Review of: Deciphering the components of regional net ecosystem fluxes following a bottom-up approach for the Iberian Peninsula
Author(s): N. Carvalhais et al.
MS No.: bg-2010-74
Reviewer: Ron Neilson, USDA Forest Service and Oregon State University, Corvallis, Oregon

The paper is an extension of Carvalhais et al. 2008, 2010 (which should be made very clear at the outset of the paper). In fact, it has taken me a long time to review this paper because I found that I had to go back and absorb much of the information previously published, as well as some basic information on CASA and Monteith's version of 'energy-use-efficiency', even though this should be my area of expertise. My review is colored by the knowledge of prior review and publication, as well as the unnecessary fact that I had to absorb the gist of those prior publications. But, I will challenge a few aspects not covered by the previous papers.

The paper assesses a methodology for reducing uncertainty to initial conditions in the use of General Vegetation Models (GVM), when calibrated across a network of trace-gas flux towers in order to scale up from towers to a grid over the entire Iberian Peninsula (IP). Since this ultimately boils down to an understanding of all (or most) of the parameters and their sensitivities, I had to go much deeper into the model structure than I normally would in reviewing a paper. I build both Equilibrium and Dynamic GVMs that have been used in U.S. Assessments, as well as in the IPCC, and I teach graduate-level classes on DGVMs in general. So, my understanding of the intrinsic assumptions is, perhaps, deeper than most. Thus, this review is also a bit deeper and far longer, with apologies, than most that I write.

Since CASA is a largely empirical model, although termed as 'bottom-up', this requires just a bit of review of the basic principles being analyzed in order to un-earth some of the unspoken assumptions. My review may sound, at times, a bit critical; however, I am generally very positive on this paper and the philosophy behind what they are doing. So, my comments are meant entirely to clarify the discussion a bit, in part for the benefit of the naïve reader, e.g., graduate students (or perhaps a few aging scientists).

The net balance of carbon entering versus leaving the biosphere is of global interest, affecting policy decisions on climate change and CO_2 management at all levels of world, nation, state and local governments and landowners. Dynamic General Vegetation Models (DGVMs) are designed to simulate these processes in order to calculate and ultimately to forecast the regional to local carbon balance and its interannual dynamics and longterm trends. Quantitative validation of these models is critical and is the ultimate goal of this and Carvalhais et al.'s previous papers. The net flux of carbon into and out of the biosphere in this paper is termed Net Ecosystem Productivity (NEP), which is the difference between the fluxes going in, Net Primary Productivity (NPP, autotrophic respiration being a fixed factor of GPP), and those leaving, Heterotrophic Respiration (R_H). The term, Net Biome Productivity (NBP) does not come up until much later in the paper. NBP is correctly defined as NEP less fluxes from disturbances (especially fire), land use and other factors.

Yet, until I had reached that point in the paper, I was under the delusion that all these factors were being subsumed in NEP. These distinctions should be clarified right up front in order to dispel confusion. However, in careful description of the tower site histories, from a previous paper, it is clear that all these sites have been heavily disturbed, but that they might have been free of disturbance during the 25 year period of the satellite record, rendering the analyses clearly of just NEP and not, inadvertently, NBP. But, this does become an issue when the authors use the site calibrations to scale up, via a grid to the entire IP, which has clearly experienced a great deal of fire and other disturbances during the satellite period. In that sense, they are scaling NEP into a reality that is in fact NBP, a point that should be made clearer.

The authors have mounted a heroic and very innovative methodology to test the model structure and parameterization against a suite of flux towers, recognizing that each is in a different state of disequilibrium relative to a simulated, equilibrium condition by the model. There is virtually no landscape that has not been disturbed at some time in the recent past. Model and observational

experiments show that the magnitude and interdecadal trends of both fluxes (NPP, RH) are strongly influenced by the amount of carbon in the soil and vegetation, slow carbon in particular, which varies with the time since the last disturbance, and which affects many processes critical to both fluxes.

However, the slow carbon pools, by definition have quite a bit of inertia and tend to track longer term, e.g. decadal to 1,000-year trends. The authors do point out, at some point in the discussion and conclusions, that the year to year variability of the two fluxes is also subject to the high-frequency variability of the weather (climate) and that these fast fluxes could be quite independent of and even contrary to the longer-term trends induced by the slower carbon pools and the tendency of the system to 'relax' back to that condition. R_H has a very rich suite of inertial constraints, or lags, whereas, NPP is much, much faster. How do these various levels of inertia respond, or 'resonate' with the increasingly rich knowledge of interannual, interdecadal, centennial and even longer climate oscillations? The paper would be far less confusing, I think, if these distinctions between slow and fast processes and fluxes could be approached and discussed much earlier in the paper, perhaps at the end of the introduction or beginning of results. Lack of this discussion makes the paper more difficult to comprehend. Prepare the reader!

