Final author response to **"Evaluation of terrestrial pan-Arctic carbon cycling using a data-assimilation system"**

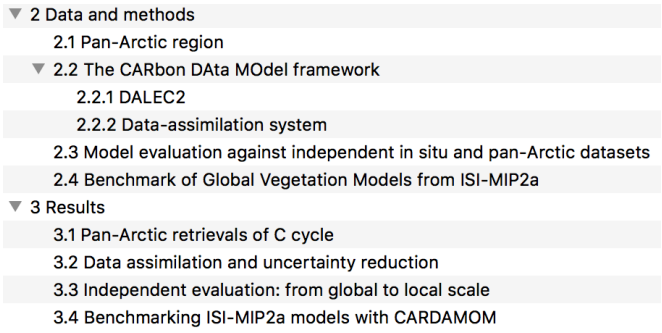
Efrén López-Blanco, Jean-François Exbrayat, Magnus Lund, Torben R. Christensen, Mikkel P. Tamstorf, Darren Slevin, Gustaf Hugelius, Anthony A. Bloom, Mathew Williams

October 24th, 2018

We thank the two reviewers for their ideas and suggestions to improve this paper. We have carefully considered them all and changed our manuscript accordingly. In the following we include a point-by-point response to the reviews, and attached a marked-up manuscript version showing the differences to the initially submitted version. Please note that the line numbers point to the non-marked manuscript.

In general lines we cautiously considered each of the **REF#1** and **REF#2** comments, paying special attention to the tundra-taiga domain split, biases with biomass and GPP, the uncertainty reduction of the model parameters, and the GVM benchmarking exercise. We have included detailed answers to the questions raised, and have performed additional implementations, tests and changes to provide more support on these items. We paid special attention to 6 issues:

- We revised the paper to specifically improve the introduction to facilitate a better understanding of the research context. We highlighted the explanation of transit time, its importance and the difference with other similar terminology used in literature.
- The structure of some sections has been modified to improve comprehension and readability. Now it reads as:



- We improved the method section with supporting equations and supplementary tables to better describe the experimental design and the data-assimilation framework.
- We implemented a new conceptual diagram in Figure 1summarizing the C flux, stocks and transit times numbers described in section 3.1. We also implemented a new Table 2 to illustrate the uncertainty reduction (priors vs posteriors) from all parameters in the model. Similarly, Figure 2, 3, 4, 5, 6, and 7, as well as Table 1, 3 were all implemented as result of the specific requirements from the Referees.
- We substituted ISI-MIP models (forced with GCMs data) for ISI-MIP2a models (forced with ERA-Interim data) in the benchmarking analysis. Apart from the fact that these new models are run with similar climatic forcing, the ISI-MIP2a models also cover a larger temporal period (2000-2010) compared to the old 2000-2004 period.
- We carefully edit the paper for clarity and communication (wrong wording and repetitions). Also, we incorporated a list of acronyms in the supplementary material to facilitate the large number of variable names, parameters and models.

Interactive comment 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 Received and published: 20 June 2018
Review of "Evaluation of terrestrial pan-Arctic carbon cycling using a data- assimilation system"

Matthias Forkel, Wien, 2018-06-20

---

1 Summary

López-Blanco et al. apply a land carbon data assimilation system to assess carbon fluxes, stocks, and turnover times in arctic and boreal regions. Within the CARDAMON system, parameters of the DALEC2 model are optimized per 1◦ grid cell against observational datasets of LAI, biomass, and soil organic carbon. From the optimized model, carbon stocks, fluxes and turnover times are computed and then compared against results from global vegetation models (GVMs). The approach is very valuable because the carbon turnover in land ecosystem is a main uncertain feature of the global carbon cycle. I really appreciate this work; however, the paper needs substantial revisions before I can recommend publication in ESD (see major comments). Also the structure of many chapters needs to be revised because information are either repeated often at several places or is not given at the appropriate places (see specific comments).

We thank Matthias Forkel for taking the time to assess our manuscript. We believe the comments have substantially improved the manuscript. We thoroughly considered each of the comments, paying special attention to the structure of the paper, to the tundra-taiga domain split, biases with biomass and GPP, the uncertainty reduction of the model parameters, and the GVM benchmarking exercise with the required ISI-MIP2a comparison

2 Major comments

2.1 Tundra-taiga transition and Mongolian grasslands

The grassland region in Mongolia is rather a "steppe" than a tundra (lines 110-112, Fig. S1). Please separate steppe and tundra by either using a latitude threshold, temperature conditions, or a biome map.

The reviewer is correct that the steppe should not be considered as tundra in lines 132-134, Fig. S1. We decided to remove the grasslands/steppes, and focus on higher latitudes and Arctic ecosystems consistent with the focus of the paper. We set the southern boundary of the taiga as our limit - that cuts out the Mongolian grasslands and looks more natural (see Figure S1).

2.2 Computation of transit times

Based on our theoretical assumptions on carbon turnover times (Carvalhais et al., 2014) [supplement], your computations of transit times are partly wrong. The turnover (or transit) time is defined by the C stock of a carbon pool and its outgoing flux. For example the transit time of vegetation is TT_vegetation = biomass / T whereby T includes all processes that remove C from vegetation (litter fall, disturbance, mortality, etc.). Under the steady state assumption (i.e. T = NPP), the transit time of the entire vegetation can be defined as TT_veg = biomass / NPP. Accordingly for the entire ecosystem, the transit time can be defined under the steady state assumption as TT_eco = (biomass + SOC) / Reco = (biomass + SOC) / GPP.
In your calculations, all transit time are computed based on NPP However, only a fraction of NPP goes into the different C pools which is in DELC well defined based on the allocation parameters. Hence the correct computation of a transit time for a certain carbon pool should be based on a fraction of NPP:
NPP_photo = a_foliage * NPP + a_labile * NPP [I assume that C_photo contains the foliar and labile C pools of DALEC2 but this is not described in the paper.]
TT_photo = C_photo / NPP_photo
TT_veg = C_veg / (NPP – NPP_photo) [should this rather be named TT_wood?] TT_soil = C_soil / litterfall [?] = C_soil / Rh [Why you name this TT_dom?]

First, we need to clarify the terminology used here: C_{photo} indeed corresponds to the sum of the foliar and labile pools, C_{veg} refers to the sum of all vegetation pools (foliar, labile, wood and roots) while C_{dom} is the sum of the litter and SOM pools.
We were not clear here - we have used C input into each pool to calculate each pool's transit time (as Matthias suggests). The text in the original paper section 2.2 was wrong in saying:

> "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 each C stock and NPP."

The last bit should be (S2.2.1, L155-158):

> "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.".

To clarify, the transit time for each C pool at grid cell is derived following Bloom et al. (2016) procedure (explained in detail in their Supplementary material, equation 8) as:

$$TT_{pool} = \frac{C_{pool}}{F_{in} - \Delta C_{pool}}$$

Where C_{pool} is the mean pool size, F_{in} is the mean annual C pool input, and ΔC_{pool} is the mean annual change in pool size through 2000-2015 used to correct the calculation for any changes in mean stocks over the study period.

2.3 GVMs with GCM climate forcing

What is the reason for using GVM results that are based on climate forcing from GCMs (lines 176-182)? ISIMIP provides also historical forcing that is based on observed climate data and at least LPJmL provides also model output based on historical data (ISIMIP2A). I assume that the historical climate data better represents climate conditions than the (even though bias-corrected) GCM outputs. Differences in climate forcing can have huge impacts in GVMs. Hence, the comparison between CARDAMON (forced with reanalysis data) and GVM outputs is per se unfair and not comparable.
I request that the comparison between CARDAMON and GVMs should be made comparable by either taking GVM outputs from the historical forcing with ERA-Interim data or by running the optimized CARDAMON with the same GCM forcing.

We understand Matthias' point about consistency between climate drivers. Therefore, we downloaded and used the ISIMIP2a simulations instead of the original ISIMIP1. This new implementation presents many advantages like a longer overlap of simulations (2001-2010) and more similar climate drivers based on the ERA-Interim reanalysis. Overall ISI-MIP2a models are closer related to CARDAMOM than ISI-MIP. ISI-MIP2a models estimated lower NPP, C_{veg} and TT_{veg} than ISI-MIP, and the uncertainties were also lower. In general, we found higher R^2 and lower RMSEs in NPP than C_{veg} and TT_{veg}. Moreover, by using ISI-MIP2a models we remove odd comparisons such as with HYBRID, which may have led to a bias in the attribution analysis we earlier preformed.

We therefore changed the method's part related to ISI-MIP2a (S2.4, L223-232):

> "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."

And the results section (S3.4, L311-338):

> "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, the assessed GVMs estimated 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, with very varied spatial patterns.

The poor 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. 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}. 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}."

And the discussion section (S4.2, L420-446):

"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 low complexity and independent of the benchmarked models.

Using CARDAMOM as a benchmarking tool for six GVMs we found major disagreements for mapping of NPP, C_{veg} and TT_{veg} across the Pan-Arctic for all models (Figure 6) 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."

2.4 Biases with biomass and GPP – wrong use of data and parameter uncertainties?
CARDAMON underestimates the biomass and FLUXCOM GPP. The overestimation of FLUXNET GPP is contradictory but the source of the mismatch is almost impossible to assess given the scale mismatch between FLUXNET sites and 1◦ grid cells. However, if we would assume that both biomass and FLUXCOM GPP are consistent; this could tell us that CARDAMON only needs a higher GPP to gain higher biomass. I'm wondering if it was actually possible to constrain both biomass and GPP within the assimilation framework. Were there any prior parameters used that constrain GPP? As far as I can understand the setup of the approach, there were no data and no parameters included that would constrain GPP (apart from LAI that, however, likely only constrains the seasonality of GPP). To better understand the assimilation results, it is necessary to show maps of reduction of uncertainty of each DALEC2 model parameter. Which parameters were mostly reduced (phenology, allocation, C pools, turnover rates)?
Please also note that the biomass map by Thurner et al. (2014) is largely in agreement with in situ observations of forest carbon density in Russia and slightly underestimates in the USA. If CARDAMON underestimates the biomass, this implies that it would even stronger underestimate the in situ observations than the biomass map. From the results, I get the feeling that the assimilation is over-confident in the SOC data and degrades the performance with the biomass map. Hence the key question is how data uncertainties were used as weights in the assimilation? The uncertainties in SOC are much larger than in biomass (Carvalhais et al., 2014); so I expect that CARDAMON should rather fit the biomass map than the SOC map if these data uncertainties were correctly used.

Matthias asks why the biomass relationship is not 1:1, i.e. why does CARDAMOM not force biomass or LAI to match the assimilated data from Carvalhais et al. (2014) and Myneni et al. (2002), when we do effectively get SOM to match the assimilated data.

It is correct that we do not have a direct constraint on the magnitude of GPP, except for the prior we provide (Table S1). The magnitude of GPP is not so well constrained in CARDAMOM as is its spatial and temporal variability by LAI. We rely on the Aggregated Canopy Model (ACM; Williams et al. 1997) global calibration which is based on SPA runs with some fixed leaf N content. The comparison here suggests that the ACM prior is biased compared to the Arctic, and hence the mismatch. We could assimilate the FLUXCOM GPP data, but here wanted to use FLUXCOM as an independent check. By not assimilating Jung et al., 2017 we have a clearer idea of model reliability and the calculation of transit times. We can conclude that the mismatch is due to using different models; (Jung et al. (2017) uses a range of machine-learning techniques to upscale flux data, we use a process-based ecosystem model DALEC2).

The likely answer to why CARDAMON underestimates FLUXCOM GPP perhaps is that our GPP estimate is biased low to Jung et al. (due to the calibration of ACM); this leads to less NPP which results in a low bias on LAI and biomass. The new implementation in Figure 2 shows that our error includes the 1:1 line, so we are not far out. However, we note that the CARDAMOM regional mean GPP estimate of 314 gC m^{-2} yr^{-1} in tundra is intermediate between the regional models' (350) and global models' (272) reported in McGuire et al. (2012) estimates, as noted in the Discussion).

We add the following text to the discussion (S4.1, L386-400):

> "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"

On top of that, the mapping of uncertainty reduction was a very good idea, which we have implemented (Table 2 and Figure S2):

Table 2. Parameter uncertainty reduction in percentage ranked from least (red) to most (blue) constrained in the pan-Arctic, tundra and taiga domains. The reduction percentage is calculated based on the difference between the 90% CI prior range and the 90% CI posterior range.

Parameter	Name	Process	Pan-Arctic	Tundra	Taiga
MR_{litter}	**Litter mineralization**	Turnover	**3.3**	**3.6**	**2.9**
TOR_{roots}	**Root turnover**	Turnover	**4.8**	**7.2**	**2.2**
TOR_{wood}	**Wood turnover**	Turnover	**9.0**	**8.5**	**9.7**
C_{litter}	**Litter C stock**	Stocks	**13.9**	**13.7**	**14.1**
D_{rate}	**Decomposition rate**	Turnover	**18.2**	**18.6**	**17.8**
f_{rau}	**Fraction of GPP respired (Autotropic respiration)**	Allocation	**30.9**	**31.7**	**30.2**
L_f	**Leaf fall duration**	Phenology	**37.3**	**25.0**	**51.1**
LMA	**Leaf mass per area**	Phenology	**42.8**	**46.3**	**38.9**
C_{roots}	**Fine root C stock**	Stocks	**52.4**	**72.1**	**30.3**
R_l	**Labile C release duration**	Phenology	**53.1**	**52.0**	**54.4**
f_{wood}	**Fraction of NPP to wood C pool**	Allocation	**65.8**	**68.1**	**63.3**
F_{day}	**Leaf fall day**	Phenology	**67.0**	**51.1**	**84.8**
MR_{som}	**Soil organic matter mineralization**	Turnover	**69.1**	**69.6**	**68.6**
f_{labile}	**Fraction of NPP to labile C pool**	Allocation	**74.2**	**75.5**	**72.8**
C_{eff}	**Canopy efficency**	Phenology	**74.7**	**75.5**	**73.7**
f_{roots}	**Fraction of NPP to roots C pool**	Allocation	**75.7**	**74.7**	**76.8**
B_{day}	**Leaf onset day**	Phenology	**76.2**	**67.4**	**86.1**
C_{SOM}	**Soil organic matter C stock**	Stocks	**80.7**	**81.4**	**80.0**
L	**Lifespan**	Turnover	**83.4**	**76.4**	**91.4**
f_{foliar}	**Fraction of NPP to foliage C pool**	Allocation	**88.0**	**88.6**	**87.4**
C_{labile}	**Labile C stock**	Stocks	**92.2**	**95.3**	**88.8**
C_{wood}	**Woody C stock**	Stocks	**92.6**	**90.1**	**95.5**
C_{foliar}	**Foliar C stock**	Stocks	**95.2**	**96.0**	**94.3**

