Literature review

Ewan Pinnington

November 30, 2016

1 State of data assimilation with models of the carbon cycle

This chapter reviews recent efforts in using data assimilation with carbon cycle models in order to improve current estimates of ecosystem carbon balance. The approach of this review is to consider different relevant models of carbon balance and their implementation with data assimilation schemes.

1.1 DALEC

- Richardson et al. [2010] uses Markov Chain Monte Carlo data assimilation with a modified version of the DALEC model running at a twice-daily time step. Data from the spruce-dominated Howland Forest AmeriFlux site, Maine, USA. Assimilate gap-filled daytime and nighttime NEE. Takes the product of data stream miss match rather than sum to avoid overweighting of NEE observations in assimilation algorithm. Woody biomass increment found to provide orthogonal constraint to NEE and reduce uncertainties in parameter estimates.

- Williams et al. [2005] original DALEC paper, uses aggregated canopy model (ACM) to predict photosynthesis [Williams et al., 1997]. Observations from young ponderosa pine stand in Oregon used. Gap-filled NEE data used. Ensemble Kalman Filter used (EnKF), uses multiple EnKF's within a quasi-Newton minimisation routine (using a finite difference approximation to the gradient) to find a guess to an initial parameter set, this is perhaps not a good idea as the derivative of a sequential DA method is not meaningful. Found that rare measurements of stocks have limited effect constraining other components of C cycle, suggested that long time-series of these stocks with be more important to constrain C pool turnovers.

- Fox et al. [2009] present the REgional Flux Estimation eXperiment (REFLEX) model-data fusion (DA) inter-comparison project. Study had nine participants using a variety of DA methods (MCMC, EnKF and a genetic algorithm) to combine both synthetic and observed NEE and LAI data with DALEC model. Observed NEE data gap-filled. No DA technique shown to perform consistently better than others. Parameters linked directly to GPP and Reco best constrained, parameters related to slower processes (allocation to and turnover of fine root/wood C pools) are poorly constrained. Suggest more data on seasonal variations in slow large C pools would add useful constraint to DA schemes, to help constrain parameters

that are poorly estimated with EC data alone. Discuss that future studies should explore the importance of prior error estimates.

- Bloom and Williams [2015] present the DALEC2 model an updated version of the DALEC model which can be optimised for a deciduous or evergreen ecosystem. Outline a set of ecological dynamical constraints on parameter inequalities, steady state proximities and C pool turnover rates. When EDC's are used in a MCMC DA scheme with the DALEC2 model showed to retrieve improved analyses due to limiting unrealistic ecosystem behaviours.

- Quaife et al. [2008] use a EnKF to assimilate MODIS at-ground spectral bidirectional reflectance factor satellite data with the DALEC model at the Metolius forest, a temperate coniferous forest site in Oregon, Northwestern US. Find a discrepancy between model estimated LAI after assimilation of MODIS data and ground based LAI estimates, with the modelled estimate being higher than measurements from the site. Prediction of NEP is improved by assimilating the EO data and significant reductions in modelled flux uncertainties are achieved.

- Bloom et al. [2016] use MCMC methods (with some prior constraints) to assimilate MODIS LAI and HWSD soil carbon with DALEC2 model for a global $1^o \times 1^o$ grid, could be issues with computing power as implemented models become more complex and have more observations available, variational methods may alleviate this problem. Uses EDC's described in Bloom and Williams [2015] to limit parameter space, could be an issue for some global ecosystems that do not adhere to these specified constraints. Use the retrieved global DALEC2 map to gain insight into ecosystem functioning. Suggest that conventional land cover maps cannot adequately describe the spatial variability of carbon states and processes. Creates a valuable set of prior model estimates which could be used with variational assimilation methods.

1.2   BETHY

- Rayner et al. [2005] implement 4Dvar with the Biosphere Energy Transport HYdrology (BETHY) model [Knorr and Heimann, 2001] to assimilate satellite observations and atmospheric CO_2 concentrations in a stepwise manner on a global scale. Assimilating observation types in separate stages has been shown by Macbean 2016 that assimilating concurrently better? Use automatic diffrentiation to find the derivative of the BETHY model which is needed in the calculation of the gradient of the 4Dvar cost function. Calculate the inverse of the hessian of the cost function to find an estimate to posterior parameter errors. Find that uncertainty in long-term soil carbon storage is the largest contributor to uncertainty in the net CO_2 flux.

