# Estimation of Landscape Carbon Budgets: Combining Geostatistical and Data Assimilation Approaches

Luke Spadavecchia

PhD by Research The University of Edinburgh 2008

# Declaration

This thesis is a presentation of my original research work. Wherever contributions of others are involved, every effort is made to indicate this clearly, with due reference to the literature, and acknowledgement of collaborative research and discussions. Contributions from other authors are acknowledged prior to the relevant sections. The work was done under the supervision of Mathiew Williams at the University of Edinburgh.

Luke Spadavecchia

Date:

# Acknowledgements

I would like to thank all of my friends and family for their patience over the last few years; without their support this would never have been possible. In particular I would like to thank my parents, whose love and encouragement kept me going. I would like to thank Tim Hill and Casey Ryan for good company, organizing me when I was hopeless, and providing a sympathetic ear for a good moan. I owe a huge debt of gratitude to Mat Williams, my supervisor, who tolerated my somewhat erratic work methods and kept me on track when I was distracted. Much of the quality of my work is due to his guidance and without him I never would have achieved this. Throughout my PhD I benefited from the friendship and expertise of everyone at the Centre for Terrestrial Carbon Dynamics (CTCD), and was lucky to be surrounded by such talented and stimulating people: In particular Martin De Kauwe and Steve Hancock. I benefited from the hospitality and kindness of many people at Oregon State University, particularly James and Kathryn Irvine and Bev Law. Finally I would like to thank my wonderful girlfriend Danielle, for putting up with me at my worst, and helping me forget work when things were tough.

# Contents

# 1. Abstract

Carbon fluxes at the site scale (~ $1km^2$) are well quantified by continuous monitoring with eddy flux covariance instruments, whilst national to continental scale fluxes may be measured by tall towers or flask measurements. Quantification of carbon (C) budgets at the landscape or catchment scale is more problematic, and is generally achieved using process-based models as scaling tools. Such models require some metric of the exchange surface capability (e.g., Leaf Area Index, LAI) and a set of rate parameters for C processing. The net C exchange is then determined by driving the model with meteorological observations. Regional fields of parameters and drivers may be derived by upscaling site level measurements, constrained using Earth Observation (EO) data such as radiance derived vegetation indices and digital elevation models (DEMs). I explore issues of error and uncertainty when upscaling C model parameters and drivers, and the effect of these uncertainties on the final analysis of the carbon budget. Two study areas, with excellent research infrastructure, focus the research: a region of tundra in Arctic Sweden and a ponderosa pine stand in Oregon. I use geostatistical techniques to develop fields of LAI and meteorology, complete with error statistics, whilst the distributions of rate parameters for a C model are derived *via* the Ensemble Kalman filter (EnKF). I report that the use of DEM data can provide LAI fields with an $r^2$ ~50% greater than those derived from EO data alone. In particular I find strong relationships between LAI, elevation and topographic exposure. I explore the use of spatio-temporal geostatistics to improve meteorological fields, but report a better interpolation skill when temporal autocorrelations are ignored. I employ simulation techniques to propagate parameter and driver uncertainty through a simple carbon dynamics model, finding that variation in parameters has a much larger effect on the uncertainty of the carbon budget (~50%) than driver uncertainty (<10%). Whilst driver uncertainty is related to the quantity and spatio-temporal arrangement of the conditioning data, we find this result to be stable in cases of extreme data scarcity (max driver uncertainty <20%). The combined uncertainty in parameterisation and meteorology may result in a 53% uncertainty in total C uptake. I conclude that improved methods to constrain vegetation surface characteristics on the regional scale should take precedence over improvements to model drivers: It is likely that data assimilation of high quality EO products may go some way to providing such constraint.

