University of Florida

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

Our main accomplishments so far were to develop the TerraC database structure, functions to enter/download data, meta data tracking, tools to integrate carbon and ancillary environmental data, project menus, and user and administrative tools to manage the data. Several submenus have been developed to manage and store carbon specific data including: below-ground carbon (soils, geology/parent material), atmospheric carbon, vegetation (biomass carbon), and ecosystem (aggregated carbon). In addition, menus to accommodate ancillary environmental data were created that complement carbon dataset. The latter set of data is usually required for synthesis analysis.

Several improvements to handling of streaming out of large data sets with several thousands of records, additions of sub-menus, and a new spectral sub-module were developed. After beta-testing is completed they will be released to TerraC users.

Several large datasets, among them forest and climate data from the PINEMAP project and soil carbon data across the State of Florida, have been uploaded into TerraC. A soil carbon synthesis project conducted in the State of Florida has been completed. This analysis was based on soil carbon data and environmental covariates where machine learning and data mining techniques were used to integrate data and model soil carbon in dependence of environmental conditions (soils, topography, geology) and drivers (land use, climate).

Funds leveraged/new partnerships created

TerraC is now providing the data infrastructure for a $20 million integrated research, education, and extension project. This large-scale 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) allows to populate TerraC and will cross-fertilize several research idea centered around carbon budgets and assessments, carbon change in dependence of global climate change and other stressors, and carbon sequestration and regulation as an ecosystem service. Many other similar synthesis projects will be facilitated through TerraC-PINEMAP.

The PINEMAP 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.

PINEMAP has a multi-tier data structure representing different scales including:

- Tier 1 (historic measurements of tree response in dependence of treatments at about 700 locations across the southeastern U.S.)
- Tier 2 (new base measurements at hundreds of sites across the southeastern U.S.)
- Tier 3 (high-intensity measurements to capture water and carbon cycle at 4 sites)

Annual Progress Report

The main components of the TerraC database structure and web interface has been completed. Several submenus have been developed to allow integration of carbon and environmental covariates including: below-ground carbon (soils, geology/parent material), atmospheric carbon, vegetation (biomass carbon), and ecosystem (aggregated carbon). Submenus had to be standardized to give users the flexibility to enter (upload) carbon data created with different methods and protocols.

TerraC is now providing the data infrastructure of the PINEMAP (NIFA funded) project described above, which required to enhance TerraC to include not only carbon data, but also other types of environmental observations / measurements. For example, the atmosphere module in TerraC previously allowed storing of only carbon-specific properties (e.g., carbon flux measurements). Over the last year we extended its capabilities to include also ancillary environmental observations. For example, users can now also upload climatic properties (e.g., temperature, precipitation, vapor pressure, etc.) that impact and relate to the carbon cycle. These extensions to allow storage of ancillary environmental datasets in submenues of TerraC (such as soils and vegetation) have also been tackled.

The PINEMAP project has been utilizing the TerraC system as its data hub. The tier 1 PINEMAP dataset, which are historic data collected by Forest Cooperatives from hundreds of sites across the southeastern United States, have been integrated into TerraC. The legacy data consists of thirteen data sets that had to be standardized and harmonized, and meta datasets had to be developed. They are now being shared throughout the PINEMAP project via TerraC. With the addition of the legacy data, new standardized methods have been added to TerraC so that vegetation data (e.g., tree height and diameter) and treatment options (e.g., fertilization level) can now be uploaded into the system and shared among researchers, students, and other users.

Ancillary environmental data that are spatially extracted to tier sites were compiled to represent historic and future climate. To accomplish this, the spatial data layers (GIS data layers) from the (i) Parameter-Elevation Regressions on Independent Slopes Model (PRISM) [800 m grid resolution], (ii) (North American Regional Climate Change Assessment Program) (NARCAAP) [50 km grid resolution], and (iii) Idaho geospatial group [4 km grid resolution] were acquired (Table 1). The spatial data were reformatted, harmonized to a common map projection, clipped and then extracted to tier site locations using GIS techniques.

Table 1. Summary of climate data complementing the tier 1 site locations.