- Scholze et al. [2007] building on Rayner et al. [2005] describes improvements to the CCDAS system and BEHTY model and run the model for a 3 year hindcast period after a 10 year assimilation. Calculate posterior parameter uncertainties and correlations from the inverse of the hessian and then shows how this uncertainty can be propagated through time for future predictions. For a more recent review of the CCDAS system see Kaminski et al. [2013].

1.3 SIPNET

- Braswell et al. [2005] Use MCMC techniques to combine gap-filled NEE data from Harvard Forest flux site with simplified Photosynthesis and Evapo-Transpiration model (SIPNET). Use data assimilation to estimate the standard deviation of EC measurements, find an estimate of approximately 1 g C m^{-2} for daily flux errors. Find that some parameters are unconstrained by NEE data as modelled observations of NEE are insensitive to these model parameters. Recommend using a simple model for DA before adding levels of complexity.

1.4 JULES

- Raoult et al. [2016]

1.5 ORCHIDEE

1.6 Reviews of DA in C cycle

- Williams et al. [2009] discuss methodologies for combining FLUXNET data with land surface models. The many sources of error in EC observations are highlighted, and the fact that NEE represents the different between the two quantities of interest GPP and Reco, therefore both quantities can still be misspecified in a model and still achieve the observed NEE this is related to the concept of equifinality. Different DA techniques are outlined. It is proposed that temporal error correlations between sub-daily fluxes are likely to be severe and states that these could be included in the observation error covariance matrix R, although it is commented that this would be a difficult task. Identified challenges: Defining improved observation operators and understanding what measurements best complement EC data, better quantification of error in DA schemes and to test current modelling concepts such as plant functional types and their ability to describe FLUXNET data. Suggests combinations of techniques can be successful, e.g. global search methods (MCMC) to get close to global minimum followed by gradient based method to get closer to this minimum [Vrugt et al., 2005].

- Trudinger et al. [2007] present the Optimisation InterComparison (OptIC) project, where participants used different DA methods to estimate the parameters of a highly simplified model representation of terrestrial carbon dynamics. The project found more variation in estimated parameters due to the definition of the cost function than choice of optimisation technique (e.g gradient-based, sequential and MCMC methods). Different errors were added to pseudo observations in 21 outlined experiments for participants to complete, these included; Gaussian, lognormal and temporally correlated distributions. None of the participants made an effort to account for added temporally correlated error and this resulted in biased results being achieved from optimisation schemes. It was stated that the main criterion for success was the accurate specification of errors.

- In a review of data assimilation for the carbon-cycle Rayner [2010] suggest that there should be more focus on the information content of different relevant observations. It is also stated that more effort should be made in developing of new observation operators and specification of their errors.

3

1.7 Misc.

- MacBean et al. [2016]

References

A. A. Bloom and M. Williams. Constraining ecosystem carbon dynamics in a data-limited world: integrating ecological "common sense" in a modeldata fusion framework. *Biogeosciences*, 12(5):1299–1315, 2015. ISSN 1726-4189. doi: 10.5194/bg-12-1299-2015. URL http://www.biogeosciences.net/12/1299/2015/.

A. A. Bloom, J.-F. Exbrayat, I. R. van der Velde, L. Feng, and M. Williams. The decadal state of the terrestrial carbon cycle: Global retrievals of terrestrial carbon allocation, pools, and residence times. *Proceedings of the National Academy of Sciences*, 113(5): 1285–1290, 2016.

B. H. Braswell, W. J. Sacks, E. Linder, and D. S. Schimel. Estimating diurnal to annual ecosystem parameters by synthesis of a carbon flux model with eddy covariance net ecosystem exchange observations. *Global Change Biology*, 11(2):335–355, 2005.

A. Fox, M. Williams, A. D. Richardson, D. Cameron, J. H. Gove, T. Quaife, D. Ricciuto, M. Reichstein, E. Tomelleri, C. M. Trudinger, et al. The reflex project: comparing different algorithms and implementations for the inversion of a terrestrial ecosystem model against eddy covariance data. *Agricultural and Forest Meteorology*, 149(10): 1597–1615, 2009.

