
Report #1 on “Evaluation of terrestrial pan-Arctic carbon cycling using a data-assimilation system” by Efrén López-Blanco et al.

M. Forkel (Referee)
matthias.forkel@geo.tuwien.ac.at
Submitted on 19 Nov 2018

The authors substantially revised the manuscript and addressed my comments appropriately.
I only disagree how uncertainties where used during the data assimilation. Specifically, the authors state at several places (lines 86, 184-85) that the (biomass) dataset lacks uncertainty or error estimates and hence they used a global uncertainty factor of 1.5 in the cost function.
It is clearly a wrong statement that the biomass maps by Carvalhais et al. (2014) miss uncertainty estimates.
In this dataset, uncertainty was provided based on an ensemble of biomass estimates. This biomass map is also based on the map of forest biomass by Thurner et al. (2014) which also includes a detailed estimate of uncertainties for various vegetation carbon pools.
Please remove the wrong statements about missing uncertainty estimates for the biomass datasets and describe why you did not use these uncertainty estimates or how a potential use could affect your results. With these changes, I'm happy to accept the manuscript for publication.

We apologise for the lack of clarity about uncertainty derivation for the analysis. Here we have adjusted the text on the Introduction section (S1) to remove the sentence

> “However, these products tend to lack clear error estimates.”

On S2.2.2, L188-195 we have adjusted the text to:

> “The reported uncertainty on biomass data from Thurner et al. (2014) was +/- 37% at pixel scale. Because of undetermined errors related to tree cover thresholds used in the upscaling, and to reflect unknown model structural error, we slightly inflate the error estimate and use a log-transform(1.5) of ×/÷1.5 (i.e. ×/÷1.5 spans 67% of the expected error). We use the same proportional error for SOC. For MODIS LAI we inflate the proportional error further to log(2) based on well reported biases in this product for evergreen forests (De Kauwe et al. 2011) and the estimated measurement and aggregation uncertainty for boreal forest LAI of 1 m^2 m^{-2} reported by Goulden et al. (2011). The uncertainty assumptions in expression 3 are chosen in lack of better knowledge about the combined uncertainties arising from model representation errors and observation errors:”

In the Discussion we also now review the challenges associated with generating observation and model errors (see response to reviewer 2).

Report #2 on "Evaluation of terrestrial pan-Arctic carbon cycling using a data-assimilation system" by Efrén López-Blanco et al.

Anonymous Referee #2

Received and published: 06 Feb 2019

I'm going to be upfront that I'm very torn about what to recommend with respect to this paper. On the one hand, I acknowledge the incredible amount of work that went into this project and believe that there is important and interesting science coming out of this project. On the other hand, based on the responses to questions raised, it is now clear there are definitely things here that I don't think were done correctly. What complicates this is that many of the things done wrong (especially with respect to model process error) were also done wrong in previous papers on the Bayesian calibration of terrestrial carbon models (both by this team and others). This helps explain such mistakes, but it doesn't justify them, and I worry that continuing to allow papers to make the same mistakes just perpetuates the situation. The crux of the issue is really in how the authors are treating the error term in their likelihood. First, they are ascribing 100% of the error as coming from the observations, and not acknowledging (statistically) that their model is imperfect (though their own Results and Discussion clearly demonstrate that the model is far from perfect). By incorrectly ascribing 100% of the error to observations, and none to process error (model misspecification, stochastic events, unaccounted for heterogeneity), the authors are also missing that (unlike observation error) process error propagates forward into model predictions. This means that modeled fluxes and pools are going to be consistently overconfident by an unknown (but potentially nontrivial) amount. Second, not only do the author ascribe all the error to observations, but they treat that observation error as a known parameter, despite acknowledging that the data products used don't have error estimates. This is a significant departure from standard statistical modeling, where the variance is an unknown fit parameter. For example, when you fit a linear regression the model has three unknown parameters (slope, intercept, sigma) and sigma is virtually never treated as an a prior known quantity. While treating sigma as a known shouldn't have large effects on the mean values of the model parameters (though this is far from guaranteed when dealing with nonlinear models; Jensen's Inequality), more important is that it can have a real effect on the uncertainties about the model parameters. By subjectively choosing the observation error, one is also subjectively choosing the confidence intervals on the parameters. And since in CARDAMOM the only uncertainties that are included in predictions are parameter uncertainties, this also means you are subjectively choosing the uncertainty in the predictive confidence intervals. Ideally, these models should be refit including an unknown, fit model process error, and then that process error should be propagated into predictions/hindcasts. This process error ideally should also be in addition to, not instead of, an observation error (which may not be a known, but may have an informative prior on it)

We recognise the reviewer's concerns about using the correct process for error characterisation in analyses such as that we present here. We agree that our model is not perfect and that identification of process error is critical. We also regret that we did not provide the necessary information on how data uncertainties were derived. We do appreciate the reviewer's concern about effective error characterisation, and have adjusted the text to reflect this, and to make recommendations about how to address this better.

We did specifically focus on identification of model process error by comparison with independent data (GPP, R_h). Thus, we identified biases in our estimates of LAI, GPP and biomass at landscape scale, and suggest that these likely reflect systematic bias in our photosynthesis model. A next step is to analyse the representation of photosynthesis process error and include this in further analyses. On the other hand, we note that independent evaluation of fluxes at site scale (FLUXNET2015) does not match the GPP bias at landscape pixel scale (FLUXCOM). New site level comparisons (see below) also suggest CARDAMOM produces reasonable or slightly high biased results. We conclude that further investigations into heterogeneity error are required, linked to process error calculation on products such as FLUXCOM as well as our GPP model.

We have adjusted the text (S2.2.2, L188-195) to clearly state that error in the biomass product is reported, and have explained why we have inflated this error in our analysis. We also note that MODIS LAI products have large reported biases, and local observations have important errors, which justifies the larger error we assigned to these data. Our point here is to report an honest overview of uncertainty assumptions used in CARDAMOM:

> "The reported uncertainty on biomass data from Thurner et al. (2014) was +/- 37% at pixel scale. Because of undetermined errors related to tree cover thresholds used in the upscaling, and to reflect unknown model structural error, we slightly inflate the error estimate and use a log-transform(1.5) of ×/÷1.5 (i.e. ×/÷1.5 spans 67% of the expected error). We use the same proportional error for SOC. For MODIS LAI we inflate the proportional error further to log(2) based on well reported biases in this product for evergreen forests (De Kauwe et al. 2011) and the estimated measurement and aggregation uncertainty for boreal forest LAI of 1 m^2 m^{-2} reported by Goulden et al. (2011). The uncertainty assumptions in expression 3 are chosen in lack of better knowledge about the combined uncertainties arising from model representation errors and observation errors:"

We note the reviewer's concerns about making forecasts without properly accounting for model process error. This paper involves an analysis of historical fluxes constrained by contemporary forcing and data. We do not make forecasts or hindcasts, so this criticism is not relevant for this paper.

We have adjusted the text in the discussion (S4.3; L487-501) to reflect the lack of robust knowledge on the interactions between random and systematic biases in the observations, model representation errors and errors in the model drivers:

> "Our approach has used estimated observation error, and inflated this to include unknown errors associated with model process representation. We currently lack any better knowledge of the combined uncertainties arising from model representation errors and observation errors. We acknowledge that all models are an imperfect representation of C dynamics, which generates irreconcilable model-data errors due to the inherent assumptions in model structure. Future analyses should investigate model structural error, using for example error-explicit Bayesian approaches (Xu et al., 2017), or comparing the likelihoods of alternate model structures, of varying complexity. Using multiple sources of data, we have highlighted systematic errors in the model at

> landscape scale (Figure 2 and 3) for LAI, GPP and biomass. However, these biases are not consistent for site-scale evaluations. Thus, a next step would be to include explicitly both random and systematic process errors for C fluxes in the data assimilation. These errors could be determined from field scale evaluation of model process representation (Table 2) using e.g. FLUXNET2015 data. We also need to understand better the error associated with landscape heterogeneity of C stocks and fluxes, to upscale from flux tower observations, or direct measurements of LAI, to landscape pixel. This could be achieved by constructing robust observation error models (Dietze, 2017) from field to pixel scale, for e.g. GPP, LAI and foliar N. Evaluation of the sensitivity of C cycling DA analyses to observation error has shown relatively low sensitivity to data gaps and random error on net ecosystem flux data (Hill et al., 2012), but further analyses of error sensitivity are required for multiple streams of stock data."

Additional points of concern:

1) Neither the DALEC2 model nor the CARDAMOM system appear to be publically archived. This means this work can't be reproduced or expanded upon by others. I don't know if such lack of openness is within the letter of the law of this journal, but it's definitely a deviation from the current norms of the community.

We agree that openness is critical to scientific advances. We have submitted the code for DALEC2 on Edinburgh DataShare. We are working to release a community version of CARDAMOM. At present we invite researchers to contact us to gain access to the code.

We have adjusted the text (L517-520):

> "**Data and software availability**
>
> CARDAMOM output used in this study is available from Exbrayat and Williams (2018) from the University of Edinburgh's DataShare service at https://doi.org/10.7488/ds/2334. The DALEC2 code is also available on Edinburgh DataShare at https://doi.org/10.7488/ds/2504. Contact MW for access to the CARDAMOM software."

2) As noted in my original review, I'm not comfortable with this system being called data assimilation, at least not with some additional qualifier being added (e.g. "parameter data assimilation") to make it clear that the outputs are deterministic model forward simulations not a reanalysis. To me, calling this data assimilation is like calling linear regression "machine learning." Sure people do it, but it makes the term pretty meaningless.

We disagree; we are using Bayesian parameter calibration of a dynamic model - which is typically referred to as data assimilation or model-data fusion; see "Ecological Forecasting" p. 168, by M. Dietze. However, we adjust our introductory text to improve clarity (S1; L100-104):

> "To address these issues we integrate model and data more formally. We apply data assimilation (DA), defined as a Bayesian calibration process for a model of a dynamic

> system. DA, through probabilistic parameterisation, supports robust model estimates of C stocks and fluxes consistent with multiple observations and their errors (Fox et al., 2009; Luo et al., 2009; Williams et al., 2005). By following Bayesian methods, the uncertainty on observations weights the degree of data constraint, and the outcome is a set of acceptable parameterisations for a given model structure linked to likelihoods."

3) After clearly diagnosing your photosynthesis scheme (ACM) as being at the root of model biases and compensating errors, the decision to not include any ACM parameters in the calibration (and toss the issue up to a lack of acclimation rather than simple miscalibration) strikes me as odd and I cannot understand why the authors are digging in their heels on this.

We do include an ACM parameter (C_{eff}) in the calibration (and so it is adjusted by the MHMCMC), according to Bloom et al. (2016). We apologise for not making this clear. We consequently have adjusted the Methods text (S2.2.1; L143-145) to read:

> "DALEC2 simulates canopy-level GPP via the Aggregated Canopy Model (ACM; Williams et al., 1997) and the most sensitive ACM parameter, related to canopy photosynthetic efficiency, is included in the CARDAMOM calibration."

4) Similar to (3), since NPP in DALEC is very tightly tied to GPP, and TT = Cstock/NPP, it sure seems like systematic biases in GPP will translate to systematic biases in TT. As noted earlier, I find some of the reported TT estimates to be implausible and don't understand the authors resistance to even considering comparing their results to independent field estimates.

We note that the mean NPP for GVMs across the region is 8% lower than in CARDAMOM, so the regional GVM-CARDAMOM NPP analyses are less different on average than the comparisons of CARDAMOM against data such as FLUXCOM (for GPP). We note that the high latitude TT estimates for CARDAMOM, GVMs (Figure 5) and reported in Carvalhais et al. (2014) are broadly similar. The critical issue we identify is that the spatial differences in NPP and C_{veg} between CARDAMOM and GVMs result in important spatial mismatches in TT estimated by both (compare Figure 5 and Figure 6).

We are confused at the statement that we have "**resistance to even considering comparing their results to independent field estimates**"; we have presented a clear evaluation against multiple independent FLUXNET site data, shown in Figure 4. Nonetheless, we add some further field-based estimates to complement these comparisons in the Discussion (S4.1, L421-435):

> "For a further independent evaluation of CARDAMOM outputs, we compare the tundra and boreal estimates to plot scale flux and stock information. For tundra, Street et al. (2012) calculate growing season GPP estimates of 263-380 g C m^{-2} for *Empetrum nigrum* communities, and 295-386 g C m^{-2} for *Betula nana* communities, which is consistent with the ranges in Figure 1 for tundra. Biomass stocks for Arctic tundra recorded in the Arctic LTER at Toolik Lake range from 105-1160 g C m^{-2} (Hobbie and Kling, 2014), which are consistent with the estimates from CARDAMOM, albeit at the lower end of the model estimates. For boreal forests, Goulden et al. (2011) report annual GPP estimates across a chronosequence of stands, and thus a variation across canopy densities, which varied from 450-720 g C m^{-2} yr^{-1}. These data are consistent with the

> span of GPP in CARDAMOM (Figure 1), again best matching the lower end of the model estimates. For the same study, the vegetation C stock estimates varied from 100-5000 g C m^{-2}, consistent with CARDAMOM, and with measurements of 10 to 40-year old boreal stands best matching the CARDAMOM median estimate of ~1500 g C m^{-2}. We conclude from comparisons against site data that CARDAMOM analyses are broadly consistent, with some tendency for CARDAMOM to have a high bias. This comparison is similar to the FLUXNET2015 evaluation of CARDAMOM. But it conflicts with the estimation of low bias from the comparison of CARDAMOM against FLUXCOM GPP and Carvalhais et al. (2014) biomass stock maps. It is possible that the scale differences between site level products and landscape estimates is confusing these comparisons, but there is clearly a need to understand better these inconsistencies in C cycle estimates."

5) The differences between DALEC and observations are greater than the differences between DALEC and the ISI-MIP models, so why are the authors so hard on the ISI-MIP models?

Our key point is that DALEC outputs match the spatial variation in independent (FLUXCOM) and assimilated data (LAI, biomass) well. There may be biases in these comparisons, indicative of model process error and/or upscaling error in the biomass and FLUXCOM products, but CARDAMOM can match the pattern in LAI, biomass, and SOC very well (Figure 2). The poor agreement with ISI-MIP models is with the spatial pattern (Table 3), not with regional median values (Figure 5). From these analyses we note that a reasonable regional estimate is not very useful if patterns are wrong, as this challenges the reliability of ISIMIP models when used for projections. Some models actually match CARDAMOM well, and we noted this clearly. We have edited the text to emphasise these points:

In Results (S3.4, L318-330):

> "We used our highest confidence retrievals of NPP, C_{veg} and TT_{veg} (i.e. retrievals including assimilated LAI, biomass and SOC) to benchmark the performance of the GVMs from the ISI-MIP2a project. In this assessment we compared not only their spatial variability across the pan-Arctic, tundra and taiga region (Figure 5), but also the degree of agreement between their mean model ensemble within the 90% confidence interval of our assimilation framework (Figure 6, Table 3). NPP estimates (RMSE = 0.1 kg C m^{-2} yr^{-1}; R^2= 0.44) are in better agreement than C_{veg} (RMSE = 1.8 kg C m^{-2}; R^2= 0.22) and TT_{veg} (RMSE = 4.1 years; R^2= 0.12). The assessed GVMs estimated on average 8% lower NPP, 16% higher C_{veg} and 22% longer TT_{veg} than CARDAMOM across the entire pan-Arctic domain (Figure 5 and 6) on average. Thus, at regional aggregation CARDAMOM analyses agreed more closely with ISI-MIP2a models than with FLUXCOM (51% difference) and with the Carvalhais et al. (2014) biomass data (28% bias).
>
> The poor spatial agreement regarding TT_{veg} between CARDAMOM and ISI-MIP2a (Table 3) is indicative of uncertainties in the internal C dynamics of these models. For instance, the slopes in Table 3 are steep and the R^2 are poor – so there is a substantial disagreement in the spatial pattern, not just a large bias. For ISI-MIP2a comparison R^2 values ranged from 0.03-0.52 for NPP; 0.00-0.31 for C_{veg}; and 0.00-0.24 for TT_{veg}."

In Discussion (S4.3, L449-451):

> “Using CARDAMOM as a benchmarking tool for six GVMs we found disagreements that varied among models for spatial estimates of NPP, C_{veg} and TT_{veg} across the Pan-Arctic (Figure 6) in comparison against CARDAMOM confidence intervals.”

Detailed comments:

L60: The authors responses suggested that a more complex calculation of TT was actually performed that relaxed the assumption of steady state. I would include that here (along with the steady state calculation) as I suspect a number of readers (myself included) would prefer to know that you’re not relying on a steady state assumption to assess a system that’s clearly not in steady state.

The residence time is calculated as per Bloom et al. (2016) equation S8 (SI text, S3 Global State and Process Variables), which specifically accounts for changes in stocks over time. We now adjust the text accordingly in the Introduction (S1) by removing “at steady state” and the Methods (S2.2.2; L202-203):

> “We calculate the transit time for C pools using the approach for non-steady state pools described in Bloom et al. (2016), supplementary information S3.”