Type	Variables (Properties)	Data source *Additional Information*
Climate	Precipitation Minimum temperature Maximum temperature Dew point temperature Mean temperature (average of Tmin and Tmax) Vapor pressure	Parameter-Elevation Regressions on Independent Slopes Model (**PRISM)** *Period: 1970 to 2010* *Monthly climatic data;* *Yearly averages of climatic data;* *Long-term (30 year) average of climatic data.* *Spatial resolution: ~4 km; 2.5 arcmin* *Spatial resolution (SE U.S.): 800 m; 30 arcsec*
Climate	Precipitation Maximum relative humidity Minimum relative humidity Mean specific humidity Mean downward shortwave radiation at surface Mean wind direction Minimum temperature Maximum temperature Mean wind speed	**Idaho Geospatial** *Period: 1971 to 2011* *Daily climatic data;* *Monthly climatic data;* *Yearly averages of climatic data;* *Long-term (30 year) average of climatic data.* *Spatial resolution: ~4 km; 2.5 arcmin*
Climate	Surface air temperature Precipitation Surface downdwelling shortwave radiation Surface pressure Surface specific humidity Num. frost days Minimum monthly temperature Maximum monthly temperature Mean daily minimum temperature Mean daily maximum temperature	**NARCAAP (North American Regional Climate Change Assessment Program)** *Historic climatic data (1971-2000);* *Climate change projections (6 different scenarios) (downscaled to the finest resolution in North America)* *Spatial resolution: 50 km*

The tier 1 (site-specific forest measurements and geographic coordinates) and climate data (PRISM, NARCAAP, and Idaho geospatial) were uploaded into TerraC and meta data created. These data are now useable for synthesis analysis to identify the effect of climate on forest treatments along geographic gradients in the southeastern U.S. (in progress).

The PINEMAP project also spurred a new development that has been named “Umbrella”. Because the project has researchers from multiple institutions that needed to upload and share their data to entire groups of people inside the project, the TerraC team developed “Umbrella” to house multiple projects from multiple owners and assign group privileges to individual project datasets. So PINEMAP (and other Umbrella research projects) is able to assign all of the modeling researchers' permissions to all of the data they need by applying the group's read permission to the project data they need. Using the PINEMAP project, this function is still in the final stages of testing and will be released to all users soon.

Carbon measurements based on spectral techniques have become popular and are now commonly used to complement lab-based analytical techniques. For example, visible-near infrared (VNIR) and mid-infrared (MIR) spectroscopy has shown success to infer on soil carbon rapidly and cost-effectively. In addition, spectral data derived from remote sensing (e.g., Landsat Thematic Mapper satellite images) have been used extensively to infer on landscape features including vegetation indices, such as the Normalized Difference Vegetation Index or Biomass, which allows relating to the carbon cycle. Thus, we developed a spectral sub-module in TerraC that can handle spectral data. Hyperspectral datasets are large and complex consisting of hundreds of repetitive blocks of reflectance values or wavelengths. The spectral sub-module uses column names with incremental wavelengths. To accommodate this kind of data, and to make the system as user-friendly as possible, we are currently beta testing a new method of adding columns that will automate the process of adding large numbers of columns by allowing the user to input variables such as beginning wavelength, ending wavelength, and the increment in order for the system to automatically generate the columns needed. Otherwise data managers would need to add each column individually, which can be quite cumbersome. The design of the spectral module has been completed and is now being coded into TerraC to allow testing.

The most telling indicator that TerraC has been growing has been that a new backend must be developed in order to access larger datasets. The current system has served the TerraC community well, but, as datasets get larger, the system needs to grow in order to keep up with the demand. The PINEMAP project is utilizing datasets that are quite large. In order to output these datasets to users, the TerraC group is moving away from the ASP programming language that could only handle 15,000 to 20,000 rows of data. The new platform, based on Microsoft's ASP.NET, will be able to handle data chunks much larger and will be able to provide them to users much faster than the previous version. The recoding is underway and will be completed in the near future after successful beta-testing.