# 2. Introduction

In the past, carbon cycle research depended on site level experiments or observational studies of ecosystems to make local assertions about carbon budgets (Grace 2004). Typically these were based around micro-meteorological eddy flux covariance methods (Baldocchi et al. 1988, Grace et al. 1995a, Grace et al. 1995b, Moncrieff et al. 1997) at the stand to forest scale, or flask sampling giving information at the continental to global scale (Keeling et al. 1996a, Keeling et al. 1996b). Later attempts to formalise this knowledge for hypothesis testing led to increasing focus on modelling studies to understand ecosystem dynamics on a global scale and typically over long time periods *via* Dynamic Global Vegetation Models (DGVMs) (Woodward et al. 1995, Sitch et al. 2003).

DGVMs are generally heuristic, and do not aim to match site level data accurately, but rather describe likely system behaviour in the event of various scenarios (e.g. IPCC 2007). As such, the system can be 'spun up', typically with synthetic meteorology generated within the model, to generate surface vegetation characteristics. This approach assumes an ecosystem in equilibrium, with vegetation settled at some 'climax community', in accordance with Clement's view of succession and plant community structure (Clements 1916, 1936): Such notions of climax communities have been challenged, and 'non-equilibrium' concepts associated with Gleason (Gleason 1927) are generally considered to be more appropriate at the regional scale. Large-scale models have been useful tools for predictions of future climate and vegetation states, and for exploring system behaviour under different sets of assumptions (Cox et al. 2000, Cramer et al. 2001), but have caveats due to the omission of key feedbacks; particularly in terms of soil nutrient dynamics, e.g. the relationship between the decomposition rate of soil organic carbon and soil nitrogen availability (Henriksen and Breland 1999).

More recently, modelling studies have focused on regional to catchment scale studies of the carbon balance (Running 1994, Williams et al. 2001). These models may be simpler in terms of the number of processes represented in the model structure, but tend to be better at matching site level observations over fairly short time periods (~3 years). At the regional scale, these models are used as a scaling tool to implement knowledge gained at the site level to a larger region of interest. Interest in local ecosystem potential as C sinks and their behaviour in response to climatic variability is increasing, particularly with a view towards sequestration and climate change mitigation.

Regional scale modelling presents a different set of challenges to global scale modelling: Local scale models must accurately reproduce the C fluxes observed at the site level, to provide both diagnostic and prognostic information on regional dynamics of C. Furthermore, at the regional scale, effects of micro-topography and vegetation surface heterogeneity which are irrelevant at the global scale have an appreciable effect on the C balance. As such the synthetic 'spin-up' methods employed in DGVMs are inappropriate; we therefore require a set of meteorological driving variables, and some conception of the vegetative surface at an appropriate scale to derive estimates of C dynamics. Typically we utilise a combination of site level observational data and earth observation (EO) products to parameterise the vegetation surface and derive fields of meteorological drivers. There are unavoidable errors inherent in the up/downscaling of observational data sources, which are often poorly quantified, or not considered in regional modelling studies (Fuentes et al. 2006).

Quantification of errors is becoming increasingly important for C modelling. Current trends in research towards data assimilation (DA) and data fusion techniques (e.g. Williams et al. 2005, Quaife et al. 2008) require some knowledge of model and data uncertainty, which are often difficult to quantify. DA techniques are the next logical step in the development of our understanding of the C cycle, as they allow the use of formalised knowledge in the form of a model to flag and correct aberrant observational data, whilst allowing better integration of site level and satellite derived ecological observations into such models in a way which optimally balances the errors of each. Such methods effectively bridge the gap between field ecologists and modellers, and provide a better analysis than either model or data alone (Maybeck 1979, Williams et al. 2005). In order to achieve an unbiased estimate of the system state, DA requires an accurate estimate of model variability, without which the results may be highly questionable (Quaife et al. 2008). It is the goal of accurate model uncertainty analysis that motivates this thesis.

