**Read About NGGI Data Sets**

Meng Lu and Pat Megonigal Dec 2017

**Introduction**

This data library supports the coastal wetland National Greenhouse Gas Inventory (NGGI). The library was developed from the data used for the 2016 NGGI, but has been considerably modified as described in the project report. Each publication in the present library is assigned a consecutive ID number, and data extracted from the library is contained in a separate folder (i.e. Set-001 to Set-105). Each folder includes:

1. The original publication (PDF file, numbered consecutively from Set-001 to Set-104), or the original data set from EPA (Set-105).

2. The word file that recorded the interpretations made by Meng Lu for data extracting (the source of the reported data, i.e., from Figures or tables from the PDFs).

3. A “calculation file” in cases where the data in the paper could not directly transferred into the data template. In such cases, any calculations that were made in translating data from the publication to the data template are documented here. For example, in the case that we only have soil organic matter from the publication, we did a pre-calculation to convert the data to soil C units (using an equation developed by James Holmquist).

4. The SAS code we used for any further calculations, such as converting between units of area or mass. Also, the SAS code transposes the data in preparation for merging it into a master file.

5. The filled data entry sheet (EXCEL file) for each publication.

6. A CSV file containing the transposed version of the data that is found in the data entry sheet. This is the file that will be merged into a master file for further calculations.

***Criteria and Methods for Extracting Data From Literature***

The following are the criteria for the papers in this folder, or rules about how the data were extracted from the publications.

Criterion 1: The library will only include data that is in a citable public repository, which includes both publications and public data libraries. Only primary literature is used, not review papers unless they contain original data. This criterion excluded a small number of data points that were used in the 2016 NGGI. However, a number of new papers were added to the dataset as well.

Criterion 2: The data library captured data at a resolution appropriate for NGGI goals, and is not meant to be a repository of all data points reported in a publication. Authors typically sample in a design that includes strata such as plant community composition or elevation, and replication within strata. As such, publications report both averages within strata, and averages across strata. We used the means of the strata identified by the authors as reported in the publication. In some cases we extracted data from a publication (i.e. the authors did not report averages for a stratum), we calculated the average only if there were clear criteria in the metadata for linking each data point to the appropriate stratum.

Criterion 3: Perhaps the biggest risk of using large spreadsheets for the NGGI is the likelihood of introducing errors when converting from one unit to another. As such, data were translated from the publication to the flat file in their native units to minimize transcription errors, and unit conversions were then calculated by code.

Criterion 4: Soil and biomass stock data are rarely reported in a form appropriate for the NGGI, and converting the data for our purposes requires complex code and data structures. In order to minimize the amount of information stored in the flat file for each publication, calculations of stocks were performed in a separate spreadsheet and document in the “read me” file. Any further extrapolations or manipulations of these data (e.g. to estimate soil carbon from organic matter) will be done separately by code.

Criterion 5. Scrub-shrub communities were classified using dimension data when such data are available. We used the forest service definition for the division between trees and lower stature woody plants used in the US carbon accounting system. For now we are assuming that trees are >4.6 m in height and 7.6 cm for DBH.

Criterion 6. All sites were in the United States.