Thrust Area 5: Carbon Capture & Nuclear

Database Infrastructure for Integrative Carbon Science Research

PI: Sabine Grunwald **Co-PI:** Timothy A. Martin
Students: C.W. Ross (M.S.); X. Xiong (Ph.D.)
Technical staff: Brandon Hoover
Post-Docs: Nichola M. Knox

Description: Rising CO_2 concentrations in the atmosphere and effects on global climate change have been well documented, and future impacts are uncertain but potentially devastating. Florida's natural and agro-forest ecosystems have much potential to sequester carbon in biomass and soils due to unique climatic and landscape conditions. However, research gaps exist to accurately assess carbon pools and fluxes at coarse scales, ranging from county to the region and larger. The overarching objective of this project is to address these obstacles by developing a terrestrial carbon information system (called "TerraC") for the carbon science community, focused on ecosystems in Florida. The information system will be administered through the UF Carbon Resources Science Center (http://carboncenter.ifas.ufl.edu), a multi-disciplinary Center dedicated to research in support of enhanced agricultural and natural resource carbon management.

Budget: $199,440
Universities: UF
External Collaborators: Natural Resources Conservation Service-U.S. Department of Agriculture

Progress Summary

1) Database
The conceptual design for the Terra Carbon Information System (Terra C) has been completed and the database structure implemented in Structured Query Language (SQL). TerraC consists of a web-accessible database, meta data editors, and project wizard. The database hosts carbon and associated environmental data from the soils/geologic, atmospheric, vegetation, water, and whole ecosystem domains. A major amount of time was spent on testing, evaluation, and improvement to upload data into the database, which requires the use of a standardized data template. To accommodate a variety of different carbon data types (e.g. labile soil carbon, recalcitrant soil carbon, total soil carbon, carbon flux, and respiration rate), units of data (e.g. concentrations, contents, stocks, fluxes, CO_2eq., etc.), and spatial and temporal resolution of data the design of the database structure is divided into core and optional fields. The core fields are standardized to avoid duplication of carbon data entered into the TerraC database by different users and projects. Optional fields contain carbon data which are specialized in terms of their analytical methods and/or data collection protocol.

2) Web Design & Tutorials
The web site http://terraC.ifas.ufl.edu has been developed which provides access to the data engine, query tools, and tutorials. The latter ones provide step-by-step instructions how to use the system implemented in form of Adobe Presenter with voice over.

3) Carbon Datasets
Several site-specific carbon datasets have been identified and acquired from various sources. These carbon data are streamlined and documented in TerraC. Among them is the large, historic (~1965 to 1996) Florida Soil Characterization Dataset which contains about 1,300 soil samples and about 8,300+ horizons with 144 different soil physical, chemical, morphological, and taxonomic data which are

georeferenced. In this dataset 7716 samples have soil organic carbon measurements, belonging to 1252 profiles. Soil carbon has been measured based on Walkley-Black modified acid-dichromate method (mineral soils) and Loss on Ignition (organic soils), respectively. Other carbon data streamlined into TerraC include data from the Santa Fe River Watershed where various carbon pools (hot-water extractable (labile) carbon, recalcitrant carbon, total carbon, mineralizable carbon, and other biogeochemical properties were collected. We will continue to populate TerraC with carbon data representing various ecosystem types and ecosystem components over the next project phase.

4) Synthesis Analysis – Carbon Assessment / Capture

We continue to work on various prototype synthesis projects which utilize TerraC. For example, the Florida Soil Characterization Data has been synthesized with spectral data derived from visible/near-infrared diffuse reflectance spectroscopy to build a spectral carbon library for the State of Florida (Vasques et al., 2010). In another synthesis project, the soil carbon data in the Santa Fe River Watershed (FL) were fused with a large set of environmental factors to model the spatial distribution of soil carbon across the watershed in dependence of climate and land use change (Vasques et al., 2010). Similar synthesis analysis will continue in the next project phase.

References:

Vasques G.M., S. Grunwald and W.G. Harris. 2010. Building a spectral library to estimate soil organic carbon in Florida. J. Environ. Qual. 39: 923-934.

Vasques G.M., S. Grunwald, N.B. Comerford and J.O. Sickman. 2010. Upscaling of dynamic soil organic carbon pools in a north-central Florida watershed. Soil Sci. Soc. Am. J. 74: 870-879.

