Please find our replies in italic below specific comments and questions.

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.

We acknowledge the importance of connecting these previous works and improved the link with the two previous papers and the current one explicitly in the introduction section, in the beginning of the last paragraph of the introduction one can read now: "This study builds directly on two previous studies that explored the impacts of initial steady state assumptions in inverse parameter optimization exercises conducted at site level [Carvalhais et al., 2008; 2010]. At the site level these studies showed that initial steady state assumptions tend to force neutral NEP estimates which tend to bias parameters and inflate parameter uncertainty. These issues can be circumvented through the relaxation of initial conditions in ecosystem pools using empirical approaches. Here, we follow a bottom-up approach"...

Indeed we could have been clearer in the methodological descriptions, despite the usual limitations on text amounts and the always avoidable repetitions of previous descriptions. We also transferred some conceptual details from the previous manuscripts and add the necessary to the radiation use efficiency description in section "2.2 The CASA model" after equation 2: ..."where T_ε follows an asymmetric bell shaped curve that peaks at an optimum temperature (T_{opt}), dropping faster above than below T_{opt}, and W_ε responds linearly to the ratio between actual and potential evapotranspiration. Here, the actual net photosynthetic capacity ($fAPAR \times \varepsilon$) considers both canopy structural components – via fAPAR – and instant climate effects – in ε. However, changes in the fAPAR also implicitly include the effects of environmental conditions in the canopy's structure, which hampers the isolation of the pure climate effects on NPP." We also complemented the description of the methodological section "2.9 Decomposition of ecosystem fluxes" in order to clarify the posterior analysis on the decomposition of the NPP into its drivers. In the end of the second paragraph of this section we added: ..."In this regard, the isolation of the climatic effects from the fAPAR time series is not mechanistically feasible. Hence, the partitioning of NPP drivers is understood as between the changes in phenology (fAPAR, which also integrate effects of climate in the canopy's structure) and the effects of climate in the instantaneous light use efficiency (ε)."

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).

We would like to say here that we deeply appreciate the constructive character of the review and made our best to consider most of these points throughout the manuscript that we hope improve its readability and clearness.

Regarding the particular comment on the understanding of the optimized parameters, in addition to the clarifications on the NPP calculations referred above, we have now after equation 3 of the methodological section "2.2 The CASA model": "where each pool i is characterized by a different turnover rate k_i *that is regulated by the effect of temperature (*T_s*) and water availability conditions (*W_s*) (Potter et al., 1993). The effect of temperature in decomposition (*T_s*) follows an exponential Q10-type response; while the effect of soil water conditions (*W_s*) is estimated as a non-linear function of the ratio between water availability (soil moisture plus precipitation) over the potential evapo-transpiration, following a linear positive response from 0 to 1, with a maximum set for values between 1 and 2, and then decreasing gradually as excess conditions of water supply increase (Potter et al., 1993)."*

The net balance of carbon entering versus leaving the biosphere is of global interest, affecting policy decisions on climate change and CO2 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 (RH). 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.

Indeed no reference to NBP, and the differences with our main variable of interest NEP are referred up front in the manuscript. To clarify this point from an earlier phase of the manuscript we explicitly refer to it in the Introduction. In the end of the 3rd paragraph now it can be read: "The full accounting of the regional ecosystem carbon fluxes would imperatively include the effects of disturbance events (e.g. fire, insect outbreaks) and management regimes (e.g. grazing, logging); which summed up to NEP would yield the net biome production (NBP) (e.g. Chapin et al., 2006). Although our approach attempts to be comprehensive, we do not consider the dynamics of NBP, which would require further model evaluation and parameterization efforts."

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.