Another major accomplishment is to utilize carbon data hosted by TerraC to facilitate synthesis analysis. This has been demonstrated in a project where soil carbon data (n = 1,080 site locations) collected throughout the State of Florida were spatially correlated with a comprehensive set of 212 environmental covariates (representing soils, climate, topography, geology, and ecology). The environmental covariates were derived from various GIS and remote sensing resources. Machine learning and data mining techniques, such as ensemble regression trees (State Vector Machines and Random Forest), were used for data integration and to model soil carbon in dependence of environmental properties. This synthesis research found that land use / land cover and hydrologic properties impart most control on the variation of soil carbon.

Peer-reviewed Publications:

1. Cao B., S. Grunwald and X. Xiong. 2012. Cross-regional digital soil carbon modeling in two contrasting soil-ecological regions in the U.S. *In* **Minasny B.**, B.P. Malone, and **A.B. McBratney (eds.).** CRC Press, Taylor and Francis, 2012. ISBN: 978-0-415-62155-7.
2. Patarasuk R., S. Grunwald, T.A. Martin and B. Hoover. 20 . Integrative modeling of tree response along geographic and ecological trajectories in the southeastern U.S. Ecological Modeling J. (in preparation).
3. Ross C.W. 2011. Spatiotemporal modeling of soil organic carbon across a subtropical region. M.S. thesis. University of Florida, Gainesville, FL.

4. Ross C.W., S. Grunwald, and D.B. Myers. 20 . Spatiotemporal modeling of soil carbon stocks across a subtropical region. Soil Sci. Soc. Am. J. (in review).
5. Xiong X., S. Grunwald, D.B. Myers, J. Kim, W.G. Harris and N.B. Comerford. 2012. Which soil, environmental and anthropogenic covariates for soil carbon models in Florida are needed? *In* **Minasny B.**, B.P. Malone, and **A.B. McBratney (eds.).** CRC Press, Taylor and Francis, 2012. ISBN: 978-0-415-62155-7.
6. Xiong X., S. Grunwald, D.B. Myers, J. Kim*, W.G. Harris and N.B. Comerford. 2012. Which soil, environmental and anthropogenic covariates for soil carbon models in Florida are needed? The 5th Global Workshop on Digital Soil Mapping 2012, Sydney, Australia, April 10-13, 2012.
7. Xiong X., S. Grunwald, D.B. Myers, J. Kim, W.G. Harris and N.B. Comerford. 20 . Optimal selection of predicting variables for soil carbon modeling in Florida, USA. Geoderma (in review).

Presentations

1. Grunwald S. 2012. Soil carbon variability across large landscapes. Soil and Water Science Research Forum, Gainesville, FL, Sept. 7, 2012.
2. Grunwald S., B. Hoover, and R. Patarasuk. 2012. Terra C and Pinemap data resources. Webinar series Pinemap project. Gainesville, FL, July 13, 2012.
3. Grunwald S. 2011. Geospatial and spectral soil carbon modeling across large regions. NRCS, National Soil Survey Center (NSSC), Lincoln, NE, May 13, 2011.
4. 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.
5. 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.
6. Hoover B., S. Grunwald, T.A. Martin, G.M. Vasques, N.M. Knox, J. Kim, X. Xiong, P. Chaikaew, J. Adewopo, B. Cao and C.W. Ross. 2011. The Terrestrial Carbon (Terra C) Information System to facilitate carbon synthesis across heterogeneous landscapes No. 264-10. Symposia Spatial Predictions in Soils, Crops and Agro/Forest/Urban/Wetland Ecosystems, ASA-CSSA-SSSA Int. Meeting, San Antonio, TX, Oct. 16-19, 2011.
7. Hoover B., N.M. Knox, S. Grunwald, T.A. Martin, X. Xiong, P. Chaikaew, J. Kim, and B. Cao. 2011. Synthesis tools for carbon assessment in ecosystems. FESC Summit, University of Florida, Gainesville, FL, Sept 28-29, 2011.
8. 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.
9. Xiong X., S. Grunwald, D.B. Myers, W.G. Harris, A. Stoppe and N.B. Comerford. 2011. Are soil carbon models transferable across distinct regions or scales in Florida? No. 262-8. Symposia Spatial Predictions in Soils, Crops and Agro/Forest/Urban/Wetland Ecosystems, ASA-CSSA-SSSA Int. Meeting, San Antonio, TX, Oct. 16-19, 2011.