Violet Buxton-Walsh

EPS229: Introduction to Climate Modeling

Prof. Inez Fung

7 May, 2021

Parameterizing Land Use for Predictive Models of Net Ecosystem Exchange

Project Scope

The goal of this project is to discover the feasibility of improving models of ecosystem carbon fluxes through the inclusion of land use history parameters. Previous work has documented the potential utility of reconstructing historical land cover, and discussed the variety of qualitative and quantitative techniques which support the creation of land use models (Yang 2014). Here, existing models of ecosystem fluxes and their relationship to flux predictions based on FLUXNET's in situ measurements are taken from Jung et al. 2011, and examined with the intention of better aligning model output with FLUXNET observations.

Methods to quantifying land use have a variety of motivations. These range from economic interests which predict land value as a proxy for likelihood of conversion to agricultural or other purposes, models which aim to quantify past effects of human influence on a landscape, and models which intend to further develop a subset of a global climate model, thereby improving the work of climate and ecosystem scientists (Chang et al. 2015; Hurtt et al. 2011; Brown 2000).

This project falls under the later, and attempts a preliminary assessment of land use consideration might improve estimation of Net Ecosystem Exchange.

Unfortunately, the incorporation of detailed land use information is unlikely to become scaleable in the way which other aspects of flux prediction and climate models are. Given the highly local nature of land use history assessment, it is not currently feasible to include exhaustive surveys of site history in every FLUXNET data set as would be required to provide information gathered for this project. However, for flux tower sites noted to measure artificially depressed NEE values as a result of night time stability effects on advection, this sort of more detailed examination could be applied. Furthermore, there may be a possibility of including a more limited set of parameters found to be most helpful for aligning model predictions with observed fluxes. Broadly, the results of Jung et al. 2011 demonstrate the need for better inclusion of land use history in models used to predict NEE.

The area of land use considered for this project is a 0.5 by 0.5 degree grid cell with respect to the tower site. Since this is the scale on which Jung et al. 2011 make their prediction of NEE, it seems necessary to scale land-use information similarly. The grid cells in this project are only established for three of the many flux tower sites included in Jung et al., and are not projected globally, so they do not necessarily overlap perfectly with the land area being considered in their results. It is also relevant that the hyperlocal land use history of a single flux footprint may affect the measured fluxes at that site, but it is unlikely to yield meaningful conclusions about the carbon storage capacity and dynamics of a broader ecosystem unless a larger area is considered.

As such, this maximum resolution of NEE and environmental data helpfully reaches the minimum scale necessary to attain a realistic historic data set (about 130-140km^2 for the sites in questions).

This approach holds cross-disciplinary interest, as paleoecologists and geographers are likely to find independently useful information in land use data sets collected for the purpose of parameterizing NEE predictions at flux tower sites. Conversely, existing land use data can be applied for this purpose, hopefully yielding new insights.

Introduction

This project explores the possibility of incorporating land-use parameters in global scale calculations of carbon and energy fluxes. Previous studies have succeeded in predicting the global values of Latent Heat (LH), Sensible Heat (S), Gross Primary Productivity (GPP), Net Ecosystem Exchange (NEE), and Terrestrial Ecosystem Respiration (TER) given by independent remote sensing methods using predictive models trained on uscaled climate and meteorological data. The results of these models yield global values of ecosystem carbon exchange metrics, which scale well with in-situ observations of carbon exchange derived from flux partitioning. Models of this type are useful for deriving global measurements of carbon fluxes, and validating methods for parameterizing such estimates with in situ observations. However, these models do not perform satisfactorily in predicting Net Ecosystem Exchange, tending to yield higher values for NEE than those given by eddy covariance (Jung et al. 2011).

NEE measures the net exchange of carbon between an ecosystem and the atmosphere per unit ground area, and gauges the strength of a particular site as a carbon sink (Kramer et al., 2002). NEE differs from GPP and TER because it requires accurate estimation of more than one ecosystem process. Importantly, GPP only considers the amount of carbon fixed by photosynthesis (NASA), and TER only represents the carbon fluxes from the ecosystem to the atmosphere which are produced by respiration. Because NEE is represented as the difference between GPP and TER, errors in the estimation of either effect NEE. These errors could result from either model overprediction of NEE, or eddy covariance underprediction, and distinguishing between these possibilities is pivotal in shaping the decision of how to address uncertainties. For this reason, I consider the possible drivers of this result below.