Quantification of regional model errors is not only an academic exercise. Political decisions to achieve binding emission reduction targets (Kyoto protocol) through offsetting have led to a growing C trading market (Grace 2004); reflected in the recent restructuring of the National Environmental Research Councils (NERC) Earth observation centres to the National Centre for Earth Observation (NCEO), which has an objective towards developing commercial deliverables to customers from research. To be truly useful, such products must have some form of error quantification (Kennedy et al. 2008), and this is likely to be a profitable area of research in the future.

In this thesis I aim to quantify and reduce the errors associated with the production of regionalised data sources for the parameterisation and driving of models. I employ geostatistical techniques to the problem of upscaling, which confer the considerable advantage of providing estimates of uncertainty to estimated fields. Furthermore, I aim to examine the effects of these errors on the state vector when propagated through a simple carbon dynamics model (DALEC). Issues related to parameterisation of the vegetation surface are tackled in chapters 3 and 4, whilst issues related to the estimation of driver fields are addressed in chapters 5 to 6. An examination of error propagation is undertaken in chapter 6.

Chapter three was published as a paper in *Global Change Biology* (Williams et al. 2008), and aims to quantify the errors associated with upscaling leaf area index (LAI) from site level harvest data to the regional scale in an arctic tundra ecosystem. Correctly specifying LAI is critical to the quantification of the carbon balance (Sitch et al. 2003) because (along with foliar nitrogen content) it dictates the rate of exchange of mass and energy between the land and atmosphere by defining the total exchange surface. LAI can be inferred from EO reflectance data *via* vegetation indices such as the Normalised Difference Vegetation Index (NDVI) (Lillesand et al. 2004). Issues of scale invariance for relationships of NDVI and LAI are explored. We found that for a relatively large range of spatial scales, the same relationship between LAI and NDVI held, with similar prediction error. However, we are only able to capture ~ 17% of the LAI variation with EO data sources, indicating the use of EO data alone may be insufficient to parameterise the vegetation surface in highly heterogeneous areas. This result motivated the chapter four, in which we attempted to find suitable topographic predictors of LAI to support or supplant the use of EO data for vegetation surface parameterisation.

Chapter four was submitted as a paper to the *Journal of Ecology*, and explores the spatial relationships between topography and variation in LAI in an arctic tundra ecosystem. We report significant scale dependent relationships between LAI, elevation and topographic position, indicating that at larger spatial scales LAI is constrained by elevation (perhaps due to temperature variation), whilst topographic exposure dominates the spatial patterns of vegetation at smaller scales. The effect of topographic exposure on LAI is likely due to wind shear, and shelter effects on snow accumulation and melt. Geostatistical techniques were used to build simple spatially explicit models of LAI variation with relevant topographical characteristics, better replicating the observed vegetation characteristics than EO sources ($r^2$ ~ 30%). Future development in this area may integrate EO derived NDVI and vegetation classifications with data on surface topography to provide more accurate LAI parameterisations complete with error statistics.

Chapter five was submitted as a paper to *Agricultural and Forest Meteorology*, and aims to address the issue of spatio-temporal autocorrelation in meteorological data sources, and how this information can be potentially exploited to improve estimation of model driver fields for a moderately large region of central Oregon, USA. The paper also explores the effects of temporal aggregation on error magnitude and bias. We employ the product-sum representation of spatiotemporal covariance (De Cesare et al. 2001) to meteorological upscaling problems for the first time. Interestingly, incorporation of temporal autocorrelation did not improve the accuracy of driver fields over utilisation of spatial data sources alone. However, we report that *post hoc* temporal aggregation of high-resolution estimates tends to reduce their bias and error. The likely consequences of this in terms of model error propagation are uncertain, as some model processes react instantaneously to driving variable, whilst others act as capacitors, integrating driver error over longer time periods. These results provide the motivation for chapter six.

Chapter six is intended for submission to *Global Change Biology*, and aims to quantify and compare model uncertainties resultant from parameter and driver uncertainties respectively. The paper utilises DA techniques to parameterise a simple model of C dynamics for an intensive observation site at Metolius, central Oregon.