L160: This line refers to DALEC2 as an ‘intermediate complexity’ model, but later arguments actually hinge on it being a simple model, and most of us would consider DALEC to really be on the simple end of the process-model spectrum

We have had internal discussions about where on the spectrum of complexity DALEC lies. We have decided that simple models would have only a handful of parameters and few state variables. DALEC has 17 parameters and 6 state variables, so it just qualifies as intermediate. We agree that this is partially a subjective categorisation (now in S2.2.2; L157). We also changed wording in L110, L447, L474, and L478 to keep consistency across the full text.

L171: MODIS LAI reports an uncertainty estimate. How did you aggregate those uncertainties when aggregated the observations? This is nontrivial as neither the MODIS products or MODIS LAI validation papers report anything about the spatial or temporal autocorrelation in the product’s errors.

We have adjusted our text to report on MODIS uncertainties (S2.2.2, L191-193):

> “For MODIS LAI we inflate the proportional error further to log(2) based on well reported biases in this product for evergreen forests (De Kauwe et al. 2011) and the estimated measurement and aggregation uncertainty for boreal forest LAI of 1 m^2 m^{-2} reported by Goulden et al. (2011).”

We have also adjusted the discussion to note the challenge for scaling these errors (S4.3, L496-499):

> “We also need to understand better the error associated with landscape heterogeneity of C stocks and fluxes, to upscale from flux tower observations, or direct measurements

> of LAI, to landscape pixel. This could be achieved by constructing robust observation error models (Dietze, 2017) from field to pixel scale, for e.g. GPP, LAI and foliar N."

L188: Table S2 looks like it just contains a bunch of uniform priors for all other parameters. I think that should be stated here so that readers don't need to find the supplement to learn that. It's perfectly fair, however, to make readers go to the supplement to see the exact numerical values of the priors.

We now include a note (S2.2.1, L143):

> "(Table S2; most priors are uniform with broad ranges)"

Moreover, we corrected a mistake with C pools units in Table S2. We replaced g C m^{-2} yr with g C m^{-2}.

L194: This sentence states that MODIS doesn't report an uncertainty estimate, but that's not accurate.

The cited statement was removed and we have adjusted (see above) our text to report on MODIS uncertainties (S2.2.2, L191-193):

> "For MODIS LAI we inflate the proportional error further to log(2) based on well reported biases in this product for evergreen forests (De Kauwe et al. 2011) and the estimated measurement and aggregation uncertainty for boreal forest LAI of 1 m^2 m^{-2} reported by Goulden et al. (2011)."

L206: I'm concerned about the way the statistics are being reported here. For example, the RMSE of a model is traditionally based on the model error (difference between the model and the observations). Here, the authors are defining the model's RMSE as the RMSE after applying both a multiplicative and additive bias correction (i.e. the predicted/observed regression). Similarly, the R2 isn't the variance explained by the model, but the variance jointly explained by the model and a linear bias correction to that model. This results in a very optimistic view of the model's actual performance.

We have calculated RMSE following the traditional approach, and we have adjusted the text to clarify this (S2.3, L207-209):

> "To assess the degree of statistical agreement we calculated linear goodness-of-fit (slope, intercept, R^2) between CARDAMOM and the two independent datasets and determined RMSE and bias from direct comparison on model-data residuals."

Following the same logic, we have also clarified this in S2.3, L221-223:

> "We performed a point-to-grid cell comparison to assess the degree of agreement between each flux magnitude and seasonality calculating the statistics of linear fit (slope, intercept, R^2) per flux and site between CARDAMOM and FLUXNET2015 datasets and determined RMSE and bias from model-data residuals comparison."

L251: Just want to continue to express my skepticism about some of these pool and flux estimates. For example, in my own experiences in Alaska, the boreal forest has WAY

more than 160% more structural tissue than the tundra. There needs to be some independent plot-scale validation of this.

Independent data from Toolik Lake (tundra) and Boreas (boreal) sites shows the general validity of the CARDAMOM outputs at these intensively studied ecological field sites.

As we presented earlier on, we included the following text in S4.1, L421-430:

> "For a further independent evaluation of CARDAMOM outputs, we compare the tundra and boreal estimates to plot scale flux and stock information. For tundra, Street et al. (2012) calculate growing season GPP estimates of 263-380 g C m^{-2} for *Empetrum nigrum* communities, and 295-386 g C m^{-2} for *Betula nana* communities, which is consistent with the ranges in Figure 1 for tundra. Biomass stocks for Arctic tundra recorded in the Arctic LTER at Toolik Lake range from 105-1160 g C m^{-2} (Hobbie and Kling, 2014), which are consistent with the estimates from CARDAMOM, albeit at the lower end of the model estimates. For boreal forests, Goulden et al. (2011) report annual GPP estimates across a chronosequence of stands, and thus a variation across canopy densities, which varied from 450-720 g C m^{-2} yr^{-1}. These data are consistent with the span of GPP in CARDAMOM (Figure 1), again best matching the lower end of the model estimates. For the same study, the vegetation C stock estimates varied from 100-5000 g C m^{-2}, consistent with CARDAMOM, and with measurements of 10 to 40-year old boreal stands best matching the CARDAMOM median estimate of ~1500 g C m^{-2}".

L258: Likewise, this stem turnover time seems much too fast and needs independent validation. I understand that grid cell to plot- or plant-scale validation isn't perfect, but it's better to report the performance explicitly, and then cushion it based on possible scale mismatch, rather than to ignore whether these estimates are consistent with prior research.

Based on comparison to Carvalhais et al. (2014) TT estimates and to the GPP and C_{veg} estimates reported above, our TT estimates are consistent with independent calculations and their component parts. We understand that TT seem short compared to concepts of stand age. However, litterfall (plant mortality) occurs throughout succession, from all live pools, which means that C turns over faster than age suggests.

L294: typo on "uncertainties"

Corrected.

L313: It would be good to have some sort of quantification of spatial coherence beyond RMSE & R2 (which are nonspatial). Look to the GIS and remote sensing literature for examples of what sort of statistics are available to do this.

There are a number of potential statistics to use. We suggest that our choice of statistics is familiar to biogeochemists and earth system scientists. Coupled with direct mapping of ratios and confidence intervals for visual assessment, we suggest our analysis provides readers with the relevant information on spatial coherence. Adding further statistics is likely to provide only marginal gains, but also increase the intricacy of an already complex paper.

L328: Don't introduce new Methods in the Results. Please document what this analysis is and why you are doing it earlier in the paper.

We agree the reviewer 2 is correct and we have adjusted the text as requested, moving material into the last part of the Methods (S2.4, L235-239):

> "To understand the sources of errors in TT_{veg} calculations, we used CARDAMOM to calculate two hypothetical TT_{veg} (i.e. EXPERIMENT A TT_{veg} = ISI-MIP2a C_{veg} / CARDAMOM NPP and EXPERIMENT B TT_{veg} = CARDAMOM C_{veg} / ISI-MIP2a NPP) and then assessed the largest difference with CARDAMOM's CONTROL TT_{veg}. We estimated the hypothetical TT_{veg} for each pixel in each model, and derived a pixel-wise measure of the contribution of biases in NPP and C_{veg} to biases in TT_{veg} by overlapping their distribution functions."

L378: Consistent with my previous concerns, DALEC appears to be running to fast. That said, this is still a comparison to other models, not to data.

We agree that biases may exist in the CARDAMOM TT estimate, but see above about difference between stand age and TT (L258 comment). Also, note that we are exploring where ISI-MIP2a models lie outside the analysis confidence intervals of CARDAMOM for TT.

L391: Here you say you had a 'strong prior on photosynthesis' but as far as I can tell the photosynthetic parameters were fixed at defaults, not assigned priors. According to Eqn 2, the only 2 parameters assigned non-uniform priors were canopy efficiency (which in Tables 2 and S2 is labeled as a phenology parameter) and autotrophic respiration

As noted before, the canopy efficiency is the calibrated parameter in CARDAMOM for the photosynthesis model ACM; we apologise for confusion in not making this clear before. Now this point is clarified in text (S2.2.2, L143-145):

> "DALEC2 simulates canopy-level GPP via the Aggregated Canopy Model (ACM; Williams et al., 1997) and the most sensitive ACM parameter, related to canopy photosynthetic efficiency, is included in the CARDAMOM calibration."

L397: If you've demonstrated a bias in your photosynthetic model, I'm not sure I agree that this could be resolved with more precise data if you're not updating the parameters in the photosynthetic submodel

Again, we have now clarified that a parameter in the photosynthesis model (canopy efficiency in ACM) is being updated by CARDAMOM.

L427: I fundamentally disagree that models should be benchmarked against highly-derived, model-based data products. But this isn't the central point of the paper and thus I won't hold up this paper over that disagreement.

Every data product used here is in some way model-derived – LAI from MODIS requires a model, biomass from radar and landcover maps also, SOC data from interpolation and machine learning approaches, even in-situ data such as GPP and R_{eco} are separated from NEE using a wide range of partitioning algorithms.

L459: While it's true that brute-force MCMC is not feasible for complex models, but there are other options available that do work with larger models, such emulators (Fer et al 2018 Biogeoscience) and ensemble or particle filters.

We agree that there are a range of alternative approaches beyond MCMC and decided to include a sentence in Discussion including reviewer 2's suggestion (S4.3; L476-477):

> "Other viable options include using emulators (Fer et al., 2018) and particle filters (Arulampalam et al., 2002), but MCMC methods provide the most detailed description of error distributions."

We also re-arranged the following sentences and merged paragraphs to improve clarity (S4.3; L477-486):

> "There remains a strong argument to utilize intermediate complexity models like DALEC2 to evaluate the minimum level of detail required to represent ecosystem processes consistent with local observations, and to allow testing of alternate model structures. And, assimilating further data products, for instance patterns in soil hydrology and snow states across the pan-Arctic from earth observation, could provide useful information on spatio-temporal controls on soil activity and microbial metabolism to constrain below ground processes. This information would need to be tied to process level information on SOM turnover generated from experimental studies, and included in updated versions of DALEC. Thus, more field observations are crucial across the pan-Arctic, specifically on decomposition and TT of SOC (He et al., 2016) and respiratory processes such as partitioning of R_{eco} into R_a and R_h (Hobbie et al., 2000; McGuire et al., 2000), across the growing season and also during wintertime (Commane et al., 2017; Zona et al., 2016)."

L477: For the record, if you didn't fit every grid cell independently then you wouldn't need to upscale/interpolate field observations.

Our point is that critical ecological processes remain poorly understood, and so further field observations are required to constrain these processes. Also, if each pixel had not been treated independently, we would have then relied on PFTs with all their problems (clearly pointed at Introduction and Discussion sections), and which is basically the opposite to what CARDAMOM framework is about.

L495: Where are the DALEC2 and CARDAMOM code repositories?

Following up on reviewer 2 initial concern, we have submitted the code for DALEC2 on Edinburgh DataShare. We are working to release a community version of CARDAMOM.

We have adjusted the text (L517-520):

> "**Data and software availability**
>
> CARDAMOM output used in this study is available from Exbrayat and Williams (2018) from the University of Edinburgh's DataShare service at https://doi.org/10.7488/ds/2334. The DALEC2 code is also available on Edinburgh

DataShare at https://doi.org/10.7488/ds/2504. Contact MW for access to the CARDAMOM software."

Table 2: I find it interesting that, given the papers focus on turnover times, turnover parameters are the least constrained part of the model.

Yes, this is the case, and reinforces the focus on TT in this analysis – we will only improve forecasts of high latitude C dynamics from better understanding TT.

Evaluation of terrestrial pan-Arctic carbon cycling using a data-assimilation system

Efrén López-Blanco[1,2], Jean-François Exbrayat[2,3], Magnus Lund[1], Torben R. Christensen[1,4], Mikkel P. Tamstorf[1], Darren Slevin[2], Gustaf Hugelius[5], Anthony A. Bloom[6], Mathew Williams[2,3]

[1] Department of Biosciences, Arctic Research Center, Aarhus University, Frederiksborgvej 399, 4000 Roskilde, Denmark
[2] School of GeoSciences, University of Edinburgh, Edinburgh, EH93FF, UK
[3] National Centre for Earth Observation, University of Edinburgh, Edinburgh, EH9 3FF, UK
[4] Department of Physical Geography and Ecosystem Science, Lund University, Sölvegatan 12, 223 62 Lund, Sweden
[5] Department of Physical Geography and Bolin Centre for Climate Research, Stockholm University, 106 91 Stockholm, Sweden
[6] Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA 91109, USA

Correspondence to: Efrén López-Blanco (elb@bios.au.dk) and Mathew Williams (mat.williams@ed.ac.uk)

Keywords: Arctic tundra, Arctic taiga, Net Ecosystem Exchange, Gross primary production, Ecosystem Respiration, carbon stocks, transit times, field observations, global vegetation models.

Abstract. There is a significant knowledge gap in the current state of the terrestrial carbon (C) budget. Recent studies have highlighted poor understanding particularly of C pool transit times, and whether productivity or biomass dominate these biases. The Arctic, accounting for approximately 50% of the global soil organic C stocks, has an important role in the global C cycle. Here, we use the CARDAMOM data-assimilation system to produce pan-Arctic terrestrial C cycle analyses for 2000-15. This approach avoids using traditional plant functional type or steady-state assumptions. We integrate a range of data (soil organic C, leaf area index, biomass, and climate) to determine the most likely state of the high latitude C cycle at a 1° x 1° resolution, and also to provide general guidance about the controlling biases in transit times. On average, CARDAMOM estimates regional mean rates of photosynthesis of 565 g C m^{-2} yr^{-1} (90% confidence interval between the 5th and 95th percentiles: 428, 741), autotrophic respiration of 270 g C m^{-2} yr^{-1} (182, 397) and heterotrophic respiration of 219 g C m^{-2} yr^{-1} (31, 1458), suggesting a pan-Arctic sink of -67 (-287, 1160) g C m^{-2} yr^{-1}, weaker in tundra and stronger in taiga. However, our confidence intervals remain large (and so the region could be a source of C), reflecting uncertainty assigned to the regional data products. We show a clear spatial and temporal agreement between CARDAMOM analyses and different sources of assimilated and independent data at both pan-Arctic and local scales, but also identify consistent biases between CARDAMOM and validation data. The assimilation process requires clearer error quantification on LAI and biomass products to resolve these biases. Mapping of vegetation C stocks and change over time, and soil C ages linked to soil C stocks is required for better analytical constraint. Comparing CARDAMOM analyses to global vegetation models (GVM) for the same period, we conclude that transit times of vegetation C are inconsistently simulated in GVMs due to a combination of uncertainties from productivity and biomass calculations. Our findings highlight that GVMs need to focus on constraining both current vegetation C stocks and net primary production, to improve process-based understanding of C cycle dynamics in the Arctic.

1 Introduction

Arctic ecosystems play a significant role in the global carbon (C) cycle (Hobbie et al., 2000; McGuire et al., 2012). Slow organic matter decomposition rates due to cold and poorly drained soils in combination with cryogenic soil processes have led to an accumulation of large stocks of C stored in the soils, much of which is currently held in permafrost (Tarnocai et al., 2009). The permafrost region soil organic C (SOC) stock is more than twice the size of the atmospheric C stock; and accounts for approximately half of the global SOC stock (Hugelius et al., 2014; Jackson et al., 2017). High latitude ecosystems are experiencing a temperature increase that is nearly twice the global average (AMAP, 2017). The expected future increase of temperature (IPCC, 2013)~~,~~ and precipitation (Bintanja and Andry, 2017)~~(Bintanja and Andry, 2017), and growing season length (Aurela et al., 2004; Groendahl et al., 2007)~~ will likely have consequences for the Arctic net C balance. As high latitudes warm, C cycle dynamics may lead to an increase of carbon dioxide (CO_2) emissions through ecosystem respiration (R_{eco}) driven by, for example, larger heterotrophic respiration (Commane et al., 2017; Schuur et al., 2015; Zona et al., 2016), drought stress on plant productivity (Goetz et al., 2005) and episodic disturbances (Lund et al., 2017; Mack et al., 2011). ~~However~~Alternatively, temperature-induced vegetation changes (Forkel et al., 2016; Graven et al., 2013; Lucht et al., 2002) may ~~counter-balance those effects by photosynthetic enhancement~~the increase ~~of~~ gross primary productivity (GPP) ~~(Forkel et al., 2016; Graven et al., 2013; Lucht et al., 2002). Two examples are the increase of gross primary productivity (GPP)~~ due to extended growing seasons (Zeng et al., 2011), ~~nutrient availability and~~ CO_2 fertilization (Zhuang et al., 2006) and ~~the~~ shifts in vegetation ~~dynamics~~ cover such as greening (Myneni et al., 1997; Zhu et al., 2016) and shrub expansion (Myers-Smith et al., 2011). Consequently, ~~phenology shifts~~ecosystem responses may feedback on climate with unclear magnitude and sign (Anav et al., 2013; Murray-Tortarolo et al., 2013; Peñuelas et al., 2009). As a result of the significant changes that are already affecting the structure and function of Arctic ecosystems, it is critical to understand and quantify the historical C dynamics of the terrestrial tundra and taiga, and their ~~responses~~ sensitivity to climate ~~change~~ (McGuire et al., 2012).

