

Proposal type: Postdoctoral

Descriptive title: Formulating the next generation of models of the terrestrial carbon cycle in the conterminous US

Short title: C Cycle Models

PI name and complete contact information: Trevon Fuller, Graduate Program in Ecology, Evolution, and Behavior, University of Texas at Austin, 1 University Station A6700, Biological Laboratories #311, Austin, TX 78712-0183. Phone: (512) 417-8258. E-mail: tfuller@mail.utexas.edu. Home address: 202 South Raymond Avenue #305, Pasadena, California 91105

Project summary: Comprehensive analysis published during the last three years has demonstrated that the magnitude of recent observed sea level rise and atmospheric warming exceed the forecasts associated with the highest CO₂ emission scenario constructed by the Intergovernmental Panel on Climate Change [1]. In order to provide more accurate forecasts of future temperature increases, climate models formulated during the next decade will require a more precise understanding of factors affecting atmospheric CO₂ concentrations, such as the terrestrial carbon cycle (hereafter “C cycle”) [2]. The proposed work will reanalyze TECO-R, a recently-published compartment model of the C cycle [3]. Previous work has used sub-optimal heuristic algorithms to solve TECO-R, resulting in imprecise estimates of some C residence times. The proposed work will first reformulate TECO-R as a quadratic mathematical program called TECO-R-QP. Next, I will solve TECO-R-QP using an optimal numerical algorithm to precisionize C residence time estimates for the conterminous US. Data products generated by the proposed work will include 1 km resolution raster maps of: (i) C residence times, (ii) C sources and sinks, and (iii) net primary productivity under present day climate and a range of IPCC scenarios in 2050. All maps will be made available to users for free on the World Wide Web. This work is best done at NCEAS because TECO-R-QP will be parameterized with an existing data set on C dynamics available from Oak Ridge National Laboratory (ORNL). Moreover, the proposed work will develop new techniques in mathematical and geospatial modeling, which is part of NCEAS’ mission. The geographical extent of the analysis will be restricted to the conterminous US because most of the ORNL data is for this region and because limiting the study to this area will make it feasible to complete the proposed work within two years. Finally, I will generalize TECO-R-QP by incorporating: (i) scenarios of 21st century land cover change in the Midwest developed by the EPA Office of Research and Development and (ii) uncertainty about C residence times in terrestrial pools using stochastic programming and info-gap theory [4, 5].

Expected date of Ph.D. completion: May 2009

Proposed start and end dates: June 2009 to June 2011

Proposed data release date: May 2011

Is this a resubmission? No

Conflicts of interest with reviewers: None

1 Problem statement:

1.1 What is to be done. Research objectives:

1.1.1 Using numerical optimization to precisionize estimates of C residence time:

Carbon (hereafter “C”) leaves the atmosphere and enters plants and soil via photosynthesis and exits back to the atmosphere as CO₂ via plant or microbial respiration [6]. The time that C atoms spend in an ecosystem between the entrance and exit events is referred to as C residence time [3]. C residence times are important because they affect C uptake, net primary productivity (NPP), and indirectly influence atmospheric CO₂. C residence times are typically studied using compartment models, which treat plants, soil, and the atmosphere as distinct pools among which C circulates. In the past, compartment models have been solved using heuristic algorithms, which may result in some imprecise residence time estimates [3]. The proposed work will use optimal algorithms to generate more accurate estimates of C residence times in the conterminous US at the 1 × 1 km resolution.

1.1.2 Extending compartment models to address land use/land cover change (LULC)

Objective (§1.1.1) will use optimal algorithms to analyze existing compartment models. This objective will construct new compartment models that incorporate LULC into the analysis of C dynamics among terrestrial C pools. Muller et al. [2] formulated coarse-grained (0.5° × 0.5°), global scale predictions of the effect of LULC in the 21st century on the terrestrial C cycle. The proposed work will extend the analysis of Muller et al. [2] by using a finer spatial resolution of 1 × 1 km and restricting the analysis to the conterminous US. Muller et al. [2] utilized LULC scenarios that were not specific to the US. I will use scenarios of LULC developed specifically for the US by the EPA Office of Research and Development.