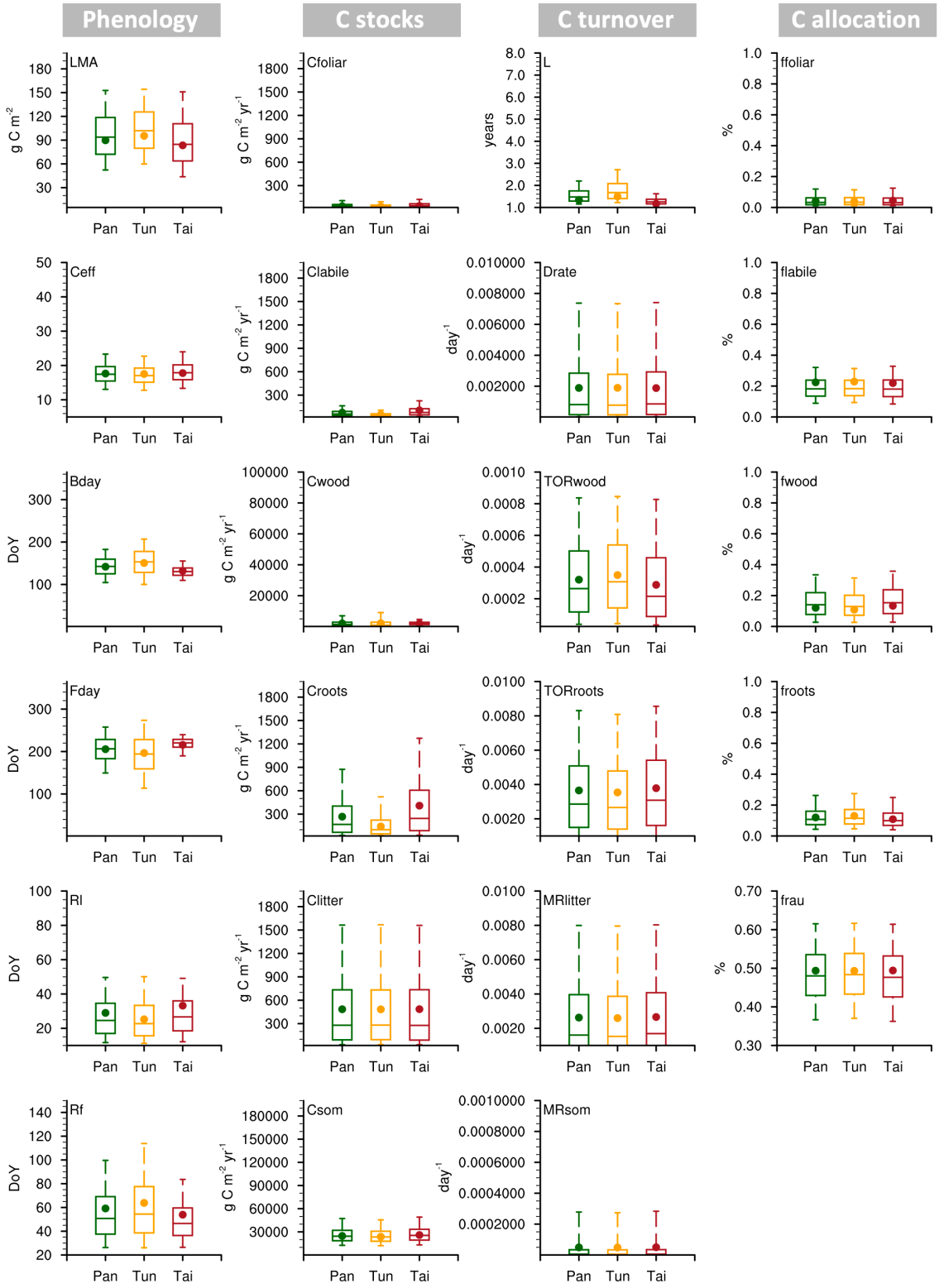


Figure S2. Posterior distributions of parameters estimated for the CARDAMOM assimilation framework in the Pan-Arctic pan-Arctic, tundra and taiga. All y-axes have been scaled to indicate prior ranges (Table S2). Whiskers indicate 90% confidence interval (P05-P95), box indicates interquartile range (P25-P75) and center line represents the median (P50). The whisker plots represents the average (spatial variability in pan-Arctic, tundra and taiga domains) for each percentile (P05, P25, P50, P75, P95).

A new paragraph in section the implemented section 3.2 rank the uncertainty reduction (prior range vs posterior range) per parameter (S3.2, L274-285):

> "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)."

In summary, please report:

1. Which data uncertainties were used and how they were included in the assimilation;

We defined uncertainties used in the likelihood function based on Bloom et al. (2016): log (1.5) for SOC and biomass, both assumed to be representative of initial conditions, and log (2) for LAI. Please see at point 3 the cost function used and how data uncertainty were included in the assimilation.

2. How parameter prior uncertainties were included in the assimilation;

In each 1º x 1º 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):

$$p(x_i|O_i) \propto p(x_i)\, p(O_i|x_i) \quad (1)$$

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

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

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.

3. How the cost function was designed and the different datasets weighted;

$p(\mathrm{O}|x_i)$ 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). 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:

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

4. What are the changes between prior and posterior parameter uncertainties.

This question has been addressed earlier on in point 2.4, referring to new Table 2, Figure S2 and Text in S3.2, L274-285.

In order to improve a solid piece all together with the above set of requests, and because the initial text in the method part referred mostly to previous work and lacked a more specific description of CARDAMON, we decided to redraft this new section (S2.2.2, L160-202) to provide some more details and equations:

"The intermediate complexity of the DALEC2 model compared to typical GVMs models facilitates computationally intense 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) for the 2000-2015 period. A Bayesian Metropolis-Hastings Markov chain Monte Carlo (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 ~130,000 1-km resolution MODIS LAI

data monthly within each 1x1 degree pixel. We aggregated biomass data at 0.5° resolution from Carvalhais et al. (2014) to 1° resolution. These are based on remotely-sensed forest biomass and upscaled GPP based on data driven estimates (Jung et al., 2011) covering the pan-Arctic domain. 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 1º x 1º 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 1º x 1º 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):

$$p(x_i|O_i) \propto p(x_i)\, p(O_i|x_i) \quad (1)$$

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

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

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, $p(\mathrm{O}|x_i)$ 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). 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:

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

For each 1º x 1º 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, $\binom{P95}{P05}$)."

2.5 Benchmarking ISIMIP with CARDAMON

At this point, I will not further comment on chapter 3.3. given the inconsistencies in climate forcing between CARDAMON and GVMs and given the fact that it is not clear how data uncertainties were treated in the assimilation and hence affect the CARDAMON results.

We both addressed new analysis with ISI-MIP2a datasets (responses to point 2.3) and the new information provided based on data uncertainties (point 2.4B).

3 Specific comments

1 Introduction: I suggest to slightly restructure the Introduction to make things a bit more clear. For example, several topics are mentioned twice: "transit times" (around lines 48 and 76), the available data (lines 46-59 and 83-86), and the specific features of the arctic carbon cycle (1. Paragraph, lines 55-59). In addition, the meaning of "transit times" is never explained. I suggest to:
- keep the first paragraph as it is,

We slightly implemented it with few more items (S1, L35-53):

> "Arctic ecosystems play a significant role in the global carbon (C) cycle (Hobbie et al., 2000; McGuire et al., 2009; 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), precipitation (Bintanja and Andry, 2017), and growing season length (Aurela et al., 2004; Groendahl et al., 2007) will likely have consequences the 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, temperature-induced vegetation changes may counterbalance those effects by photosynthetic enhancement (Forkel et al., 2016; Graven et al., 2013; Lucht et al., 2002; Zhou et al., 2001; Zhu et al., 2016). Two examples are the increase of gross primary productivity (GPP) due to extended growing seasons, nutrient availability and CO_2 fertilization (Abbott et al., 2016; Myers-Smith et al., 2015; Myneni et al., 1997) and the shifts in vegetation dynamics such as shrub expansion (Myers-Smith et al., 2011). Consequently, phenology shifts 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 C dynamics of the terrestrial tundra and taiga and their responses to climate change (McGuire et al., 2012)."

- to rewrite the second paragraph: define "transit" and/or "turnover" and /or "residence" time and why it is important,

In S1, L54-71:

> "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 for C transit time dynamics (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., 2015b)."

- to write in the third paragraph about the available in situ and satellite-based data to assess "turnover" times and the associated uncertainties,

In S1, L72-87:

> "There are currently efforts to incorporate both in-situ and satellite-based datasets to assess C cycle retrievals and to reduce 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 (Grøndahl et al., 2008; 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. ”

- to write in the fourth paragraph about the inabilities and uncertainties of GVMs with respect to turnover times,

In S1, L88-103:

“Global Vegetation Models (GVM) have been developed to determine global terrestrial C cycles and represent vegetation ecosystem processes including the structural (i.e. growth, competition, and turnover) and biogeochemical (i.e. water, carbon, and nutrients cycling) responses to climate variability (Clark et al., 2011; Fisher et al., 2014; Friend and White, 2000; Ito and Inatomi, 2012; Pavlick et al., 2013; Sitch et al., 2003; Smith et al., 2001; 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 spin-up processes until 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 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.”

- and finally to present model-data integration and CARDAMON as the potential “solution” in the last paragraph including the definition of your objectives.

In S1, L104-122:

“An approach to address these issues is to integrate models and data more formally. Data assimilation quantifies how model parameters can be adjusted to estimate C stocks and fluxes consistent with multiple observations (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 can map parameter estimation and uncertainties.

Here, we use the CARbon DAta MOdel framework (CARDAMOM) (Bloom and Williams, 2015; Bloom et al., 2016; Smallman et al., 2017) to retrieve the pan-Arctic terrestrial carbon cycle at 1º resolution for the 2000-2015 period in agreement with gridded observations of LAI, biomass and SOC stocks. 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 sites included in the FLUXNET2015 dataset, and (c) 6 extensively used GVMs from the ISI-MIP2a comparison project (Warszawski et al., 2014). 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 system, (2) quantify the degree of agreement between the CARDAMOM product with local to global scale sources of available data, and (3) use CARDAMOM as a benchmarking tool for the ISI-MIP2a models to provide general guidance towards GVM improvements in transit time simulations, 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 cycling modelling."

Line 37: Use either "warming" or "temperature increase" but not "warming increase" because this would be an acceleration in temperature increase.

This has been changed to "temperature increase" (S1, L40).

Line 40-41: In addition to Lucht et al. and Myneni et al., you could also cite more recent related publications (Forkel et al., 2016; Graven et al., 2013; Zhu et al., 2016) [I don't request to include my paper!]

Thanks for the recommendation, all the suggested publications have been added (S1, L46-47).

Line 48 and lines 75-82: "transit times" – Carvalhais et al. use "turnover" time, Friend et al. "residence" time, and Thurner et al. (2016) "turnover rate". Is there a reason why you use "transit time" and why you are not using one of the other terms? Please provide a short definition of these terms or the term that you are using and how they differ.

The "transit time" terminology instead of others arose from a paper by Sierra et al. (2017): https://onlinelibrary.wiley.com/doi/abs/10.1111/gcb.13556

The new paragraph two of the introduction (S1, L54-71) has been restructured based on your previous comment and we clarified why we used transit time and not the other terms. We also provided a short definition and how they differ.

Line 61: "PFT or spin-up": The "or" should be replaced by "and".

Changed accordingly (S1, L94).

Line 74: Relevant is also the work by Thurner et al. (2016)

Reference included now (S1, L65).

Lines 84-85: Please provide references.

References have been provided accordingly (S1, L82-85):

> "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)."

Line 87: A reference to a general overview paper on model-data integration might be useful.

An updated text, including extra information and general references regarding model-data integration, have been implemented (S1, L104-110):

> "An approach to address these issues is to integrate models and data more formally. Data assimilation quantifies how model parameters can be adjusted to estimate C stocks and fluxes consistent with multiple observations (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 can map parameter estimation and uncertainties."

Line 95-96: It is not clear to me how your analysis will provide further insight into GVMs that goes beyond the work of Friend et al. (2014), Carvalhais et al. (2014), and Thurner et al. (2017). Please make clear what kind of additional knowledge you are expecting from your analysis on the problems of GVMs.

Our system is constrained by both model structure and varied observations, using the EDCs to make sure that processes are realistic and ecologically viable. The previous studies have largely focused either on model analyses (Friend et al) or on combining data products (Carvalhais, Thurner et al) to generate TT. Here we use CARDAMOM to combine the information contained in model structure with independent observational data to produce a consistent, robust analysis. Our approach avoids using PFTs and steady state assumptions (typical in the GVMs). By including a mass balance constraint on the C cycle we evaluate consistency among different data sets (e.g. SOM, biomass, LAI, climate) using our model structure to generate TT. Thus the novelty of this study is a data-constrained descriptions of C cycling for numerous live

and dead pools, and their transit times, with errors at pixel scale. We use the complete assessments to assess better GVMs, to identify how to produce more constrained forecasts of this sensitive region.

We refined the aim (3) in the following (S1, L116-121):

> "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 system, (2) quantify the degree of agreement between the CARDAMOM product with local to global scale sources of available data, 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."

Lines 106-107: Please define which classes you used to separate forest and non- forest.

This new line has been included in the text to better refer to the different classes used to separate tundra and taiga domains (S2.1, L129-130):

> "A complete description of the classes included in each domain can be found in Figure S1 and caption."

Figure S1 caption states now:

> "Figure S1. Spatial domain defined by the Northern Circumpolar Soil Carbon Database version 2 (NCSCDv2) region. The tundra- taiga regions were separated based on the presence-absence of forested areas using the GlobCover map (http://due.esrin.esa.int/page_globcover.php). Forested areas (taiga) included: closed to open broadleaved evergreen or semi-deciduous forest (>5m), closed (>40%) broadleaved deciduous forest (>5m), open (15-40%) broadleaved deciduous forest/woodland (>5m), closed (>40%) needleleaved evergreen forest (>5m), open (15-40%) needleleaved deciduous or evergreen forest (>5m) and closed to open (>15%) mixed broadleaved and needleleaved forest (>5m). Non-forested areas (tundra) included the rest of classes: mosaic forest or shrubland (50-70%) / grassland (20-50%),mosaic grassland (50-70%) / forest or shrubland (20-50%), closed to open (>15%) (broadleaved or needleleaved, evergreen or deciduous) shrubland, closed to open (>15%) herbaceous vegetation (grassland, savannas or lichens/mosses), sparse (<15%) vegetation, closed to open (>15%) broadleaved forest regularly flooded (semi-permanently or temporary), closed (>40%) broadleaved forest or shrubland permanently flooded, and closed to open (>15%) grassland or woody vegetation on regularly flooded or waterlogged, post-flooding or irrigated croplands (or aquatic), rainfed croplands, mosaic cropland (50-70%) / vegetation (grassland/shrubland/forest) (20-50%), mosaic vegetation (grassland/shrubland/forest) (50-70%) / cropland (20-50%), bare areas and permanent snow and ice. On top of that, latitudes lower than 52°N within the tundra domain were neglected to focus on higher latitudes."

Section 2.2: The description of CARDAMON refers mostly to previous work. However, to understand _this_ paper, I suggest to provide some more details or equations with respect to the following questions:

- LAI, biomass, and SOC are used as data sets in a cost function for parameter estimation and not as forcing data. Is this correct?

This is correct, we now better describe this in the methods (see our latter comment in point 2.4 (major comments) including the new section 2.2.2 from the main text). On top of this, we included a new Table S3 describing which datasets were used as forcing, data constraints and independent validation.

- What is the cost function? How are the differences in the number of data points weighted (LAI is a time series, SOC and biomass only single values per grid cell)?

The cost function was already described before in point 2.4 (major comments) including the new section 2.2.2 from the main text.
Regarding the second question, basically Equations 2 and 3 now included in text answer this: we did not weight for differences in the number of data points.

- Why is the MHMCMC algorithm used three times? Does it not explore the full parameter space if it is applied only once? Or are there difference in initial values?

The process is repeated to make sure independent chain converge to the same posterior distribution. This is a standard procedure in MCMC analyses, see for example Vrugt et al. (2003) and Kuczera and Parent (1998). In CARDAMOM we only keep parameters from the chains which have converged.

Vrugt, J. A., H. V. Gupta, W. Bouten, and S. Sorooshian (2003), A Shuffled Complex Evolution Metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters, Water Resour. Res., 39, 1201, doi: 10.1029/2002WR001642, 8.

Kuczera, G., and Parent, E.: Monte Carlo assessment of parameter uncertainty in conceptual catchment models: the Metropolis algorithm, Journal of Hydrology, 211, 69-85, https://doi.org/10.1016/S0022-1694(98)00198-X, 1998.