Although the land surface is estimated to offset ~30% of anthropogenic emissions of CO_2 (Canadell et al., 2007; Le Quéré et al., 2018), the terrestrial C cycle is currently the least constrained component of the global C budget and large uncertainties remain (Bloom et al., 2016). Despite the importance of Arctic tundra and taiga biomes in the global land C cycle, our understanding of interactions between the allocation of C from net primary productivity (NPP), C stocks (C_{stock}), and transit times (TT), is deficient (Carvalhais et al., 2014; Friend et al., 2014; Hobbie et al., 2000). The TT is a concept that represents the time it takes for a particle of C to persist in a specific C stock and it is defined by the C stock and its outgoing flux, here addressed as TT = C_{stock} / NPP ~~at steady state~~. According to a recent study by Sierra et al. (2017), TT is an important diagnostic metric of the C cycle and a concept that is independent of model internal structure and theoretical assumptions for its calculation. Terms such as residence time (Bloom et al., 2016; Friend et al., 2014), turnover time (Carvalhais et al., 2016), and turnover rate (Thurner et al., 2016; TT = 1/turnover rate) are used in the literature to represent the concept of TT (Sierra et al., 2017). Studies have focused more on the spatial variability with climate of ecosystem productivity rather than C transit times (Friend et al., 2014; Nishina et al., 2015; Thurner et al., 2016; Thurner et al., 2017). Friend et al. (2014) detailed that transit time dominates uncertainty in terrestrial vegetation responses to future climate and atmospheric CO_2. They found a 30% larger variation in modelled vegetation C change than response of NPP. Nishina et al. (2015) also suggested that long term C dynamics within ecosystems (vegetation turnover and soil decomposition) are more critical factors than photosynthetic processes (i.e. GPP or NPP). The respective contribution of bias from biomass and NPP to biases in transit times remains unquantified. Without an appropriate understanding of current state and dynamics of the C cycle, its feedbacks to climate change remains highly uncertain (Hobbie et al., 2000; Koven et al., 2015).

There are currently efforts to incorporate both in-situ and satellite-based datasets to assess C cycle retrievals and to reduce their uncertainties. At local scale, the net ecosystem exchange (NEE) of CO_2 between the land surface and the atmosphere is usually measured using eddy covariance (EC) techniques (Baldocchi, 2003). International efforts have led to the creation of global networks such as FLUXNET (http://fluxnet.fluxdata.org/) and ICOS (https://www.icos-ri.eu/), to harmonise data and support the reduction of uncertainties around the C cycle and its driving mechanisms. However, upscaling field observations to estimate regional to global C budget presents important challenges due to insufficient spatial coverage of measurements and heterogeneous landscape mosaics (McGuire et al., 2012). Furthermore, harsh environmental conditions in high latitude ecosystems and their remoteness complicates the collection of high-quality data (Lafleur et al., 2012). Given the lack of continuous, spatially distributed in situ observations of NEE in the Arctic, it remains a challenging task to calculate with certainty whether or not the Arctic is a net C sink or a net C source, and how the net C balance will evolve in the future (Fisher et al., 2014). Over the past decade, regional to global products generated from in situ networks and/or satellite observations have improved our understanding of the terrestrial C dynamics. These range from machine-learning based upscaling of FLUXNET data (Jung et al., 2017), remotely-sensed biomass products (Carvalhais et al., 2014; Thurner et al., 2014) and the creation of a global soil database (FAO/IIASA/ISRIC/ISSCAS/JRC, 2012). ~~However, these products tend to lack clear error estimates.~~ Due to a reliance on interpolation and upscaling with other spatial data, it is challenging to evaluate these products for inherent biases.

Global Vegetation Models (GVM) have been developed to determine global terrestrial C cycling,~~es~~ ~~and~~ through representing vegetation and ~~ecosystem~~ soil processes, including ~~the structural~~vegetation dynamics (i.e. growth, competition, and turnover) and biogeochemical (i.e. water, carbon, and nutrients cycling) responses to climate variability (Koven et al., 2011; Sitch et al., 2003; Woodward et al., 1995). The advantage of using process-based models to characterise C dynamics is that processes which drive ecosystem-atmosphere interactions can be simulated and reconstructed when data is scarce. However, C cycle modelling in GVMs typically relies on ~~pre-arranged~~ parameters retrieved from literature, prescribed plant-functional-type (PFT) and a spin-up process~~es~~ ~~until~~ ensuring~~the~~ C stocks (biomass and SOC) reach ~~their~~ steady state. Further, inherent differences of model structure contribute more significantly to GVM uncertainties (Exbrayat et al., 2018; Nishina et al., 2014), than ~~from~~ do differences in climate projections (Ahlström et al., 2012). Many model inter-comparison projects have demonstrated a lack of coherence in future projections of terrestrial C cycling (Ahlström et al., 2012; Friedlingstein et al., 2014). Recent studies have used simulations from the first phase of the Inter-Sectoral Impact Model Inter-comparison Project (ISI-MIP) (Warszawski et al., 2014) to evaluate the importance of key elements regulating vegetation C dynamics, but also the estimated magnitude of their associated uncertainties (Exbrayat et al., 2018; Friend et al., 2014; Nishina et al., 2014; Nishina et al., 2015; Thurner et al., 2017). An important insight is that TTs in GVMs are a key uncertain feature of the global C cycle simulation. Further, GVMs tend not to report uncertainties in their estimates of stocks and fluxes, which weakens their analytical value.

~~An approach t~~To address these issues ~~is to~~we integrate model~~s~~ and data more formally. We~~. We us~~apply~~e~~ data assimilation (DA), defined as a Bayesian calibration process for a model of a dynamic system. DA, through probabilistic parameterisation, supports robust model estimates of C stocks and fluxes consistent with multiple observations and their errors (Fox et al., 2009; Luo et al., 2009; Williams et al., 2005). By following Bayesian methods, the uncertainty on observations weights the degree of data constraint, and the outcome is a set of acceptable parameterisations for a given model structure linked to likelihoods. ~~Data assimilation quantifies how model parameters can be calibrated probabilistically using Bayesian methods to estimate C stocks and fluxes consistent with multiple observations and their errors (Fox et al., 2009; Luo et al., 2009; Williams et al., 2005). By following Bayesian methods, the uncertainty on observations weights the degree of data constraint, and the outcome is a set of acceptable parameterisations linked to likelihoods.~~ Overall, this approach determines whether model structure, observations and forcing are (in)consistent, and thus assesses validity of model structure. By assimilating co-located climatic, ecological and biogeochemical data from remote sensing observations at a specific grid scale across landscapes and regions ~~we~~ DA can map parameter estimation and uncertainties.

Here, we use the CARbon DAta MOdel framework (CARDAMOM) (Bloom et al., 2016; Bloom and Williams, 2015; Smallman et al., 2017) to ~~retrieve~~ analyse the pan-Arctic terrestrial carbon cycle at 1° resolution for the 2000-2015 period ~~in agreement~~, We assimilate ~~with~~ gridded observations of LAI, biomass and SOC stocks at these spatio-temporal scales into an intermediate complexity C model (DALEC2, which is less complex than GVMs). We compare analyses of C dynamics of Arctic tundra and taiga against (a) global products of GPP (Jung et al., 2017) and heterotrophic respiration (R_h) (Hashimoto et al., 2015); (b) NEE, GPP and R_{eco} field observations from 8 ~~sub- and high- Arctic~~high latitude sites included in the FLUXNET2015 dataset, and (c) 6 ~~extensively used~~ GVMs from the ISI-MIP2a comparison project (Akihiko et al., 2017). Our objectives are to (1) present and evaluate the analyses and uncertainties of the current state of the pan-Arctic terrestrial C cycling using a ~~model-data fusion~~ DA system, (2) quantify the degree of agreement between the CARDAMOM product with local to global scale sources of available data to assess analytical bias;~~,~~ and (3) use CARDAMOM as a benchmarking tool for the ISI-MIP2a models, to provide general guidance towards GVM improvements in transit time simulation~~, taking the advantage that this assimilation system produces error estimates, and is constrained by observations~~. Finally, we suggest future work to be done in the context of advancing pan-Arctic C cycle~~ing~~ modelling.

2 Data and methods

2.1 Pan-Arctic region

The spatial domain we considered in this study (Figure S1) corresponds to the extent of the Northern Circumpolar Soil Carbon Database version 2 (NCSCDv2) dataset (Hugelius et al., 2013a; Hugelius et al., 2013b), bounded by latitudes 42°N - 80°N and longitudes 180°W - 180°E, and at a spatial resolution of 1º x 1º. This area of study totals 18.0 million km^2 of land area. We used the GlobCover vegetation map product developed by the European Space Agency (Bontemps et al., 2011) to separate regions dominated by non-forested (tundra) and forested (taiga) land cover types. A complete description of the classes included in each domain can be found in Figure S1 and caption. The differentiation between tundra and taiga grid cells is in agreement with the tree line delimitated by Brown et al. (1997) together with the tundra domain defined from the Regional Carbon Cycle Assessment and Processes Activity reported by McGuire et al. (2012). The extensive grasslands without presence of trees in some areas such as the in South Russia, Mongolia and Kazakhstan were neglected to focus on higher latitudes. This classification of tundra and taiga totals 8.1 and 9.9 million km^2 of land area, respectively.

2.2 The CARbon DAta MOdel framework

Here we use the CARbon DAta MOdel framework (CARDAMOM; Bloom et al., 2016) (list of acronyms can be found in Table S1) to retrieve terrestrial C cycle dynamics, including explicit confidence intervals, in the pan-Arctic region. CARDAMOM consist of two key components: (1) an ecosystem model, the Data Assimilation Linked Ecosystem Carbon version 2 (DALEC2) (Bloom and Williams, 2015; Williams et al., 2005), constrained by observations and (2) a data-assimilation system (Bloom et al., 2016). This framework reconciles observational datasets as part of a representation of the terrestrial C cycle in agreement with ecological theory.

2.2.1 DALEC2

The DALEC2 ~~ecosystem~~ecosystem model simulates monthly land-atmosphere C fluxes and the evolution of six C stocks (foliage, labile, wood, roots, soil organic matter (SOM) and surface litter) and corresponding fluxes. DALEC2 includes 17 parameters controlling the processes of plant phenology, photosynthesis (GPP), allocation of primary production to respiration and vegetation carbon stocks, plant and organic matter turnover rates, all established within specific prior ranges based on ecologically viable limits (Table S2; most priors are uniform with broad ranges). DALEC2 simulates canopy-level GPP via the Aggregated Canopy Model (ACM; Williams et al., 1997)~~)~~ and the most sensitive ACM parameter, related to canopy photosynthetic efficiency, is included in the CARDAMOM calibration. ~~DALEC2 simulates canopy-level GPP via the Aggregated Canopy Model (ACM; Williams et al., 1997) and its allocation to the four plant stocks (foliage, labile, wood and roots) and autotrophic respiration (R_a) as time-invariant fractions of GPP.~~ DALEC ~~-~~allocates net primary production to the four plant stocks (foliage, labile, wood and roots) and autotrophic respiration (R_a) as time-invariant fractions of GPP. Plant C decays into litter and soil stocks where microbial decomposition generates heterotrophic respiration (R_h). Turnover of litter and soil stocks is simulated using temperature dependent first-order kinetics. For practical purposes we aggregated the different C stocks into photosynthetic (C_{photo}; leaf and labile), vegetation (C_{veg}; leaf, labile, wood and roots), soil (C_{dom}; litter and SOM) and total ($C_{tot} = C_{photo} + C_{veg} + C_{dom}$) C stocks. The Net Ecosystem Exchange (NEE) is calculated as the difference between GPP and the sum of the respiration fluxes ($R_{eco} = R_a + R_h$), while Net Primary Productivity (NPP) is the difference between GPP and R_a. Only NEE follows the standard micrometeorological sign convection presenting the uptake of C as negative (sink), and the release of C as positive (source); both GPP and R_{eco} are reported as positive fluxes. In this study, we addressed C turnover rates and decomposition processes as their inverse rates, this is the C transit time (TT_{photo}, TT_{veg} and TT_{dom}), represented as the ratio between the mean C stock and the mean C input into that stock during the simulation period.

2.2.2 Data-assimilation system

The intermediate complexity of the DALEC2 model compared to typical GVMs facilitates computationally intense Monte-Carlo (MC) data-assimilation to optimize the initial stock conditions and the 17 process parameters that shape C dynamics. CARDAMOM is forced with climate data from the European Centre for Medium-Range Weather Forecast Reanalysis interim (ERA-interim) dataset (Dee et al., 2011) monthly for the 2000-2015 period. A Bayesian Metropolis-Hastings Markov chain ~~Monte Carlo~~MC (MHMCMC) algorithm is used to retrieve the posterior distributions of the process parameters according to observational constraints and Ecological and Dynamic constraints (EDCs; Bloom and Williams, 2015). EDCs ensure that DALEC2 simulations of the terrestrial carbon cycle are realistic and ecologically viable and help to reduce the uncertainty in the model parameters by rejecting estimations that do not satisfy different conditions applied to C allocation and turnover rates as well as trajectories of C stocks.

Observational constraints include monthly time series of Leaf Area Index (LAI) from the MOD15A2 product (Myneni et al., 2002), estimates of vegetation biomass (Carvalhais et al., 2014) and soil organic carbon content (Hugelius et al., 2013a; Hugelius et., 2013b) (Table S3). We aggregated ~130000 1-km resolution MODIS LAI data monthly within each 1°x 1° pixel. ~~We aggregated biomass data at 0.5° resolution from Carvalhais et al. (2014) to 1° resolution. These are~~Biomass based ~~was ~~on remotely-sensed forest biomass (Thurner et al., 2014) and upscaled GPP ~~based on data driven estimates ~~(Jung et al., 2011) covering the pan-Arctic domain~~,~~ was aggregated to 1° resolution (Carvalhais et al., 2014). We used the NCSCD spatial explicit product (Hugelius et al., 2013a; Hugelius et al., 2013b) which was generated from 1778 soil sample locations interpolated to a 1° grid.~~ There is significant uncertainty for these data, due to the models involved in generating LAI and biomass, and the interpolation process for soils. Hence we apply broad confidence intervals commensurate with this uncertainty (Equation 3).~~

We apply the setup described above to 3304 1° x 1° pixels (1686 in tundra; 1618 in taiga) using a monthly time step. Each pixel is treated independently without assuming a prior land cover or plant functional type and we assume no spatial correlation between uncertainties in all pixels. In each 1° x 1° pixel, we applied the MHMCMC algorithm to determine the probability distribution of the optimal parameter set and initial conditions (x_i; Table S2) given observational constraints (O_i; LAI, SOC and biomass, Table S3) using the same Bayesian inference approach described in Bloom et al. (2016):

$$p(x_i|O_i) \propto p(x_i)\, p(O_i|x_i) \tag{1}$$

First, in the expression 1, $p\,(x_i)$ represents the prior probability distribution of each DALEC2 parameter (x_i) and is expressed as:

$$p(x_i) = p_{EDC}(x_i)\ e^{-0.5\left(\frac{\log(f_{auto})-\log(0.5)}{\log(1.2)}\right)^2}\ e^{-0.5\left(\frac{\log(C_{eff})-\log(17.5)}{\log(1.2)}\right)^2} \tag{2}$$

where $p_{EDC}(x_i)$ is the prior parameter probability according to the EDCs included in Table S2 and described in Bloom and Williams (2015). In addition, prior values for two parameters and their uncertainties (canopy efficiency [C_{eff}] and fraction of GPP respired [f_{auto}]) are imposed with a log-normal distribution following Bloom et al. (2016) to be consistent with the global GPP range estimated in Beer et al. (2010) and f_{auto} ranges specified by DeLucia et al. (2007) respectively.