1.1.3 Incorporating uncertainty into compartment models of the terrestrial C cycle

Objective (§1.1.1) will formulate and solve deterministic models of the terrestrial C cycle. The present objective will generalize those models to include uncertainty. For example, deterministic C cycle models assume that C resides in woody biomass for no more than 500 years [3]. The upper bound of 500 years is based on expert opinion and literature review but may be open to the charge of arbitrariness. The proposed work will analyze the effect of treating such parameters as random variables and will solve the resulting stochastic models using stochastic mathematical programming and info-gap decision theory [4, 5].

1.1.4 Forecasting C uptake at a fine spatial scale in response to future increases in atmospheric CO₂

Recent work has investigated spatial and temporal variation in NPP in the Great Plains of the US using linear regression to quantify the effect of climate on NPP [7, 8]. This objective will extend the analysis of Bradford et al. [7, 8], which analyzed present day climate, by forecasting NPP under future climate scenarios formulated by the IPCC. Bradford et al. used multiple linear regression to analyze the effect of climate and soil conditions on NPP. The proposed work will refit the regression models of Bradford et al. with the IPCC’s A2, B1, and B2 scenarios of temperature and precipitation in 2050.

1.2 Why it is important:

Terrestrial carbon dynamics influence atmospheric CO₂, which influences climate [2]. Climate change is likely to have important societal impacts during the 21st century, such as affecting agri-

cultural production [9]. In light of these impacts, it is important to develop more precise forecasts of climate change. One aspect of more precise forecasting is to develop a more accurate understanding of the terrestrial C cycle.

Zhou and Lou [3] have recently developed a nuanced model of the terrestrial C cycle (Fig. 1). The contribution of the proposed work to ecological theory will be to formulate the model of Zhou and Lou [3] as a mathematical program. A large body of computational techniques exist for solving mathematical programs optimally [10]. By formulating a C cycle model as a mathematical program, the proposed work will make it possible to apply those techniques to the study of C dynamics in terrestrial pools. I have already developed a preliminary version of the mathematical program, which can be downloaded from <http://uts.cc.utexas.edu/~consbio/Cons/TECO-R-QP.pdf>.

1.3 How it will be accomplished:

1.3.1 Data sources

I will reanalyze existing data sets on the terrestrial C cycle that are available from the Oak Ridge National Laboratory Distributed Active Archive Center (ORNL DAAC). Current climate data will be obtained from the PRISM data set of Oregon State University [11]. Future climate forecasts will be based on the new CO₂ emission scenarios published in 2007 as part of the IPCC's Fourth Assessment Report [12]. The University of Maryland Global Land Cover Facility's 1 km land cover raster, which is based on satellite imagery from 1981-1994, is typically used to model the dynamics of terrestrial C [13]. My analysis will use the European Space Agency's new GlobCover version 2.2 300 m land cover raster, which is based on 2004-2006 imagery [14]. All of the data sets that will be utilized in the proposed work are already available for free on the World Wide Web.

1.3.2 Using numerical optimization to precisionize estimates of C residence time (see §1.1.1)

Inverse modeling is a statistical technique that attempts to minimize an objective function (whose ideal value is zero), which represents the difference between observed data and the parameters of a model that explains the data [15]. In the context of the terrestrial C cycle, inverse modeling estimates parameters of a compartment model given observed data on plant biomass, precipitation, soil organic C, and solar radiation. Due to the computational difficulty of the estimation, parameter estimates are typically obtained using a heuristic algorithm, which provides good but possibly sub-optimal estimates. For example, Zhou and Lou [3] used a genetic algorithm to estimate spatial patterns of C residence times in the conterminous US. They solved the inverse model 30 times and reported the mean and standard deviation of the estimated residence times. A shortcoming of this approach is that different runs of the heuristic algorithm may return significantly different estimates. For instance, the mean residence time of C in coarse litter in the shrubland biome was estimated to be 3 years, but the standard deviation of the estimate was also 3 years [3]. The proposed work will use optimal numerical algorithms rather than heuristics to provide more precise estimates of C influx and efflux in compartment models. Optimal numerical algorithms have already proved effective for modeling oceanic carbon cycles [16]. I have previously used optimal algorithms to design biodiversity conservation areas and am familiar with open source C++ software implementations of such algorithms [17, 18, 19]. In the proposed work, I will use optimal algorithms to estimate terrestrial C residence times in compartment models in order to develop more accurate estimates of spatially heterogeneous C uptake.