In fact, at regional scales a full carbon budget estimate would require considering NBP, and not NEP. It was also our understanding that the model evaluation performed at site level would at most support the CASA estimates of NEP, hence the option for regional estimates of NEP, in detriment of NBP. The modeled dynamics do not explicitly account for the effects of disturbances in the carbon cycle – neither in terms of fast (e.g. instantaneous emissions in the case of fire events) nor in terms of slower (e.g. reductions in carbon inputs to the soil pools resulting from logging activities) fluxes. Despite the fact that at regional scale NBP would be the most appropriate variable to model and compare to regional observations of C fluxes, given the underlying model dynamics only NEP can be considered for analysis. We should also say here that the upscaling of NEP from towers to regions is hardly feasible – due to lack of knowledge on the ecosystem carbon pools. With the current approach we only demonstrate the feasibility in determining anomalies and trends in these fluxes. In addition, we should refer that reductions in the vegetation state variables that result from disturbance events can be implicitly prescribed in the simulations through the NDVI time series inputs; which embed the effects of disturbance events in the vegetation components if these represent a significant portion of the pixel's 8x8km area. In these cases, the modeled reductions in photosynthetic activity originating from the reductions in the NDVI time series implicitly include the effects of disturbance. Consequently, the phenological driver of the model includes the effects of disturbances and climate (as referred to previously) and the phenology driven trends cannot be seen as completely independent from the effects of disturbances and climate.

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. RH 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 different levels of inertia (or characteristic levels of occurrence) that characterize (and sum up to) R_H are embedded in the model structure through the consideration of different turnover rates in the soil level carbon pools that span from monthly to centennial scales. The effects of temperature and soil water availability – T_s and W_s, respectively, as defined in Potter et al. (1993) and summarized above – are explicitly considered in the on the actual turnover rates and consequently in the respiratory fluxes. It is in the model assumptions that these effects equally affect the different carbon pools. The changes in the inertia of the soil C pools are then dependent on its intrinsic turnover rates that are directly influenced by the different time scales at which the abiotic input variables vary. We added some text in this regard just before the last line of the methodological section "2.2 The CASA model": "CASA assumes nine soil carbon pools (P=9) that circulate carbon inputs from vegetation debris and have different turnover rates, ranging from monthly (metabolic litter pools) to centennial (old pools) scales. According to the model structure, temperature (T_s) and water availability (W_s) controls on the instantaneous turnover rates equally affect all soil C pools. Higher inertia is seen in the slower turnover soil carbon pools, which take longer to respond to changes in climate conditions and changes in vegetation states – and vice versa. Hence, these pools take longer to reflect a perturbation as well as to recover from it." We further clarify in the "2.3 Inverse model parameter optimization" section that "In this regard, an equilibrium situation would be easily achieved by the faster pools while non-equilibrium conditions would be sustained when lower turnover carbon pools are adjusted. Therefore the strategy to explore non-equilibrium conditions in the recalcitrant pools was chosen. The inclusion of microbial pools in the set of pools affected by η was motivated by the effective role on model performance and the structural dependencies from the slower pools."

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.

The use of parameter eta is different at site level or at regional level, more in detail:

1. *At site level, the non equilibrium premise serves mostly the purpose of obtaining sets of parameters unbiased regarding any initial conditions assumptions of equilibrium. Unbiased sets of parameters that are later prescribed throughout the Iberian Peninsula in the regional model runs.*
2. *At regional level, η serves the purpose of forcing different initial conditions as distances to the initial equilibrium solution.*

The uncertainties stemming from the initial conditions are only removed in the regional simulations by subtracting carbon fluxes estimated under constant climate conditions. Since we do not simulate the explicit dynamics of carbon under ecosystem disturbances, we argue that this should be seen as reductions in the effects of initial conditions in estimates of decadal trends in NEP time series. Given the caveats towards a comprehensive estimate of regional ecosystem fluxes, the current approach could be adapted – and the results compared – to simulations with a dynamic vegetation model or to simulations with prescribed level of disturbance.

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?

We reviewed the manuscript for these issues and performed improvements accordingly:

- *Regarding the model drivers, we now start the section "2.5 Data for spatial runs" with: "CASA requires drivers related to climate (temperature, precipitation and solar radiation) as well as to biophysical properties of vegetation (NDVI or fAPAR and LAI)."*
- *We replaced: "For consistency in the bottom-up approach between site level and regional simulations we estimated fAPAR from NDVI according to Los et al. (2000) and leaf area index (LAI) estimates followed (Sellers et al., 1996)."; by: "For site level and regional simulations we estimated fAPAR as the mean of the two linear scaling methods with the simple ratio and the NDVI, individually per plant functional*

type (PFT), according to Los et al. (2000). We computed leaf area index (LAI) as varying exponentially with fAPAR, or linearly for clustered PFTs (e.g. coniferous trees and shrubs) or as a combination of the previous two for clustered and evenly distributed vegetation types, following Sellers et al. (1996).”