Second, $p(O|x_i)$ from expression 1 represents the likelihood of x_i with respect to O_i, and it is calculated based on the ability of DALEC2 to reproduce (1) biomass (Carvalhais et al., 2014), (2) SOC (Hugelius et al., 2013a, Hugelius et al., 2013b), and (3) MODIS LAI (Myneni et al., 2002). ~~“~~The reported uncertainty on biomass data from Thurner et al. (2014) was +/- 37% at pixel scale. Because of undetermined errors related to tree cover thresholds used in the upscaling, ~~,~~and to reflect unknown model structural error, we slightly inflate the error estimate and use a log-transform(1.5) of ×/÷1.5 ~~–~~(i.e. ×/÷1.5 spans 67% of the expected error). We use the same proportional error for SOC. For MODIS LAI we inflate the proportional error further to log(2) based on well reported biases in this product for evergreen forests (De Kauwe et al., 2011) ~~(de Kauwe et al. 2011)~~ and the estimated measurement and aggregation uncertainty for boreal forest LAI of 1 m^2 m^{-2} reported by Goulden et al. (2011). The uncertainty assumptions in expression 3 are chosen in lack of better knowledge about the combined uncertainties arising from model representation errors and observation errors:. ~~Because MODIS LAI, SOC and biomass data lack specific uncertainty estimates, we used the same broad uncertainty factors as per Bloom et al. 2016: log transformed (1.5) for SOC and biomass (i.e. ×/÷ 1.5 spans 67% of the expected error), both assumed to be representative of initial conditions, and log(2) for LAI:~~

$$p(O_i|x_i) = e^{-0.5\left(\frac{\log(O_{biomass})-\log(M_{biomass,0})}{\log(1.5)}\right)^2} e^{-0.5\left(\frac{\log(O_{SOC})-\log(M_{SOC,0})}{\log(1.5)}\right)^2} e^{-0.5\left(\frac{\log(O_{LAI,t})-\log(M_{LAI,t})}{\log(2)}\right)^2} \qquad (3)$$

For each 1° x 1° pixel we run three MHMCMC chains with 10^7 accepted simulations each until convergence of at least two chains. We use 500 parameter sets sampled from the second half of each chain to describe the posterior distribution of parameter sets. We produce confidence intervals of terrestrial C fluxes and stocks from the selected parameter sets. In the following we report highest confidence results (median; P50) and the uncertainty represented by the 90% confidence interval (5th percentile to 95th percentile, $\binom{P95}{P05}$). We calculate the transit time for C pools using the approach for non-steady state pools described in Bloom et al. (2016), supplementary information S3~~)~~.

2.3 Model evaluation against independent in situ and pan-Arctic datasets

At the pan-Arctic scale, we compared CARDAMOM GPP with FLUXCOM dataset from Jung et al. (2017). We also compared our CARDAMOM R_h with the global spatiotemporal distribution of soil respiration from Hashimoto et al. (2015) calculated by a climate-driven empirical model. To assess the degree of statistical agreement we calculated linear goodness-of-fit (slope, intercept, R^2) between CARDAMOM and the two independent datasets and determined RMSE and bias from direct comparison on model-data residuals. ~~To assess the degree of statistical agreement we calculated linear goodness-of-fit (slope, intercept, R^2, RMSE, and bias) between CARDAMOM and the two independent datasets.~~ The mapping includes stipples representing locations where the independent datasets are within CARDAMOM's 90% confidence interval.

At a local scale, we compare CARDAMOM NEE and its partitioned components GPP and R_{eco} estimates against monthly aggregated values from the FLUXNET2015 sites. We selected 8 sites (Belelli Marchesini et al., 2007; Bond-Lamberty et al., 2004; Goulden et al., 1996; Ikawa et al., 2015; Kutzbach et al., 2007; López-Blanco et al., 2017; Lund et al., 2012; Sari et al., 2017) located across sub- and high-Arctic latitudes, covering locations with different climatic conditions and dominating ecotypes (Table S4). For this evaluation, we compared the same years for both observations and CARDAMOM, and we selected data using day-time method (Lasslop et al., 2010) due to the absence of true night-time period during Arctic summers in some locations. Additionally, we selected a variable u^* threshold to identify insufficient turbulence wind conditions from year to year similar to López-Blanco et al. (2017). In this data-model comparison we included the median (P50) ± the 90% confidence interval (percentile 5th to 95th; $\binom{P95}{P05}$) including both random and u^* filtering uncertainty following the method described in Papale et al. (2006). Some of the sites lack wintertime measurements and we filtered out data for months with less than 10% observations. We performed a point-to-grid cell comparison to assess the degree of agreement between each flux magnitude and seasonality calculating the statistics of linear fit (slope, intercept, R^2~~, RMSE, and bias~~) per flux and site between CARDAMOM and FLUXNET2015 datasets and determined RMSE and bias from model-data residuals comparison.~~.~~

2.4 Benchmark of Global Vegetation Models from ISI-MIP2a

We compared CARDAMOM analyses of pan-Arctic net primary production (NPP), vegetation biomass carbon stocks (C_{veg}) and vegetation transit times (TT_{veg}) against six participating GVMs in the ISI-MIP2a comparison project (Akihiko et al., 2017). In this study we have considered DLEM (Tian et al., 2015), LPJmL (Schaphoff et al., 2013; Sitch et al., 2003), LPJ-GUESS (Smith et al., 2014), ORCHIDEE (Guimberteau et al., 2018), VEGAS (Zeng et al., 2005), and VISIT (Ito and Inatomi, 2012). The specific properties and degree of complexity of each ISI-MIP2a model are summarized in Table S5. The comparisons have been performed under the same spatial resolution as the CARDAMOM spatial resolution (1° x 1°) for the 2000-2010 period. Also, the chosen GVMs from the ISI-MIP2a phase have their forcing based on ERA-Interim climate data, similar to the forcing used in CARDAMOM. We estimated the degree of agreement using the statistics of linear fit (slope, intercept, R^2, RMSE, and bias) per variable and model between CARDAMOM and GVMs, but also their spatial variability including stipples where the GVM datasets are within the CARDAMOM's 90% confidence interval.

To understand the sources of errors in TT_{veg} calculations, w~~W~~e used CARDAMOM to calculate two hypothetical TT_{veg} (i.e. EXPERIMENT A TT_{veg} = ISI-MIP2a C_{veg} / CARDAMOM NPP and EXPERIMENT B TT_{veg} = CARDAMOM C_{veg} / ISI-MIP2a NPP) and then assessed the largest difference with CARDAMOM's CONTROL TT_{veg}. We estimated the hypothetical TT_{veg} for each pixel in each model, and derived a pixel-wise measure of the contribution of biases in NPP and C_{veg} to biases in TT_{veg} by overlapping their distribution functions.

3 Results

3.1 Pan-Arctic retrievals of C cycle

Overall, we found that the pan-Arctic region (Figure 1 and Table 1) acted as a small sink of C (area-weighted P50) over the 2000-2015 period with an average of -67.4 $\binom{1159.9}{-286.7}$ g C m^{-2} yr^{-1}, P50 $\binom{P95}{P05}$, although the 90% confidence intervals remain large (and so the region could be a source of C). Tundra regions NEE was estimated at -14.9 $\binom{1116.1}{-163.4}$ g C m^{-2} yr^{-1}, a weaker sink compared to taiga regions, -110.4 $\binom{1195.8}{-387.7}$ g C m^{-2} yr^{-1}. The photosynthetic inputs exceeded the respiratory outputs (GPP > R_{eco}; Table 1), although the much larger uncertainties stemming from R_{eco}, and more specifically from R_h, compared with GPP, complicate the net C sink/source estimate beyond the median's average ensembles. In the pan-Arctic region approximately half of GPP is autotrophically respired resulting in an NPP of 290.3 $\binom{410.7}{196.4}$ g C m^{-2} yr^{-1}. Carbon use efficiency (NPP/GPP) averages 0.51$\binom{0.55}{0.46}$, and marginally varied across tundra 0.51$\binom{0.54}{0.46}$ and taiga 0.52$\binom{0.56}{0.46}$. Despite these apparent small variations, tundra photosynthesized and respired (respectively 327.2$\binom{463.3}{236.8}$ and 310.0$\binom{1536.8}{124.3}$ g C m^{-2} yr^{-1}) approximately half as much as the Taiga region (759.8$\binom{967.9}{584.1}$ and 635.3$\binom{2114.0}{285.3}$ g C m^{-2} yr^{-1}).

The total size of the pan-Arctic soil C stock (C_{dom}) averaged 24.4$\binom{47.5}{10.3}$ kg C m^{-2}, 16-fold greater than the vegetation C stock (C_{veg}), 1.5$\binom{5.8}{0.5}$ kg C m^{-2}. The soil C stock (fresh litter and SOM) is dominated by C_{som}, accounting for the 99%, which also dominates the total terrestrial C stock in the pan-Arctic. Among the living C stocks, 93% of the C (88% in tundra and 90% in taiga) is allocated to the structural stocks (wood and roots; 1.4$\binom{5.6}{0.4}$ kg C m^{-2}) compared to 7% (12% in tundra and 10%

in taiga) to the photosynthetic stock (leaves and labile; $0.1\binom{0.2}{0.1}$ kg C m^{-2}). On average, the total ecosystem C stock is $26.3\binom{51.0}{11.8}$ kg C m^{-2} in the pan-Arctic region, with slightly lower stocks in tundra ($24.6\binom{50.6}{10.8}$ kg C m^{-2}) than taiga ($27.7\binom{51.2}{12.7}$ kg C m^{-2}). In general, the taiga region holds on average ~100 % more photosynthetic tissues, ~160 % more structural tissue and ~9 % more soil C stocks, than tundra. In other words, taiga holds ~12 % more total C than tundra. The greater living stock of C in taiga ($2.1\binom{5.1}{0.8}$ kg C m^{-2}) than tundra ($0.8\binom{6.8}{0.3}$ kg C m^{-2}) means that the relative size of R_a and R_h in the two regions differs. Thus in tundra R_a accounts for 51% of total ecosystem respiration, while in taiga this fraction is 57%. R_a is 4% larger than R_h in tundra, but 24% greater in taiga, reflecting the greater rates of C cycling in taiga. Uncertainties in estimates of soil C stock are notably higher than for living C stocks, highlighting the lack of observational and mechanistic constraint on heterotrophic respiration.

The global mean C transit time is $1.3\binom{2.1}{0.8}$ years in leaves and labile plant tissue (TT_{photo}), $4.5\binom{15.7}{1.7}$ years in stems and roots (TT_{veg}), and $120.5\binom{822.6}{9.8}$ years in litter and SOM (TT_{dom}). The total C transit time (TT_{tot}) ($133.1\binom{1013.6}{11.5}$ years) is clearly dominated by the soil C stock, highlighting the very long periods of times that C persists in Arctic soils. CARDAMOM calculated 62% longer TT_{dom} in tundra compared to taiga, likely linked to lower temperatures, but uncertainties are large due to the limitations of data constraints.

3.2 Data assimilation and uncertainty reduction

The CARDAMOM framework generated an analysis broadly consistent with the combination of SOC, biomass and LAI in each grid cell (Figure 2), and the errors assigned to these data products. The agreement for the SOC dataset by Hugelius et al. (2013a) is a 1:1 relationship ($R^2 = 1.0$; RMSE = 0.95 kg C m^{-2}), reflecting a straightforward model parameterisation. The biomass product from Carvalhais et al. (2014), was well correlated ($R^2 = 0.97$; RMSE = 0.46 kg C m^{-2}), but CARDAMOM was consistently biased ~28% low. MODIS LAI data were also well correlated ($R^2 = 0.79$; RMSE = 0.42 kg C m^{-2}), but ~28% higher than CARDAMOM analyses. These biases (Figure 2) likely arise due to a low estimate in the photosynthesis model (ACM) used in CARDAMOM which propagates through the C cycle. CARDAMOM balances uncertainty in data products and the models (ACM photosynthesis model and DALEC2), to generate a weighted analysis, typical of Bayesian approaches. The CARDAMOM analysis 90% confidence interval (CI) includes the 1:1 line for biomass and LAI (Figure 2), indicating that the likelihoods on C cycle analyses include the expected value of the observations.

The degree to which posterior distributions were constrained from the prior distributions in each of the 17 model parameters and 6 initial stock sizes (Table S2) varied considerably depending on the parameters in question and their related processes (Table 2 and Figure S2). The 90% CI posterior range of foliar, wood, labile and SOM C stocks (C_{foliar}, C_{wood}, C_{labile} and C_{som}) as well as parameters such as allocation to foliage (f_{fol}) and lifespan (L) were considerably reduced (>80% uncertainty reduction compared to priors) most likely controlled by the information on LAI, biomass and SOC constraints. Contrarily, parameters that have not been regulated in any way in the MHMCMC algorithm, i.e. turnover processes such as litter mineralization (MR_{litter}), roots turnover (TOR_{roots}), wood turnover (TOR_{wood}), decomposition rates (D_{rate}) and initial C stock such as litter (C_{litter}) were found poorly constrained (<20% uncertainty reduction). Overall, the uncertainty reduction classified by processes and ranked from most to least constrained estimated a 71% reduction for C stocks, 67% reduction for C allocation, 59% for plant phenology and 31% for C turnover related parameters. Although there are not substantial differences between tundra and taiga, C_{roots} was better constrained in tundra regions (42%), while leaf onset day (B_{day}), leaf fall day (F_{day}), and leaf fall duration (L_f) were better constrained in taiga regions (>18% or more).

3.3 Independent evaluation: from global to local scale

We compared our estimates of GPP and R_h with independent datasets to evaluate the model performance (Figure 3). We found GPP to be well correlated ($R^2 = 0.81$; RMSE = 0.43 kg C m^{-2}), but biased lower (~53%) compared to Jung et al.

(2017)’s GPP estimates. There are in general very few pixels where FLUXCOM product falls within CARDAMOM’s 90% confidence interval. Additionally, the R_h product from Hashimoto et al. (2015) is less consistent with our estimates ($R^2 = 0.40$; RMSE = 0.09 kg C m^{-2}), presenting a tendency towards lower values in tundra pixels and higher values in taiga pixels. The spatial variability of R_h is considerably smaller in Hashimoto et al. (2015) compared to our CARDAMOM estimates. R_h falls within the 90% confidence interval of CARDAMOM in most of the pan-Arctic region due to the fact that the R_h uncertainties are significant (Figure 3). This finding confirms the ~~uncetartanties~~uncertainties previously noted in modelled respiratory processes (Table 1) where the upper P95 in R_h dominated NEE’s uncertainties, but also the soil C stocks and transit times.

For comparison with direct ground observations from the FLUXNET2015 dataset, we report here monthly aggregated P50 ± P05-95 estimates of NEE, GPP and R_{eco} to show timing and magnitudes, but also to diagnose whether CARDAMOM is in general ~~agrremment~~agreement with flux tower data. Overall, CARDAMOM performed well in simulating observed NEE ($R^2 = 0.66$; RMSE = 0.51 g C m^{-2} $month^{-1}$; Bias = 0.16 g C m^{-2} $month^{-1}$), GPP ($R^2 = 0.85$; RMSE = 0.89 g C m^{-2} $month^{-1}$; Bias = 0.5 g C m^{-2} $month^{-1}$) and R_{eco} ($R^2 = 0.82$; RMSE = 0.63 g C m^{-2} $month^{-1}$; Bias = 0.35 g C m^{-2} $month^{-1}$) across 8 sub-Arctic and high-Arctic sites from the FLUXNET2015 dataset (Figure 4; Table S6). CARDAMOM NEE is ~25% lower than FLUXNET2015, while GPP and R_{eco} are ~30% and ~10% higher, respectively. This mismatch is important in the context of the FLUXCOM GPP upscaling, 50% higher than CARDAMOM GPP. At some sites such as Hakasia, Samoylov, Poker Flat and Manitoba (NEE $R^2 = 0.73$; GPP $R^2 = 0.92$ and R_{eco} $R^2 = 0.88$) CARDAMOM better matches the seasonality and the magnitude of the C fluxes than the rest, i.e. Tiksi, Kobbefjord, Zackenberg and UCI-1998 (NEE $R^2 = 0.58$; GPP $R^2 = 0.67$ and R_{eco} R^2=0.67). In general, CARDAMOM captured the beginning and the end of the growing season well (Figure 4), although the assimilation system has some bias due to (1) difference in timing (e.g. earlier shifts of peak of the growing season in Manitoba GPP and R_{eco} and earlier end of the growing season in Poker Flat NEE) and (2) differences in flux magnitudes (such as in Hakasia GPP and R_{eco} and Kobbefjord NEE).