Error Source 1: FLUXNET Data

There are several ways which eddy covariance data might yield underpredictions of Net Ecosystem Exchange, as the processing and quality assurance of FLUXNET data is a complex process. Generally errors in flux data fall into two categories, one being random errors produced by instrumentation, biased measurements, or uncertainty, and the other being systematic errors which occur evenly across a system and show bias, but do not affect net flux measurements (Baldocchi 2021- ESPM228 Lecture: Eddy Covariance pt. III).

NEE estimates in particular can be affected by selective systematic errors which result from periods with low turbulence and limited air mixing at night, yielding consistently low measurements of TER that produce underestimates of NEE. In such cases, CO2 can be either removed by advection or stored in the canopy air, where it may not be measured unless a CO2 profile is calculated in addition to measurements at tower height (Papale et al. 2006). Previous work has shown instances of NEE underestimation resulting from nighttime absorption of CO_2 by open water in wetland ecosystems, or varying by forest age (ESPM 228 Lecture 5).

Potential for systematic error also arises during gap-filling, which is regularly performed on flux data in order to increase the number of points of the time series for which flux values are known. In fact, there is debate on whether CO_2 exchange data should be regularly gap-filled using corrections based on the friction velocity of the atmosphere u* at all, as most flux variables are (Falge 2001). Generally u* is auto correlated with CO_2 exchange, and higher values of u* produce similar trends with overall higher values for NEE (ESPM 228 Lecture 6).

In considering possible sources of model underestimation of NEE, it's important to acknowledge the complexity of flux measurements and data processing. Through instrumentation, experimental setup, and data processing, eddy flux measurements can underestimate carbon fluxes, particularly at night, leading to artificially lower values for NEE. June et al. 2011 make note of this fact, and are explicit in their selection and processing of flux data, being conscious to exclude gap-filled data with low confidence, aggregate data monthly to reduce random errors, and perform energy balance closure corrections based on the Bowen ratio (Monteith 2013).

Error Source 2: Selecting and Training MTEs

A second cause of poor model performance for NEE maybe the structure of the model itself. June et al. use a model tree ensemble (MTE) which readily accepts categorical and continuous variables and is designed to minimize Mean Squared Error (MSE). Seeing as the MTE uses the Schwarz to minimize overfitting and yields extremely high R^2 values when interpolating across sites, this seems least likely to be the source of error in the NEE estimates (Jung et al. 2009).

Error Source 3: Climate/Meteorological data

The final source of error in NEE predictions is the meteorological and climate data on which the model is trained to estimate NEE, and compared to flux measurements. The values predicted from models trained on this type of data in Jung et al. tend to be higher than those predicted by flux data for previously discussed reasons.

Jung et al. uses the explanatory variables of photosynthetically active radiation fractions (fPAR), monthly climate variable means, and vegetation cover (IGBP class), and this project uses a simplified subset of these. While Jung et al. finds this is effective at producing accurate empirical results for GPP and TER, they struggle to predict NEE, and would perhaps find it better represented by variables which affect climatic trends rather than climatic means (Piao et al. 2009). Additionally, known determinants of NEE including legacy carbon pool, assimilation rates, biomass storage, disturbance, ecosystem age, management activity, and land use are not included in the explanatory variables (Blasko 2015). Complete heterotrophic respiration may also be underestimated in this model (Jung et al. 2011).

Study Design

I consider the land use and flux measurements from three flux tower sites across two types of ecosystem. Vaira and Tonzi Ranches are central California agricultural sites with similar climates, while ME4 represents an aged coniferous forest in Oregon. The first two sites share many ecosystem similarities including local history of agriculture and ranching. Data availability was a constraining factor for site selection. By considering two agricultural sites, I hope to be able to compare the effects of agricultural management practices on NEE predictions. By including a forested site, I hope to observe the difference in NEE driven by carbon pools resulting from ecosystem type.

Understandably, I cannot reproduce the multi-decadal global model simulations given in Jung et al. due to time and computational limits, so the results of my inclusion of land use cannot be fairly compared to previous NEE modeling work. For this reason, I establish a simplified model of ecosystem attributes surrounding flux sites based on the model of Jung et al. 2011, and use it as an independently established benchmark based on which I can compare the effect of parameterizing land use. This model is derived based on one year of data at monthly resolution from the 0.5 by 0.5 degree region surrounding the four flux sites I consider. The attributes included are listed in Table 1, and are inspired by those in Jung et al. 2011, but exclude parameters only possessing yearly variability. I also use monthly NEE data from the LPJml global biosphere simulation model. The model scripts, output, and detailed metrics used to score and improve the model can be found in the concurrently submitted model scripts.