Parameters are derived *via* the Ensemble Kalman filter (ENKF) (Evensen 2003), to construct optimal parameter distributions. The variation in C fluxes due to parameter uncertainty is derived, and compared with the uncertainties resultant from meteorological driver uncertainty. Driver uncertainty is quantified using geostatistical simulation techniques (Sequential Gaussian Simulation, SGS) (Goovaerts 2001), whereby an ensemble of 1000 weather scenarios is produced. We also undertake a series of experiments to disaggregate the errors resultant from temperature and precipitation uncertainty.

We find that parameter error dominates the total C sink strength uncertainty, despite the considerable uncertainties associated with upscaling meteorology. In order to assess the robustness of this conclusion we examine the effect of conditioning the simulated meteorology on increasingly remote sets of stations. We report that in cases of extreme data sparsity, conditioning the meteorology on stations over 100km from the study site, the effect parameter uncertainty still exceeds the effect of meteorological uncertainty on NEE by ~50%. Disaggregation of the driver uncertainty reveals that temperature variability has a larger impact on total C sink uncertainty than precipitation. Interestingly we find that biases in simulated meteorological drivers appear to cancel out over model runs, although further research at other sites is needed to rule out the possibility of this occurring by chance. We conclude that producing reasonable parameterisations over the study area is of greater importance than reducing driver uncertainty.

All geostatistical analyses presented in this thesis were conducted using a set of software tools developed specifically for this thesis. This set of software tools, the *Edinburgh Spatio-Temporal Geostatistics* package is documented in chapter seven, which serves as a technical paper and manual for the software, and also provides an exposition and justification of some of the modelling choices made in the thesis. The Fortran 90 code is provided in the appendix in digital format.

## 2.1 References

Baldocchi, D. D., B. B. Hincks, and T. P. Meyers. 1988. Measuring biosphere-atmosphere exchanges of biologically related gases with micrometeorological methods. Ecology 69:1331-1340.

Clements, F. E. 1916. Plant Succession: An analysis of the development of vegetation. Carnegie Institute Washington Publication 242, Washington D.C.

Clements, F. E. 1936. Nature and structure of the climax. The Journal of Ecology 24:252-284.

Cox, P. M., R. A. Betts, C. D. Jones, S. A. Spall, and I. J. Totterdell. 2000. Acceleration of global warming due to carbon-cycle feedbacks in a coupled climate model. Nature 408:184-187.

Cramer, W., A. Bondeau, F. I. Woodward, I. C. Prentice, R. A. Betts, V. Brovkin, P. M. Cox, V. Fisher, J. Foley, A. D. Friend, C. Kucharik, M. R. Lomas, N. Ramankutty, S. Sitch, B. Smith, A. White, and C. Young-Molling. 2001. Global response of terrestrial ecosystem structure and function to CO2 and climate change: results from six dynamic global vegetation models. Global Change Biology 7 357-373.

De Cesare, L., D. E. Myers, and D. Posa. 2001. Estimating and modeling space-time correlation structures. Statistics & Probability Letters 51:9-14.

Evensen, G. 2003. The ensemble Kalman filter: theoretical formulation and practical implementation. Ocean Dynamics 53:343-367.

Fuentes, M., T. G. F. Kittel, and D. Nychka. 2006. Sensitivity of ecological models to their climate drivers: Statistical ensembles for forcing. Ecological Applications 16:99-116.

Gleason, H. A. 1927. Further views on the succession-concept. Ecology 8:299-326.

Goovaerts, P. 2001. Geostatistical modelling of uncertainty in soil science. Geoderma 103:3-26.

Grace, J. 2004. Understanding and managing the global carbon cycle. Journal of Ecology 92:189-202.

Grace, J., J. Lloyd, J. Mcintyre, A. Miranda, P. Meir, H. Miranda, J. Moncrieff, J. Massheder, I. Wright, and J. Gash. 1995a. Fluxes of Carbon-Dioxide and Water-Vapor over an Undisturbed Tropical Forest in South-West Amazonia. Global Change Biology 1:1-12.