- *We replaced: “The carbon assimilated by vegetation is partitioned between the different vegetation pools according to the dynamic allocation scheme of Friedlingstein et al. (1999).”; by: “The carbon assimilated by vegetation is partitioned between the different vegetation pools according to the dynamic allocation scheme of Friedlingstein et al. (1999) which adjusts the investments in leaves, stems and roots towards the most limiting resource (water, light and mineral nitrogen).”*
- *We replaced: “Every dataset was spatially interpolated to the GIMMS NDVI 8km by 8 km grid following Zhao et al. (2005).”; by: “Every dataset was spatially interpolated to the GIMMS NDVI 8km by 8 km grid following a weighted average considering a non-linear distance to the four neighboring pixels (Zhao et al., 2005).”*
- *In the beginning of the section “2.6 Regional model runs for a range of initial conditions” one can now read: “The CASA model estimates of ecosystem fluxes for the Iberian Peninsula are performed on a PFT basis: the parameter vector used per pixel per PFT originates from the upscaling exercise above. The fraction of each PFT within a grid cell is specified according to the land cover map and remote sensing datasets, for which the carbon fluxes are simulated independently from each other, with no interactions competing for any limiting resources (light, water or nutrients).”*

The water conditions stressor (W_{ε}*) is not equal between the different PFTs, since* $B_{w\varepsilon}$ *is optimized per PFT. However, the index of water availability is considered constant per pixel, so it will be the same for both the woody and the grass PFTs. In the end one can say that the availability is the same but the sensitivities are different. However, from Table 3 one could not say that there is a generalized higher or lower sensitivity to soil water availability in grasses, compared to woody PFTs.*

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.

Indeed, we agree that most of these processes and consequent impacts on ecosystem states are relevant for landscape dynamics at longer time scales, often simulated within the context of a dynamic vegetation model. However, within the context of remote sensing-based production efficiency modeling, it is common to observe a trade-off between such mechanistic detail and spatially explicit information regarding the different PFTs fractional coverage and the temporal dynamics of biophysical properties of vegetation. The current model does not account for the community level interactions that are referred here. In this regard, we could have only simulated those dynamics by prescribing changes in the land cover inputs to the model. Given CASA's characteristics – and the characteristics of production efficiency modeling in general –, we agree that the current experimental design would not detect climate driven threshold dynamics and state changes in ecosystem structure. With such goals in mind, we agree that the option towards DVMs would be clearly more appropriate.

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).

Also, the year to year variability may be a biasing factor in determining the ecosystem's distance to equilibrium, hence the need for longer records (Carvalhais et al., 2008). In addition to the prior clarifications on the relaxation of the steady state mentioned above, we added next to the end of the "2.3 Inverse model parameter optimization" section: "However, due to inter-annual variability in climate and ecosystem fluxes, the accurate estimation of non-equilibrium states may be hampered because of the short time series of fluxes, or periods that do not represent the development stage of the observed ecosystem, yielding a non representative η. In principle this problem is not exclusive to the determination of η since such representativeness issues can be in the basis of biases in the estimation of other parameters as well."

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 CO2 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.

It is true, the η is chosen in terms of the 'target' for recovery that corresponds to the observed vegetation cover at site level. Such can correspond to a current 'target' vegetation type that is different from the previously existent vegetation cover. In this case, the empirical character of η is a strength of the approach, since it renders 'free' estimates of initial estimates of these carbon pools depending only on the current observations of net ecosystem fluxes. The possible disconnection between the fast and slow pools is partially allowed in our approach, since η 'attacks' both the slower and microbial pools.

The consideration of a constant land cover map it is a strong assumption for regional simulations. Indeed then, the target for recovery could be completely different then the actually present due to land cover transitions, although we observed a maximum of 9.3% of land cover change for the IP region based on the CORINE census of 1990 and 2000. DVMs tend to embed such dynamics by design, although we would assume it's application would require the need for further developments and evaluation of transition dynamics (e.g. against land cover maps).