I have already formulated a preliminary version of the optimization model that will be analyzed during the proposed work, which is called TECO-R-QP. TECO-R-QP has 11 data parameters that are set by the user prior to solving the model, 25 state variables whose values are determined by solving the model, and 40 constraints that model properties of the terrestrial C cycle. Due to the

word limit for NCEAS proposals, TECO-R-QP is not included in this proposal. However, a pdf document describing TECO-R-QP can be downloaded from my web site at the University of Texas: <http://uts.cc.utexas.edu/~consbio/Cons/TECO-R-QP.pdf>.

1.3.3 Extending compartment models to address LULC (see § 1.1.2)

A consensus has emerged that terrestrial C dynamics depend on LULC as well as plant photosynthesis and respiration (reviewed in [2]). The TECO-R model of the C cycle will be extended to incorporate LULC by adding state variables that model land-use patterns at geographic sites. I will analyze a set of spatially-explicit LULC scenarios that have already been formulated by the US Department of Agriculture [20]. Data on land cover change associated with these scenarios are already available in GIS-format from the Environmental Protection Agency's Future Midwestern Landscapes Environmental Decision Toolkit (<http://www.waratah.com/fmledt>). First, I will determine the effect of each USDA scenario on C residence times and NPP at each site in the conterminous US at the 1 km resolution. Second, I will compare the USDA scenarios with scenarios derived from LPJmL, a process-based model of vegetation dynamics that has seen frequent use in the literature on land-use change [2].

1.3.4 Incorporating uncertainty into compartment models (see § 1.1.3)

TECO-R-QP is a model of C circulation among terrestrial pools comprised of soil, roots, and leaves. TECO-R-QP is deterministic insofar as no data parameter has any uncertainty explicitly associated with it. I will extend TECO-R-QP by converting it into a stochastic mathematical program in which some parameters are treated as random variables [4]. Initial work will assume that each random variable has a known distribution function but that the exact value of the parameter will be determined by a random experiment. For example, TECO-R-QP assumes that C resides in deep soil for at most 500 years (see constraint (A39) of TECO-R-QP). I will relax this assumption by modeling the upper bound on C residence time in deep soil as a normally-distributed random variable with a mean of 500 years.

1.3.5 Forecasting C uptake at a fine spatial scale in response to future increases in atmospheric CO₂ (see § 1.1.4)

Bradford et al. [7] formulated a multiple regression model that predicts NPP using data on temperature, precipitation, soil, and land use as explanatory variables. I will construct fine-scale predictions of temperature and precipitation in the conterminous US in 2050 by downscaling IPCC scenarios from 0.5° to 1 km using ordinary kriging interpolation [21]. Next, I will use the future climate forecasts to estimate future NPP and the residence time of C in terrestrial pools. Finally, I will compare how C residence times are affected by (i) increases in atmospheric CO₂ only and (ii) increases in atmospheric CO₂ accompanied by LULC (see § 1.3.3).

2 Proposed activities:

2.1 Brief description of methods:

A mathematical program is an optimization model that takes as input data parameters [22]. Solving the mathematical program amounts to selecting values for state variables that optimize an objective function while obeying one or more constraints. In the context of compartment models of the terrestrial C cycle, the state variables represent estimates of the residence time of C in the compartments. The data parameters include information about temperature, precipitation, and soil characteristics in geographic sites. Typically, the objective function minimizes the difference between the observed data and model estimates.

The proposed work will reformulate the C cycle model of Zhou and Lou [3] as a mathematical program called TECO-R-QP. A preliminary version of TECO-R-QP is available from

<http://uts.cc.utexas.edu/~consbio/Cons/TECO-R-QP.pdf>. TECO-R-QP will be solved optimally using a modified version of the simplex algorithm for linear programs [10, 23]. The proposed work will use an implementation of the simplex algorithm in C++ that is already available as part of the COIN-OR library [24].