Funds leveraged/new partnerships created: By utilizing TerraC as a core structure for data warehousing and synthesis, it helped to leverage a new project funded by the United State Department of Agriculture (USDA) – National Institute of Food and Agriculture (NIFA) – Agriculture and Food Research Initiative (AFRI) Regional Project "Pinemap: Integrating Research, Education and Extension for Enhancing Southern Pine Climate Change" (2011-2016). This is a large-scale $20 million project with 50+ Co-PIs and Collaborators from 13 institutions (UF is the lead institution; PI: T.A. Martin; Co-PI: S. Grunwald) which supports 25+ students, 10+ Post-Docs, and staff members. The project goals are to create, synthesize, and disseminate the necessary knowledge to enable southern forest landowners to:

- harness pine forest productivity to mitigate atmospheric carbon dioxide
- more efficiently utilize nitrogen and other fertilizer inputs
- adapt their forest management approaches to increase resilience in the face of changing climate.

The TerraC system provides the data infrastructure for the Pinemap project. Since Pinemap funds a comprehensive monitoring program of ecosystem properties, many measurements related to the carbon cycle, will be streamed into TerraC populating the database and allow sharing of carbon and ecosystem data widely (constraint by the data sharing policy implemented in the Pinemap project).

2011 Annual Report

The overall goals of this project are to create a database infrastructure for the carbon science community, focused on ecosystems in Florida and the southeastern United States. The availability of this database will enable the analysis and synthesis of carbon data at multiple spatial and temporal scales. A short description of the TerraC Information System is provided.

What is Terra C?

The **Terrestrial Carbon (TerraC) Information System** is dedicated to advance terrestrial carbon science through sharing of carbon and environmental data. It provides information about carbon cycling as it relates to global climate and land use change. TerraC offers tools to upload, store, manage, query, analyze, and download data characterizing terrestrial carbon dynamics from various sources, including soils, plants/biomass, atmosphere, water, and whole ecosystems. The purpose of Terra C is three-fold: (i) advance terrestrial carbon science through sharing of carbon and environmental data; (ii) facilitate environmental synthesis; and (iii) enhance collaboration among researchers, scientists, and extension specialists through shared resources. Research projects involving carbon and related properties are many and spread across multiple disciplines and spatial and temporal scales. TerraC aims to integrate data from these widespread sources in a shared information system to facilitate cross-cutting analysis of carbon and environmental data, synthesis of carbon research, and enhance communication and collaboration among researchers. Data stored in TerraC conform to quality standards and can be shared privately among selected users or publicly with any user. Detailed information about the data sharing options available in TerraC is listed in the data sharing and usage policy.

Objectives of the Terra C Project

(1) *Build the Terra C Information System*: (i) Develop a coherent, searchable, and expandable database that integrates terrestrial carbon and associated environmental datasets, and (ii) provide information about carbon related to environmental stressors such as climate and land use change.
(2) *Data synthesis*: Synthesize multiple large carbon datasets to gain insight into carbon cycling and dynamics across various spatial and temporal scales; upscaling of site - specific carbon observations to landscape scales.
(3) *Geospatial outreach*: Build a 'GoogleEarth' application to deliver and share terrestrial carbon data in form of a Google carbon application.

Motivation

Rising CO_2 emissions in the atmosphere and effects on global climate change have been well documented, and future impacts are uncertain but potentially devastating. Florida's natural and agro-forest ecosystems have much potential to sequester carbon in biomass and soils due to unique climatic and landscape conditions. However, research gaps exist to accurately assess carbon pools and fluxes at coarse scales, ranging from county to the region and larger scales. The overarching objective of this project is to address these obstacles by creating a database infrastructure for the carbon science community, focused on ecosystems in Florida and the southeastern United States. The database engine of TerraC is administered through the UF Carbon Resources Science Center, a multi-disciplinary Center dedicated to research in support of enhanced agricultural and natural resource carbon management. The TerraC project aims to provide the framework to synthesize carbon and environmental data to facilitate meta-analysis, modeling of carbon dynamics and biogeochemical cycles, and to conduct applied and cutting-edge carbon science research.

TerraC in a Nutshell

TerraC provides a data engine which allows managing, archiving, sharing, editing, modifying, and querying carbon and associated environmental data. These data are derived from various projects and sources; thus, provide a wide array of different carbon measurements, in various ecosystems and geographic regions, and spatial and temporal scales. The Terra C data engine facilitates synthesis and modeling to gain better insight into carbon cycling from micro, plot, field, watershed, basin, large region, and global scales.

Data Sharing and Usage Policy

Data users submitting data to or use data from the TerraC Information System agree to abide by the terms and conditions explained in this document. Data users may be held responsible for any misuse that is caused or encouraged by failure to abide by this agreement.

Definitions

Project: Set of one or more datasets that contain carbon (and related) environmental data.
Dataset: Set of data comprised of one or more data fields that contain carbon (and related) data that is part of a single project.