The author's approach begins with the premise that each of a suite of tower sites is at some point of disequilibrium, with respect to a potential equilibrium condition, but that it is on a trajectory of approach to that equilibrium. Their goal is to remove the uncertainty of recovery as a collective, statistical estimation (η) of perturbations to carbon pools, largely slow carbon, essentially removing the sensitivity to initial conditions, in order to analyze any trends of change in the equilibrium condition itself and then to scale this up through a gridded analysis of the entire Iberian Peninsula (IP). The results, presenting enormous reductions in uncertainty, are clearly impressive. But, without disturbance, I am not entirely sure how to interpret this.

I also have some concerns about terminology and unclear or fuzzy definitions, as well as poorly defined 'drivers.' However, I think my concerns can largely be reconciled by some duly-considered caveats and explanations. Briefly, the methods section is populated with many references to process descriptions in the form of "following" so and so, or "after" 'x' et al. Given that this is a 'test of methods' paper and sensitivity analysis of the parameters, these rather glib descriptions are insufficient. I don't need detailed descriptions, but simply stating the dependencies, such as 'process xyz is a curvilinear function of soil water and temperature.' Or, 'PFTs are simulated independently from each other, but compete for space within a gridcell'; or are '…specified to cover a specific percentage of a gridcell, based on remote sensing, but there is no direct competition among the PFTs for any limiting factors, light, water or nutrients.' Do woody and grass PFTs have the same available root water?

A known feature of complex systems is their sensitivity to 'initial conditions' and to 'boundary' or 'driving' conditions, which can send them into one 'attractor basin' or another. This is particularly true in semi-arid systems where the same 'average' climate and soil might support a savanna, chaparral, shrubland, or grassland, each of which could be quasi-stable, but 'moveable' with some energy input to a different 'state'. Given that much of the IP is semi-arid (Fig. 3 PFT distributions), this is a critical observation to bear in mind. To which alternative state is each site returning? This is an important question, given that these different 'states' have quite different above and belowground carbon pools. And, depending on site history, there may even be a hysteretic carryover in the soil carbon pools from previous states. Our experience has also shown that without true canopy light competition, root-weighted water competition and drought-induced fire many of these complex dynamics cannot be revealed. Yet, the water factor is not clearly enough defined to know how differently-rooted PFTs will respond. These competitive and feedback processes are particularly prominent in semi-arid ecosystems. In our own modeling we have found that direct PFT competition has proven to be critically important in capturing threshold dynamics and state changes and can be quite sensitive to the frequencies of variability of the climate. Yet, I don't believe that this experimental design would notice this type of model structural sensitivity.

When the biosphere is in balance (stationary climate), NPP and RH are nearly equal. But for any given year, the likelihood of the two fluxes being exactly in balance is extremely low and one or the other

will dominate; thus, the concept of equilibrium, as one might think of it in a chemostat, is essentially impossible (for the fast pools, but not necessarily for the slow carbon pools).

Thus, the biosphere is ultimately a 'complex', possibly even 'chaotic' system (in the mathematical sense), with different regions being susceptible to rapid growth spurts or threshold collapses. The recognition of this complexity, even under a 'stationary, but variable' climate, led to the development of 'state-and-transition' models (STM), which can suddenly shift from one state to another, conditional on various probabilities and each with very different carbon states and dynamics. However, this becomes much more problematical under a non-stationary climate, driven by increases in CO_2 concentrations and other greenhouse gases, and STMs are not very useful.

The two timeframes of change of the very slow soil carbon pools and the considerably more volatile carbon pools is a critical point in this suite of papers and under a state transition, the 'targets' could change. However, only the slow carbon pool equilibrium condition is considered in terms of the 'target' for recovery, based on the η statistic. Considering the slow carbon pools and the potential for multiple 'states' in semi-arid conditions, and considering the very complex management history of these sites, as described in some depth in the previous papers, it is difficult to know what the true 'target' for recovery really is. In other words, the equilibrium state of the slow carbon pools could be quite disconnected from the potential equilibrium state of the faster pools, which are more closely aligned with the fluxes. In fact, the results in this paper demonstrate this in concluding that the 'states' and the 'fluxes' might be of opposite trends and produce somewhat independent parameter sensitivities.

I am not suggesting that the authors re-iterate all of their prior discussion in this paper, but I do think that briefly summarizing their previously published results and conclusions, as well as some greater depth of the true complexity of these site dynamics, would tremendously benefit the paper.