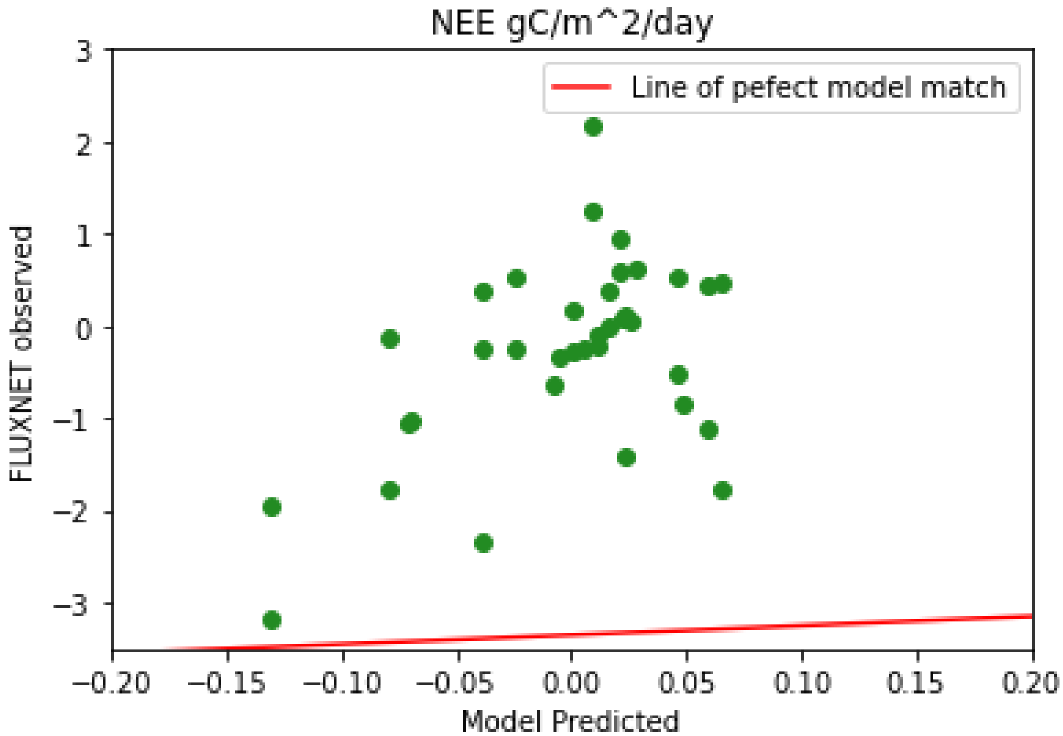
TABLE 1	Variable type	TABLE 2	Variable type
x-vars	------	added x-vars	------
Maximum fPAR	yearly	Years since disturbance	yearly, continuous
Minimum fPAR	yearly	Agricultural	binary
Average Monthly fPar	yearly	Ecosystem age	yearly, continuous
Monthly temperature	monthly	Fire adapted	binary
Monthly precipitation	monthly	Land value USD/m^2	yearly, continuous
IGBP vegetation type	static	------	------
y-vars	------	y-vars	------
NEE	monthly	NEE	monthly

Table 1 & 2. Data used to train Random Forest Regressor Model in NEE predictions. Table 1 represents the baseline model and Table 2 represents additional parameters designed to account for land use history.