3.4 Benchmarking ISI-MIP2a models with CARDAMOM

We used our highest confidence retrievals of NPP, C_{veg} and TT_{veg} (i.e. retrievals including assimilated LAI, biomass and SOC) to benchmark the performance of the GVMs from the ISI-MIP2a project. In this assessment we compared not only their spatial variability across the pan-Arctic, tundra and taiga region (Figure 5), but also the degree of agreement between their mean model ensemble within the 90% confidence interval of our assimilation framework (Figure 6, Table 3). ~~Overall, ISI-MIP2a models are in poor agreement with CARDAMOM across the pan-Arctic.~~ NPP estimates (RMSE = 0.1 kg C m^{-2} yr^{-1}; R^2= 0.44) are in better agreement than C_{veg} (RMSE = 1.8 kg C m^{-2}; R^2= 0.22) and TT_{veg} (RMSE = 4.1 years; R^2= 0.12). ~~Moreover, t~~The assessed GVMs estimated on average 8% lower NPP, 16% higher C_{veg} and 22% longer TT_{veg} than CARDAMOM across the entire pan-Arctic domain (Figure 5 and 6) on average. Thus, at regional aggregation CARDAMOM analyses agreed more closely with ISI-MIP2a models than with FLUXCOM (51% difference) and with the Carvalhais et al. (2014) biomass data (28% bias).~~, with very varied spatial patterns.~~

The poor spatial agreement regarding TT_{veg} between CARDAMOM and ISI-MIP2a (Table 3) is indicative of uncertainties in the internal C dynamics of these models. For instance, the slopes in Table 3 are steep and the R^2 are poor – so there is a substantial disagreement in the spatial pattern, not just a large bias. For ISI-MIP2a comparison R^2 values ranged from 0.03-0.52 for NPP; 0.00-0.31 for C_{veg}~~vegetation carbon~~; and 0.00-0.24 for TT_{veg}. Spatially, the stippling in Figure 6 indicates areas where the GVMs are within the 90% CI of CARDAMOM; agreement is best over the taiga domain rather than in tundra for TT_{veg}. The benchmark area of consistency (stippling) is more extensive for C_{veg} and TT_{veg} than for NPP. Thus, while there is a stronger spatial correlation for NPP between CARDAMOM and GVMs (Table 3), this is a clearer bias for NPP. Some models (LPJ-GUESS and ORCHIDEE) systematically calculate lower values in all the assessed variables while others (LPJmL and VISIT) calculate higher estimates. The models in closer agreement with CARDAMOM were DLEM (5%

difference) and LPJ-GUESS (17%) while VEGAS (44%) and ORCHIDEE (56%) were the models with larger discrepancies (Table 3; Figure 5 and 6).

The attribution analysis to identify the origin of bias from ISI-MIP2a models indicated a joint split between NPP and C_{veg} for TT_{veg} error simulated in GVMs (Figure 7). ~~We used CARDAMOM to calculate two hypothetical TT_{veg} (i.e. EXPERIMENT A TT_{veg} = ISI-MIP2a C_{veg} / CARDAMOM NPP and EXPERIMENT B TT_{veg} = CARDAMOM C_{veg} / ISI-MIP2a NPP) and then assessed the largest difference with CARDAMOM's CONTROL TT_{veg}. We estimated the hypothetical TT_{veg} for each pixel in each model, and derived a pixel-wise measure of the contribution of biases in NPP and C_{veg} to biases in TT_{veg} by overlapping their distribution functions (Figure 7).~~ The distribution of the differences relative to CARDAMOM revealed that the higher error (i.e. the lower overlapped area, and by extension the largest contributor to TT_{veg} biases) comes from ISI-MIP2a NPP with a 69% agreement in the distribution, while C_{veg} agrees 72%. In fact, the TT_{veg} R^2 for each model (Table 3) is very close to the product of the NPP R^2 and C_{veg} R^2 for that model, i.e. the uncertainty on the TT_{veg} is a direct interaction of NPP and C_{veg} uncertainty (R^2 of the correlation = 0.71). This finding supports Figure 6, which shows TT_{veg} error derives equally from both NPP and C_{veg}.

4 Discussion

4.1. Pan-Arctic retrievals of C cycle

The CARDAMOM framework has been used to evaluate the terrestrial pan-Arctic C cycle in tundra and taiga at coarse spatio-temporal scale (at monthly and annual time steps for the 2000-2015 period and at 1° x 1° grid cells). Overall, we found that the pan-Arctic region~~ (1)~~ was most likely a consistent sink of C (weaker in tundra and stronger in taiga), although the large uncertainties derived from respiratory processes (Table 1) strongly increase the 90% confidence interval uncertainty~~; (2)~~. We estimate that tundra experienced 62% longer transit times in litter and SOM C stocks~~ in tundra compared to~~ than taiga ecosystems~~; and (3)~~. Further, the contribution of R_a and R_h to total ecosystem respiration was similar in tundra (51%, 49%) but dominated by R_a in taiga (57% compared to ~~,~~43% in tundra).

CARDAMOM retrievals are consistent with outcomes from relevant papers such as the (I) C flux observations and model estimates reported in McGuire et al. (2012); (II) C stocks and transit times described by Carvalhais et al. (2014), and (III) NPP, C stocks and turnover rates stated in Thurner et al. (2017):

I. The CARDAMOM NEE estimates reported in this study for the tundra domain are inside the variability comparison of values compiled by McGuire et al. (2012) considering field observation, regional process-based models, global-process based models and inversion models. The authors reported that Arctic tundra was a sink of CO_2 of -150 Tg C yr^{-1} (SD=45.9) across the 2000-2006 period over an area of 9.16 x 10^6 km^2. Here, CARDAMOM NEE estimated -129 Tg C yr^{-1} over an area of 8.1 x 10^6 km^2 for the same period. This exhaustive assessment of the C balance in Arctic tundra included approximately 250 estimates using the chamber and eddy covariance method from 120 published papers (McGuire et al., 2012; Supplement 1) with an area-weighted mean of means of -202 Tg C yr^{-1}. The regional models, including runs from LPJ-Guess WHyMe (Wania et al., 2009a, b), Orchidee (Koven et al., 2011), TEM6 (McGuire et al., 2010), and TCF model (Kimball et al., 2009), reported a NEE of -187 Tg C yr^{-1} and GPP, NPP, R_a and R_h of 350, 199, 151 and 182 g C $m^{-2}y^{-1}$, respectively. GVMs applications such as CLM4C (Lawrence et al., 2011), CLM4CN (Thornton et al., 2009), Hyland (Levy et al., 2004), LPJ (Sitch et al., 2003), LPJ- Guess (Smith et al., 2001), O-CN (Zaehle and Friend, 2010), SDGVM (Woodward et al., 1995), and TRIFFID (Cox, 2001) estimated a NEE of -93 Tg C yr^{-1} and GPP, NPP, R_a and R_h of 272, 162, 83 and 144 g C $m^{-2}yr^{-1}$. For the same period, CARDAMOM has estimated 330, 167, 160 and 154 g C m^{-2} yr^{-1} respectively for the same gross C fluxes.

II. Carvalhais et al. (2014) estimated a total ecosystem carbon (C_{tot}) of $20.5\binom{52.5}{8.0}$ kg C m^{-2} for tundra and $24.8\binom{58.0}{15.2}$ kg C m^{-2} for taiga, while values from CARDAMOM were $24.6\binom{50.6}{10.8}$ kg C m^{-2} for tundra, and $27.7\binom{51.2}{12.7}$ kg C m^{-2} in taiga (Figure 5; Table 1) for the same area.Thus, Carvalhais et al. (2014)'s C_{tot} product stored 20% and 12% less carbon in tundra and taiga respectively than CARDAMOM. Overall, CARDAMOM calculated 20% and 6% longer transit times for tundra and taiga respectively, with average values of $80.8\binom{195.2}{21.8}$ years in tundra and $51.2\binom{109.3}{22.1}$ years in taiga (Table 1) compared to the $64.4\binom{259.8}{25.7}$ years in tundra and $48.2\binom{111.6}{24.9}$ years in taiga in Carvalhais et al. (2014). These numbers have been retrieved from the same biome classification and they include the 90% confidence interval of the assessed spatial variability. Also, we applied a correction factor of $TT_{gpp} = TT_{npp}$*(1-fraction of GPP respired) to be comparable with Carvalhais et al. (2014) TT. Both datasets agree on the fact that high (cold) latitudes, first tundra, and second taiga have the longest transit times in the entire globe (Bloom et al., 2016; Carvalhais et al., 2014).

III. A recent study from Thurner et al. (2017) assessed temperate and taiga-related TTs presenting a 5-year average NPP dataset applying both MODIS (Running et al., 2004; Zhao et al., 2005) and BETHY/DLR (Tum et al., 2016) products and an inovative biomass product (Thurner et al., 2014) accounting for both forest and non-forest vegetation. Our estimate of TT_{veg} for the exact same period is $5.3\binom{18.2}{1.9}$ years, compared to Thurner et al. (2017)'s TT, $8.2\binom{11.5}{5.5}$ years using MODIS and $6.5\binom{8.7}{4.2}$ years using BETHY/DLR. A note of caution here, the number reported by the authors are turnover rates, which are inferred to transit times by just applying the inverse of turnover rates (TT_{veg}=1/turnover rates). Additionally, their NPP estimates, 0.35 and 0.45 kg C m^{-2} yr^{-1} from both MODIS and BETHY/DLR, is only 5% more productive as average than CARDAMOM NPP estimate, $0.4\binom{0.6}{0.3}$ kg C m^{-2} yr^{-1}; and the biomass derived from Thurner et al. (2014), 3.0 ±1.1 kg C m^{-2}, is ~30% lower than CARDAMOM C_{veg}, $2.2\binom{5.0}{1.1}$kg C m^{-2}, calculated for the same period and for the same taiga domain.

In general, we found a reasonable agreement between CARDAMOM and assimilated and independent data at pan-Arctic scale. CARDAMOM retrievals of assimilated data are in good agreement with the SOC (Figure 2). The simulation of TT_{dom} is weakly constrained (Table 1) - our analysis adjusts TT to match mapped stocks, hence the strong match of modelled to mapped SOC. So, independent data on TT_{dom} data (e.g. ^{14}C) is required across the pan-Arctic region to provide stronger constraint on process parameters and reduce the very broad confidence intervals of CARDAMOM analyses. The low bias in mean estimates of LAI and biomass (Figure 2) likely relates to the strong prior on photosynthesis estimates from the ACM model, which lacks a temperature acclimation for high latitudes in this implementation. However, the uncertainty in the biomass and LAI analyses spans the magnitude of the bias. So, CARDAMOM generates some parameters sets that are consistent with observations. CARDAMOM produces analyses that reproduce the pattern of LAI, GPP, biomass and SOC (Figure 2 and 3) – this demonstrates that the DALEC model structure can be calibrated to simulate the links between these variables as a function of mass balance constraints, and realistic process interactions and climate sensitivities. ~~Biases could be reduced by assimilation of data with better resolved errors. Greater confidence in LAI and biomass data would increase the weight on their assimilation, and result in analyses closer to these data, overriding model priors by adjusting photosynthesis upwards. Further experiments can evaluate this possibility. Certainly the need for robust characterisation of error for data products is of critical importance for improved analyses.~~

There are clear biases in CARDAMOM analyses compared to independent global upscaled GPP (Jung et al., 2017) and R_h products (Hashimoto et al., 2015) (Figure 3). However, CARDAMOM resolves the spatial pattern in GPP effectively, while the spatial mismatch in R_h estimates is marked~~clear~~ (Figure 3), echoing the large uncertainty found in NEE (Figure 1, Table 1). One difference with Hashimoto et al. (2015)'s R_h model is the lack of moisture limitation on respiration in CARDAMOM. Conversely, GPP is relatively well-constrained in space through the assimilation of LAI and a prior for productivity (Bloom et al., 2016), although an important mismatch has been found: CARDAMOM GPP is 50% lower than FLUXCOM, but 30% higher than FLUXNET2015 EC data.

The agreement between CARDAMOM analyses and EC data is ~~good~~ high given the scale difference. A direct point-to-grid cell comparison with local observations derived from the FLUXNET2015 dataset (Figure 4, Table S6) is challenging and always difficult. CARDAMOM outputs covers 1° x 1° grid cells, whereas local eddy covariance flux measurements are in the order of 1-10 ~~hectares~~ha. Thus, for observational sites located in areas with complex terrain, such as Kobbefjord in coastal Greenland, the agreement can be expected to be low. For inland forest sites, such as Poker Flat in Alaska, there may be less differences in vegetation characteristics and local climatology between the local scale measurement footprint and the corresponding CARDAMOM grid cell. This scaling issue is likely to have a larger impact on flux magnitudes compared with seasonal dynamics. In general, CARDAMOM captured the seasonal dynamics in NEE, GPP and R_{eco} well (Figure 4, Table S6), although the monthly model time-step does reduce skill in shoulder seasons. There was ~~, however,~~ a consistent timing-mismatch in early season flux increase, where CARDAMOM predicts earlier growing season onset compared with observations. This is likely due to the impact of snow cover, which is not explicitly included in the CARDAMOM framework.

For a~~n~~ further independent evaluation~~e~~ of CARDAMOM outputs, we compare the tundra and boreal estimates to plot scale flux and stock information. For tundra, Street et al. (2012) calculate growing season GPP estimates of 263-380 g C m^{-2} for *Empetrum nigrum* communities, and 295-386 g C m^{-2} for *Betula nana* communities, which is consistent with the ranges in Figure 1 for tundra. Biomass stocks for Arctic tundra recorded in the Arctic LTER at Toolik Lake range from 105-1160 g C m^{-2} (Hobbie and Kling, 2014), which are consistent with the estimates from CARDAMOM, albeit at the lower end of the model estimates. For boreal forests, Goulden et al. (2011) report annual GPP estimates across a chronosequence of stands, and thus a variation across canopy densities, which varied from 450-720 g C m^{-2} yr^{-1}. These data are consistent with the span of GPP in CARDAMOM (Figure 1), again best matching the lower end of the model estimates. For the same study, the vegetation ~~carbon~~C stock estimates varied from 100-5000 g C m^{-2}, consistent with CARDAMOM, and with measurements of 10 to 40-year old boreal stands best matching the CARDAMOM median estimate of ~1500 g C m^{-2}. We conclude from comparisons against site data that CARDAMOM analyses are broadly consistent, with some tendency for CARDAMOM to have a high bias. This comparison is similar to the FLUXNET2015 evaluation of CARDAMOM. But it conflicts with the estimation of low bias from the comparison of CARDAMOM against FLUXCOM GPP and Carvalhais et al. (2014) biomass stock maps. It is possible that the scale differences between site level products and landscape estimates is confusing these comparisons, but there is clearly a need to understand better these inconsistencies in C cycle estimates.

4.2. CARDAMOM as a model benchmarking tool

An ideal benchmarking tool for GVMs would compare model state variables and fluxes against multiple, independent, unbiased, error-characterised measurements collected repeatedly at the same temporal and spatial resolution. Of course direct measurements of key C cycle variables like these are not available. Even at FLUXNET sites GPP and R_{eco} must be inferred, and NEE data often gap-filled. Satellite data can provide continuous fields, but do not directly measure ecological variables like biomass or LAI, so calibrated models are required to generate ecological products. Atmospheric conditions can introduce biases and data gaps into optical data that are poorly quantified. Upscaling of FLUXNET data requires other spatial data, e.g. MODIS LAI, which challenges the characterisation of error and generates complex hybrid products. We suggest that CARDAMOM provides some of the requirements of the ideal benchmark system – an error-characterised, complete analysis of the C cycle that is based on a range of observational products. CARDAMOM includes its own C cycle model; this has the advantage of evaluating the observational data for consistency (e.g. with mass balance), propagating error across the C cycle, and generating internal model variables such as TT. Further the model is of intermediate ~~low~~ complexity and independent of the benchmarked models.

Using CARDAMOM as a benchmarking tool for six GVMs we found ~~major~~ disagreements that varied among models for ~~mapping~~ spatial estimates of NPP, C_{veg} and TT_{veg} across the Pan-Arctic ~~for all models~~ (Figure 6) in comparison against CARDAMOM confidence intervals. GVM NPP estimates had a higher correlation than TT_{veg} and C_{veg} with CARDAMOM analyses (Table 3), but because CARDAMOM confidence intervals on NPP were relatively narrow (Figure 1) the benchmarking scores from GVM NPP were relatively poor (Figure 6). Consequently, we used CARDAMOM to calculate the relative contribution of productivity and biomass to the transit times bias by applying a simple attribution analysis (Figure 7). We concluded that the largest bias to transit times originated not by a deficient understanding of one single component, but by an equal combination of both productivity and biomass errors together. Therefore, this study partially agrees with previous studies (Friend et al., 2014; Nishina et al., 2014; Thurner et al., 2017) highlighting the deficient representation of transit times/turnover dynamics, but we further suggest that GVM and ESM modellers need to focus on the productivity and vegetation C stocks dynamics to improve inner C dynamics. A major challenge for GVMs is the spin-up problem (Exbrayat et al., 2014). GVMs need to find a way to ensure that the spin-up process produces biomass estimates consistent with the growing availability of biomass maps from earth observations. CARDAMOM solves this problem by avoiding spin-up. Its fast run time allows the biomass maps to act as a constraint on the probability distribution of model parameters. There may be opportunities to use CARDAMOM style approaches to assist the GVM community address this problem.

4.3 Outlook

Although CARDAMOM estimates for pan-Arctic C cycling are in moderately good agreement with observations and data constraints, we have not included important components controlling ecosystem processes that could potentially improve our understanding on C feedbacks, and with emphasis for high latitude ecosystems. For example, thaw and release of permafrost C is not represented in CARDAMOM, but the influence on vegetation dynamics, permafrost degradation and soil respiration is critical in high latitudes (Koven et al., 2015; Parazoo et al., 2018). Also, Koven et al. (2017) shown that soil thermal regimes are key to resolving the long-term vulnerability of soil C. Moreover, we have not characterized snow dynamics ~~and~~ nor the insulating effect of snow affecting respiratory losses across wintertime periods ~~either~~ (López-Blanco et al., 2018). Further, methane emissions, another important contributor to total C budget (Mastepanov et al., 2008; Zona et al., 2016), was neglected from this modelling exercise since it is not easy to model due to its three complex transport mechanisms (Walter et al., 2001).