- Can you make a conceptual figure that shows which data sets go into the assimilation and which are only used as independent evaluation data?

We implemented a new Table S3 (S2.2.2, L171) instead, since we also implemented a conceptual Figure 1 (S3.1, L235) representing key C fluxes, stocks and transit times which is already complex enough:

Table S3. Forcing dataset, observational constraints and independent validation datasets used in this study's experimental design.

Dataset	Source	Forcing	Constraint	Validation
ERA-Interim	Dee et al. (2011)	X		
MODIS - LAI	Myneni et al. (2002)		X	
NCSCD - SOC	Hugelius et al. (2003)		X	
Biomass	Carvalhais et al. (2014)		X	
NEE, GPP, R_{eco}	FLUXNET2015			X
GPP	Jung et al. (2017)			X
R_h	Hashimoto et al. (2015)			X

Line 147: What is the difference between "photosynthetic" and "vegetation" C stocks? Is photosynthesis not vegetation?

Text now has been updated to explicitly define "photosynthetic" and "vegetation" C stocks (S2.2.1, L150-152):

> "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."

Line 159: Did you directly compare the 1◦ grid cell with the FLUXNET site? If yes, how are the FLUXNET sites representative for the 1◦ grid cell? If no, did you run CARDAMON with the site meteorological data? Add: Only from the discussion (lines 318-238), I now learn that you did a grid cell to point comparison. This should be already mentioned in the methods and be recalled at the appropriate place in the results section.

We included a whole paragraph discussion on how FLUXNET2015 sites are representative for the 1◦ grid cell in the S4.1. However, the referee is right that it was only mentioned in the discussion part, so we implemented the following text in S2.3, L219-221:

> "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."

Lines 166-168: This is a repetition from lines 123-125. Please merge the two sentences.

We have merged the two sentences as requested (S2.2.1, L152-155):

> "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."

Line 171: TT_veg was already mentioned at line 148. I suggest to remove both occurrence of TT_veg and to already define TT_veg in the new second paragraph of the introduction.

We followed this request, and we restructured the introduction to define transit times (TT) in the second paragraph as you suggested above.

Line 171: At the end CARDAMON is also just only a GVM but with grid cell-specific parameters. I don't see how CARDAMON then serve as a benchmark for the other GVMs. Would it be not enough to directly benchmark the GVMs against the reference data? You should try to better motivate already in the Introduction why you can use CARDAMON as a benchmark for GVMs.

We implemented already aim (3) in the introduction, see above. We use CARDAMOM because it provides estimates of TT with errors that are consistent with theories of C cycling and with multiple observational data. CARDAMOM avoids the significant problem of spin-up associated with GVMs.

Moreover, we implemented the following text in the discussion (S4.2, L420-431):

> "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 low complexity and independent of the benchmarked models."
> "

Line 173: LPJmL (capital L). Please indicate which version of LPJmL was used. Is it the most recent version (LPJmL4) (Schaphoff et al., 2018a)? LPJmL4 includes also a new permafrost module (Schaphoff et al., 2013) and a data-constrained phenology module (Forkel et al., 2014) and hence better reproduces boreal and arctic carbon stocks and carbon cycling than the previous versions (Forkel et al., 2016; Schaphoff et al., 2018b).

Thanks for the correction, we now refer to the model as "LPJmL". LPJmL4 seems a very good candidate to be compared against CARDAMOM. However, the LPJmL version (version name is missing) available in the ISI-MIP2a database only includes the permafrost module (Schaphoff et al., 2013) you mentioned with runs dated from 2016: https://www.isimip.org/impactmodels/details/81/#tab_isimip2a

Lines 184-194: I got really confused by this paragraph because initially I got the impression that you "jump" across all results without explanation. Please make clear that this paragraph is a summary of all results by either using a heading or a suitable topic sentence.

Following Matthias comment, together with the one from **REF#2**, we decided to drop the summary paragraph in the results section. We believe this also contributes to a lighter text.

Line 202: Which of the used data sets constrain the separation between GPP and NPP?

We only have a prior of Ra:GPP = 0.5

Line 216: Do you find spatial pattern in TT_photo that would resample the distribution of evergreen and deciduous trees?

In this initial study using CARDAMOM over the pan-Arctic region we decided to partition results into tundra (grass/shrub dominated) and taiga (forest dominated) areas. Given the fact that the text is already quite dense, we believe further analysis such as the one you proposed (partition evergreen and deciduous trees) would add an extra layer of complexity to this manuscript. However, we find this suggestion quite relevant and attractive, and so it would be wise to have it addressed in coming papers.

Line 219: "Interestingly" – Please tell me why this is "interestingly".

We removed the wording "Interestingly", plus we rephrased the sentence to (S3.1, L260-262):

> "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."

Line 257: "as noted " – Please check.

We rephrased the sentence to (S3.3, L302-303):

> "This mismatch is important in the context of the FLUXCOM GPP upscaling, 50% higher than CARDAMOM."

Lines 330-331: This sentence should be merged with the numbers given at lines 349- 352.

We restructured the section 4.1. to improve readability and clarity (L347-349):

> "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):"

This paragraph is now followed by a list of 3 numbers (I, II, III) focusing on each of the previous relevant papers addressed above.

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., 2009b, a), 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 $\binom{47.5}{10.3}$ 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.

Line 341: Are you sure to use the right reference for LPJ-GUESS-WhyMe?

Matthias is right, we changed this reference to (S4.1, L358):

"LPJ-Guess WHyMe (Wania et al., 2009a, b)"

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.

Interactive comment on Earth Syst. Dynam. Discuss., https://doi.org/10.5194/esd-2018-19, 2018.

Interactive comment 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: 7 August 2018

In this paper the authors take the CARDOMOM + DALEC Bayesian calibration system an apply it specifically to the arctic using a number of regional-scale data products. Once the model is fit to data, it is then used to assess carbon pools and benchmark global vegetation models. The scale and scope of the analysis is quite impressive – building up their system to this point was clearly a lot of work and the attempt to synthesize multiple data constraints at a regional scale is really important, especially for a highly influential and understudied region like the arctic.

We are thankful for the reviewer's insightful and thorough comments. We believe this review has substantially improved the manuscript, spotting incomplete areas and highlighting convincing areas that perhaps we could further emphasize. We have carefully considered the reviewer's remarks and clarified our manuscript accordingly.

That said, I do have a few high level concerns about what the authors have done. The easiest of these to address is that the details of what was actually done was insufficient and teasing out important high-level facets of CARDOMOM are left to the reader tracking down earlier papers. Particularly important is to clarify whether DALEC is calibrated independently for every pixel, in some sort of spatially correlated manner, or with a single parameterization for the two PFTs across the whole region. My recollection from earlier papers made me think the first (independent fits), but in reading the results it is hard to distinguish parameter uncertainty from parameter spatial heterogeneity. The authors need to be more explicit about this. Likewise, the authors need to be more clear about whether this is really a data assimilation system, or if it's just a calibration system. This matters because in DA (e.g. EnKF) the analysis provides a formal synthesis of observations and process understanding, but in a calibration system your estimated states are ultimately just a forward model run. To me, it feels like the authors are treating a forward model run as if it were a reanalysis product. If this is true what the authors did is still valuable but they should be more open about this and the limitations of this approach.

We fully understand this point, agreeing with the **REF#1** (one of the major critic comments). DALEC2 is independently calibrated in each pixel. We tried to do a better job in showing the within pixel parameter error. In order to do so, we added to the main text, figures, tables and SI:

1) An implemented section 2.2 (The CARbon Data Model framework) in methods (L136-141)

> "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) An implemented section 2.2.1 (DALEC2) in methods (L143-158)

"DALEC2 ecosystem model simulates 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, 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). 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 fraction 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 each C stock and the NPP allocated into that stock."

3) An implemented section 2.2.2 (Data-assimilation) in methods (L160-202) to be more explicit about the experimental design including equations as noted above in responding to the **REF#1**,

4) A new section in the results section regarding Data assimilation and uncertainty reduction (S3.2, L274-285) together with a new Table 2 and Figure S2 as noted above in responding to the **REF#1**.

Second, in light of the earlier point about reanalysis vs forward simulation, I am really uncomfortable about the author's use of their model as benchmark for other models. This is particularly true given the non-trivial biases in some of the verification (biomass) and validation (GPP, Rh) analyses and the lack of independent validation of a number of the other processes in the model (e.g. turnover). I think this manuscript could stand alone without the GVM component.

In this manuscript we want to recognise that the uncertainties are large – uncertainties are rarely if ever calculated and presented, and this means analyses have been overconfident. We agree that the full descriptions of C cycling with errors are novel. We aim to go further to use the

complete assessments of C cycling to assess better GVM, to allow more constrained forecasts for this region.

As noted above in responding to the **REF#1**, CARDAMOM reconciles observational datasets as part of a representation of the terrestrial C cycle in agreement with ecological theory – CARDAMOM provides observationally-constrained estimates of C dynamics. A key reason to use CARDAMOM as a benchmark for GVMs is that CARDAMOM produces parameter likelihoods for each pixel based on data, and it does not assume PFTs or steady states, hence it is much more strongly data constrained that a GVM. CARDAMOM takes data and the model to produce parameter maps, whereas GVMS take PFT maps of parameters to produce flux/stock outputs. For this reason, we used our data assimilation system as a benchmarking tool for six GVMs.

Our analyses of the C cycle are independent of the GVMs. As the referee notes there are biases in the verification and validation. However, because our approach takes into account the error in assimilated data, we produce uncertainties on our C cycle estimates. Our analytical uncertainties on e.g. biomass include the data within their confidence intervals, so the potential for this bias is explicit within our outputs, and propagated into e.g. TT estimates. Independent estimates of TT do not exist – one purpose of this study is to provide robust estimates of TT from CARDAMOM to compare with GVMs. The novelty here is that we can locate which GVMs and for which regions the TT estimates of models are outside CARDAMOM confidence intervals (stippling in Figure 6).

Third, I'm really concerned about how the authors assimilate these derived data products. There's not really any discussion of how the observation errors in the data and process error in the model are treated. There's not any discussion of how the authors handled the non-independence of spatial pixels in these data products. Indeed the authors seem to treat data products as if they are truly data, which likely results in an overestimation of the true information content in the data. For example, if I have 10 observations I can Krige a map that has 10k grid cells, but my true sample size remains 10 not 10k and any data assimilation system needs to reflect that.

As noted above in responding to the **REF#1**, errors from the observational products are not available from the data providers but we still defined them in the likelihood function based on Bloom et al. (2016). Therefore, there were errors attached to each observation, e.g. MODIS LAI, biomass for each pixel. We assumed independent data in each grid cell for LAI and biomass deriving from satellite products. We recognise that the algorithms used to produce e.g. LAI may include spatial assumptions that generate correlated biases. We acknowledge that the soil C data are interpolated using machine learning approaches. Our response is to include a large error on the data for each pixel in the absence of a detailed, spatially defined error.

As we pointed earlier (**REF#1 response**), we included a full new description about how these data were assimilated and which uncertainties were considered in each dataset (S2.2.2, L178-202). More specifically in this new section we mention that "Each pixel is treated independently without assuming a prior land cover type and we assume no spatial correlation between uncertainties in all pixels."

Detailed Comments:

L126: 1) Is calibration really data assimilation? 2) inclusion of process error?

The correct term for CARDAMOM is data assimilation or model-data fusion. Again, new sections 2.2.1 (DALEC2) and 2.2.2 (Data-assimilation) were implemented in the methods section to better address parameter uncertainty reduction. Data assimilation explicitly propagates data uncertainty into model calibration.

L135: What is the actual underlying sample size? Derived data products can massively conflate the actual information content. Errors in these data products are hugely autocorrelated and that observation uncertainty is not captured correctly in these products. Also, many of these constraints are not data (GPP, LAI, biomass) but just different models.

We have updated section 2.2.2 with the following text (L169-177):

> "Observational constraints include monthly time series of Leaf Area Index (LAI) from the MOD15A2 product (Myneni et al., 2002), estimates of vegetation biomass and soil organic carbon content (Table S3). We aggregated ~130,000 1-km resolution MODIS LAI data monthly within each 1x1 degree pixel. We aggregated biomass data at 0.5° resolution from Carvalhais et al. (2014) to 1° resolution. These are based on remotely-sensed forest biomass and upscaled GPP based on data driven estimates (Jung et al., 2011) covering the pan-Arctic domain. 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)."

L144: 500 samples per chain? That's way too small. Also, what's the effective sample size after accounting for autocorrelation? I'd recommend the authors shoot for an effective sample size around ~5000 total, which likely will require a much larger total number of samples given their reliance on Metropolis-Hastings. Not stated explicitly whether this is one global parameter set or one per grid cell? My memory from Bloom et al 2016 is the latter.

We apologise that we did not explain our process more clearly; we ran three chains accumulating 10^7 accepted parameters We ensured convergence in at least 2 chains and gathered 500 sampled parameter sets from each chain. This sampling was used to generate a distribution of model analyses that we then analysed and plotted, see section 2.2.2.

L146: A 90% CI is typical. Reason for not 95% norm?

A 90% CI is a widely used in the literature so we decided to assess this specific uncertainty range in our analysis.

L154: This isn't independent of the calibration product

Indeed, this is independent of the calibration product. We do not assimilate GPP, but LAI.
The new text in Section 2.2.2 hopefully clarifies this, together with the new Table S3.

L164: should really include the 95% CI in addition to the interquartile

We changed from 50% CI to 90% CI. Figure 4 has been updated accordingly, and uncertainties represent the 25th and 50th percentiles (darker shade) and the 5th and 95th percentiles (lighter shade) of both field observations and the CARDAMOM framework.

L169: You can't compare a complex model against a (mis)calibrated simple model and call it a benchmark. Especially true if you're looking at the marginal distributions of indirectly inferred latent variables.

This point is arguable – what do we mean by a benchmark? As the referee has acknowledged, all global/gridded data products involve some degree of modelling (e.g. FLUXCOM, MODIS LAI), and hence are not direct measurements. Thus, a comparison against a data product is open to this same criticism. In response we make the following addition to the text in the discussion (S4.2, L420-431):

> "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 low complexity and independent of the benchmarked models"

L177: If looking at the historical period, why weren't models run under reanalysis meteorology rather than GCMs?

We have adjusted our analysis in response to this comment. A complete new GVM exercise was performed since both **REF#1** and **REF#2** raised the same criticism. As we pointed earlier (**REF#1 response**), we used the ISI-MIP2a simulations instead of the original ISI-MIP1. This new implementation presents many advantages like the full overlap of the studied period (2001-2010) and more similar climate drivers. Earlier we also specified that we changed the method's part related to ISI-MIP2a (S2.4, L223-232), the results section (S3.4, L311-338) and discussion (S4.2, L420-446).

L184: Drop this whole paragraph – it's a bit confusing to give a summary of the results before presenting the results without making it clear that this is a summary of highlights. Right now it just feel like you're going though the results really quickly without much explanation.

This summary has been deleted accordingly since **REF#1** suggested that too.

L187: A 28% bias against the data that the model was calibrated to seems like a pretty big problem.

As discussed for **REF#1**, this bias likely arises from a lower predicted rate of photosynthesis in Arctic systems in the ACM photosynthesis model. The newly implemented Figure 2 shows that the 90% CI includes the 1:1 line. Thus, our analysis is not inconsistent with the data. We should note that the data may also be biased. Further, at regional scale our GPP estimates lie within the ranges of GVMs and regional models.

L190: "This mismatch is important in the context of FLUXCOM, as noted" what do you mean "as noted" you never noted anything

We rephrased the sentence to (S3.3, L302-303):

> "This mismatch is important in the context of the FLUXCOM GPP upscaling, 50% higher than CARDAMOM GPP."