Methods

To compare the effect of including land use in a predictive model of NEE to the results of Jung et al. 2011, a similar predictive model was first replicated using the explanatory variables found in Table 1. This entailed the interpretation of a number of data sets. First, information on the fraction of photosynthetically active radiation, from which a number of features were computed on the basis of Jung et al. including yearly average, maximum, and minimum fractions for each site (Vermote 2019). Additionally, model derived predictions of NEE were taken from 2013 observations under the North American Carbon Project (Wei 2013). Finally, in situ observations of precipitation and air temperature were derived directly from FLUXNET data sites. While the use of FLUXNET in situ data observations in training the model is questionable, seeing as FLUXNET also provides the target for model predictions, this strategy was adhered to in accordance with the practice of Jung et al 2011. The alternative for obtaining in situ environmental variables and avoiding downscaling would require appropriately located measurement stations across the grid cell over which one can interpolate. This information is not usually available for FLUXNET sites, and interpolation was not a realistic possibility for those sites included in this investigation based on the limited availability data. Having derived several explanatory variables for NEE, a Random Forest Regression model was implemented based on the training data set in Table 1. with the ability to bootstrap resample the data, and hyperparameter adjustments to avoid overfitting. The resulting model explained 92% of the variance in the training dataset, and gave 70% predictive accuracy. However, model predictions had a correlation of approximately 0.3 with the FLUXNET predicted values of NEE compared to a low value of 0.6 in Jung et al. While this model yielded lower accuracy and worse correlation with FLUXNET values than Jung et al. 2011, the predicted trend is the same, indicating the replicability of their model result and the need for refinement of NEE predictions. In a second iteration of the model, addition parameters were included, as seen in Table 2. Results of this implementation and a comparison of the two models is given below.

Figure 1. Model prediction of NEE in comparison to FLUXNET observed NEE values, first model iteration not including parameters for land use from Table 2.



To define parameters for the second model iteration relevant land use history was conducted

covering the last 500 years, noting average fire return intervals, the most recent regional disturbances, and the apparent ecological age of the site (Baldocchi, Field Notes; Law, Field Notes; Argawal 2002; Verburg 2006). Noted but eventually discarded parameters including the dominant photosynthetic pathway and size of soil carbon storage. Each was particularly difficult to estimate for the sites selected, or, was completely collinear with another variable (e.g. vegetation type) due to the small sampling of sites. In a larger study, the inclusion of more land use variables and parameters would likely be beneficial, and not result in excessive overfitting.

Results

After incorporation of several land use variables and adjusting the model, the final result was not dissimilar from the initial model, indicating a poor accounting of factors affecting NEE. Inclusion of the additional parameters in Table 2. led to a final correlation of 0.35, increased by 0.05 from the original model. One probable reason for this result is the amount of data used to train and fit the model. Aggregation of three flux tower sites into monthly means resulted in 36 data points with 14 attributes each. Ultimately, this is a relatively small dataset for which to implement a random forest, and better results would likely be attained running the same analysis for more flux tower sites over a wider range of years to decrease the probability (Paper 2019).

Future improvement to modeling results should include examination of model behavior across all ecosystem fluxes, and not only NEE. This single flux was examined for the sake of model simplicity, but excluding these other carbon flux behaviors limits the extent to which the baseline model developed by this project can be used as a reference for model performance which includes land use. The conclusion that these results resemble the lack of strong correlation between modeled NEE and FLUXNET predicted NEE found by Jung et al. could be better validated if findings for LE, H, GPP, and TER were also replicated. Naturally, improved data filtering with the intensity of Jung et al. would likely also improve model performance (Jung et al. 2009; Jung et al. 2016).

It would be especially interesting to consider the possibilities opened by access to more computational power, seeing as speed and operability of netCDF files is commonly noted as a bottleneck for researchers utilizing local machines with inefficient processors and limited memory.

Outcome

My goal in conducting this project was to gain practical familiarity with processing FLUXNET data and to practice computational modeling. After taking Prof. Dennis Baldocchi's Micrometeorology course I found this project to be a useful application of my newfound knowledge of the theory and systems supporting carbon flux monitoring. As a student in the Geography department, I also thought my familiarity with qualitative elements of land use history in the American West would be an asset to the modeling process. Practical outcomes of this work were a newfound familiarity with processing netCDF files in Python, and appreciation of the immense computational intensity of climate modeling problems.

Acknowledgements

I would like to thank D. Krasheninnikov and F. Walsh for running my final model code on their faster processors. Additional thanks to Prof. Dennis Baldocchi (University of California, Berkeley) and Prof. Bev Law (Oregon State University) for the use of their eddy covariance data sets, as available through AmeriFlux and FLUXNET. Lastly, many thanks to Prof. Inez Fung for the opportunity to further study climate science and modeling.

References

Agarwal, Chetan; Green, Glen M.; Grove, J. Morgan; Evans, Tom P.; Schweik, Charles M. 2002. A review and assessment of land-use change models: dynamics of space, time, and human choice. Gen. Tech. Rep. NE-297. Newton Square, PA: U.S. Department of Agriculture, Forest Service, Northeastern Research Station.