However, our approach to use a ~~low~~n intermediate complexity model has the strong advantage of allowing very large (10^7) model ensembles per pixel, and thus a thorough exploration of model-parameter interactions, that is not feasible with typical ~~complex models~~GVMs. O~~There are o~~ther viable options include using~~available that work with larger models, such~~ emulators (Fer et al., 2018) and particle filters (Arulampalam et al., 2002), but MCMC methods provide the most detailed description of error distributions. There remains a strong argument to utilize intermediate ~~low~~ complexity models like DALEC2 to evaluate the minimum level of detail required to represent ecosystem processes consistent with local observations, and to allow testing of alternate model structures. And, assimilating further data products, for instance patterns in soil hydrology and snow states across the pan-Arctic from earth observation, could provide useful information on spatio-temporal controls on soil activity and microbial metabolism to constrain below ground processes. This information would need to be tied to process level information on SOM turnover generated from experimental studies, and included in updated versions of DALEC.

~~Thus, in order to reduce uncertainties on the balance between photosynthetic inputs and respiratory outputs, we must devise intermediate low complexity model representations of SOC decomposition by microbial activity (Xenakis and Williams, 2014), nutrient interactions with carbon (Thomas and Williams, 2014), and mechanisimsmechanisms driving carbon use efficiency (Manzoni et al., 2018), and or photosynthesis-respiration couplingdrivers of gross flux coupling (López-Blanco et al., 2017). There are opportunities to constrain such modelling using data on plant trait relationships across pan-Arctic~~

~~regions (Reichstein et al., 2014; Sloan et al., 2013). We also need to assimilate data describing annual biomass maps, and landscape disturbances such as fires and moth outbreaks at the pan-Arctic scale. From a modelling perspective, w~~Thus, m~~e consider that m~~ore field observations are crucial across the pan-Arctic, specifically on decomposition and TT ~~from plant and soil~~ ~~C stocks~~ ~~decomposition~~of SOC ~~(C stocks turnover rates)~~(He et al., 2016) and respiratory processes such as ~~(~~partitioning of R_{eco} into R_a and R_h~~)~~ (Hobbie et al., 2000; McGuire et al., 2000), ~~not only~~ across the growing season~~, but~~ and also during wintertime (Commane et al., 2017; Zona et al., 2016). ~~These data could be upscaled using machine learning, following the approaches used for creating SOM maps, with uncertainty attribution, as further assimilation data sets for frameworks like CARDAMOM. An improved data-model integration will move us towards enhanced analytical robustness and a decrease of model uncertainties.~~

Our approach has ~~ascribed~~used estimated observation error, ~~to observations only, and we have ignored~~and inflated this to include unknown errors associated with model process representation. We currently lack any better knowledge of the combined uncertainties arising from model representation errors and observation errors~~, stochastic events, and unaccounted for heterogeneity, for example~~. We acknowledge ~~both~~ that ~~any~~all models ~~is~~ are an imperfect representation of C dynamics, which generates irreconcilable model-data errors due to the inherent assumptions in model structure~~and the need to quantify model process error~~. Future analyses should investigate model structural error, using for example error-explicit Bayesian approaches (Xu et al., 2017), or comparing the likelihoods of alternate model structures, of varying complexity. ~~In our case, u~~Using multiple sources of data, we have highlighted systematic errors in the model at landscape scale (Figure 2 and 3) for LAI, GPP and biomass~~, which may be related to the prior from the photosynthesis we used~~. However, these biases are not consistent for site-scale evaluations. Thus, ~~There is, therefore, irreconcilable model-data errors due to the inherent assumptions in model structure, but attempts have been made pointing to potential methods for optimizing uncertainty choices (Caldararu et al., 2012) and error models in model-data fusion systems (Schoups & Vrugt, 2010). We certainly need to consider these options as feasible ways to account for model structural errors in future implementations of CARDAMOM. a~~A next step ~~for our analysis~~ ~~is to~~would be to include explicitly both random and systematic ~~photosynthesis~~ process errors for C fluxes in the data assimilation. These errors could be determined from field scale evaluation of model process representation (Table 2) using e.g. FLUXNET2015 data. ~~However, w~~We also need to ~~first~~ understand better the error associated with landscape heterogeneity of C stocks and fluxes, to upscale from flux tower observations, or direct measurements of LAI, to landscape pixel. This could be achieved by constructing robust observation error models (Dietze, 2017) from field to pixel scale, for e.g. GPP, LAI and foliar N. Evaluation of the sensitivity of C cycling DA analyses to observation error has shown relatively low sensitivity to data gaps and random error on net ecosystem flux data (Hill et al., 2012), but further analyses of error sensitivity are required for multiple streams of stock data.~~We need to explain the contrasting photosynthesis biases at landscape (Figure. 3) and flux site scales (Figure. 4) in order to understand and scale the C flux process error.~~

5 Conclusions

The Arctic is experiencing rapid environmental changes, which will influence the global C cycle. Using a data-assimilation framework we have evaluated the current state of key C flux, stocks and transit times for the pan-Arctic region, 2000-15. We found that the pan-Arctic was a likely sink of C, weaker in tundra and stronger in taiga, but uncertainties around the respiration losses are still large, and so the region could be a source of C. Comparisons with global and local scale datasets demonstrate the capabilities of CARDAMOM for analysing the C cycle in the Arctic domain. CARDAMOM is a data-constrained and data-integrated analysis, evaluated for internal consistency, and is therefore a good candidate to benchmark performance of global vegetation/ecosystem models. We conclude that a GVM bias found in transit time of vegetation C is the result of a joint combination of uncertainties from productivity processes and biomass in GVMs, and thus these are a major component of error in their forecasts. While spatial patterns in GVM predictions of NPP are reasonable, particularly in taiga, they have significant biases against the CARDAMOM benchmark. Improved mapping of vegetation and soil C stocks and change over time is required for better analytical constraint. Moreover, future work is required on assimilating data on soil hydrology, permafrost and snow dynamics to improve accuracy and decrease uncertainties on belowground processes. This work establishes the baseline for further process-based ecological analyses using the CARDAMOM data-assimilation system as a technique to constrain the pan-Arctic C cycle.

Data and software availability

CARDAMOM output used in this study is available from Exbrayat and Williams (2018) from the University of Edinburgh's DataShare service at ~~http://dx.doi.org/10.7488/ds/2334. The DALEC2~~https://doi.org/10.7488/ds/2334. The DALEC2 ~~-~~code ~~-~~is also available on Edinburgh Data~~s~~Share at ~~(~~https://doi.org/10.7488/ds/2504~~datashare.is.ed.ac.uk)~~. Contact MW for access to the CARDAMOM software.

Acknowledgements

This work was supported in part by a scholarship from the Aarhus-Edinburgh Excellence in European Doctoral Education Project and by the eSTICC (eScience tools for investigating Climate Change in Northern High Latitudes) project, part of the Nordic Center of Excellence. This work was also supported by the Natural Environment Research Council (NERC) through the National Center for Earth Observation. Data-assimilation procedures were performed using the Edinburgh Compute and Data Facility resources. This work used eddy covariance data acquired and shared by the FLUXNET community, including these networks: AmeriFlux, AfriFlux, AsiaFlux, CarboAfrica, CarboEuropeIP, CarboItaly, CarboMont, ChinaFlux, Fluxnet-Canada, GreenGrass, ICOS, KoFlux, LBA, NECC, OzFlux-TERN, TCOS-Siberia, and USCCC. The ERA-Interim reanalysis data are provided by ECMWF and processed by LSCE. The FLUXNET eddy covariance data processing and harmonization was carried out by the European Fluxes Database Cluster, AmeriFlux Management Project, and Fluxdata project of FLUXNET, with the support of CDIAC and ICOS Ecosystem Thematic Center, and the OzFlux, ChinaFlux and AsiaFlux offices. We thank Nuno Carvalhais for discussion that helped to focus our ideas. For their roles in producing, and making available the ISI-MIP model output, we acknowledge the modelling groups and the ISI-MIP coordination team.

References~~`~~

Ahlström, A., Schurgers, G., Arneth, A., and Smith, B.: Robustness and uncertainty in terrestrial ecosystem carbon response to CMIP5 climate change projections, Environmental Research Letters, 7, 044008, 2012.

Akihiko, I., Kazuya, N., Christopher, P. O. R., Louis, F., Alexandra-Jane, H., Guy, M., Ingrid, J., Hanqin, T., Jia, Y., Shufen, P., Catherine, M., Richard, B., Thomas, H., Jörg, S., Sebastian, O., Sibyll, S., Philippe, C., Jinfeng, C., Rashid, R., Ning, Z., and Fang, Z.: Photosynthetic productivity and its efficiencies in ISIMIP2a biome models: benchmarking for impact assessment studies, Environmental Research Letters, 12, 085001, 2017.

AMAP: Snow, water, ice and permafrost in the Arctic (SWIPA) 2017. Arctic Monitoring and Assessment Programme (AMAP). Oslo, Norway. xiv + 269 pp., xiv + 269 pp, 2017.

Anav, A., Murray-Tortarolo, G., Friedlingstein, P., Sitch, S., Piao, S., and Zhu, Z.: Evaluation of Land Surface Models in Reproducing Satellite Derived Leaf Area Index over the High-Latitude Northern Hemisphere. Part II: Earth System Models, Remote Sensing, 5, 3637, 2013.

Arulampalam, M. S., Maskell, S., Gordon, N., and Clapp, T.: A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking, IEEE Transactions on Signal Processing, 50, 174-188, 10.1109/78.978374, 2002.

Baldocchi, D. D.: Assessing the eddy covariance technique for evaluating carbon dioxide exchange rates of ecosystems: past, present and future, Global Change Biology, 9, 479-492, 10.1046/j.1365-2486.2003.00629.x, 2003.

Beer, C., Reichstein, M., Tomelleri, E., Ciais, P., Jung, M., Carvalhais, N., Rödenbeck, C., Arain, M. A., Baldocchi, D., Bonan, G. B., Bondeau, A., Cescatti, A., Lasslop, G., Lindroth, A., Lomas, M., Luyssaert, S., Margolis, H., Oleson, K. W., Roupsard, O., Veenendaal, E., Viovy, N., Williams, C., Woodward, F. I., and Papale, D.: Terrestrial Gross Carbon Dioxide Uptake: Global Distribution and Covariation with Climate, Science, 10.1126/science.1184984, 2010.

Belelli Marchesini, L., Papale, D., Reichstein, M., Vuichard, N., Tchebakova, N., and Valentini, R.: Carbon balance assessment of a natural steppe of southern Siberia by multiple constraint approach, Biogeosciences, 4, 581-595, 10.5194/bg-4-581-2007, 2007.

Bintanja, R., and Andry, O.: Towards a rain-dominated Arctic, Nature Climate Change, 7, 263, 10.1038/nclimate3240, 2017.

Bloom, A. A., and Williams, M.: Constraining ecosystem carbon dynamics in a data-limited world: integrating ecological "common sense" in a model–data fusion framework, Biogeosciences, 12, 1299-1315, 10.5194/bg-12-1299-2015, 2015.

Bloom, A. A., Exbrayat, J.-F., van der Velde, I. R., Feng, L., and Williams, M.: 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, 1285-1290, 10.1073/pnas.1515160113, 2016.

Bond-Lamberty, B., Wang, C., and Gower, S. T.: Net primary production and net ecosystem production of a boreal black spruce wildfire chronosequence, Global Change Biology, 10, 473-487, doi:10.1111/j.1529-8817.2003.0742.x, 2004.

Bontemps, S., Defourny, P., Bogaert, E., Arino, O., Kalogirou, V., and Perez, J.: GLOBCOVER 2009 - Products description and validation report, 2011.

Brown, J., Ferrians Jr, O. J., Heginbottom, J. A., and Melnikov, E. S.: Circum-Arctic map of permafrost and ground-ice conditions, Report 45, 1997.

Canadell, J. G., Le Quéré, C., Raupach, M. R., Field, C. B., Buitenhuis, E. T., Ciais, P., Conway, T. J., Gillett, N. P., Houghton, R. A., and Marland, G.: Contributions to accelerating atmospheric CO2 growth from economic activity, carbon intensity, and efficiency of natural sinks, Proceedings of the National Academy of Sciences, 104, 18866-18870, 10.1073/pnas.0702737104, 2007.

Carvalhais, N., Forkel, M., Khomik, M., Bellarby, J., Jung, M., Migliavacca, M., u, M., Saatchi, S., Santoro, M., Thurner, M., Weber, U., Ahrens, B., Beer, C., Cescatti, A., Randerson, J. T., and Reichstein, M.: Global covariation of carbon turnover times with climate in terrestrial ecosystems, Nature, 514, 213-217, 10.1038/nature13731, 2014.

Commane, R., Lindaas, J., Benmergui, J., Luus, K. A., Chang, R. Y.-W., Daube, B. C., Euskirchen, E. S., Henderson, J. M., Karion, A., Miller, J. B., Miller, S. M., Parazoo, N. C., Randerson, J. T., Sweeney, C., Tans, P., Thoning, K., Veraverbeke, S., Miller, C. E., and Wofsy, S. C.: Carbon dioxide sources from Alaska driven by increasing early winter respiration from Arctic tundra, Proceedings of the National Academy of Sciences, 114, 5361-5366, 10.1073/pnas.1618567114, 2017.

Cox, P. M.: Description of the “TRIFFID” Dynamic Global Vegetation Model. Hadley Centre technical note 24, Met Office, UK, 2001.

De Kauwe, M. G., Disney, M. I., Quaife, T., Lewis, P., and Williams, M.: An assessment of the MODIS collection 5 leaf area index product for a region of mixed coniferous forest, Remote Sensing of Environment, 115, 767-780, https://doi.org/10.1016/j.rse.2010.11.004, 2011.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J. J., Park, B. K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J. N., and Vitart, F.: The ERA-Interim reanalysis: configuration and performance of the data assimilation system, Quarterly Journal of the Royal Meteorological Society, 137, 553-597, 10.1002/qj.828, 2011.

DeLucia, E. H., Drake, J. E., Thomas, R. B., and Gonzalez-Meler, M.: Forest carbon use efficiency: is respiration a constant fraction of gross primary production?, Global Change Biology, 13, 1157-1167, doi:10.1111/j.1365-2486.2007.01365.x, 2007.

Dietze, M. C.: Ecological Forecasting, Princeton University Press, 2017.

Exbrayat, J.-F., Pitman, A. J., and Abramowitz, G.: Response of microbial decomposition to spin-up explains CMIP5 soil carbon range until 2100, Geosci. Model Dev., 7, 2683-2692, 10.5194/gmd-7-2683-2014, 2014.

Exbrayat, J. F., Bloom, A. A., Falloon, P., Ito, A., Smallman, T. L., and Williams, M.: Reliability ensemble averaging of 21st century projections of terrestrial net primary productivity reduces global and regional uncertainties, Earth Syst. Dynam., 9, 153-165, 10.5194/esd-9-153-2018, 2018.

Exbrayat, J. F., and Williams, M.: CARDAMOM panarctic retrievals 2000-2015. 2000-2015 [Dataset], National Centre for Earth Observation and School of GeoSciences. University of Edinburgh, http://.doi.org/10.7488/ds/2334, 2018.

FAO/IIASA/ISRIC/ISSCAS/JRC: Harmonized World Soil Database (version 1.21). FAO, Rome, Italy and IIASA, Laxenburg, Austria, 2012.

Fer, I., Kelly, R., Moorcroft, P. R., Richardson, A. D., Cowdery, E. M., and Dietze, M. C.: Linking big models to big data: efficient ecosystem model calibration through Bayesian model emulation, Biogeosciences, 15, 5801-5830, 10.5194/bg-15-5801-2018, 2018.

Fisher, J. B., Sikka, M., Oechel, W. C., Huntzinger, D. N., Melton, J. R., Koven, C. D., Ahlström, A., Arain, M. A., Baker, I., Chen, J. M., Ciais, P., Davidson, C., Dietze, M., El-Masri, B., Hayes, D., Huntingford, C., Jain, A. K., Levy, P. E., Lomas, M. R., Poulter, B., Price, D., Sahoo, A. K., Schaefer, K., Tian, H., Tomelleri, E., Verbeeck, H., Viovy, N., Wania, R., Zeng, N., and Miller, C. E.: Carbon cycle uncertainty in the Alaskan Arctic, Biogeosciences, 11, 4271-4288, 10.5194/bg-11-4271-2014, 2014.