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.

It is our impression that the additions under section "2.3 Inverse model parameter optimization" already clarify much of the two previous works. We find that the description of site level information concerning the 33 sites that are considered in this study would be overwhelming, although we agree that such information could corroborate the interpretation of all the individually estimated η. But site level optimizations aim at estimating the parameter vectors governing NPP and R_H dynamics and the empirical nature of η allows for an unbiased estimation of such parameters regarding the simulated initial conditions. Having in mind the objectives of the optimization exercise for the regional parameterization of the IP, we find that we can leave aside such information. For clarity, we added also the following text close to the end of the section "2.3 Inverse model parameter optimization": "Also the different historical backgrounds within the selection of sites imply different past dynamics behind the current ecosystem states and fluxes observations. However, the application of this empirical approach – through η – has shown robust model performances and reduced biases in parameter estimation for different sites with different historical backgrounds (Carvalhais et al., 2008)."

Theoretically, there is nothing that suggests that NPP and RH should be nearly equal. Consider a thought-model of the relationships between NPP and RH. At high latitudes, where temperatures are cold, but there is still sufficient light and warmth to drive NPP, then NPP should exceed RH 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 RH 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 RH. 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.

We discussed these issues above where, overall, the trends in fAPAR are associated to trends in the canopy's structure while trends in ε are associated to trends in the climate effects on instant light use efficiency.

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.

Clearly these are crucial factors for consideration under more mechanistic *prescriptions of ecosystem dynamics simulations. However, having fAPAR as a model input (such as in our case), water availability and competition factors are unrelated to our estimations of LAI. Any developments in the soil water module estimates would change the model's ability to estimate* W_ε *and* W_s*, which ultimately could lead to improvements in the estimates of fast fluxes and yield better model parameterizations.*

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.

Theoretically, one could assume that if a higher amount of energy would be absorbed by the canopy such would be reflected in fAPAR – through the lower retrieved energy in the PAR band. This example would imply an improvement in the photosynthetic capacity of this system that would be due to a structural development not quantified by LAI but by fAPAR, through biological factors. The circumvention of a detailed description of the canopy structure given the input of fAPAR is a significant advantage of remote sensing-based production efficiency models over DVMs in diagnostic exercises.

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.

In this manuscript the 'bottom-up' terminology is used to define the overall approach: from site level optimization to regional flux estimations (contrasting with 'top-down' approaches that rely on atmospheric inversions to perform such estimates). Here, 'bottom-up' is not to intended to characterize the model structure as from first *principles. Such convention is used when referring to "ecosystem to region" estimates in the context of the CarboEurope-IP and the Global Carbon Project, for instance, which we believe to be of appropriate usage here.*

Nevertheless, the authors are quite clever in dissecting the constraints on NPP and RH, 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.

We believe the confusion between the roles of fAPAR and ε has been clarified with the discussions above and the subsequent changes performed in the manuscript.

The current analysis aimed at evaluating the drivers of ecosystem carbon fluxes at the light of model structure. This approach allows for an objective evaluation of the different factors behind the decadal dynamics of ecosystem carbon fluxes as estimated through model simulations. It is our understanding that this would be a more appropriate approach since NEP estimates are based on model computations and the responses to climate are seldom linear. Such mechanistic approach allows us to infer and quantify causality in our current estimates. Given the thought model above, the ultimate constraints – or drivers – of the underlying processes for NEP would always be climate variables. Our approach hampers the establishment of causal links between canopy phenological developments (LAI and fAPAR) and climate variables; although it allows associating changes in the substrate availability of R_H *to changes in the allocation strategies driven by changes in water availability regimes.*

We would ultimately like to say that we agree with the comment that not considering disturbances in the simulation exercises impedes us from reaching a carbon cycle regional budget (NBP). And such accounting attempt is not our present goal with this study. We would like to enforce that the effort behind appropriate NBP estimates clearly requires superior efforts regarding both model development and evaluation approaches, then those presented in the this study.

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