L203: "and marginally varied across tundra" I don't understand what you mean here

Wrong wording, we meant (S3.1, L242):

> "and marginally varied between tundra $0.50\binom{0.54}{0.46}$ and taiga $0.52\binom{0.56}{0.46}$."

L209: Distinguish tundra and taiga. These numbers don't seem plausible for tundra

As requested, we implemented to (S3.1, L247-249):

> "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})."

L211: That the tundra numbers are so close to the taiga numbers doesn't seem correct. How well do these numbers validate against direct field data (not derived data products)?

The relatively small differences between tundra and taiga are a result of the coarse scale of the data and the relatively low values of biomass in the assimilated product (Figure 2). It is not appropriate to evaluate biomass products against field data due to the scale mismatch. As the other referee has noted:
"**…the biomass map by Thurner et al. (2014) is largely in agreement with in situ observations of forest carbon density in Russia and slightly underestimates in the USA.**"

L216: A transit time of 4.3 years in the woody tissues of a spruce tree seems really fast give their lifespan. How does this compare to field data (e.g. isotopes)

The 4.3 years does not differentiate between branch size of woody components. This value represents the median's of all woody material in the entire pan-Arctic region, including twigs, branches, stems and coarse roots.

L217: The CI on the SOM is really large (essentially 10-1000 years). Is this just the prior?

Yes, due to the lack of data on SOM turnover, the TT for SOM remains poorly constrained.

L228: This results needs additional explanation with regards to what this test statistic applies to. You calibrated a mechanistic model via MCMC, this isn't a t-test. What specifically changed that much?

This entire paragraph has been dropped. Please check answer to **REF#2** in L295 (three points below).
This change arose mainly due to the fact that the presented sensitivity analysis did not included a version without a soil C constraint. The fair analogy should be done by removing soil C as a constraint, or by using other biomass product as constraint. Otherwise, the information content in both data-streams cannot be comparable by this exercise that eliminates one data stream and compares two of them.

L234: What do you mean priors, isn't this the data?

Confusing wording here, we implemented the text to (S3.2, L264-265):

> "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 (Figure 2)."

L257: This is almost the exact same sentence as L190

Since the initial summarizing paragraph in the results section has been dropped following **REF#1** and **REF#2** suggestions, this sentence now can be kept, but rephrased to (S3.3, L302-303):

> "This mismatch is important in the context of the FLUXCOM GPP upscaling, 50% higher than CARDAMOM GPP."

L295: Is this statement that CARDOMOM is more sensitive biomass than soil C really fair? In one case you're comparing whether a data constraint is included at all, while in the other your comparing different derived data products, which are likely relying on similar underlying raw data. I think for this to be fair you would want to include a version where you don't have any soil C constraint.

The referee is completely right, the presented sensitivity analysis did not include a version without a soil C constraint. The fair analogy should be done by removing soil C as a constraint, or by using other biomass product as constraint. Otherwise, the information content in both data-streams cannot be comparable by this exercise that eliminates one data stream and compares two of them. Therefore, and because the manuscript is complex enough, this entire paragraph has been dropped.

L308: There's a 28% bias in biomass, how is that "good agreement". The Discussion seems to be missing the critical point that if a model is faced with multiple constraints and can't reconcile them then there's either inconsistencies in the data, structural errors in the model, or both. And why is there no comparison to LAI and GPP constraints? Also, there seems to be no discussion of how observations error in the data are derived/treated and how you're handling the process error in the model (is this a fit parameter or just ignored).

As we pointed earlier (**REF#1 response**), the magnitude of biomass is not so well constrained in CARDAMOM (although the 90% CI include the 1:1 line, see new Figure 2) as is its spatial and temporal variability (well constrained by LAI). We do not have a direct constraint on the magnitude of GPP, except for the prior we provide (Table S1). We rely on the Aggregated Canopy Model (ACM; Williams et al. 1997) calibration which is based on SPA runs with some fixed leaf N content. Perhaps ACM is biased compared to the Arctic, lacking a temperature acclimation, and hence the mismatch. We now include reference to this issue in the text, and include comparison to GPP explicitly earlier in the results and discussion. We note the importance of data error in the resulting biases in both calibration and validation, and the need for more robust error characterisation of data products.
We note that the Calvalhais et al. (2014) biomass data relies on various assumptions in areas of low tree cover, covering much of the high latitudes. In these areas GPP data from Jung et al. are used for calculating herbaceous biomass. Thus, there is a dependence between these data products. We need fully independent, error characterised data to make the next steps forward. Hence the broad errors set on data products in this analysis.

A new comparison set of assimilated LAI has been implemented in Figure 2 as was requested by **REF#2**.
We have adjusted the text in 3.2 to read (L264-273):

> "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 = 0.97 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, but ~28% higher than CARDAMOM analyses. These biases (Figure 3) 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 capacity of DALEC2 to reproduce the patterns in LAI, SOM, biomass and GPP is a strong indicator that the model structure is valid. What we notice is bias on the estimation of some of these parameters across the region. We relate this bias to a strong model prior on photosynthesis and large uncertainty on the LAI and biomass data. An increase in data confidence would resolve this problem.

In the discussion 4.1 (L386-400) we now state:

"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) relates to the strong prior on photosynthesis estimation 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 DALEC2 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 smaller 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 sensitivity. Certainly, the need for robust characterisation of error for data products is of critical importance for improved analyses."

L312: I haven't looked into the details of the Jung 2011 product vs the Jung 2017 product, but I'm skeptical that these are independent. Would be good to state more explicitly what each product is upscaling to generate GPP (FLUXNET? SIF?). If they're both FLUXNET-based then they're not independent if they're just applying different algorithms to upscale the same underlying data.

We only used Jung et al 2017's GPP product as independent validation, thus we do not see why we should state more explicitly what each product is upscaling to generate GPP. However, it seems that there is a misunderstanding - **REF#2** believes that GPP has been assimilated in CARDAMOM. We hope now that the new section 2.2.2 in the method section had clarified this point.

L314: "One difference between these two models is. . ." What two models?

This sentence has been changed accordingly (4.1, L403-404):

"One difference with Hashimoto et al. (2015)'s R_h model is the lack of moisture limitation on respiration in CARDAMOM."

L317: I'd recommend making this sentence the start of the next paragraph

This sentence has been also changed accordingly to properly conclude the paragraph (4.1, L405-407):

"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."

L319: How do you know that the issue is only one of scale difference, and not some other error in the model or DA system? What could you do to confirm this (e.g. run with local drivers)?

We cannot be certain here, but mismatch on LAI due to local variability seems the most likely cause of mismatch. We could confirm this by running using LAI data directly determined from the study site at an appropriate scale.

L326: This error in timing is an example of why it might be better to run a system that performs both state and parameter data assimilation, rather than just parameters.

We believe parameter estimation should be acceptable – given that the parameters we incorporate for calibration include those that drive phenology, and thus state changes.

L328: It's a bit surprising that you're running a model in the arctic that doesn't include snow or permafrost. I see that this point is in the Discussion, but it seems really important to be more upfront about this earlier in the paper, as it's a pretty limiting assumption and should lead to greater caution in how confidently you interpret the results. It also begs the question as to why you didn't couple CARDOMOM to a more sophisticated land model for this analysis.

Our goal is to use a C model as simple as possible, with strong data constraint. DALEC has direct state variable linkage to LAI, biomass, SOM. Key processes are climate-sensitive (GPP, R_h). For forecasts of the future C cycle we agree that simulating the changes to permafrost and hydrology would be vital, but for analyses of the current C cycle and its internal dynamics, snow and permafrost are secondary factors.

L365: But is there any direct field constraint (e.g. isotope data)

We recognise the value of using independent data on turnover rates (e.g. from isotopes) to evaluate analytical estimates. Such comparison was beyond the scope of this paper, but is the target of current and future research.

Interactive comment on Earth Syst. Dynam. Discuss., https://doi.org/10.5194/esd-2018-19, 2018.

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)

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. ~~The Arctic accounts for approximately 50% of the global soil organic C stock, emphasizing the important role of Arctic regions in the global C cycle.~~ Recent studies have ~~pointed to the~~highlighted poor understanding particularly of C pool~~s turnover~~transit times, ~~although remain unclear as to~~and whether productivity or biomass dominate these biases. The Arctic, accountings for approximately 50% of the global soil organic C stocks, ~~emphasizing~~has a ~~the~~n important role ~~of Arctic regions~~ in the global C cycle. Here, we use ~~an improved version of~~ the CARDAMOM data-assimilation system~~,~~ to produce pan-Arctic terrestrial C ~~-related~~ cycle ~~variables~~ analyses for 2000-15. This approach ~~without~~ avoids using traditional plant functional type or steady-state assumptions. ~~Our approach~~We integrate~~s~~ 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 ~~for the first 15 years of the 21st century~~, ~~but~~ and also to provide general guidance about the controlling biases in ~~the turnover~~ transit ~~dynamics~~times. ~~As~~ On average, CARDAMOM estimates ~~513 (456, 579), 245 (208, 290) and 204 (109, 427) g C m^{-2} yr^{-1} (90% confidence interval) from~~regional mean rates of photosynthesis of ~~513551~~565 g C m^{-2} yr^{-1} (90% confidence interval between the 5th and 95th percentiles:~~confidence interval: 45617~~428, ~~579722~~741), autotrophic respiration of ~~24563~~270 g C m^{-2} yr^{-1} (~~208~~177~~82~~, ~~290386~~397) and heterotrophic respiration of ~~2015~~219 ~~4~~ g C m^{-2} yr^{-1} (~~1093031~~, ~~427141958~~)~~respectively~~, suggesting ~~that the~~ a pan-Arctic ~~region acted as a likely~~ sink of ~~-55~~ 67 (~~-152280~~7, ~~157112860~~) g C m^{-2} yr^{-1}, weaker in tundra and stronger in taiga~~, but~~. However, our confidence intervals remain large (and so the region could be a source of C), reflecting uncertainty assigned to the regional data products. ~~In general, w~~We show ~~find a good balanced~~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. T~~-~~he 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. ~~Using~~ Comparing CARDAMOM ~~as a benchmarking tool for~~analyses to global vegetation models (GVM) for the same period, we ~~also~~ conclude that ~~turnover~~ transit times of vegetation C ~~is~~ are ~~weakly~~ inconsistently simulated in GVMs~~vegetation models and is a major component of error in their forecasts~~ due to a ~~joint~~ combination of uncertainties from ~~their~~ productivity and biomass calculations. Our findings highlight that GVMs ~~modellers~~ need to focus on ~~the~~ constraining both current vegetation C stock~~s~~ ~~dynamics, but~~ and ~~also~~ net primary production~~their respiratory losses~~, to improve ~~our~~ process-based understanding of ~~internal~~ 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., 2009; 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 ~~soil organic C~~SOC stock (Hugelius et al., 2014; Jackson et al., 2017). High latitude ecosystems are experiencing a ~~warming~~ temperature increase that is nearly twice the global average (AMAP, 2017). The expected future increase of temperature (IPCC, 2013), precipitation~~s~~ (Bintanja and Andry, 2017), and growing season length (Aurela et al., 2004; Groendahl et al., 2007) will likely have conse~~c~~quences ~~in~~for ~~the sink/source d of~~ the Arctic net C balance. ~~SOC mineralisation~~As high latitudes warm, C cycle dynamics ~~may increase rapidly in response to warming, which~~ may lead to an increase of carbon dioxide (CO_2) emissions through ecosystem respiration (R_{eco}) ~~in release of carbon dioxide (CO_2) through~~ 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) ~~and thus increased emissions of CO_2 through ecosystem respiration (R_{eco})~~. However, temperature-induced vegetation changes may counter-balance those effects by photosynthetic enhancement (Forkel et al., 2016; Graven et al., 2013; Lucht et al., 2002; Zhou et al., 2001; Zhu et al., 2016) ~~may mitigate those effects by photosynthetic enhancement~~. Two examples are the- increase of gross primary productivity (GPP) due to extended growing seasons, nutrient availability and CO_2 fertilization (Abbott et al., 2016; Myers-Smith et al., 2015; Myneni et al., 1997) and the shifts in vegetation dynamics such as shrub expansion (Myers-Smith et al., 2011). Consequently, phenology shifts 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 C dynamics of the terrestrial tundra and taiga and their responses 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 ~~vast~~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 ~~controls~~ interactions~~g~~ between the ~~C~~ allocation of C from~~storage~~ net primary productivity (NPP), ~~and C turnover~~C stocks (C_{stock}), and transit times (TT), ~~-~~is deficient (Carvalhais et al., 2014; Friend et al., 2014; Hobbie et al., 2000)~~(Carvalhais et al., 2014). There are large gaps of knowledge in the current state of the pan-arctic terrestrial C budget that is mainly due to the vast uncertainties in C allocation, C stocks and transit times (Bloom et al., 2016; Carvalhais et al., 2014; Friend et al., 2014).~~ The ~~transit time~~TT ~~(TT) here addressed as TT = C_{stock} / NPP~~ 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 ~~of a C stock~~ and its outgoing flux, here addressed as ~~, this is~~ TT = C_{stock} / NPP at steady state~~a variable that represents the time it takes for a particle of C to persist in a specific C stock~~. 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 ~~terminology~~ and theoretical assumptions ~~(steady state for example)~~ for its calculation. -Terms such as residence time (Bloom et al., 2016; Friend et al., 2014), turnover time (Carvalhais et al., 2016), ~~residence time (Friend et al., 2014),~~ and turnover rate (Thurner et al., 2016; TT = 1/turnover rate~~s~~) are used in the literature to represent the concept of TT (Sierra et al. 2017). Studies have focused more on t~~T~~he spatial variability with climate ~~has been more studied for~~of ecosystem productivity rather than ~~for~~ C transit times ~~dynamics~~ (Friend et al., 2014; Nishina et al., 2015; Thurner et al., 2016; Thurner et al., 2017).

~~There are large gaps of knowledge in the current state of the pan-arctic terrestrial C budget that is mainly due to the vast uncertainties in C allocation, C stocks and transit times at global scales (Bloom et al., 2016; Carvalhais et al., 2014; Friend~~

et al., 2014). For example, the interactions between the land C cycling and climate is primarily determined by C turnover and productivity (Carvalhais et al., 2014). However, the spatial variability with climate has been more studied for NPP than for C turnover processes (Thurner et al., 2016; Thurner et al., 2017). Global scale vegetation model development has intensely concentrated on ecosystem productivity whereas the dynamics of C turnover, here addressed as transit time (Ceballos-Núñez et al., 2017; Sierra et al., 2017), has been less studied (Friend et al., 2014). 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 basic components of the C cycle, the understanding of C cycleits feedbacks to climate change remains highly uncertain (Hobbie et al., 2000; Koven et al., 2015).

There are currentlyare these days important efforts to incorporatinge 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 high-quality data (Grøndahl et al., 2008; Lafleur et al., 2012). Given the lack of continuous, spatially distributed ground-based scalein 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, an increasing number of regional to global datasetsproducts generated from in situ networks and/or satellite observations haves improved our understanding of the terrestrial C dynamics, including global scale vegetation dynamics. T (REF). 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) andor the creation of a harmonized global soil databases (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. important limitations in data availability still remain in difficult measured drivers such as soil-plant turnover and soil respiration responses to climate, especially for a highly influential and understudied region like the Arctic.