Forkel, M., Carvalhais, N., Rödenbeck, C., Keeling, R., Heimann, M., Thonicke, K., Zaehle, S., and Reichstein, M.: Enhanced seasonal CO2 exchange caused by amplified plant productivity in northern ecosystems, Science, 351, 696-699, 10.1126/science.aac4971, 2016.

Fox, A., Williams, M., Richardson, A. D., Cameron, D., Gove, J. H., Quaife, T., Ricciuto, D., Reichstein, M., Tomelleri, E., Trudinger, C. M., and Van Wijk, M. T.: 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, 1597-1615, https://doi.org/10.1016/j.agrformet.2009.05.002, 2009.

Friedlingstein, P., Meinshausen, M., Arora, V. K., Jones, C. D., Anav, A., Liddicoat, S. K., and Knutti, R.: Uncertainties in CMIP5 Climate Projections due to Carbon Cycle Feedbacks, Journal of Climate, 27, 511-526, 10.1175/jcli-d-12-00579.1, 2014.

Friend, A. D., Lucht, W., Rademacher, T. T., Keribin, R., Betts, R., Cadule, P., Ciais, P., Clark, D. B., Dankers, R., Falloon, P. D., Ito, A., Kahana, R., Kleidon, A., Lomas, M. R., Nishina, K., Ostberg, S., Pavlick, R., Peylin, P., Schaphoff, S., Vuichard, N., Warszawski, L., Wiltshire, A., and Woodward, F. I.: Carbon residence time dominates uncertainty in terrestrial vegetation responses to future climate and atmospheric CO2, Proceedings of the National Academy of Sciences, 111, 3280-3285, 10.1073/pnas.1222477110, 2014.

Goetz, S. J., Bunn, A. G., Fiske, G. J., and Houghton, R. A.: Satellite-observed photosynthetic trends across boreal North America associated with climate and fire disturbance, Proceedings of the National Academy of Sciences of the United States of America, 102, 13521-13525, 10.1073/pnas.0506179102, 2005.

Goulden, M. L., Munger, J. W., Fan, S.-M., Daube, B. C., and Wofsy, S. C.: Exchange of Carbon Dioxide by a Deciduous Forest: Response to Interannual Climate Variability, Science, 271, 1576-1578, 10.1126/science.271.5255.1576, 1996.

Goulden, M. L., McMillan, A. M. S., Winston, G. C., Rocha, A. V., Manies, K. L., Harden, J. W., and Bond-Lamberty, B. P.: Patterns of NPP, GPP, respiration, and NEP during boreal forest succession, Global Change Biology, 17, 855-871, 10.1111/j.1365-2486.2010.02274.x, 2011.

Graven, H. D., Keeling, R. F., Piper, S. C., Patra, P. K., Stephens, B. B., Wofsy, S. C., Welp, L. R., Sweeney, C., Tans, P. P., Kelley, J. J., Daube, B. C., Kort, E. A., Santoni, G. W., and Bent, J. D.: Enhanced Seasonal Exchange of CO2 by Northern Ecosystems Since 1960, Science, 341, 1085-1089, 10.1126/science.1239207, 2013.

Guimberteau, M., Zhu, D., Maignan, F., Huang, Y., Yue, C., Dantec-Nédélec, S., Ottlé, C., Jornet-Puig, A., Bastos, A., Laurent, P., Goll, D., Bowring, S., Chang, J., Guenet, B., Tifafi, M., Peng, S., Krinner, G., Ducharne, A., Wang, F., Wang, T., Wang, X., Wang, Y., Yin, Z., Lauerwald, R., Joetzjer, E., Qiu, C., Kim, H., and Ciais, P.: ORCHIDEE-MICT (v8.4.1), a land surface model for the high latitudes: model description and validation, Geosci. Model Dev., 11, 121-163, https://doi.org/10.5194/gmd-11-121-2018, 2018.

Hashimoto, S., Carvalhais, N., Ito, A., Migliavacca, M., Nishina, K., and Reichstein, M.: Global spatiotemporal distribution of soil respiration modeled using a global database, Biogeosciences, 12, 4121-4132, 10.5194/bg-12-4121-2015, 2015.

He, Y., Trumbore, S. E., Torn, M. S., Harden, J. W., Vaughn, L. J. S., Allison, S. D., and Randerson, J. T.: Radiocarbon constraints imply reduced carbon uptake by soils during the 21st century, Science, 353, 1419-1424, 10.1126/science.aad4273, 2016.

Hill, T. C., Ryan, E., and Williams, M.: The use of CO2 flux time series for parameter and carbon stock estimation in carbon cycle research, 18, 179-193, doi:10.1111/j.1365-2486.2011.02511.x, 2012.

Hobbie, J. E., and Kling, G. W.: Alaska's changing Arctic: Ecological consequences for tundra, streams, and lakes, Oxford University Press, 2014.

Hobbie, S. E., Schimel, J. P., Trumbore, S. E., and Randerson, J. R.: Controls over carbon storage and turnover in high-latitude soils, Global Change Biology, 6, 196-210, 10.1046/j.1365-2486.2000.06021.x, 2000.

Hugelius, G., Bockheim, J. G., Camill, P., Elberling, B., Grosse, G., Harden, J. W., Johnson, K., Jorgenson, T., Koven, C. D., Kuhry, P., Michaelson, G., Mishra, U., Palmtag, J., Ping, C. L., O'Donnell, J., Schirrmeister, L., Schuur, E. A. G., Sheng, Y., Smith, L. C., Strauss, J., and Yu, Z.: A new data set for estimating organic carbon storage to 3 m depth in soils of the northern circumpolar permafrost region, Earth Syst. Sci. Data, 5, 393-402, 10.5194/essd-5-393-2013, 2013a.

Hugelius, G., Tarnocai, C., Broll, G., Canadell, J. G., Kuhry, P., and Swanson, D. K.: The Northern Circumpolar Soil Carbon Database: spatially distributed datasets of soil coverage and soil carbon storage in the northern permafrost regions, Earth Syst. Sci. Data, 5, 3-13, 10.5194/essd-5-3-2013, 2013b.

Hugelius, G., Strauss, J., Zubrzycki, S., Harden, J. W., Schuur, E. A. G., Ping, C. L., Schirrmeister, L., Grosse, G., Michaelson, G. J., Koven, C. D., O'Donnell, J. A., Elberling, B., Mishra, U., Camill, P., Yu, Z., Palmtag, J., and Kuhry, P.: Estimated stocks of circumpolar permafrost carbon with quantified uncertainty ranges and identified data gaps, Biogeosciences, 11, 6573-6593, 10.5194/bg-11-6573-2014, 2014.

Ikawa, H., Nakai, T., Busey, R. C., Kim, Y., Kobayashi, H., Nagai, S., Ueyama, M., Saito, K., Nagano, H., Suzuki, R., and Hinzman, L.: Understory CO2, sensible heat, and latent heat fluxes in a black spruce forest in interior Alaska, Agricultural and Forest Meteorology, 214-215, 80-90, https://doi.org/10.1016/j.agrformet.2015.08.247, 2015.

IPCC: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2013.

Ito, A., and Inatomi, M.: Water-Use Efficiency of the Terrestrial Biosphere: A Model Analysis Focusing on Interactions between the Global Carbon and Water Cycles, Journal of Hydrometeorology, 13, 681-694, 10.1175/jhm-d-10-05034.1, 2012.

Jackson, R. B., Lajtha, K., Crow, S. E., Hugelius, G., Kramer, M. G., and Piñeiro, G.: The Ecology of Soil Carbon: Pools, Vulnerabilities, and Biotic and Abiotic Controls, Annual Review of Ecology, Evolution, and Systematics, 48, 419-445, 10.1146/annurev-ecolsys-112414-054234, 2017.

Jung, M., Reichstein, M., Margolis, H. A., Cescatti, A., Richardson, A. D., Arain, M. A., Arneth, A., Bernhofer, C., Bonal, D., Chen, J., Gianelle, D., Gobron, N., Kiely, G., Kutsch, W., Lasslop, G., Law, B. E., Lindroth, A., Merbold, L., Montagnani, L., Moors, E. J., Papale, D., Sottocornola, M., Vaccari, F., and Williams, C.: 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, 116, doi:10.1029/2010JG001566, 2011.

Jung, M., Reichstein, M., Schwalm, C. R., Huntingford, C., Sitch, S., Ahlström, A., Arneth, A., Camps-Valls, G., Ciais, P., Friedlingstein, P., Gans, F., Ichii, K., Jain, A. K., Kato, E., Papale, D., Poulter, B., Raduly, B., Rödenbeck, C., Tramontana, G., Viovy, N., Wang, Y.-P., Weber, U., Zaehle, S., and Zeng, N.: Compensatory water effects link yearly global land CO2 sink changes to temperature, Nature, 541, 516, 10.1038/nature20780, 2017.

Kimball, J. S., Jones, L. A., Zhang, K., Heinsch, F. A., McDonald, K. C., and Oechel, W.: A Satellite Approach to Estimate Land CO2 Exchange for Boreal and Arctic Biomes Using MODIS and AMSR-E, IEEE Transactions on Geoscience and Remote Sensing, 47, 569-587, 10.1109/TGRS.2008.2003248, 2009.

Koven, C. D., Ringeval, B., Friedlingstein, P., Ciais, P., Cadule, P., Khvorostyanov, D., Krinner, G., and Tarnocai, C.: Permafrost carbon-climate feedbacks accelerate global warming, Proceedings of the National Academy of Sciences, 108, 14769-14774, 10.1073/pnas.1103910108, 2011.

Koven, C. D., Schuur, E. A. G., Schädel, C., Bohn, T. J., Burke, E. J., Chen, G., Chen, X., Ciais, P., Grosse, G., Harden, J. W., Hayes, D. J., Hugelius, G., Jafarov, E. E., Krinner, G., Kuhry, P., Lawrence, D. M., MacDougall, A. H., Marchenko, S. S., McGuire, A. D., Natali, S. M., Nicolsky, D. J., Olefeldt, D., Peng, S., Romanovsky, V. E., Schaefer, K. M., Strauss, J., Treat, C. C., and Turetsky, M.: A simplified, data-constrained approach to estimate the permafrost carbon–climate feedback, Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 373, 10.1098/rsta.2014.0423, 2015.

Koven, C. D., Hugelius, G., Lawrence, D. M., and Wieder, W. R.: Higher climatological temperature sensitivity of soil carbon in cold than warm climates, Nature Climate Change, 7, 817, 10.1038/nclimate3421, 2017.

Kutzbach, L., Wille, C., and Pfeiffer, E.-M.: The exchange of carbon dioxide between wet arctic tundra and the atmosphere at the Lena River Delta, Northern Siberia, Biogeosciences, 4, 869-890, 10.5194/bg-4-869-2007, 2007.

Lafleur, P. M., Humphreys, E. R., St. Louis, V. L., Myklebust, M. C., Papakyriakou, T., Poissant, L., Barker, J. D., Pilote, M., and Swystun, K. A.: Variation in Peak Growing Season Net Ecosystem Production Across the Canadian Arctic, Environmental Science & Technology, 46, 7971-7977, 10.1021/es300500m, 2012.

Lasslop, G., Reichstein, M., Papale, D., Richardson, A. D., Arneth, A., Barr, A., Stoy, P., and Wohlfahrt, G.: Separation of net ecosystem exchange into assimilation and respiration using a light response curve approach: critical issues and global evaluation, Global Change Biology, 16, 187-208, 10.1111/j.1365-2486.2009.02041.x, 2010.

Lawrence, D. M., Oleson, K. W., Flanner, M. G., Thornton, P. E., Swenson, S. C., Lawrence, P. J., Zeng, X., Yang, Z. L., Levis, S., Sakaguchi, K., Bonan, G. B., and Slater, A. G.: Parameterization improvements and functional and structural advances in Version 4 of the Community Land Model, Journal of Advances in Modeling Earth Systems, 3, doi:10.1029/2011MS00045, 2011.

Le Quéré, C., Andrew, R. M., Friedlingstein, P., Sitch, S., Pongratz, J., Manning, A. C., Korsbakken, J. I., Peters, G. P., Canadell, J. G., Jackson, R. B., Boden, T. A., Tans, P. P., Andrews, O. D., Arora, V. K., Bakker, D. C. E., Barbero, L., Becker, M., Betts, R. A., Bopp, L., Chevallier, F., Chini, L. P., Ciais, P., Cosca, C. E., Cross, J., Currie, K., Gasser, T., Harris, I., Hauck, J., Haverd, V., Houghton, R. A., Hunt, C. W., Hurtt, G., Ilyina, T., Jain, A. K., Kato, E., Kautz, M., Keeling, R. F., Klein Goldewijk, K., Körtzinger, A., Landschützer, P., Lefèvre, N., Lenton, A., Lienert, S., Lima, I., Lombardozzi, D., Metzl, N., Millero, F., Monteiro, P. M. S., Munro, D. R., Nabel, J. E. M. S., Nakaoka, S.-I., Nojiri, Y., Padin, X. A., Peregon, A., Pfeil, B., Pierrot, D., Poulter, B., Rehder, G., Reimer, J., Rödenbeck, C., Schwinger, J., Séférian, R., Skjelvan, I., Stocker, B. D., Tian, H., Tilbrook, B., Tubiello, F. N., van der Laan-Luijkx, I. T., van der Werf, G. R., van Heuven, S., Viovy, N., Vuichard, N., Walker, A. P., Watson, A. J., Wiltshire, A. J., Zaehle, S., and Zhu, D.: Global Carbon Budget 2017, Earth Syst. Sci. Data, 10, 405-448, 10.5194/essd-10-405-2018, 2018.

Levy, P. E., Friend, A. D., White, A., and Cannell, M. G. R.: ‘The Influence of Land Use Change On Global-Scale Fluxes of Carbon from Terrestrial Ecosystems’, Climatic Change, 67, 185-209, 10.1007/s10584-004-2849-z, 2004.

López-Blanco, E., Lund, M., Williams, M., Tamstorf, M. P., Westergaard-Nielsen, A., Exbrayat, J. F., Hansen, B. U., and Christensen, T. R.: Exchange of CO2 in Arctic tundra: impacts of meteorological variations and biological disturbance, Biogeosciences, 14, 4467-4483, 10.5194/bg-14-4467-2017, 2017.

López-Blanco, E., Lund, M., Christensen, T. R., Tamstorf, M. P., Smallman, T. L., Slevin, D., Westergaard-Nielsen, A., Hansen, B. U., Abermann, J., and Williams, M.: Plant Traits are Key Determinants in Buffering the Meteorological Sensitivity of Net Carbon Exchanges of Arctic Tundra, Journal of Geophysical Research: Biogeosciences, 123, https://doi.org/10.1029/2018JG004386, 2018.

Lucht, W., Prentice, I. C., Myneni, R. B., Sitch, S., Friedlingstein, P., Cramer, W., Bousquet, P., Buermann, W., and Smith, B.: Climatic Control of the High-Latitude Vegetation Greening Trend and Pinatubo Effect, Science, 296, 1687-1689, 10.1126/science.1071828, 2002.

Lund, M., Falk, J. M., Friborg, T., Mbufong, H. N., Sigsgaard, C., Soegaard, H., and Tamstorf, M. P.: Trends in CO2 exchange in a high Arctic tundra heath, 2000-2010, Journal of Geophysical Research: Biogeosciences, 117, 2012.

Lund, M., Raundrup, K., Westergaard-Nielsen, A., López-Blanco, E., Nymand, J., and Aastrup, P.: Larval outbreaks in West Greenland: Instant and subsequent effects on tundra ecosystem productivity and CO(2) exchange, AMBIO, 46, 26-38, 10.1007/s13280-016-0863-9, 2017.

Luo, Y., Weng, E., Wu, X., Gao, C., Zhou, X., and Zhang, L.: Parameter identifiability, constraint, and equifinality in data assimilation with ecosystem models, Ecological Applications, 19, 571-574, doi:10.1890/08-0561.1, 2009.

Mack, M. C., Bret-Harte, M. S., Hollingsworth, T. N., Jandt, R. R., Schuur, E. A. G., Shaver, G. R., and Verbyla, D. L.: Carbon loss from an unprecedented Arctic tundra wildfire, Nature, 475, 489, 10.1038/nature10283, 2011.

Manzoni, S., Čapek, P., Porada, P., Thurner, M., Winterdahl, M., Beer, C., Brüchert, V., Frouz, J., Herrmann, A. M., Lindahl, B. D., Lyon, S. W., Šantrůčková, H., Vico, G., and Way, D.: Reviews and syntheses: Carbon use efficiency from organisms to ecosystems – definitions, theories, and empirical evidence, Biogeosciences, 15, 5929-5949, 10.5194/bg-15-5929-2018, 2018.

Mastepanov, M., Sigsgaard, C., Dlugokencky, E. J., Houweling, S., Strom, L., Tamstorf, M. P., and Christensen, T. R.: Large tundra methane burst during onset of freezing, Nature, 456, 628-630, 10.1038/nature07464, 2008.

