree species (“Species”), the senor number (“Sensor”), the column number of the sap flux density data file (“Column”), the canopy status of the tree being monitored (“Status”; 1 for dominant and 2 for suppressed), the depth of the probe (“Depth”; cm), whether or not fertilization treatments were applied (“Fert”; currently not in use), the tree number (“Tree”), the name of the column in the sap flux density data file (“colname”), and the tree diameters for different years (e.g., “DBH2003”).

**3. Prior Specification**

Model priors are stored in “SF.Priors.r” in the “Code” directory. Please refer to Bell et al. *in* review for full details regarding prior specification. Full discussion of prior specification is coming soon.

**4. Controlling and Initiating StaCC Model Runs**

The StaCC model code is located in four R scripts: “SF.Input.r”, “SF.Setup.r”, “SF.functions.r”, and “SF.Gibbs.r”. For many users, the StaCC model can be used with relatively minor modifications to the “SF.Input.r” file. This file allows the user to identify input data, modify the study period length, and activate or deactivate certain modeling features. In this section, we describe the use of the “SF.Input.r”.

dir (line 8): This object holds the working directory path for the StaCC model run and must be set by the user. The working directory is the directory holding the “Code”, “Data”, and “Output” folders for a given StaCC model project.

DATA (line 12) and RUN (line 13): These objects are naming conventions for the purposes of saving R workspaces after model fitting. We recommend that users use DATA to represent the data source being used and use RUN to identify different implementations of the StaCC model (e.g., different priors)

FLAT (line 18): This object controls the implementation of the sap flux density depth profile. If set to TRUE, then the model assumes that *α*3 = 1. As a result, sap flux density is assumed to be constant with depth across tree sapwood. If set to FALSE, the model allows sap flux to decline with sapwood depth.

RAND (line 19): This object controls the use of sap flux sensor random effects If set to TRUE, then the model includes random effects to capture sensor-level biases in the sap flux density data. As a result, some sap flux sensors are assumed to differ consistently (i.e., a given sensor exhibits higher or lower sap flux densities compared to other sensors). If set to FALSE, the model assumes all sensors are unbiased. In general, setting this object to FALSE is an unrealistic assumption, though it may be informative in determining how important random effects are for statistical inference.

BS (line 20): This object controls the use of canopy status (XX in data table) in the StaCC model. If set to TRUE, then the model predicts the effect of canopy status on sap flux density. As a result, it is assumed that sap flux density differs between canopy dominant and suppressed trees. If set to FALSE, the model assumes canopy status has no impact. Only set this object to TRUE if the user’s sampling design provides sufficient information to test the impact of canopy status.

DPAR (line 8):

**5. StaCC Output**

Coming Soon.

Literature Cited

Bell et al. *in review*.

Clark et al. 2011

Ward et al. 2013