Conveniently, Global Vegetation Models (GVM) have been developed to determine global terrestrial C cycles and represent vegetation and ecosystem processes including the structural (i.e. growth, competition, and turnover) and biogeochemical (i.e. water, carbon, and nutrients cycling) responses to climate variability (Clark et al., 2011; Fisher et al., 2014; Friend and White, 2000; Ito and Inatomi, 2012; Pavlick et al., 2013; Sitch et al., 2003; Smith et al., 2001; Woodward et al., 1995). The advantage of using process-based models to measurecharacterise C dynamics is that processes which drive ecosystem-atmosphere interactions can be simulated and reconstructed when data is scarce. moderatelyHowever, Beyond direct ground observations, C cycle modelling in process-based global vegetation models (GVMs) typically relies on pre-arranged parameters retrieved from literature, prescribed plant-functional-type (PFT) andor spin-up processes until the C stocks (biomass and SOC) reach their steady state (Clark et al., 2011; Friend and White, 2000; Ito and Inatomi, 2012; Pavlick et al., 2013; Sitch et al., 2003; Smith et al., 2001; Woodward et al., 1995). Further, inherent differences of model structure contribute more significantly to GVM uncertainties (Exbrayat et al., 2018; Nishina et al., 2014), than from differences in climate projections (Ahlström 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 (Bloom et al., 2016).~~ 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~~ly, many~~ studies have used simulations from the first phase of the Inter-Sectoral Impact Model Intercomparison Project (~~ISIMIP~~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 ~~the~~ TTs in ~~land ecosystem~~GVMs are a ~~main~~key uncertain feature of the global C cycle simulation ~~and. For example, the interactions between the land C cycling and climate is primarily determined by C turnover and productivity (Carvalhais et al., 2014). However, the spatial variability with climate has been more studied for NPP than for C turnover processes (Thurner et al., 2017). Global scale vegetation model development has intensely concentrated on ecosystem productivity whereas the dynamics of C turnover, here addressed as transit time (Ceballos-Núñez et al., 2017; Sierra et al., 2017), has been less studied (Friend et al., 2014). 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 remain unquantified. Without an appropriate understanding of current state of basic components of the C cycle, the understanding of C cycle feedbacks to climate change remains highly uncertain (Hobbie et al., 2000; Koven et al., 2015a).~~

~~Over the past decade, an increasing number of datasets has improved our understanding of the terrestrial C dynamics, including global scale vegetation dynamics. These range from machine-learning based upscaling of FLUXNET data, remotely-sensed biomass products or the creation of a harmonized soil databases.~~ d. ~~Despite the increasing volume of C cycling related products, they~~ GVMs ~~do not provide estimates of the internal dynamics which regulate the C cycle and its response to changes.~~ Further, GVMs tend not to report uncertainties in their estimates of stocks and fluxes, which weakens their analytical value.

An approach to ~~circumvent~~ address these issues is to integrate models and data more formally. Data assimilation quantifies how ~~to~~ model parameters can be adjusted to estimate ~~these~~ C ~~dynamics~~ stocks and fluxes ~~in agreement~~consistent with multiple observations (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 ~~Integrating~~assimilating ~~available~~co-located climatic, ecological and biogeochemical data from ~~field experiments, flux towers and~~ remote sensing observa~~r~~tions ~~using data assimilation system (e.g. ensemble Kalman filter, Bayesian technique) can~~ at a specific grid scale across landscapes and regions ~~reduce~~we can map ~~uncertainties in~~ parameter estimation and uncertainties ~~and improve calculation of C cycle dynamics (Fox et al., 2009; Luo et al., 2009; Williams et al., 2005)~~.

Here, we use the CARbon DAta MOdel framework (CARDAMOM) (Bloom et al., 2016; Bloom and Williams, 2015; Smallman et al., 2017) to retrieve the pan-Arctic terrestrial carbon cycle at 1º resolution for the 2000-2015 period in agreement with gridded observations of LAI, biomass and SOC stocks.

~~In this paper, w~~We compare analyses of C dynamics of Arctic tundra and taiga ~~for the period 2000-2015 for~~ 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 sites included in the FLUXNET2015 dataset ~~(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)~~, and (c) 6 extensively used GVMs ~~(HYBRID, JeDi, JULES, LPJmL, SDGVM and VISIT)~~ from the ~~ISI-MIP~~ISI-MIP2a comparison project ~~-~~ (Akihiko et al., 2017)~~(Warszawski et al., 2014)~~. Our objectives are to (1) ~~to~~ present and evaluate the ~~retrievals~~ analyses and uncertainties of the current state of the pan-Arctic terrestrial C cycling using a model-data fusion system, (2) ~~to~~ quantify the degree of agreement between ~~our better constrained~~the CARDAMOM product

with ~~several~~ local to global scale sources of available data ~~available, from local to global scale~~, and (3) ~~to~~ use CARDAMOM as a benchmarking tool for the ~~ISIMIP~~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~~is~~ constrained by observations ~~and the ecological and dynamic constraints make sure that processes are realistic and ecologically viable~~. ~~-~~ Finally, we suggest future work to be done in the context of advancing pan-Arctic C cycling modelling ~~at the global scale~~.

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~~34~~°N - 80°N and longitudes 180°W - 180°E, and at a spatial resolution of 1º x 1º. This area of study totals ~~18.7~~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 ~~(hereafter referred as tundra and taiga, respectively)~~. A complete description ~~about~~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 ~~(RECCAP)~~ Activity reported by McGuire et al. (2012). The extensive grasslands ~~However, the tundra region extents into taiga regions~~ without presence of trees in some areas such as the ~~extensive grasslands~~ in South Russia, ~~and~~ Mongolia and Kazakhstan were neglected to focus on higher latitudes. ~~(Figure S1).~~ This classification of tundra and taiga totals ~~9.0~~8.1 and 9.~~7~~ 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 ~~. CARDAMOM is centred around 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

~~to~~ DALEC2 ecosytem model simulates 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 ~~(Bloom and Williams, 2015; Williams et al., 2005)~~. ~~-~~ DALEC2 includes 17 parameters controlling the processes of plant phenology, photosynthesis, 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 S1~~Table S2). DALEC2 simulates canopy-level GPP ~~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. Plant C decays into litter and soil stocks where microbial decomposition generates heterotrophic respiration (R_h). ~~In each~~Turnover of ~~plant,~~ litter and soil stocks ~~turnover rate~~is simulated using temperature dependent first-order kinetics. For practical purpo~~u~~ses 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. ~~represented as the ratio between each C stock and NPP.~~

2.2.2 Data-assimilation system

~~**2.2.2 Data-assimilation system**~~

The ~~condensed~~intermediate complexity of the DALEC2 model compared to ~~other ecosystem process based~~typical GVMs ~~models~~ facilitates computationally intense data-assimilation~~s~~ to optimize the ~~uncertainty in the~~ initial stock conditions and the ~~20~~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) for the 2000-2015 period. ~~CARDAMOM is driven by climate data from the European Centre for Medium-Range Weather Forecast (ECMWF) Reanalysis interim (ERA-nterim) dataset (Dee et al., 2011) for the 2000-2015 period.~~ A Bayesian Metropolis-Hastings Markov chain Monte Carlo (MHMCMC) algorithm is used to retrieve the posterior distribution~~s~~ of ~~17~~ the process parameters according to observational constraints and Ecological and Dynamic constraints (EDCs; Bloom and Williams, 2015). EDCs ensure that DALEC2 ~~simulates~~ simulations of the terrestrial carbon cycle ~~in agreement with ecological theory~~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~~XXX(insert number here)~~ 1-km resolution MODIS LAI data monthly within each 1°x 1° ~~degree~~ pixel. ~~-~~We aggregated b~~XXX b~~iomass data at ~~X m~~0.5° resolution from Carvalhais et al. (2014) to 1° resolution. ~~degree biomass estimates from Carvalhais et al. (2014).~~ These ~~which~~ are based on remotely-sensed forest biomass ~~-~~and upscaled GPP based on data driven estimates (Jung et al., 2011) covering the pan-Arctic domain. We used the NCSCD spatial explicit product (Hugelius et al., 2013a; Hugelius et al., 2013b) ~~-~~which was generated from 1778 ~~XXX~~ soil sample locations interpolated to a 1º 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 (E~~e~~quation 3).

~~Observational constraints include monthly time series of Leaf Area Index (LAI) from the MOD15A2 product (Myneni et al., 2002), estimates of vegetation biomass and soil organic carbon content (Table S3). In this paper, there are two main differences with the global approach described in Bloom et al. (2016). First, the biomass constraints used by Bloom et al. (2016) only cover tropical regions. Instead, here we use global biomass estimates from Carvalhais et al. (2014) which are based 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. Second, we constrained the storage of soil organic carbon (SOC) in the 0-1 m topsoil from the Circum-Arctic permafrost region (Brown et al., 1997) using the NCSCD spatial explicit product (Hugelius et al., 2013a; Hugelius et al., 2013b) instead of the Harmonized World Soil Database (HWSD) (FAO/IIASA/ISRIC/ISSCAS/JRC, 2012). While we report results using this configuration~~biomass and NCSCD SOC, ~~we provide estimates of the sensitivity of retrievals to the choice of SOC database and the inclusion or omission of C_{veg} prior (global biomass product) in the Supplement.~~

We apply the setup described above to ~~3433~~ 33084 1º x 1º pixels (~~1815~~ 16896 in tundra; 161988 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 pixel~~In each 1° x 1° pixel~~s~~, ~~the~~we applied the MHMCMC algorithm to determine the probability distribution of the~~an~~ optimal parameter set and initial conditions (x_i; Table S2) given observational ~~datasets~~constraints (O_i; LAI, SOC and biomass, Table S3) using the same Bayesian inference approach described in Bloom et al. (2016):

$$p(x_i|O_i) \propto p(x_i) \cdot p(O_i|x_i) \quad (1)$$

First, in the expression 1,

~~where~~ $p(x_i)$ represents the prior probability distribution of each DALEC2 parameter (x_i) and is expressed as:

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

~~with~~where $p_{EDC}(x_i)$ is ~~representing~~ the prior parameter probability according to the EDCs included in Table S2~~1~~ and described in Bloom and Williams (2015). In addition, ~~Bloom et al. (2016).~~ p~~P~~rior values for two parameters and their uncertainties (~~fraction of GPP respired~~ [f_{auto}] ~~and~~ canopy efficiency[C_{eff}] and fraction of GPP respired [f_{auto}]~~, both derived by Bloom et al. (2016)~~) ~~present~~are imposed with~~are set~~ a log-normal distribution following Bloom et al. (2016) to be consistent with ~~f_{auto} stated by DeLucia et al. (2007) and~~ the global GPP ~~values~~range estimated in Beer et al. (2010) and f_{auto} ranges specified by DeLucia et al. (2007) ~~DeLucia et al. (2007)~~ respectively. ~~according to Bloom et al. (2016).~~

Second, $p(O|x_i)$ from expression 1 represents the~~The~~ likelihood of x_i with respect to O_i, ~~$p(O|x_i)$,~~ and it is calculated ~~according to~~based on the ability of DALEC2 to reproduce (1) biomass (Carvalhais et al., 2014), (2) SOC (Hugelius et al., 2013a, Hugelius et al., 2013b~~b~~), and (3) MODIS LAI (Myneni et al., 2002). ~~We~~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:

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

For each 1° x 1° pixel we run three~~The~~ MHMCMC chains with $10^{~~000,000~~7}$ accepted simulations each ~~is performed three times~~ until convergence of at least two chains. ~~ConsistingUsing of~~We use 500 parameter sets sampled from the ~~2~~ndsecond half of each ~~accepted~~ chain to describe the posterior distribution of parameter sets.~~ we re-run the model 1500 times and derive the posterior distribution of parameter sets which allow .~~ We produc~~ing~~e ~~corresponding density functions~~confidence intervals of terrestrial C fluxes and stocks from the selected parameter sets~~with attached confidence intervals~~. ~~and a total of 1500 parameter sets is sampled from the posterior distribution of parameter sets which allow producing corresponding density function of all C fluxes and stocks.~~This process is repeated to make sure independent chain converge to the same posterior distribution.~~ In the following we report highest confidence results (median; P50) and the uncertainty represented by the 90% confidence interval (5th percentile to 95th ~~percencile~~percentile, $\binom{P95}{P05}$). ~~We aggregated the different C stocks into photosynthetic (C_{photo}; leaf and labile), vegetation (C_{veg}; leaf, labile, wood and roots) and soil (C_{dom}; litter and SOM) C stocksC stocks into photosynthetic (C_{photo}), vegetation (C_{veg}) and soil (C_{dom}) C stocks. 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 each C stock and NPP. Transit time is an important diagnostic metric that is independent of model internal structure, and good candidate to compare among models.~~

2.3 Model evaluation against independent in situ ~~at local~~ and pan-Arctic ~~scales~~datasets

At the pan-Arctic scale, we compared ~~our~~ CARDAMOM GPP with FLUXCOM dataset from Jung et al. (2017). ~~FLUXCOM is based on a machine learning approach to upscale local GPP data from eddy covariance towers and provide gridded estimates of monthly fluxes at 0.5º x 0.5º resolution. FLUXCOM has been used in previous studies as a benchmark for simulated GPP (Exbrayat et al., 2018; Slevin 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, RMSE, and bias) between CARDAMOM and the two independent datasets. The mapping includes stipples representing locations where the independent datasets are within ~~the~~ CARDAMOM's 90% confidence interval.

At a local scale, ~~s,~~ we compare CARDAMOM NEE and its partitioned components GPP and R_{eco} estimates against monthly aggregated values from the FLUXNET2015 sites.~~dataset.~~ 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 S2~~Table S4~~3~~). 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). ~~For readability purposes, in~~In this data-model comparison we included the median (P50) ± the ~~50~~90~~5~~% confidence interval (percentile ~~25~~th-5th to ~~75~~th95th; $\left(\frac{P_{75}^{95}}{P_{25}^{05}}\right)$) 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.

2.4 Benchmark of Global Vegetation Models from ISI-MIP2a

~~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.~~

~~2.4 Benchmark of Global Vegetation Models from ISI-MIP2a~~

We ~~examined~~ compared CARDAMOM ~~the~~ analyses of pan-Arctic ~~annual changes in~~ net primary production (NPP), vegetation biomass carbon stocks (C_{veg}) and vegetation transit times (TT_{veg}~~; $TT_{veg} = C_{veg}/NPP$~~) ~~using CARDAMOM as benchmark tool for~~ against six participating GVMs in the ISI-MIP2a comparison project (Akihiko et al., 2017) ~~(Warszawski et al., 2014)~~. 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). ~~HYBRID4 (Friend and White, 2000), JeDi (Pavlick et al., 2013), JULES (Clark et al., 2011), LPJmLl (Sitch et al., 2003), SDGVM (Woodward et al., 1995), and VISIT (Ito and Inatomi, 2012).~~ The specific properties and degree of

complexity of each ISI--MIP2a model are summarized in ~~Table S3~~Table S5~~4, and more detailed information can be found in Friend et al. (2014) and Thurner et al. (2017)~~.

~~In this study, each model simulation has been conducted under multiple five different General Circulation Models (GCM). Here we included HadGEM2-ES (Collins et al., 2011), IPSLCM5A-LR (Dufresne et al., 2013), MIROC-ESM-CHEM (Watanabe et al., 2011), GFDL-ESM2M (Dunne et al., 2012), and NorESM1-M (Bentsen et al., 2013) GCMs from the fifth phase of the Coupled Model Intercomparison Project (CMIP5) experiment (Arora et al., 2013; Taylor et al., 2012), which is temperature and precipitation bias corrected following Hempel et al. (2013). The comparisons with CARDAMOM for each GVM included the mean ensemble of all GCM forcings.~~ The comparisons have been performed under the same spatial resolution as the CARDAMOM spatial resolution (1° x 1°) for the 2000-2010~~04~~ period ~~(1° x 1° resolution)~~. Also, the chosen GVMs from the ISI-MIP2a phase ~~provides historical reanalysis~~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.