2.2 Why the methods are appropriate:

I have already produced peer-reviewed articles about GIS-based ecological modeling, including publications on the use of mathematical programming for biodiversity conservation planning [17, 18, 19]. I have seven years' experience with C++ programming and GIS using the ArcGIS software package. I also have experience modeling the responses of ecological communities to climate change with a focus on terrestrial vertebrates [18]. The proposed work will apply the approach of [18] to vascular plants. I have already produced an open source C++ software package for conservation planning called LQGraph (available from <http://uts.cc.utexas.edu/~consbio/Cons/ResNet.html>) [25]. My work on LQGraph demonstrates that I am capable of developing new C++ software for ecological modeling, such as the software that will be required to solve TECO-R-QP optimally. The proposed work will also require analysis of satellite imagery such as the GlobCover land cover data set [14]. I already have experience with the such analysis using the ERDAS Imagine software package. As part of my dissertation, I have utilized ERDAS to analyze satellite images (Landsat 7 scenes) in order to construct a supervised classification of five vegetation communities in central Texas using a decision tree classifier.

The proposed work will also require developing a Web portal from which users can download raster maps of C residence times in the US. I have experience developing Web sites for computational biology such as the BBCL Software Google Group (http://groups.google.com/group/bbcl_software), the Systematic Conservation Planning Primer (<http://www.consnet.org/primer>), and the ConsNet Portal (<http://consnet.org/>), which has received more than 14,000 visits since 2005. The ConsNet Web portal allows conservation practitioners worldwide to use computers at the University of Texas to run conservation planning software through a Web browser. The Systematic Conservation Planning Primer is a Web-based course on conservation biology, which has been used in undergraduate courses at the University of Texas. The Primer consists of 14 modules about different topics in conservation planning such as viability analysis, multiple criteria analysis, and place prioritization. Each module takes approximately 50 minutes and is geared toward students with a working knowledge of the basics of biology, ecology, and evolution. Each module is followed by an online quiz.

Timetable of activities:

| Activity | Summer'09 | Fall'09 | Winter'10 | Spring'10 | Summer'10 | Fall'10 | Winter'11 | Spring'11 |
|----------|-----------|---------|-----------|-----------|-----------|---------|-----------|-----------|
| §1.1.1 | X | X | | | | | | |
| §1.1.2 | | | X | X | | | | |
| §1.1.3 | | | | | X | X | | |
| §1.1.4 | | | | | | | X | X |

3 Anticipated results and benefits:

The proposed work will generate a manuscript submitted to the journal *Global Biogeochemical Cycles* as well as the following data products, which will be made available to users for free on the World Wide Web in the form of 1×1 km raster maps in May 2011: (i) a map of C sources and sinks in the US under present day climate, (ii) maps of NPP in the US in 2050 under IPCC scenarios of

atmospheric CO_2 concentrations and EPA scenarios of LULC, (iii) a map of C residence times in the plant and soil pools in the conterminous US.

The benefit of constructing maps that quantify the effect of climate change and LULC on NPP is that these maps provide important guidance to policy makers. For example, the EPA's Region 7 office is studying the possible effects of increased biofuel production in the Midwestern US. The maps generated by the proposed work could provide decision makers at EPA with forecasts of how LULC associated with increased cultivation of biofuel stocks is likely to affect the terrestrial C cycle.

C++ software for solving TECO-R-QP optimally will be made available for free on the Internet. I have experience with the development of open source software and have already written a C++ program for the design of connectivity areas to provide decision support for systematic conservation planning in Mexico [25, 26].

The proposed work is suitable for NCEAS because the development of new mathematical and geospatial techniques is part of the NCEAS mission. The proposed work will analyze existing data on terrestrial C pools. Research at NCEAS focuses on the use of cutting-edge techniques to analyze existing ecological data. The proposed work is cutting-edge because it would be the first application of optimal numerical algorithms to analyze C cycle models. The proposed work will also synthesize several existing C data sets and use modeling to garner ecological insights from these data sets.

How I heard about NCEAS Call for Proposals: The NCEAS Web site (www.nceas.ucsb.edu/).

Letters of Recommendation: Letters will be sent directly as email attachments to proposal@nceas.ucsb.edu by Chris Margules of Conservation International, Víctor Sánchez-Cordero of the National Autonomous University of Mexico, and Sahotra Sarkar of the University of Texas at Austin.