Roles of users

Project owner (leader): Principal Investigator or person with similar credentials responsible for collecting and managing the original, quality controlled data generated by a specific project. The project leader needs to initiate a project before a dataset can be submitted to TerraC and is responsible for the quality of all datasets under his/her projects. The project leader controls the levels of data sharing and can assign one or more data managers to each of his/her projects.

Data contributor (or manager): User that has read/write access to a dataset in TerraC. The data manager has privileges to submit a new dataset to a project and access and modify existing ones in part or as a whole. The project leader needs to assign a user manager status before he/she can submit a new dataset or modify an existing one in a project.

Data user: User that can view a dataset in TerraC. The data user can read public datasets and also private datasets as long as he/she has been granted access to them by the project leader. The data user cannot submit a new dataset or modify existing ones unless he/she receives manager status from the project leader to a project.

Data sharing: Data stored in TerraC can be shared at three access levels. The access levels are chosen by the project leader to control access to their projects by different users. Different access levels can be assigned to different users, the level being project- and user-specific. Levels 1 and 2 mirror the roles of data user and data manager, respectively. Level 3 in the most restricted access level. *Levels of data sharing*:

Level 1 – Public with read-only access: Access to the data is open to all TerraC users. Any person that has a TerraC user account (i.e. data users) can view the data, but not modify it directly from the TerraC database. Only the project leader can modify/edit data.
Level 2 – Private read/write access: Access to the data is open to data managers who were assigned (approved) by the project leader to have permissions to view and modify/edit data directly from TerraC. *Private read/write access* is password-protected.
Level 3 – Private read-only access: Access to the data is restricted to the project leader and users selected by the project leader. Users can only view the data, but not modify it directly from TerraC.

The project leader controls the sharing of data in TerraC. He/she provides leadership for collaboration with new partners on behalf of the project teams. The project leader can switch sharing levels from Level 3 to 2 and 1, but not vice versa, meaning if the data are released to other users or the general public this right cannot be reversed.

Data users who are interested in to gain access to a specific protected dataset can contact the project

leader and negotiate agreement of data use of a specific project. The project leader may agree to share data with the data user to collaborate on a joint project, work on a co-authored research publication, or use them for other purposes.

Data usage: Data users are expected to use data obtained from TerraC to the highest level of professional integrity and ethics. Data users must abide by the following guidelines when distributing or publishing data obtained from TerraC:

Data sharing and usage in TerraC is governed by the **Attribution Non-Commercial Share Alike** license provided by Creative Commons (summary: http://creativecommons.org/licenses/by-nc-sa/3.0; legal code: http://creativecommons.org/licenses/by-nc-sa/3.0/legalcode), which observes the following rules:

Attribution: The data user must give credit to the project leader (or project) in the manner specified by him/her (but not in any way that suggests that the project leader endorses the data user or his/her use of the data);

Noncommercial: The data user may not use TerraC data for commercial purposes; data should be used for reserach and non-profit applications;

Share Alike: If the data is modified in any manner or used to derive other products, the data user may distribute the resulting work only under the same or similar license to this one;

Credits and publications derived from TerraC usage:

- The data user must inform or consult the project leader about his/her intentions to use the data for publication well in advance of submission of the publication; the project leader should be given the opportunity to read the manuscript and, if appropriate, be offered co-authorship;
- The data user must give credit to the project leader (or project), which can be in the form of co-authorship, citation, or acknowledgement, according to the requirements imposed by the project leader; any deviation from this rule must be formally agreed between the data user and project leader;
- The data user must cite or acknowledge TerraC as the data host used to obtain the data;
- Any modification to the data originally obtained from TerraC by the data user must be fully documented.

Carbon Data and Associated Environmental Data

(1) Core Data Fields:

- Identification number for each observation (SN)
- X coordinate (X) {Geographic Coordinate format (latitude/longitude in decimal degrees) with World Geographic Datum 1984, WGD 1984}
- Y coordinate (Y) {Geographic Coordinate format (latitude/longitude in decimal degrees) with World Geographic Datum 1984, WGD 1984}
- Sample date (DATE) {MM/DD/YYYY}
- Height or depth of measurement (Z) {in cm; below the soil surface negative numbers; above the soil surface positive numbers}
- Carbon measurements (variable names, data values, and meta data: analytical methods & units of measurement in Standard International Units)
- Biogeochemical or other environmental data (variable names, data values, and meta data: analytical methods & units of measurement in Standard International Units)

(2) Project Elements (meta data):