3 Results

~~The CARDAMOM framework quantified C fluxes, C stocks and transit times in plant, litter and soil carbon stocks over the pan-Arctic land region (including the tundra and taiga partitioning) for the period 2000-2015 (Table 1). The system is likely a small C sink, although the 90% confidence intervals remain large (and so the region could be a source of C) (Table 1). Our analysis indicates that SOC is successfully assimilated (1:1 agreement), biomass stocks are 28% lower than earth observation mapping (Figure 1), and that GPP is 50% lower than FLUXCOM (Figure 2). We also suggest that R_h is lower in the tundra and higher in the taiga than upscaled estimates (Figure 2). We note that independent tests at EC locations suggests that CARDAMOM's GPP is 30% biased high (Figure 3). This mismatch is important in the context of FLUXCOM, as noted (Figure 2). We benchmarked six GVMs to compare to not only their spatial variability across the pan-Arctic, tundra and taiga region (Figure 4), but also the degree of agreement between their mean model (6 GVMs – 5 GCMs) ensemble within the 90% confidence interval of our framework (Figure 5, Table 3). We finally found that turnover time of vegetation C is weakly simulated in GVMs and is a major component of error in their forecasts (Figure 6).~~

3.1 Pan-Arctic retrievals of C cycle

~~The CARDAMOM framework quantified C fluxes, C stocks and transit times in plant, litter and soil carbon stocks over the pan-Arctic land region (including the tundra and taiga partitioning) for the period 2000-2015 (Table 1). The system is likely a small C sink, although the 90% confidence intervals remain large (and so the region could be a source of C) (Table 1).~~Overall, we found that the pan-Arctic region (Figure 1 and Table 1) acted as a ~~consistent~~ small sink of C (area-weighted P50) over the 2000-2015 period with an average of -6~~55~~7.1~~48~~ $\left(^{11258.5}_{-286.7}{}^{1159.91}_{7960.75}\right)$ g C m^{-2} yr^{-1}, P50 $\left(^{P95}_{P05}\right)$, although the 90% confidence intervals remain large (and so the region could be a source of C). ~~However,~~ Ttundra regions NEE was estimated at ~~presented a weaker sink compared to taiga regions, this is~~ -~~13~~114.9~~6.0~~ $\left(^{114067.81116.17}_{-1578.6163.45}\right)$ ~~and~~g C m^{-2} yr^{-1}, a weaker sink compared to taiga regions, -~~104~~110.4~~.1~~ $\left(^{11965215.820}_{-37875.74}\right)$ ~~g C m^{-2} yr^{-1}~~g C m^{-2} yr^{-1}~~respectively, but also lower uncertainties with nearly 1265.5 g C m^{-2} yr^{-1} between P05 and P95 in tundra~~. ~~In general, t~~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 ~~263~~283~~90.3~~43~~ $\left(^{37641000.786}_{1779}{}^{}_{16.494}\right)$ g C m^{-2} yr^{-1}. Carbon use efficiency (NPP/GPP) averages 0.51$\left(^{0.55}_{0.46}\right)$, and marginally varied across tundra 0.51~~0~~$\left(^{0.54}_{0.46}\right)$ and taiga 0.52$\left(^{0.56}_{0.46}\right)$.

Despite these apparent small variations, tundra photosynthesized and respired (respectively ~~313~~3275~~.2~~.2~~70~~($^{~~450.1~~452.~~1~~463.3}_{~~226.9~~225.~~0~~236.8}$) and 3~~00.03~~10.0($^{~~15~~1545.~~6~~36.~~80~~}_{~~1198.~~39~~124.3}$) g C m^{-2} yr^{-1}) approximately half as much as the Taiga region (7~~36.5~~59.~~68~~($^{~~967.89~~940.9}_{~~564.9~~584.~~01~~}$) and ~~618.98~~635.~~23~~($^{~~2114.30~~06.3}_{~~276.6~~285.~~23~~}$) g C m^{-2} yr^{-1}).

~~The total size of the pan-Arctic vegetation C stock (C_{veg}) averaged 1.4($\frac{6.0}{0.5}$) kg C m^{-2}, an estimate 94% smaller than the soil C stock (C_{dom}), 24.4($\frac{47.6}{10.3}$) kg C m^{-2}.~~ The total size of the pan-Arctic soil C stock (C_{dom}) averaged 24.4($^{47.5}_{10.3}$) kg C m^{-2}, 16-fold greater than the vegetation C stock (C_{veg}), 1.5($^{5.8}_{0.5}$) 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($^{5.6}_{0.4}$) kg C m^{-2}) compared to 7% (12% in tundra and 10% in taiga) to the photosynthetic stock (leaves and labile; 0.1($^{0.2}_{0.1}$) kg C m^{-2}). On average, the total ecosystem C stock is 26.3($^{51.0}_{11.8}$) kg C m^{-2} in the pan-Arctic region, with slightly lower stocks in tundra (24.6($^{50.6}_{10.8}$) kg C m^{-2}) than taiga (27.7($^{51.2}_{12.7}$) 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($^{5.1}_{0.8}$) kg C m^{-2}) than tundra (0.8($^{6.8}_{0.3}$) 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 total size of the pan-Arctic soil C stock (C_{dom}) averaged 23.7($\frac{47.6}{10.3}$) kg C m^{-2}, 16-fold greater than the vegetation C stock (C_{veg}), 1.5($\frac{5.6}{0.5}$) 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 (89% in tundra and 91% in taiga) is allocated to the structural stocks (wood and roots; 1.4($\frac{5.4}{0.4}$) kg C m^{-2}) compared to 7% (11% in tundra and 9% in taiga) to the photosynthetic stock (leaves and labile; 0.1($\frac{0.2}{0.1}$) kg C m^{-2}). On average, the total ecosystem C stock is 25.6($\frac{49.5}{11.5}$) kg C m^{-2} in the pan-Arctic region, with slightly lower stocks in tundra (25.5($\frac{52.7}{11.2}$) kg C m^{-2}) than taiga (27.7($\frac{51.3}{12.7}$) kg C m^{-2}). In general, the taiga region holds on average ~50 % more photosynthetic tissues, ~133 % more structural tissue and ~5 % more soil C stocks, than tundra. In other words, taiga holds ~9 % more total C than tundra. The greater living stock of C in taiga (2.1($\frac{5.1}{0.8}$) kg C m^{-2}) than tundra (0.9($\frac{7.4}{0.3}$) 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 3% larger than R_h in tundra, but 23% 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.~~

~~17-foldThe soil C stock (fresh litter and soil organic matter, SOM) is clearly dominated by C_{som}, accounting for the 998.8%, which also dominates the terrestrial C stock in the pan-Arctic. Among the living C stocks, 91% of the C is allocated to the structural stocks (wood and roots; 1.3 kg C m^{-2}) compared to 9% to the photosynthetic stock (leaves and labile; 0.1 kg C m^{-2}). On average, the total ecosystem carbon density in the pan-Arctic region is 26.2 kg C m^{-2}, with slightly lower stocks in tundra (24.6 kg C m^{-2}) than taiga (28.0 kg C m^{-2}). In general, the taiga region accumulated holds on average ~44 % more photosynthetic tissues, ~55 % more structural tissue and ~10 % more C than tundra region in photosynthetic, structural and soil C stocks, respectivelythan tundra. In other words, taiga accumulates holds ~12 % more total C than tundra. The greater living stock of C in taiga than tundra 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 tundra this fraction is 57%. R_a is 3% larger than R_h in tundra, but 23% 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 1.43($^{2.21}_{0.98}$) years in leaves and labile plant tissue (TT_{photo}), 4.3 45 ($^{15.473}_{1.67}$) years in stems and roots (TT_{veg}), and 129118.3520.5($^{981.1798.3822.6}_{10.99.78}$) years in litter and SOM (TT_{dom}). The total C transit time (TT_{tot}) (142131.4133.1.5($^{1089.4997.41013.6}_{11.95}$) years) is clearly dominated by the soil C stock, highlighting the very long periods of times that C particles can persists in Arctic soils. Interestingly, CARDAMOM estimates longer TT_{veg} in Taiga compared to tundra with 5.9 ($\frac{8.6}{3.2}$) years vs 3.5 ($\frac{13.0}{1.4}$) years (for tundra), but also shorter transit times in TT_{photo} and TT_{dom} with 1.6 ($\frac{2.0}{1.3}$) and 157.3 ($\frac{329.9}{63.1}$) years respectively compared to 1.1 ($\frac{1.6}{0.7}$) years and 97.9 ($\frac{660.2}{8.9}$) years in tundra. CARDAMOM calculated 62% longer TT_{dom} in tundra compared to taiga, likely linked to lower temperatures, but also significantly greatebutr uncertainties likely echoiare largeng from the caviates found in C_{dom} and R_h due to the limitations of data constraints. Longer transit times are likely affected at some degree by low temperatures, wet soils, and thus slower decomposition processes.The turnover rates estimated by CARDAMOM in tundra regions suggest a tendency towards longer transit times in the photosynthetic and soil C stocks, likely affected at some degree by low temperatures, wet soils, and thus slower decomposition processes.

Additionally, we performed a sensitivity analysis using four different experiments to assess the impact of using different soil C datasets (NCSCD and HWSD) and with without biomass constraints (Table S4). These results show that the differences introduced by biomass constraints were significantly larger compared to the differences introduced by different soil C databases. Also, the difference between soil C databases resulted in shifts of 22% up to 20% in the total mean carbon stock (C_{tot}) and transit time (TT_{tot}). Our results showed significant ($p < 0.01$) differences whether biomass constraint was used in the framework or not. In simulations with biomass constraints, C_{photo}, C_{veg}, and C_{dom} stocks decreased the C stored ~ 78%, 62% and 10% respectively. Moreover, introducing biomass shortens retrievals the TT_{photo} and TT_{veg} by 87% and 60%, but also their uncertainties (97.3% and 66% respectively). While using biomass constraints only has a negligible effect on productivity, increased respiration fluxes induce weaker C sink strengths.

3.2 Data assimilation and evaluation: from global to local scaleand uncertainty reducition

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 = 0.957 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- (Figure 3) which propagates through the C cycle. CARDAMOM balances uncertainty in data products and, the models (ACM photosynthesis model and DALEC2), to generate a weighted analysies, 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 observationsOur analysis indicates that the assimilated data by the CARDAMOM framework retrieved the terrestrial C cycleare in good good agreement with priors of of SOC and, in a lesser degree, and biomass and LAI (Figure 2). The agreement for the SOC dataset by Hugelius et al. (2013a) is a 1:1 relationship ($R^2 = 1.0$; RMSE = 0.97 kg C m^{-2}) (Figure 2). while for The biomass product from Carvalhais et al. (2014), although well correlated ($R^2 = 0.97$; RMSE = 0.46 kg C m^{-2}), led to a tendency towards ~28% higher accumulation of C in the vegetation C stock (leaf, labile, wood and roots) C stocks compared to the assimilated biomass in CARDAMOM is close to 1:1 ($R^2 = 0.97$; RMSE = 0.46 kg C m^{-2}) (Figure 1). MODIS LAI presented ~28% higher leaf area estimates compared to our data assimilation system. Using Carvalhais et al. (2014)'s biomass product led to a tendency towards larger accumulation (~28%) of C in the structural (wood and roots) stocks (Table S4) compared to the assimilated CARDAMOM's biomass. These results suggest that thee understanding of ecological dynamics implemented in CARDAMOM cannot fully resolve Carvalhais et al. (2014) biomass and Myneni et al. (2002) LAIbiomass in agreement with other products of LAI andsuch as SOC.

~~However, the data assimilations in CARDAMOM are not far out since the 90% confidence interval (CI) uncertianties includes the 1:1 line (Figure 2).~~

The degree to which posterior distributions were constrained from the prior distributions in each of the 1~~8~~7 model parameters and ~~5~~6 initial stock sizes (Table S~~1~~2) 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 ~~The parameter uncertainty reduction estimated from most to least constrained, C stocks, C allocation, plant phenology and C turnover related parameters, as average 74%, 70%, 63% and 38% respectively. In the one hand,~~ C stocks~~- such as~~ (C_{foliar}, C_{wood}, ~~-and~~ C_{labile} and C_{som})~~,~~ as well as parameters such as allocation to foliage (~~allocation to foliage~~f_{fol}) ~~,~~and lifespan (~~leafspan~~L) were ~~highly~~considerably ~~constrained~~reduced (>80% uncertainty reduction compared to priors) ~~(>85% uncertainty reduction) because~~most likely controlled by the information on LAI, biomass and SOC ~~of the usage of the MODIS LAI~~ constraints. ~~constraint, but also C_{som} (>80%) due to the NCSCD SOC constraint. On the other hand,~~Contrarily, parameters that have not been ~~not contrained~~regulated ~~at all~~in any way in the MHMCMC algorithm, i.e. ~~such as~~ turnover processes such as litter mineralization (MR_{litter}), roots turnover (TOR_{roots}) ~~and~~, wood turnover (TOR_{wood}), decomposition rates (D_{rate}) ~~or~~and initial C stock such as litter (C_{litter}) were found poorly constrained (<20% uncertainty reduction)~~(<30% uncertaintiy reduction)~~. Overall, the uncertainty reduction classified by processes and ranked from most to least constrained estimated a 71%~~, 67%, 59% and 31%~~ reduction ~~respectively~~ 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

~~On the other hand, we~~We compared our estimates of GPP and R_h with independent datasets to evaluate the model performance (~~Figure 2~~Figure 3). We found GPP to be well correlated (R^2 = 0.81~~01~~; RMSE = 0.43 kg C m^{-2}), but biased ~~significantly~~ lower (~53~~1~~%) compared to Jung et al. (2017)'s GPP estimates. There are in general very few pixels where ~~The areas with larger agreement, this is where~~ FLUXCOM product falls within CARDAMOM's 90% confidence interval. Additionally, ~~, are~~is in taiga regions ~~, rather than in tundra (Figure 2~~Figure 3~~). Also, we note evidence of a higher spatial variability in CARDAMOM (Figure 2). We found~~ the R_h product from Hashimoto et al. (2015) is less consistent with our estimates (R^2 = 0.40~~38~~; RMSE = 0.09 kg C m^{-2}), ~~with~~ 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 ~~only~~ within the 90% confidence interval of CARDAMOM in most of the pan-Arctic region~~in Central Northen Canada and Eurasia as well as the grasslands in South Russia and Mongolia~~ due to the fact that the R_h uncertainties are significant (Figure 3~~).~~). ~~TMoreover, the spatial variability of R_h is considerably smaller in Hashimoto et al. (2015)'s R_h compared to our CARDAMOM estimates, for example in central Eurasia.~~ This findings confirms~~s~~ the uncetartanties 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.

~~In order to get also a~~For comparison with direct ground observations from the FLUXNET2015 dataset, we report here monthly aggregated P50 ± P05~~25~~ ~~75~~ 95 estimates of NEE, GPP and R_{eco} to show timing and magnitudes, but also to diagnose whether CARDAMOM is in general agremment 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 3~~Figure 4; Table ~~2~~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% ~~lower~~higher than CARDAMOM GPP~~, as noted~~. ~~At s~~Some sites such as Hakasia, Samoylov, Poker Flat and Manitoba (NEE R^2 = ~~0.85, 0.87, 0.81, 0.7~~0.73; GPP R^2 = 0.92~~0.98, 0.98, 0.93, 0.93~~ and R_{eco} R^2 = 0.88~~0.98, 0.89, 0.89, 0.80 respectively~~) CARDAMOM ~~represent~~ 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 = ~~0.41, 0.58, 0.54, 0.65~~0.58; GPP R^2 = ~~0.87, 0.80, 0.84, 0.82~~0.67 and R_{eco} R^2=0.67~~0.76, 0.86, 0.86, 0.83 respectively~~). In general, CARDAMOM captured the beginning and the end of the growing season well (~~Figure 3~~Figure 4), although~~. However,~~ the assimilation system ~~showed~~ has~~ve important~~some bias due to (1) difference in timing (e.g. earlier shifts of peak of the growing season in Manitoba~~, and UCI-1998,~~ GPP and R_{eco} ~~or~~ and earlier end of the growing season in Poker Flat NEE) and (2) differences in flux magnitudes (such as in Hakasia ~~and~~ GPP and R_{eco} and Kobbefjord NEE).