McGuire, A. D., Melillo, J. M., Randerson, J. T., Parton, W. J., Heimann, M., Meier, R. A., Clein, J. S., Kicklighter, D. W., and Sauf, W.: Modeling the effects of snowpack on heterotrophic respiration across northern temperate and high latitude regions: Comparison with measurements of atmospheric carbon dioxide in high latitudes, Biogeochemistry, 48, 91-114, 10.1023/a:1006286804351, 2000.

McGuire, A. D., Hayes, D., Kicklighter, D. W., Manizza, M., Zhuang, Q., Chen, M., Follows, M. J., Gurney, K. R., Mcclelland, J. W., Melillo, J. M., Peterson, B. J., and Prinn, R. G.: An analysis of the carbon balance of the Arctic Basin from 1997 to 2006, Tellus B, 62, 455-474, doi:10.1111/j.1600-0889.2010.00497.x, 2010.

McGuire, A. D., Christensen, T. R., Hayes, D., Heroult, A., Euskirchen, E., Kimball, J. S., Koven, C., Lafleur, P., Miller, P. A., Oechel, W., Peylin, P., Williams, M., and Yi, Y.: An assessment of the carbon balance of Arctic tundra: comparisons among observations, process models, and atmospheric inversions, Biogeosciences, 9, 3185-3204, 10.5194/bg-9-3185-2012, 2012.

Murray-Tortarolo, G., Anav, A., Friedlingstein, P., Sitch, S., Piao, S., Zhu, Z., Poulter, B., Zaehle, S., Ahlström, A., Lomas, M., Levis, S., Viovy, N., and Zeng, N.: Evaluation of Land Surface Models in Reproducing Satellite-Derived LAI over the High-Latitude Northern Hemisphere. Part I: Uncoupled DGVMs, Remote Sensing, 5, 4819, 2013.

Myers-Smith, I. H., Forbes, B. C., Wilmking, M., Hallinger, M., Lantz, T., Blok, D., Tape, K. D., Macias-Fauria, M., Sass-Klaassen, U., Lévesque, E., Boudreau, S., Ropars, P., Hermanutz, L., Trant, A., Siegwart, C., L., Weijers, S., Rozema, J., Rayback, S. A., Schmidt, N. M., Schaepman-Strub, G., Wipf, S., Rixen, C., Ménard, C. B., Venn, S., Goetz, S., Andreu-Hayles, L., Elmendorf, S., Ravolainen, V., Welker, J., Grogan, P., Epstein, H. E., and Hik, D. S.: Shrub expansion in tundra ecosystems: dynamics, impacts and research priorities, Environmental Research Letters, 6, 045509, 10.1088/1748-9326/6/4/045509, 2011.

Myneni, R. B., Keeling, C. D., Tucker, C. J., Asrar, G., and Nemani, R. R.: Increased plant growth in the northern high latitudes from 1981 to 1991, Nature, 386, 698, 10.1038/386698a0, 1997.

Myneni, R. B., Hoffman, S., Knyazikhin, Y., Privette, J. L., Glassy, J., Tian, Y., Wang, Y., Song, X., Zhang, Y., Smith, G. R., Lotsch, A., Friedl, M., Morisette, J. T., Votava, P., Nemani, R. R., and Running, S. W.: Global products of vegetation leaf area and fraction absorbed PAR from year one of MODIS data, Remote Sensing of Environment, 83, 214-231, https://doi.org/10.1016/S0034-4257(02)00074-3, 2002.

Nishina, K., Ito, A., Beerling, D. J., Cadule, P., Ciais, P., Clark, D. B., Falloon, P., Friend, A. D., Kahana, R., Kato, E., Keribin, R., Lucht, W., Lomas, M., Rademacher, T. T., Pavlick, R., Schaphoff, S., Vuichard, N., Warszawaski, L., and Yokohata, T.: Quantifying uncertainties in soil carbon responses to changes in global mean temperature and precipitation, Earth Syst. Dynam., 5, 197-209, 10.5194/esd-5-197-2014, 2014.

Nishina, K., Ito, A., Falloon, P., Friend, A. D., Beerling, D. J., Ciais, P., Clark, D. B., Kahana, R., Kato, E., Lucht, W., Lomas, M., Pavlick, R., Schaphoff, S., Warszawaski, L., and Yokohata, T.: Decomposing uncertainties in the future terrestrial carbon budget associated with emission scenarios, climate projections, and ecosystem simulations using the ISI-MIP results, Earth Syst. Dynam., 6, 435-445, 10.5194/esd-6-435-2015, 2015.

Papale, D., Reichstein, M., Aubinet, M., Canfora, E., Bernhofer, C., Kutsch, W., Longdoz, B., Rambal, S., Valentini, R., Vesala, T., and Yakir, D.: Towards a standardized processing of Net Ecosystem Exchange measured with eddy covariance technique: algorithms and uncertainty estimation, Biogeosciences, 3, 571-583, 10.5194/bg-3-571-2006, 2006.

Parazoo, N. C., Koven, C. D., Lawrence, D. M., Romanovsky, V., and Miller, C. E.: Detecting the permafrost carbon feedback: talik formation and increased cold-season respiration as precursors to sink-to-source transitions, The Cryosphere, 12, 123-144, 10.5194/tc-12-123-2018, 2018.

Peñuelas, J., Rutishauser, T., and Filella, I.: Phenology Feedbacks on Climate Change, Science, 324, 887-888, 10.1126/science.1173004, 2009.

Reichstein, M., Bahn, M., Mahecha, M. D., Kattge, J., and Baldocchi, D. D.: Linking plant and ecosystem functional biogeography, Proceedings of the National Academy of Sciences, 111, 13697-13702, 10.1073/pnas.1216065111, 2014.

Running, S. W., Nemani, R. R., Heinsch, F. A., Zhao, M., Reeves, M., and Hashimoto, H.: A Continuous Satellite-Derived Measure of Global Terrestrial Primary Production, BioScience, 54, 547-560, 10.1641/0006-3568(2004)054[0547:ACSMOG]2.0.CO;2, 2004.

Sari, J., Tarmo, V., Vladimir, K., Tuomas, L., Maiju, L., Juha, M., Johanna, N., Aleksi, R., Juha-Pekka, T., and Mika, A.: Spatial variation and seasonal dynamics of leaf-area index in the arctic tundra-implications for linking ground observations and satellite images, Environmental Research Letters, 12, 095002, 2017.

Schaphoff, S., Heyder, U., Ostberg, S., Gerten, D., Heinke, J., and Lucht, W.: Contribution of permafrost soils to the global carbon budget, Environmental Research Letters, 8, 014026, 2013.

Schuur, E. A. G., McGuire, A. D., Schadel, C., Grosse, G., Harden, J. W., Hayes, D. J., Hugelius, G., Koven, C. D., Kuhry, P., Lawrence, D. M., Natali, S. M., Olefeldt, D., Romanovsky, V. E., Schaefer, K., Turetsky, M. R., Treat, C. C., and Vonk, J. E.: Climate change and the permafrost carbon feedback, Nature, 520, 171-179, 10.1038/nature14338, 2015.

Sierra, C. A., Müller, M., Metzler, H., Manzoni, S., and Trumbore, S. E.: The muddle of ages, turnover, transit, and residence times in the carbon cycle, Global Change Biology, 23, 1763-1773, 10.1111/gcb.13556, 2017.

Sitch, S., Smith, B., Prentice, I. C., Arneth, A., Bondeau, A., Cramer, W., Kaplan, J. O., Levis, S., Lucht, W., Sykes, M. T., Thonicke, K., and Venevsky, S.: Evaluation of ecosystem dynamics, plant geography and terrestrial carbon cycling in the LPJ dynamic global vegetation model, Global Change Biology, 9, 161-185, doi:10.1046/j.1365-2486.2003.00569.x, 2003.

Sloan, V. L., Fletcher, B. J., Press, M. C., Williams, M., and Phoenix, G. K.: Leaf and fine root carbon stocks and turnover are coupled across Arctic ecosystems, Global Change Biology, 19, 3668-3676, 10.1111/gcb.12322, 2013.

Smallman, T. L., Exbrayat, J.-F., Mencuccini, M., Bloom, A. A., and Williams, M.: Assimilation of repeated woody biomass observations constrains decadal ecosystem carbon cycle uncertainty in aggrading forests, Journal of Geophysical Research: Biogeosciences, 122, 528-545, doi:10.1002/2016JG003520, 2017.

Smith, B., Prentice, I. C., and Sykes, M. T.: Representation of vegetation dynamics in the modelling of terrestrial ecosystems: comparing two contrasting approaches within European climate space, Global Ecology and Biogeography, 10, 621-637, doi:10.1046/j.1466-822X.2001.t01-1-00256.x, 2001.

Smith, B., Wårlind, D., Arneth, A., Hickler, T., Leadley, P., Siltberg, J., and Zaehle, S.: Implications of incorporating N cycling and N limitations on primary production in an individual-based dynamic vegetation model, Biogeosciences, 11, 2027-2054, https://doi.org/10.5194/bg-11-2027-2014, 2014.

Street, L. E., Stoy, P. C., Sommerkorn, M., Fletcher, B. J., Sloan, V. L., Hill, T. C., and Williams, M.: Seasonal bryophyte productivity in the sub-Arctic: a comparison with vascular plants, Functional Ecology, 26, 365-378, 10.1111/j.1365-2435.2011.01954.x, 2012.

Tarnocai, C., Canadell, J. G., Schuur, E. A. G., Kuhry, P., Mazhitova, G., and Zimov, S.: Soil organic carbon pools in the northern circumpolar permafrost region, Global Biogeochemical Cycles, 23, GB2023, 10.1029/2008GB003327, 2009.

Thomas, R. Q., and Williams, M.: A model using marginal efficiency of investment to analyze carbon and nitrogen interactions in terrestrial ecosystems (ACONITE Version 1), Geosci. Model Dev., 7, 2015-2037, 10.5194/gmd-7-2015-2014, 2014.

Thornton, P. E., Doney, S. C., Lindsay, K., Moore, J. K., Mahowald, N., Randerson, J. T., Fung, I., Lamarque, J. F., Feddema, J. J., and Lee, Y. H.: Carbon-nitrogen interactions regulate climate-carbon cycle feedbacks: results from an atmosphere-ocean general circulation model, Biogeosciences, 6, 2099-2120, 10.5194/bg-6-2099-2009, 2009.

Thurner, M., Beer, C., Santoro, M., Carvalhais, N., Wutzler, T., Schepaschenko, D., Shvidenko, A., Kompter, E., Ahrens, B., Levick, S. R., and Schmullius, C.: Carbon stock and density of northern boreal and temperate forests, Global Ecology and Biogeography, 23, 297-310, doi:10.1111/geb.12125, 2014.

Thurner, M., Beer, C., Carvalhais, N., Forkel, M., Santoro, M., Tum, M., and Schmullius, C.: Large-scale variation in boreal and temperate forest carbon turnover rate related to climate, Geophysical Research Letters, 43, 4576-4585, 10.1002/2016GL068794, 2016.

Thurner, M., Beer, C., Ciais, P., Friend, A. D., Ito, A., Kleidon, A., Lomas, M. R., Quegan, S., Rademacher, T. T., Schaphoff, S., Tum, M., Wiltshire, A., and Carvalhais, N.: Evaluation of climate-related carbon turnover processes in global vegetation models for boreal and temperate forests, Global Change Biology, 23, 3076-3091, 10.1111/gcb.13660, 2017.

Tian, H., Chen, G., Lu, C., Xu, X., Hayes, D. J., Ren, W., Pan, S., Huntzinger, D. N., and Wofsy, S. C.: North American terrestrial CO2 uptake largely offset by CH4 and N2O emissions: toward a full accounting of the greenhouse gas budget, Climatic Change, 129, 413-426, 10.1007/s10584-014-1072-9, 2015.

Tum, M., Zeidler, J. N., Günther, K. P., and Esch, T.: Global NPP and straw bioenergy trends for 2000–2014, Biomass and Bioenergy, 90, 230-236, https://doi.org/10.1016/j.biombioe.2016.03.040, 2016.

Walter, B. P., Heimann, M., and Matthews, E.: Modeling modern methane emissions from natural wetlands: 1. Model description and results, 106, 34189-34206, doi:10.1029/2001JD900165, 2001.

Wania, R., Ross, I., and Prentice, I. C.: Integrating peatlands and permafrost into a dynamic global vegetation model: 1. Evaluation and sensitivity of physical land surface processes, Global Biogeochemical Cycles, 23, doi:10.1029/2008GB003412, 2009a.

Wania, R., Ross, I., and Prentice, I. C.: Integrating peatlands and permafrost into a dynamic global vegetation model: 2. Evaluation and sensitivity of vegetation and carbon cycle processes, Global Biogeochemical Cycles, 23, doi:10.1029/2008GB003413, 2009b.

Warszawski, L., Frieler, K., Huber, V., Piontek, F., Serdeczny, O., and Schewe, J.: The Inter-Sectoral Impact Model Intercomparison Project (ISI–MIP): Project framework, Proceedings of the National Academy of Sciences, 111, 3228-3232, 10.1073/pnas.1312330110, 2014.

Williams, M., Rastetter, E. B., Fernandes, D. N., Goulden, M. L., Shaver, G. R., and Johnson, L. C.: Predicting Gross Primary Productivity in Terrestrial Ecosystems, Ecological Applications, 7, 882-894, 10.2307/2269440, 1997.

Williams, M., Schwarz, P. A., Law, B. E., Irvine, J., and Kurpius, M. R.: An improved analysis of forest carbon dynamics using data assimilation, Global Change Biology, 11, 89-105, 10.1111/j.1365-2486.2004.00891.x, 2005.

Woodward, F. I., Smith, T. M., and Emanuel, W. R.: A global land primary productivity and phytogeography model, Global Biogeochemical Cycles, 9, 471-490, doi:10.1029/95GB02432, 1995.

Xenakis, G., and Williams, M.: Comparing microbial and chemical kinetics for modelling soil organic carbon decomposition using the DecoChem v1.0 and DecoBio v1.0 models, Geosci. Model Dev., 7, 1519-1533, 10.5194/gmd-7-1519-2014, 2014.

Xu, T., Valocchi, A. J., Ye, M., and Liang, F.: Quantifying model structural error: Efficient Bayesian calibration of a regional groundwater flow model using surrogates and a data-driven error model, 53, 4084-4105, doi:10.1002/2016WR019831, 2017.

Zaehle, S., and Friend, A. D.: Carbon and nitrogen cycle dynamics in the O-CN land surface model: 1. Model description, site-scale evaluation, and sensitivity to parameter estimates, Global Biogeochemical Cycles, 24, doi:10.1029/2009GB003521, 2010.

Zeng, H., Jia, G., and Epstein, H.: Recent changes in phenology over the northern high latitudes detected from multi-satellite data, Environmental Research Letters, 6, 045508, 10.1088/1748-9326/6/4/045508, 2011.

Zeng, N., Mariotti, A., and Wetzel, P.: Terrestrial mechanisms of interannual CO2 variability, Global Biogeochemical Cycles, 19, doi:10.1029/2004GB002273, 2005.

Zhao, M., Heinsch, F. A., Nemani, R. R., and Running, S. W.: Improvements of the MODIS terrestrial gross and net primary production global data set, Remote Sensing of Environment, 95, 164-176, https://doi.org/10.1016/j.rse.2004.12.011, 2005.

Zhu, Z., Piao, S., Myneni, R. B., Huang, M., Zeng, Z., Canadell, J. G., Ciais, P., Sitch, S., Friedlingstein, P., Arneth, A., Cao, C., Cheng, L., Kato, E., Koven, C., Li, Y., Lian, X., Liu, Y., Liu, R., Mao, J., Pan, Y., Peng, S., Peñuelas, J., Poulter, B., Pugh, T. A. M., Stocker, B. D., Viovy, N., Wang, X., Wang, Y., Xiao, Z., Yang, H., Zaehle, S., and Zeng, N.: Greening of the Earth and its drivers, Nature Climate Change, 6, 791, 10.1038/nclimate3004, 2016.

Zhuang, Q., Melillo, J. M., Sarofim, M. C., Kicklighter, D. W., McGuire, A. D., Felzer, B. S., Sokolov, A., Prinn, R. G., Steudler, P. A., and Hu, S.: CO2 and CH4 exchanges between land ecosystems and the atmosphere in northern high latitudes over the 21st century, 33, doi:10.1029/2006GL026972, 2006.

Zona, D., Gioli, B., Commane, R., Lindaas, J., Wofsy, S. C., Miller, C. E., Dinardo, S. J., Dengel, S., Sweeney, C., Karion, A., Chang, R. Y.-W., Henderson, J. M., Murphy, P. C., Goodrich, J. P., Moreaux, V., Liljedahl, A., Watts, J. D., Kimball, J. S., Lipson, D. A., and Oechel, W. C.: Cold season emissions dominate the Arctic tundra methane budget, Proceedings of the National Academy of Sciences, 113, 40-45, 10.1073/pnas.1516017113, 2016.