- Project title
- Project description (description of sampling design, sampling protocol, quality assessment, data constraints such as below detection limit treatment, missing values, etc.)
- Project owner (typically Principal Investigator of a research project; or Project Leader for agency lead project)
- Project contributor (optional)
- Project user (optional)
- Contact information (Project Owner)
- Funding source
- Project location (description of geographic location of project; size of project area)
- Project period (YYYY to YYYY)
- Link to project homepage
- Publications from project
- Acknowledgements

Data Quality and Standards

Data format: TerraC focuses on terrestrial carbon and related environmental data. Data submitted to TerraC must contain carbon data and have the following format:

- Be oriented in rows and columns, with cases (observations) listed in the rows and properties (attributes) listed in the columns;
- Carbon and other measured properties must be presented as variables in specific columns:
    - Each column must only contain properties measured using the same method; if the same property was measured using more than one method (e.g. total carbon vs. carbon fractions), each method must be presented as a separate column;
- Spatial coordinates (horizontal and vertical) and time stamps must be presented, whenever available, as variables in specific columns;
- Repeated measures (e.g., the same property collected at different times or replicated) must be treated as separate cases (i.e. listed in separate rows):
    - A column indicating that the cases are repeated measures of the same property must be included (e.g. using the same sample identifier for the repetitions);
    - A column indicating the number of the repetition (i.e. 1, 2, 3…) must be included;
- Quality assurance/quality control (QA/QC) data must not be included in the dataset, but instead in the metadata of the property it pertains to.

Metadata: Since the objective of TerraC is to share data among multiple users, it is critical that metadata are provided in detail for every project, dataset, and variable in a dataset. Upon creation of a new project, the project leader needs to provide information (i.e. metadata) describing project detail, including location, sampling design, contact information, objectives, and others. Upon submission of a new dataset, the project leader or data manager needs to provide metadata for dataset and for every variable in the dataset.

Data quality: It is the responsibility of the project leader to ensure that all data listed under a project in TerraC have passed QA/QC. The project leader provides information for each project's data about the type of QA/QC and adopted standards. The data managers can assist the project leader to meet QA/QC requirements. The TerraC team may quarantine suspicious data and request information from the project leader and/or data manager to assure quality of the data before making them available online. TerraC cannot be held responsible for mistakes in the data or inadequate data usage. Data that for some reason are restricted by funding agencies or imposed proprietary or legal rights (e.g. military projects, pending patents, projects funded by private companies, or other) should not be included in TerraC.

Fig. 1. Snapshot of the Terrestrial Carbon (TerraC) Information System website.

Fig. 2. Project setup in TerraC.

Fig. 3. Data setup in TerraC.

Fig. 4. User administration tools.

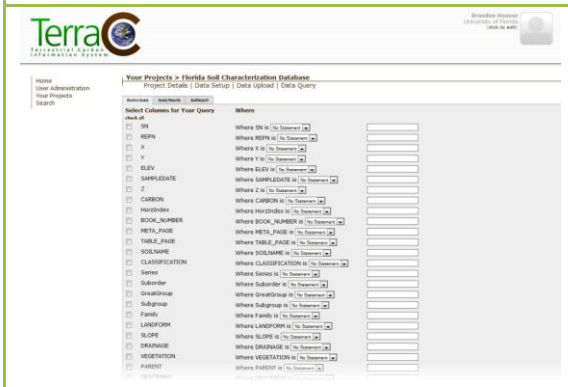

Fig. 5. Data query.

Fig. 6. TerraC tutorials.

Publications and Presentations from this Project:

Hoover B., N.M. Knox, S. Grunwald, T.A. Martin, X. Xiong, P. Chaikaew, J. Kim, B. Cao. 2011. Synthesis Tools for Carbon Assessment in Ecosystems. 2011. Florida Energy Systems Consortium (FESC) Summit, Gainesville, FL, Sept. 27-28, 2011.

Grunwald S., T. A. Martin, B. Hoover, G.M. Vasques, B. Zhong, and D.L. DePatieJr. 2010. Terrestrial carbon (TerraC) information system. 2010 Florida Energy Systems Consortium (FESC) Summit, Orlando, FL, Sep. 27-29, 2010.

Hoover B., G.M. Vasques, B. Zhong, S. Grunwald, T. A. Martin, and D.L. DePatieJr. 2010. The terrestrial carbon (TerraC) information system Vers. 1.0. 11th Annual Soil and Water Science Research Forum, Gainesville, FL, Sep. 10, 2010.

Grunwald S., T.A. Martin, G.M. Vasques and B. Hoover. 2009. Database infrastructure for integrative carbon science research. Florida Energy Systems Consortium Summit, Tampa, FL, Sept. 29-30, 2009.