Grace, J., J. Lloyd, J. Mcintyre, A. C. Miranda, P. Meir, H. S. Miranda, C. Nobre, J. Moncrieff, J. Massheder, Y. Malhi, I. Wright, and J. Gash. 1995b. Carbon-Dioxide Uptake by an Undisturbed Tropical Rain-Forest in Southwest Amazonia, 1992 to 1993. Science 270:778-780.

Henriksen, T. M., and T. A. Breland. 1999. Nitrogen availability effects on carbon mineralization, fungal and bacterial growth, and enzyme activities during decomposition of wheat straw in soil. Soil Biology and Biochemistry 31:1121-1134.

IPCC. 2007. IPCC Fourth Assessment Report: Climate Change 2007. Published for the Intergovernmental Panel on Climate Change [by] Cambridge University Press, Cambridge.

Keeling, C. D., J. F. S. Chin, and T. P. Whorf. 1996a. Increased activity of northern vegetation inferred from atmospheric CO2 measurements. Nature 382:146-149.

Keeling, R. F., S. C. Piper, and M. Heimann. 1996b. Global and hemispheric CO2 sinks deduced from changes in atmospheric O-2 concentration. Nature 381:218-221.

Kennedy, M., C. Anderson, A. O'Hagan, M. Lomas, I. Woodward, J. P. Gosling, and A. Heinemeyer. 2008. Quantifying uncertainty in the biospheric carbon flux for England and Wales. Journal of the Royal Statistical Society Series a-Statistics in Society 171:109-135.

Lillesand, T. M., R. W. Kiefer, and W. Chipman Jonathan. 2004. Remote sensing and image interpretation, 5th edition. John Wiley & Sons, New York.

Maybeck, P. S. 1979. Stochastic Models, Estimation and Control. Academic Press, New York.

Moncrieff, J. B., J. M. Massheder, H. deBruin, J. Elbers, T. Friborg, B. Heusinkveld, P. Kabat, S. Scott, H. Soegaard, and A. Verhoef. 1997. A system to measure surface fluxes of momentum, sensible heat, water vapour and carbon dioxide. Journal of Hydrology 189:589-611.

Quaife, T., P. Lewis, M. D. Kauwe, M. Williams, B. E. Law, M. Disney, and P. Bowyer. 2008. Assimilating canopy reflectance data into an ecosystem model with an ensemble Kalman filter. Remote Sensing of Environment 112:1347-1364.

Running, S. W. 1994. Testing forest-bgc ecosystem process simulations across a climatic gradient in Oregon. Ecological Applications 4:238-247.

Sitch, S., B. Smith, I. C. Prentice, A. Arneth, A. Bondeau, W. Cramer, J. Kaplan, S. Levis, W. Lucht, M. Sykes, K. Thonicke, and S. Venevsky. 2003. Evaluation of ecosystem dynamics, plant geography and terrestrial carbon cycling in the LPJ dynamic vegetation model. Global Change Biology 9:161–185.

Williams, M., R. Bell, L. Spadavecchia, L. E. Street, and M. T. van Wijk. 2008. Upscaling leaf area index in an Artctic landscape using multi-scale reflectance observations. Global Change Biology: DOI: 10.1111/j.1365-2486.2008.01590.x.

Williams, M., E. B. Rastetter, G. R. Shaver, J. E. Hobbie, E. Carpino, and B. L. Kwiatkowski. 2001. Primary production of an arctic watershed: An uncertainty analysis. Ecological Applications 11:1800-1816.

Williams, M., P. A. Schwarz, B. E. Law, J. Irvine, and M. R. Kurpius. 2005. An improved analysis of forest carbon dynamics using data assimilation. Global Change Biology 11:89-105.

Woodward, F. I., T. M. Smith, and W. R. Emanuel. 1995. A global primary productivity and phytogeography model. Global Biogeochemical Cycles 9:471–490.