3.~~3~~ 4 Benchmarking ISI-MIP2a models with CARDAMOM

~~Our analysis indicates that SOC is successfully assimilated (1:1 agreement), biomass stocks are 28% lower than earth observation mapping (Figure 1), and that GPP is 50% lower than FLUXCOM (Figure 2). We also suggest that R_h is lower in the tundra and higher in the taiga than upscaled estimates (Figure 2). We note that independent tests at EC locations suggests that~~ We used our highest confidence retrievals of NPP, C_{veg} and TT_{veg} (i.e. ~~median~~ retrievals including assimilated LAI, biomass and ~~soil organic~~ SOC ~~from NCSCD~~) to benchmark the performance of the GVMs from the ISI-MIP2a project~~ISI-MIP's NPP, C_{veg} and TT_{veg} estimates~~. 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). ~~In Table 3 we quantified the degree of agreement and bias per assessed variable, spatial domain, and forward model.~~ Overall, ~~ISIMIP~~ISI-MIP2a models are ~~more~~ in poor agreement with CARDAMOM~~'s~~ across the pan-Arctic. NPP estimates (~~for tundra and taiga respectively~~ RMSE = 0.~~3~~1 ~~0.3 and 0.3~~ kg C m^{-2} yr^{-1}; ~~Bias~~R^2= ~~0.1022 and 0.1 kg C m^{-2} yr^{-1}~~0.44) are in better agreement than C_{veg} (RMSE = ~~2.4 and 2.82~~1.8 kg C m^{-2}; ~~Bias~~R^2= ~~1.3 and 2.21~~0.~~37 kg C m^{-2}~~22) and TT_{veg} (RMSE = ~~15~~49.0~~1 and 6.1~~ years; ~~Bias~~R^2= ~~3.0 and 4.13~~1.~~52 years~~0.12). Moreover, the assessed ~~Also~~GVMs~~, the ISIMIP models consistently~~ estimated on average 8% lower NPP, 16% higher ~~36% and 53% higher NPP and~~ C_{veg} and ~~45~~22% ~~larger~~ longer~~median's mean NPP, C_{veg} and~~ TT_{veg} than CARDAMOM ~~(36%, 53% and 45% increase respectively)~~ across the entire pan-Arctic domain (Figure 5~~4~~ and ~~6~~5) on average, with very varied spatial patterns.

The ~~low~~poor 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. 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~~, so the dis; best over the taiga domain~~ rather than in tundra for TT_{veg}~~than in taiga region.~~. ~~–~~ 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.~~Interestingly, HYBRID overestimated CARDAMOM's NPP and C_{veg} more clearly than any other model by 70% and 75%, but only 36 % for TT_{veg}. This is a representative example of compensating errors that may reduce the apparent bias in the inner dynamics. In other word, HYBRID TT_{veg} is simply the result of a systematic overestimation of NPP and C_{veg}. On the other hand, Overall, JeEDiI and SDGVM are~~T ~~the models in closer agreement with CARDAMOM were DLEM (5% difference) and LPJ-GUESS (17%) ~~(Table 3; Figure 5~~4 and 5~~6)~~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}. ~~Finally, we apportioned the error contribution to TT_{veg} by applying an attribution analysis in order to identify the origin of the bias from ISI-MIP2a models, this is NPP or C_{veg}. Interestingly, The separation between non-forested (tundra) and forested (taiga) areas also exposed some similarities and inconsistencies in model performance. For example, the ISIMIP models successfully assimilate the larger productivity, biomass and longer vegetation transit times in taiga compared to tundra (Figure 4). However, as average, the ISIMIP models are on average 56% more productive than CARDAMOM in tundra (0.16 vs 0.36 kg C m^{-2} yr^{-1} respectively), but only a 19% in taiga (0.39 vs 0.48 kg C m^{-2} yr^{-1}). Also, the tendency for C_{veg} and TT_{veg} coincide, estimating more biomass and transit times in tundra (55% and 46% respectively) than in taiga (51% and 44% respectively). In general, the overall agreement is closer for tundra than for taiga (Table 3).~~

~~we found that a joint split between NPP and C_{veg} for TT_{veg} erroris weakly simulated in GVMs and is a major component of error in their forecasts (Figure 7). For vegetation transit times, the pattern is more complex (Figures 4 and 5). The spatial variability patterns are similar to biomass, although larger (C_{veg} RMSE = 2.8 kg C m^{-2} and TT_{veg} RMSE = 4.3 kg C m^{-2}). We apportioned the error contribution to TT_{veg} by applying an attribution analysis. We used CARDAMOM to calculate two hypothetical TT_{veg} estimates ((i.e. EXPERIMENT A i.e. calculate TT_{veg} = with CARDAMOMISI-MIP2a C_{veg} from CARDAMOM, / Iand SI-MIP CARDAMOM NPP from ISIMIP models, and and EXPERIMENT B with TT_{veg} = ISI-MIPCARDAMOM C_{veg} from ISIMIP models and/ CARDAMOMISI-MIP2a NPP from CARDAMOM) and then to calculateidentify the largest difference with CARDAMOM's reference 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 6Figure 7). The distribution of the differences relative to CARDAMOM revealed that the highest error (i.e. the lower overlapped area, and by extension the largest contributor to TT_{veg} biases) comes from ISI-MIP2a C_{veg} NPP with only a 2969% agreement in the distribution while C_{veg} agrees 72% (Figure 6). 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 blend of NPP and C_{veg} uncertainty (R^2 of the correlation = 0.71). This finding supports Figure 6, which shows TT_{veg} error comes from both NPP and C_{veg}:~~

~~, while NPP agrees 76%.~~

4 Discussion

~~The CARDAMOM framework has been used to evaluate the terrestrial pan-Arctic C cycling 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). The sensitivity analysis suggests that the C cycle retrieved by CARDAMOM is more sensitive to differences introduced by biomass constraints compared to the differences introduced by different soil C databases (Table S4). This result indicates that data on C stocks with shorter turnover (biomass) have a higher impact in the overall pan-Arctic dynamics than stocks with slower turnover (SOC). 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) accumulated most of the C in the soil C stock (both fresh litter and SOM, but dominated by the latter with a contribution of about 97%); and 3) experienced longer transit times in leaves, labile, litter and SOM C stock located in tundra compared to taiga. In general, we found a good agreement between CARDAMOM and different sources of assimilated and independent data at both pan-Arctic and local scale. Finally, we used CARDAMOM as a benchmarking tool for six GVMs and we found that productivity processes are more in agreement with CARDAMOM than biomass, and thus biomass is the largest contributor to the bias influencing transit times. This finding suggest that there is a need to improve simulations of vegetation C stocks in Earth System models to get the inner dynamics right.~~

4.1. Pan-Arctic retrievals of C cycle

The CARDAMOM framework has been used to evaluate the terrestrial pan-Arctic C cycle~~ing~~ 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) ~~accumulated most of the C in the soil C stock (both fresh litter and SOM, but dominated by the latter with a contribution of about 97%); and (3)~~ experienced 62~~0~~% longer transit times in litter and SOM C stocks in tundra compared to taiga; and (3) the contribution of R_a and R_h to total ecosystem respiration was similar in tundra (51%, 49~~9~~%) but dominated by R_a in taiga (57%, 43%).

~~At broader scales~~ CARDAMOM ~~demonstrates that an assimilation system can certainly be a robust and useful tool to assess large scale C cycle dynamics in the Arctic - its~~ retrievals are ~~in agreemen~~consistent~~t~~ with ~~several~~ 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~~134~~ Tg C yr^{-1} over an area of ~~9~~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~~27~~, 16~~6~~7, 160~~58~~ and 154~~151~~ 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 ~~tundra was~~ 24.6$\binom{50.6}{10.8}$ kg C m^{-2} for tundra, ~~while~~and 27.7$\binom{51.2}{12.7}$ kg C m^{-2} in taiga (Figure 5; Table 1) for the same area. ~~In other words~~Thus, Carvalhais et al. (2014)'s C_{tot} product stored ~~only~~ 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($^{195.2}_{21.8}$) years in tundra and 51.2($^{109.3}_{22.1}$) years in taiga (Table 1) compared to the 64.4($^{259.8}_{25.7}$) years in tundra and 48.2($^{111.6}_{24.9}$) 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($^{18.2}_{1.9}$) years, ~~very close~~compared to Thurner et al. (2017)'s TT, 8.2($^{11.5}_{5.5}$) years using MODIS and 6.5($^{8.7}_{4.2}$) 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($^{0.5}_{0.3}$) 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.~~1~~2($^{5.0}_{1.1}$~~0.8~~)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 ~~different sources of~~ assimilated and independent data at ~~both~~ pan-Arctic ~~and local~~ scale. CARDAMOM retrievals of assimilated data are in ~~more~~good ~~good~~ agreement with the SOC ~~constraints than the, and biomass and LAI constraints~~ (~~Figure 1~~Figure 2). The simulation of ~~SOC~~ TT_{dom} ~~turnover~~ is ~~also~~ weakly constrained (Table 1) - our analysis adjusts ~~turnover~~ TT~~time~~ to match mapped stocks, hence the strong match of modelled to mapped SOC. So, independent data on TT_{dom} data ~~SOC transit time~~ (e.g. ^{14}C) ~~data~~ 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.

~~Further, the retrievals presented here exhibited a modest coherent performance of compared to~~There are clear biases in CARDAMOM analyses compared to ~~an~~ independent global GPP ~~product~~ (Jung et al., 2017) ~~, but weaker agreement with~~ and R_h products (Hashimoto et al., 2015) (~~Figure 2~~Figure 3). However, CARDAMOM resolves the spatial pattern in GPP effectively, while~~Indeed, the~~~~The~~ spatial ~~uncertainties~~ mismatch in ~~CARDAMOM~~ R_h estimates ~~are~~ is ~~substantially larger~~clear ~~than for CARDAMOM GPP~~ (Figure 3), ~~and that~~ echo~~es~~ ing the large uncertainty found in NEE (Figure 1, Table 1). One difference with Hashimoto et al. (2015)'s R_h model ~~between these two models~~ 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 ~~. A~~an important mismatch has been found ~~with regards GPP though.~~; CARDAMOM GPP is 50% lower than FLUXCOM, but 30% higher than FLUXNET2015 EC data.

~~Interestingly,~~

~~T~~tThe agreement between ~~earth observation data~~CARDAMOM analyses and EC data is ~~surprisingly~~ good given the ~~vast~~ scale difference. ~~However, a~~A direct point-to-grid cell comparison with local observations derived from the FLUXNET2015 dataset (~~Figure 3~~Figure 4, Table ~~2~~S6) is challenging and always difficult. CARDAMOM outputs ~~C stocks and fluxes in~~covers 1° x 1° grid cells, whereas local eddy covariance flux measurements are in the order of 1-10 hectares. 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 3~~Figure 4, Table ~~2~~S6). 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.

~~At broader scales CARDAMOM estimates reported here are in good agreement with C flux observations and estimates reported from McGuire et al. (2012) for the tundra domain, but also with the C stocks and transit times described by Carvalhais et al. (2014) in tundra and taiga, and the turnover rates (inverse of the transit times calculated here) supported by Thurner et al. (2017) in taiga regions. This performance compared to independently related C cycle components demonstrates that CARDAMOM is a robust and useful tool to assess large scale C cycle dynamics in the Arctic.~~

~~First, our NEE estimates reported in this study from Arctic tundra are inside the variability comparison of estimates among field observation, regional process based models, global process based models and inversion models reported by McGuire et al. 2012. McGuire et al. 2012 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's NEE estimate for the same period estimated -125 Tg C yr^{-1} over an area of 9 x 10^6 km^2. For example, 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 estimate of -202 Tg C yr^{-1}. The regional applications reported by McGuire et al., 2012 estimated NEE in -187 Tg C yr^{-1}, including LPJ-Guess-WHyMe (Wania et al., 2009a, b)(Smith et al., 2001), Orchidee (Koven et al., 2011), version 6 of the Terrestrial Ecosystem Model (TEM6) (McGuire et al., 2010) and the Terrestrial Carbon Flux (TCF) model (Kimball et al., 2009). These models also calculated GPP, NPP, R_a and R_h to be 350, 199, 151 and 182 g C m^{-2} y^{-1}, respectively. The DGVM applications included in McGuire et al., 2012 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 NEE in -93 Tg C yr^{-1} and GPP, NPP, R_a and R_h in 272, 162, 83 and 144 g C m^{-2} yr^{-1} respectively. For the same period CARDAMOM has estimated the gross C fluxes in 318, 161, 154 and 148 g C m^{-2} yr^{-1} respectively.~~

~~Second, Carvalhais et al. (2014) estimated total ecosystem carbon (C_{tot}) of 20.5$\left(\frac{52.5}{8.0}\right)$ kg C m^{-2} for tundra and 24.8$\left(\frac{58.0}{15.2}\right)$ kg C m^{-2} for taiga, while CARDAMOM tundra was 24.6$\left(\frac{33.0}{18.3}\right)$ kg C m^{-2}, while 28.0$\left(\frac{36.1}{21.5}\right)$ kg C m^{-2} in taiga (Figure 4; Table 1). Therefore, Carvalhais et al. (2014)'s C_{tot} product stored only 16.8 and 11.5% less carbon in tundra and taiga respectively than CARDAMOM. Also, we noted that the numbers are in line with Carvalhais et al. (2014) thanks to the use of the biomass constraint dataset, where C_{photo}, C_{veg}, and C_{dom} stocks decreased the C stored ~ 78%, 62% and 10% respectively (Table S4). Overall, CARDAMOM estimated 20.3 and 5.7% longer transit times for tundra and taiga respectively, with average values of 80.8$\left(\frac{195.2}{21.8}\right)$ years in tundra and 51.2$\left(\frac{109.3}{22.1}\right)$ years in taiga (CARDAMOM)(Table 1) compared to the 64.4$\left(\frac{259.8}{25.7}\right)$ years in tundra and 48.2$\left(\frac{111.6}{24.9}\right)$ years in taiga in Carvalhais et al. (2014). These numbers have been retrieved from the same biome~~

~~classification and the confidence intervals include the 90% confidence interval of the calculated spatial variability. Both datasets agree on the fact that high (cold) latitudes, first tundra, and second taiga have the longest transit times of the entire globe (Bloom et al., 2016; Carvalhais et al., 2014).~~

~~Third, a recent study from Thurner et al. (2017) assessed temperate and taiga-related transit times, NPP and C_{veg} taking into consideration the current poor process understanding of decomposition dynamics. Thurner et al 2017 used of 5-year average NPP for the time period 2000-2004, applying both MODIS (Running et al., 2004; Zhao et al., 2005) and BETHY/DLR (Tum et al., 2016) products. They use a biomass product (Thurner et al., 2014) at 0.5 degree while accounting for both forest and non-forest vegetation. Our estimates of TT_{veg} for the exact same period are very close to Thurner et al. (2017), where their $8.2\left(\frac{11.5}{5.5}\right)$ years using MODIS and $6.5\left(\frac{8.7}{4.2}\right)$ years using Bethy/DLR are close to the $5.2\left(\frac{17.8}{1.9}\right)$ years estimated by our data assimilation approach. 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 estimated NPP (0.35 (MODIS) and 0.45 kg C m^{-2} yr^{-1} (BETHY/DLR), without any uncertainty reported) is only 5% more productive as average than CARDAMOM's NPP estimate of $0.4\left(\frac{0.5}{0.3}\right)$ kg C m^{-2} yr^{-1}; while the biomass derived from Thurner et al. (2014), 3.0 (±1.1) kg C m^{-2}, is 23% lower than our estimates of 2.3 $\left(\frac{4.9}{1.2}\right)$ kg C m^{-2} for the same period and for the taiga region. Again, here, the presence of biomass as data constraint has helped to decrease TT_{veg} and its uncertainty significantly.~~ ~~This performance compared to independently related C cycle components demonstrates that CARDAMOM is a robust and useful tool to assess large scale C cycle dynamics in the Arctic.~~

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 low complexity and independent of the benchmarked models.