Albani, M., Medvigy, D., Hurtt, G.C. and Moorcroft, P.R. (2006), The contributions of land−use change, CO2 fertilization, and climate variability to the Eastern US carbon sink. Global Change Biology, 12: 2370-2390.

Baldocchi, Dennis, ESPM 229: Advanced Methods in Biometeorology & Micrometeorology, UC Berkeley, 2021.

Blaško, Róbert , Lisbet Holm Bach, Stephanie A. Yarwood, Susan E. Trumbore, Peter Högberg, Mona N. Högberg, Shifts in soil microbial community structure, nitrogen cycling and the concomitant declining N availability in ageing primary boreal forest ecosystems, Soil Biology and Biochemistry, Volume 91, 2015, Pages 200-211.

D.G Brown, B.C Pijanowski, J.D Duh, Modeling the relationships between land use and land cover on private lands in the Upper Midwest, USA, Journal of Environmental Management, Volume 59, Issue 4, 2000, Pages 247-263.

Chang-Martínez, Laura, et al. "Modeling Historical Land Cover and Land Use: A Review From Contemporary Modeling." ISPRS International Journal of Geo-Information, vol. 4, no. 4, 2015, pp. 1791–1812., doi:10.3390/ijgi4041791.

Hurtt, G.C., Chini, L.P., Frolking, S. et al. Harmonization of land-use scenarios for the period 1500–2100: 600 years of global gridded annual land-use transitions, wood harvest, and resulting secondary lands. Climatic Change 109, 117 (2011).

Jung, Martin, et al. "Correction to 'Global Patterns of Land-Atmosphere Fluxes of Carbon Dioxide, Latent Heat, and Sensible Heat Derived from Eddy Covariance, Satellite, and Meteorological Observations.'" Journal of Geophysical Research: Biogeosciences, vol. 117, no. G4, 2011.

Jung, M., et al. "Supplementary Material to 'Towards Global Empirical Upscaling of FLUXNET Eddy Covariance Observations: Validation of a Model Tree Ensemble Approach Using a Biosphere Model.'" 2009.

Jung, M., M. Reichstein, and Alberte Bondeau. "Towards global empirical upscaling of FLUXNET eddy covariance observations: validation of a model tree ensemble approach using a biosphere model." Biogeosciences 6.10 (2009): 2001-2013.

Paper, David. "Scikit-Learn Regression Tuning." Hands-on Scikit-Learn for Machine Learning Applications, 2019, pp. 189–213., doi:10.1007/978-1-4842-5373-1_7.

"Review on 'Understanding the Representativeness of FLUXNET for Upscaling Carbon Flux from Eddy Covariance Measurements'." 2016, doi:10.5194/essd-2016-36-rc2.

Verburg, P.H. Simulating feedbacks in land use and land cover change models. Landscape Ecology 21, 1171–1183 (2006).

Vermote, Eric; NOAA CDR Program. (2019): NOAA Climate Data Record (CDR) of AVHRR Leaf Area Index (LAI) and Fraction of Absorbed Photosynthetically Active Radiation (FAPAR), Version 5. [indicate subset used]. NOAA National Centers for Environmental Information.

Wei, Y., W.M. Post, R.B. Cook, D.N. Huntzinger, P.E. Thornton, A.R. Jacobson, I. Baker, J. Chen, F. Chevallier, F.M. Hoffman, A. Jain, S. Liu, R. Lokupitiya, D.A. McGuire, A. Michalak, G.G. Moisen, R.P. Neilson, P. Peylin, C. Potter, B. Poulter, D. Price, J. Randerson, C. Rödenbeck, H. Tian, E. Tomelleri, G. van der Werf, N. Viovy, T.O. West, J. Xiao, N. Zeng, and M. Zhao. 2013. NACP Regional: Gridded 1-deg Observation Data and Biosphere and Inverse Model Outputs. Data set. Available on-line [http://daac.ornl.gov] from Oak Ridge National Laboratory Distributed Active Archive Center, Oak Ridge, Tennessee, U.S.A.

Williams, Mathew, et al. "Improving land surface models with FLUXNET data." Biogeosciences 6.7 (2009): 1341-1359.

Yang, Y., Zhang, S., Yang, J. et al. A review of historical reconstruction methods of land use/land cover. J. Geogr. Sci. 24, 746–766 (2014).