Theoretically, there is nothing that suggests that NPP and R_H should be nearly equal. Consider a thought-model of the relationships between NPP and R_H. At high latitudes, where temperatures are cold, but there is still sufficient light and warmth to drive NPP, then NPP should exceed R_H and soil carbon builds up over the years, producing deep Peatlands. However, in tropical wet latitudes, there is sufficient radiation, warmth and water. In those situations R_H should exceed NPP. However, decomposition does require a substrate, so NPP does become limiting, but only as an indirect factor. Thus, in tropical forests as soon as any litter hits the ground, it is rapidly consumed by decomposers, releasing any nutrients, which are then quickly re-assimilated by the vegetation. There is very little carbon or nutrients in the soil with most of it being suspended in the above-ground vegetation. However, in the subtropics, there is sufficient light for NPP, but water limits both NPP and R_H. This creates the curious situation where small litter decomposes quickly during short wet events, but large litter can remain for centuries. But, soils, given their inherent retention of water, create an environment where litter of most sizes decomposes fairly rapidly and there remains little carbon in the soil and much like the wet tropics, most of the nutrients are held in the live biomass. In the mid-latitudes, where most ecologists live, it happens that NPP and RH do tend to be more equal and can quickly shift to one side or the other with respect to NEP. The IP is mostly in the subtropics, but with elements extending into the temperate zone, mostly in the mountains.

This thought problem provides a context within which to judge this study. Given that decomposition cannot exceed substrate, it is generally limited by NPP in the IP, as concluded by the authors. But, the real limits are water and the thermal regime, largely water. This mental construct also frames my views toward fAPAR and energy-use-efficiency (ε). Monteith breaks ε down into several components, most of which are fixed, but is rather fuzzy on the biotic components of ε. However, he does acknowledge that a large part of ε must be structural, in part via leaf area index (LAI), or the physiognomy of the canopy. Likewise, LAI is a large component of fAPAR, which is prescribed from satellite measurements. So, in the final analysis of parameter sensitivities, the authors weigh these contributing factors from ε and fAPAR and conclude that the latter must be the sensitive factor. However, since LAI is a component of both parameters, the sensitivity could just as easily reside in either or both.

This brings me to the construction of the original 3 equilibrium GVMs. All three, BIOME, DOLY and MAPSS were constructed under the same, nearly axiomatic assumption, which is: Nearly all

ecosystems will attain the maximum leaf area that can just barely be supported by the available soil water. This observation might have been originally published by the father of American hydrology, Robert E. Horton, 1933. So, all three models are LAI calculators, but all 3 also require a savanna structure, competing woody and ephemeral plants with different rooting depths, and hence different available water. Solving this problem also requires a reasonably good and general soil water model that appropriately partitions precipitation into various components, ultimately deriving the amount and timing of infiltration, as well as its percolation through the various soil layers.

Having built the MAPSS biogeography model, I discovered that the most sensitive parameter, and the only one I generally tune, is the k parameter, directly analogous to the Beer-Lambert k used to calculate light extinction through canopies. In most GVMs, which *a priori* specify LAI, the k is simply 'set' to 0.5, as I believe it is in most fAPAR calculations. However, it varies great in reality, but is one of those 'immeasurable' entities; although it does appear to have certain bounds of variability for different PFTs. Consider that if a canopy stretches slightly higher, with no change in absolute LAI, then the lower leaves will receive more light and transpire more water. In effect, it has a higher energy-use-efficiency. Is this small adaptive response detectable by remote imagery? I do not know.

This paper defines two axes of DGVM categorization: prognostic (forward) vs. diagnostic models; and, top-down vs. bottom-up models. CASA is defined as a bottom-up, diagnostic model. Forward or prognostic models are suitable for forecasting, given only a suite of initial conditions and a timeseries of 'forecast' future climate; and 'diagnostic' models, are partially driven by observed behavior of the biosphere to produce 'hindcasts', which can then be tested against other characteristics of the biosphere that have been observed, but not used to 'drive' the model behavior. Bottom-up' models infer that they are process-based; while, top-down models, inferring that they are more correlational, or of a more simple structure and may or may not provide inference of causation. However, there is sometimes confusion between these two classification systems and I believe some of this confusion creeps into this paper. Being spectral in origin, fAPAR is necessarily a correlational model, which is used to prescribe the timeseries of leaf area and its phenology, and indirectly, its energy-use-efficiency. Thus, even though I agree that CASA contains algorithms for many processes in ecosystems, the prescription of LAI renders it ultimately a 'top-down' model.

Nevertheless, the authors are quite clever in dissecting the constraints on NPP and R_H, but I do believe that they might be a bit confused on the roles of fAPAR and ε. I also believe that although the term 'substrate limited' does accurately describe the picture in the IP, it is not particularly enlightening with regard to ultimate constraints, given the thought model above. And, I do not believe that substrate limitation is necessarily a general property of NEP. And finally, the lack of consideration of drought stress, fire and infestation limit the utility of the scaling up process.

Still, in conclusion, I do think the conceptual framework of gauging the current carbon status against some hypothetical 'attractor basin' is useful for reducing uncertainties and inferring trends.

With Respect,

/s/ Ronald P. Neilson