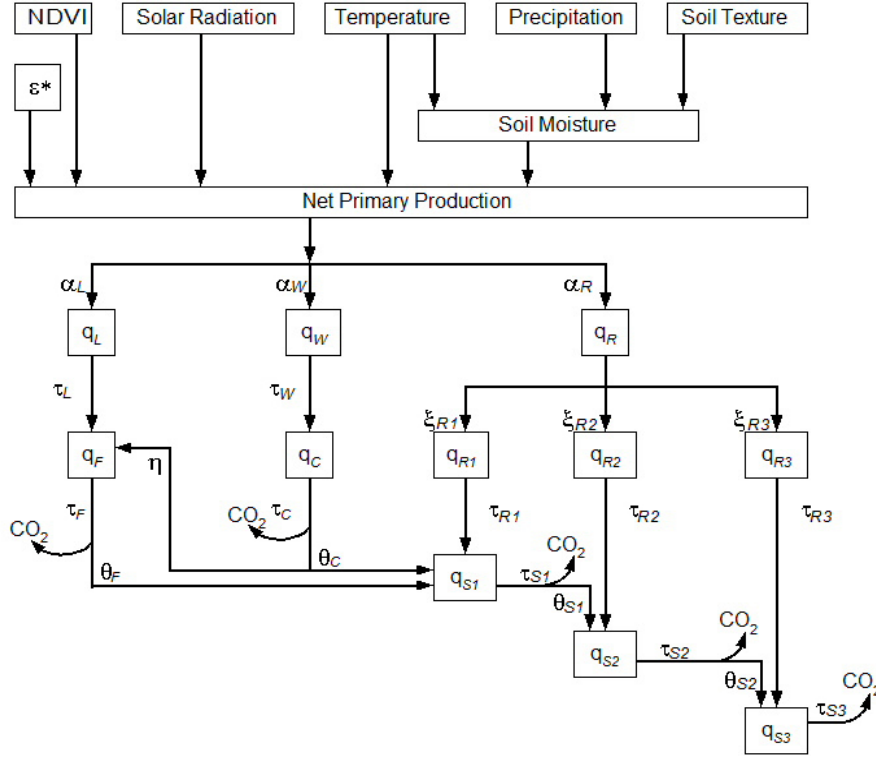


Figure 1: Terrestrial Ecosystem Regional (TECO-R) compartment model [3]. Symbols: NDVI: Normalized Difference Vegetation Index, a measure of whether a site contains live green vegetation; q_L : carbon pool in leaves; q_W : carbon pool in wood; q_R : carbon pools in roots ($R1$: roots 0–20 cm, $R2$: 20–50 cm, $R3$: 50–100 cm); q_F : carbon pool in fine litter; q_C : carbon pool in coarse litter; q_S : carbon pools in soil ($S1$: 0–20 cm, $S2$: 20–50 cm, $S3$: 50–100 cm). For definitions of the state variables α , ϵ^* , θ , ξ , and τ see the TECO-R-QP model. For legibility, the Figure omits the subscripts i , j , and k used in TECO-R-QP.

References

- [1] S. Rahmstorf, A. Cazenave, J. A. Church, J. E. Hansen, R. F. Keeling, D. E. Parker, and R. C. J. Somerville. Recent climate observations compared to projections. *Science*, 316:709, 2006.
- [2] C. Müller, B. Eickhout, S. Zachle, A. Bondeau, W. Cramer, and W. Lucht. Effects of changes in CO_2 , climate, and land use on the carbon balance of the land biosphere during the 21st century. *Journal of Geophysical Research*, 112:G02032, 2007.
- [3] T. Zhou and Y. Lou. Spatial patterns of ecosystem carbon residence time and NPP-driven carbon uptake in the conterminous United States. *Global Biogeochemical Cycles*, 22:GB3032, 2008.