T. Kaminski, W. Knorr, G. Schürmann, M. Scholze, P. J. Rayner, S. Zaehle, S. Blessing, W. Dorigo, V. Gayler, R. Giering, N. Gobron, J. P. Grant, M. Heimann, a. Hooker-Stroud, S. Houweling, T. Kato, J. Kattge, D. Kelley, S. Kemp, E. N. Koffi, C. Köstler, P. P. Mathieu, B. Pinty, C. H. Reick, C. Rödenbeck, R. Schnur, K. Scipal, C. Sebald, T. Stacke, a. T. Van Scheltinga, M. Vossbeck, H. Widmann, and T. Ziehn. The BETHY/JSBACH Carbon Cycle Data Assimilation System: Experiences and challenges. *Journal of Geophysical Research: Biogeosciences*, 118(4):1414–1426, 2013. ISSN 21698961. doi: 10.1002/jgrg.20118.

W. Knorr and M. Heimann. Uncertainties in global terrestrial biosphere modeling: 1. a comprehensive sensitivity analysis with a new photosynthesis and energy balance scheme. *Global Biogeochemical Cycles*, 15(1):207–225, 2001.

N. MacBean, P. Peylin, F. Chevallier, M. Scholze, and G. Schürmann. Consistent assimilation of multiple data streams in a carbon cycle data assimilation system. *Geoscientific Model Development*, 9(10):3569, 2016.

T. Quaife, P. Lewis, M. De Kauwe, M. Williams, B. E. Law, M. Disney, and P. Bowyer. Assimilating canopy reflectance data into an ecosystem model with an Ensemble Kalman Filter. *Remote Sensing of Environment*, 112(4):1347–1364, 2008. ISSN 00344257. doi: 10.1016/j.rse.2007.05.020.

N. M. Raoult, T. E. Jupp, P. M. Cox, and C. M. Luke. Land-surface parameter optimisation using data assimilation techniques: the adjules system v1. 0. *Geoscientific Model Development*, 9(8):2833, 2016.

P. Rayner, M. Scholze, W. Knorr, T. Kaminski, R. Giering, and H. Widmann. Two decades of terrestrial carbon fluxes from a carbon cycle data assimilation system (cc-das). *Global Biogeochemical Cycles*, 19(2), 2005.

P. J. Rayner. The current state of carbon-cycle data assimilation. *Current Opinion in Environmental Sustainability*, 2(4):289–296, 2010.

A. D. Richardson, M. Williams, D. Y. Hollinger, D. J. Moore, D. B. Dail, E. A. Davidson, N. A. Scott, R. S. Evans, H. Hughes, J. T. Lee, et al. Estimating parameters of a forest ecosystem c model with measurements of stocks and fluxes as joint constraints. *Oecologia*, 164(1):25–40, 2010.

M. Scholze, T. Kaminski, P. Rayner, W. Knorr, and R. Giering. Propagating uncertainty through prognostic carbon cycle data assimilation system simulations. *Journal of Geophysical Research: Atmospheres*, 112(D17), 2007.

C. M. Trudinger, M. R. Raupach, P. J. Rayner, J. Kattge, Q. Liu, B. Pak, M. Reichstein, L. Renzullo, A. D. Richardson, S. H. Roxburgh, et al. Optic project: An intercomparison of optimization techniques for parameter estimation in terrestrial biogeochemical models. *Journal of Geophysical Research: Biogeosciences*, 112(G2), 2007.

J. A. Vrugt, C. G. Diks, H. V. Gupta, W. Bouten, and J. M. Verstraten. Improved treatment of uncertainty in hydrologic modeling: Combining the strengths of global optimization and data assimilation. *Water Resources Research*, 41(1), 2005.

M. Williams, E. B. Rastetter, D. N. Fernandes, M. L. Goulden, G. R. Shaver, and L. C. Johnson. Predicting gross primary productivity in terrestrial ecosystems. *Ecological Applications*, 7(3):882–894, 1997.

M. Williams, P. A. Schwarz, B. E. Law, J. Irvine, and M. R. Kurpius. An improved analysis of forest carbon dynamics using data assimilation. *Global Change Biology*, 11 (1):89–105, 2005.

M. Williams, A. Richardson, M. Reichstein, P. Stoy, P. Peylin, H. Verbeeck, N. Carvalhais, M. Jung, D. Hollinger, J. Kattge, et al. Improving land surface models with fluxnet data. *Biogeosciences*, 6(7):1341–1359, 2009.