- [4] J. R. Birge and F. Louveaux. *Introduction to Stochastic Programming*. Springer–Verlag, New York, 1997.
- [5] Y. Ben-Haim. *Info-Gap Theory: Decisions Under Severe Uncertainty. Second Edition*. Academic Press, London, 2006.
- [6] R. E. Ricklefs. *The Economy of Nature. Fifth Edition*. W. H. Freeman and Company, New York, 2001.
- [7] J. B. Bradford, W. K. Lauenroth, and I. C. Burke. The impact of cropping on primary production in the U.S. Great Plains. *Ecology*, 86:1863–1872, 2005.
- [8] J. B. Bradford, W. K. Lauenroth, I. C. Burke, and J. M. Paruelo. The influence of climate, soils, weather, and land use on primary production and biomass seasonality in the US Great Plains. *Ecosystems*, 9:934–950, 2006.
- [9] S. M. Howden, J.-F. Soussana, F. N. Tubiello, N. Chhetri, M. Dunlop, and H. Meinke. Adapting agriculture to climate change. *Proceedings of the National Academy of Sciences*, 104:19691–19696, 2007.
- [10] I. Maros. *Computational Techniques of the Simplex Method*. Kluwer, Boston, 2003.
- [11] C. Daly. Guidelines for assessing the suitability of spatial climate data sets. *International Journal of Climatology*, 26:707–721, 2006.
- [12] Intergovernmental Panel on Climate Change. *Working Group I Report: The Physical Science Basis*. Cambridge University Press, Cambridge, UK, 2007.
- [13] M. Hansen, R. DeFries, J. R. G. Townshend, and Sohlberg R. Global land cover classification at 1km resolution using a decision tree classifier. *International Journal of Remote Sensing*, 21:1331–1365, 2000.
- [14] European Space Agency. European Space Agency Ionia GlobCover Portal. <http://ionia1.esrin.esa.int/index.asp>, 2008.
- [15] K. J. Daun, J. R. Howell, and D. P. Morton. Design of radiant enclosures using inverse and non-linear programming techniques. *Inverse Problems in Engineering*, 11:541–560, 2003.
- [16] C. Wunsch and J.-F. Minster. Methods for box models and ocean circulation tracers: mathematical programming and non-linear inverse theory. *Journal of Geophysical Research*, 87:5647–5662, 1982.
- [17] T. Fuller, D. P. Morton, and S. Sarkar. Incorporating uncertainty about species potential distributions under climate change into the selection of conservation areas with a case study from the Arctic Coastal Plain of Alaska. *Biological Conservation*, 141:1547–1559, 2008.
- [18] T. Fuller, D. P. Morton, and S. Sarkar. Planning for biodiversity conservation using stochastic programming. In A. Deutsch, R. Bravo de la Parra, R. de Boer, O. Diekmann, P. Jagers, E. Kisdi, M. Kretschmar, P. Lansky, and H. Metz, editors, *Mathematical Modeling of Biological Systems, Volume II*, pages 101–107. Birkhäuser, Boston, 2008.
- [19] T. Fuller, V. Sánchez-Cordero, P. Illoldi-Rangel, M. Linaje, and S. Sarkar. The cost of postponing biodiversity conservation in Mexico. *Biological Conservation*, 134:593–600, 2007.

- [20] US Department of Agriculture. Effects of an expansion in biofuel demand on US agriculture. Technical report, The Office of Economic Research/The Office of the Chief Economist, USDA, 2007.
- [21] H. Wackernagel. *Multivariate Geostatistics : An Introduction with Applications. Third Edition.* Springer, Berlin, 2003.
- [22] S. P. Bradley, A. C. Hax, and T. L. Magnanti. *Applied Mathematical Programming.* Addison-Wesley, Reading, Massachusetts, 1977.
- [23] P. A. Jensen and J. F. Bard. *Operations Research Models and Methods.* Wiley, Hoboken, New Jersey, 2008.
- [24] R. Lougee-Heimer. The Common Optimization INterface for Operations Research. *IBM Journal of Research and Development*, 47:57–66, 2003.
- [25] T. Fuller and S. Sarkar. LQGraph: a software package for optimizing connectivity in conservation planning. *Environmental Modeling and Software*, 21:750–755, 2006.
- [26] T. Fuller, M. Munguía, M. Mayfield, V. Sánchez-Cordero, and S. Sarkar. Incorporating connectivity into conservation planning: a multi-criteria case study from central Mexico. *Biological Conservation*, 133:131–142, 2006.