~~The CARDAMOM framework has proven capable of effectively simulating Arctic C cycling dynamics in the entire pan-Arctic region, but also to partition it in its two main biomes, i.e. tundra and taiga (Figure 2, 3, 4, Table 1, Discussion 4.1). CARDAMOM reconcile observational datasets as part of a representation of the terrestrial C cycle in agreement with ecological theory – it all comes down to CARDAMOM providing observationally-constrained estimates of C dynamics. A key reason to use CARDAMOM as a benchmark for GVM is that CARDAMOM produces parameter likelihoods for each pixel based on data, and it does not assume PFTs or steady states, hence it is much more strongly data constrained that a GVM. CARDAMOM takes data and the model to produce parameter maps, whereas GVMS take PFT maps of parameters to produce flux/stock outputs. For this reason, Explain why CARDAMOM can be succesfuly used as a benchmark tool~~

~~The CARDAMOM framework has proven capable of effectively simulating Arctic C cycling dynamics in the entire pan-Arctic region, but also to partition it in its two main biomes, i.e. tundra and taiga (Table 1, Figure 2, 3, 4, Table 1, Discussion 4.1).~~ ~~wWe used~~Using ~~our data assimilation system~~CARDAMOM as a benchmarking tool for six GVMs ~~and~~we ~~we~~

found major disagreements for mapping of NPP, C_{veg} and TT_{veg} across the Pan-Arctic for all models (Figure 6) 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~~ly~~ (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.~~that a combination of uncertainties from productivity (NPP) and biomass (C_{veg}) is the responsible for the bias governing C transit time (TT). This finding suggests also that there is a need to improve simulations of vegetation NPP, respiration losses and C stocks in Earth System models and to calculate more robust inner 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. ~~Recent studies used the same GVM inter-comparison models we used here raising strong arguments about the differences in model formulations and their impact on calculations, significant uncertainties and poor representation of C stocks dynamics at global scale (Exbrayat et al., 2018; Friend et al., 2014; Nishina et al., 2014; Nishina et al., 2015; Thurner et al., 2017). Here, we have present a considerably more data-constrained and data-integrated approach than traditional GVMs to calculate C dynamics. Consequently, we believe CARDAMOM is a good candidate to use as benchmark approach to pinpoint caveats of model performance. For example, Exbrayat et al. (2018) found that ISIMIPISI-MIP models are less in agreement for NPP in boreal latitudes compared to global CARDAMOM retrievals addressed in Bloom et al. (2016), lacking that did not includethe biomass constraints in boreal regions. In this study, we incorporated two new and critical layers of data constraints suitable for high latitudes (Carvalhais et al., 2014; Hugelius et al., 2013b) compared to the previous CARDAMOM ir version of CARDAMOM, and we found that NPP hasd one of the best agreements among the assessed variables (compared to C_{veg} and TT_{veg}), but also slightly better performance in tundra than in taiga (Bias = 9 and 1519 g C m^{-2} yr^{-1} respectively) (Figure 54 and 56; Table 3).~~

~~FurthermoreAlso, recent studies have emphasized the significance and need for of model comparison of variables such as turnover rates/ transit times (Ceballos-Núñez et al., 2017; Sierra et al., 2017), as diagnostic metrics independent of model internal structure. From a modelling point of view, it remains unclear why transit times differ (Figure 54 and 65) and whether NPP or C_{veg} dominates the biases. Based on Figure 4 and 5, biases in biomass C stocks likely dominate the error in turnover times. Consequently, We we used CARDAMOM to calculate the relative contribution of productivity and biomass to the transit times bias by applying a simple attribution analysis (Figure 6Figure 7). We concluded that The the largest bias to transit times are originated not byby a modest deficient understanding of one single component, but by the combination of both the productivity and biomass componenttogether. 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 turnover transit times/turnover dynamics, but we further suggest that GVM and ESM modellers need to focus on the productivity and vegetation C stocks dynamics calculations to improve inner C dynamics.~~

4.3 Outlook

Although CARDAMOM estimates for pan-Arctic C cycling are in ~~good~~ moderat~~iv~~ely good agreement with observations and data constraints, ~~; however,~~ 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~~a~~; Parazoo et al., 2018). Also, Koven et al. (2017) shown that soil thermal regimes are key to ~~getting~~ resolving the long-term vulnerability of soil C ~~right~~. Moreover, we have not characterized snow dynamics and the insulating effect of snow affecting respiratory losses across wintertime periods either (Essery, 2015; López-Blanco et al., 2018). Further, methane emissions, another important contributor to total C budget (Mastepanov et al., 2008; Mastepanov et al., 2012; Zona et al., 2016), was neglected from this modelling exercise~~,~~ ~~although~~ since it is not easy to model due to its three complex transport mechanisms (Kaiser et al., 2017; Walter et al., 2001).

However, our approach to use a low 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 complex models. There remains a strong argument to utilize low complexity models to evaluate the minimum level of detail required to represent ecosystem processes consistent with local observations. 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, i~~I~~n order to ~~decrease~~ reduce uncertainties ~~around~~ on the balance ~~of~~ between photosynthetic inputs and respiratory outputs, ~~future explorations on~~ we must devise low complexity model representations of SOC decomposition by microbial activity (Xenakis and Williams, 2014), nutrient interactions with carbon (Thomas and Williams, 2014), ~~plant traits relationships across pan-arctic~~ Arctic ~~regions (Reichstein et al., 2014; Sloan et al., 2013), the~~ mechanisims driving carbon use efficiency (Bradford and Crowther, 2013; Street et al., 2013), ~~and~~ and ~~the~~ drivers of gross flux coupling (López-Blanco et al., 2017)~~, or and the effect of~~. 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 ~~fine-scale~~ disturbances such as fires and moth outbreaks (Heliasz et al., 2011; López-Blanco et al., 2017; Lund et al., 2017) ~~should be addressed~~ at the ~~coarse~~ pan-Arctic scale. From a modelling perspective, we consider that more field observations are crucial across the pan-Arctic, specifically on plant and soil decomposition (C stocks turnover rates)(He et al., 2016) and respiratory processes (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 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 ~~model~~ analytical robustness and ~~the~~ a decrease of model~~its~~ uncertainties. ~~Field and model researchers should work on data model integration, not just driven based on data availability.~~

5 Conclusions

The Arctic is experiencing rapid environmental changes, which ~~are expected to~~will ~~significantly~~ influence the global C cycle. Using a data-assimilation framework we have evaluated the current state of key C flux, stocks and transit times ~~variables~~ for the pan-Arctic region, 2000-15. ~~We successfully assimilated SOC and biomass, and showed that biomass constraints have a greater importance for contemporary C dynamics (such as C stocks and transit times) than soil C data, which suggest that short term processes are more important than the long term for monthly to annual C dynamics. However, on the~~

~~longer term, the much greater stock of C in slower stocks means that e.g. the permafrost carbon-climate feedback, will likely have important consequences for arctic and global carbon cycling (Koven et al., 2015a).~~ We found that the pan-Arctic ~~region~~ 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. ~~We found that our pan-Arctic estimates of C cycling retrievals are consistent with previous studies.~~ Comparisons with global and local scale datasets demonstrate the ~~advantageus~~ capabilities of CARDAMOM for ~~assessing~~ analysing the C cycle~~ing~~ in the Arctic domain. ~~Moreover~~ ~~,~~ CARDAMOM is a ~~more~~ data-constrained and data-integrated ~~approach than any GVMs available~~analysis, evaluated for internal consistency, and is therefore a~~thus data-assimilation systems are~~ good candidate~~s~~ to benchmark ~~a forward model's~~ performance~~,~~ of global vegetation/ecosystem models~~and pinpoint issues that need attention~~. We conclude that ~~the~~a GVM bias found in transit time of vegetation C ~~are~~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.~~We found better agreement for NPP estimates than for biomass, which is the main contributor to transit time bias.~~ 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 ~~with modelling~~ assimilating data on~~f~~ soil ~~thermal regimes~~hydrology, permafrost and snow dynamics to improve accuracy and decrease uncertainties on belowground processes. This work establishes the baseline for ~~more~~ further process-based ecological analyses using the CARDAMOM data-assimilation system as a ~~promising~~ technique to constrain the pan-Arctic C cycle.

Data 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.

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

Abbott, B. W., Jones, J. B., Schuur, E. A. G., Chapin, F. S., Bowden, W. B., Bret-Harte, M. S., Epstein, H. E., Flannigan, M. D., Harms, T. K., Hollingsworth, T. N., Mack, M. C., McGuire, A. D., Natali, S. M., Rocha, A. V., Tank, S. E., Turetsky, M. R., Vonk, J. E., Wickland, K. P., Aiken, G. R., Alexander, H. D., Amon, R. M. W., Benscoter, B. W., Bergeron, Y., Bishop, K., Blarquez, O., Bond-Lamberty, B., Breen, A. L., Buffam, I., Cai, Y., Carcaillet, C., Carey, S. K., Chen, J. M., Chen, H. Y. H., Christensen, T. R., Cooper, L. W., Cornelissen, J. H. C., De Groot, W. J., Deluca, T. H., Dorrepaal, E., Fetcher, N., Finlay, J. C., Forbes, B. C., French, N. H. F., Gauthier, S., Girardin, M. P., Goetz, S. J., Goldammer, J. G., Gough, L., Grogan, P., Guo, L., Higuera, P. E., Hinzman, L., Hu, F. S., Hugelius, G., Jafarov, E. E., Jandt, R., Johnstone, J. F., Karlsson, J., Kasischke, E. S., Kattner, G., Kelly, R., Keuper, F., Kling, G. W., Kortelainen, P., Kouki, J., Kuhry, P., Laudon, H., Laurion, I., MacDonald, R. W., Mann, P. J., Martikainen, P. J., McClelland, J. W., Molau, U., Oberbauer, S. F., Olefeldt, D., Paré, D., Parisien, M. A., Payette, S., Peng, C., Pokrovsky, O. S., Rastetter, E. B., Raymond, P. A., Raynolds, M. K., Rein, G., Reynolds, J. F., Robards, M., Rogers, B. M., Schdel, C., Schaefer, K., Schmidt, I. K., Shvidenko, A., Sky, J., Spencer, R. G. M., Starr, G., Striegl, R. G., Teisserenc, R., Tranvik, L. J., Virtanen, T., Welker, J. M., and Zimov, S.: Biomass offsets little or none of permafrost carbon release from soils, streams, and wildfire: an expert assessment, Environmental Research Letters, 11, 034014, 2016.

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.

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.

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.

Bradford, M. A., and Crowther, T. W.: Carbon use efficiency and storage in terrestrial ecosystems, New Phytologist, 199, 7-9, 10.1111/nph.12334, 2013.

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.

Clark, D. B., Mercado, L. M., Sitch, S., Jones, C. D., Gedney, N., Best, M. J., Pryor, M., Rooney, G. G., Essery, R. L. H., Blyth, E., Boucher, O., Harding, R. J., Huntingford, C., and Cox, P. M.: The Joint UK Land Environment Simulator (JULES), model description – Part 2: Carbon fluxes and vegetation dynamics, Geosci. Model Dev., 4, 701-722, 10.5194/gmd-4-701-2011, 2011.

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.

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.

Essery, R.: A factorial snowpack model (FSM 1.0), Geosci. Model Dev., 8, 3867-3876, 10.5194/gmd-8-3867-2015, 2015.

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://dx.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.

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., and White, A.: Evaluation and analysis of a dynamic terrestrial ecosystem model under preindustrial conditions at the global scale, Global Biogeochemical Cycles, 14, 1173-1190, doi:10.1029/1999GB900085, 2000.

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.

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.

Grøndahl, L., Friborg, T., Christensen, T. R., Ekberg, A., Elberling, B., Illeris, L., Nordstrøm, C., Rennermalm, Å., Sigsgaard, C., and Søgaard, H.: Spatial and Inter-Annual Variability of Trace Gas Fluxes in a Heterogeneous High-Arctic Landscape, in: Advances in Ecological Research, edited by: Hans Meltofte, T. R. C. B. E. M. C. F., and Morten, R., Academic Press, 473-498, 2008.

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.

Heliasz, M., Johansson, T., Lindroth, A., Mölder, M., Mastepanov, M., Friborg, T., Callaghan, T. V., and Christensen, T. R.: Quantification of C uptake in subarctic birch forest after setback by an extreme insect outbreak, Geophysical Research Letters, 38, n/a-n/a, 10.1029/2010GL044733, 2011.

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.

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.

Kaiser, S., Göckede, M., Castro-Morales, K., Knoblauch, C., Ekici, A., Kleinen, T., Zubrzycki, S., Sachs, T., Wille, C., and Beer, C.: Process-based modelling of the methane balance in periglacial landscapes (JSBACH-methane), Geosci. Model Dev., 10, 333-358, 10.5194/gmd-10-333-2017, 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, 2015a.

~~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, 2015b.~~

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 https://www.nature.com/articles/nature10283 - supplementary-information, 2011.

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.

Mastepanov, M., Sigsgaard, C., Tagesson, T., Ström, L., Tamstorf, M. P., Lund, M., and Christensen, T. R.: Revisiting factors controlling methane emissions from high-arctic tundra, Biogeosciences Discuss., 9, 15853-15900, 10.5194/bgd-9-15853-2012, 2012.

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., Anderson, L. G., Christensen, T. R., Dallimore, S., Guo, L., Hayes, D. J., Heimann, M., Lorenson, T. D., Macdonald, R. W., and Roulet, N. T.: Sensitivity of the carbon cycle in the Arctic to climate change, Ecological Monographs, 79, 523-555, 10.1890/08-2025.1, 2009.

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., Elmendorf, S. C., Beck, P. S. A., Wilmking, M., Hallinger, M., Blok, D., Tape, K. D., Rayback, S. A., Macias-Fauria, M., Forbes, B. C., Speed, J. D. M., Boulanger-Lapointe, N., Rixen, C., Lévesque, E., Schmidt, N. M., Baittinger, C., Trant, A. J., Hermanutz, L., Collier, L. S., Dawes, M. A., Lantz, T. C., Weijers, S., Jørgensen, R. H., Buchwal, A., Buras, A., Naito, A. T., Ravolainen, V., Schaepman-Strub, G., Wheeler, J. A., Wipf, S., Guay, K. C., Hik, D. S., and Vellend, M.: Climate sensitivity of shrub growth across the tundra biome, Nature Climate Change, 5, 887, 10.1038/nclimate2697, 2015.

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.

Pavlick, R., Drewry, D. T., Bohn, K., Reu, B., and Kleidon, A.: The Jena Diversity-Dynamic Global Vegetation Model (JeDi-DGVM): a diverse approach to representing terrestrial biogeography and biogeochemistry based on plant functional trade-offs, Biogeosciences, 10, 4137-4177, 10.5194/bg-10-4137-2013, 2013.

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.

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., Subke, J.-A., Sommerkorn, M., Sloan, V., Ducrotoy, H., Phoenix, G. K., and Williams, M.: The role of mosses in carbon uptake and partitioning in arctic vegetation, New Phytologist, 199, 163-175, 10.1111/nph.12285, 2013.

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: 2. Interannual variations 1982–1993, Journal of Geophysical Research: Atmospheres, 106, 34207-34219, 10.1029/2001JD900164, 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.

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, 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.

Zhou, L., Tucker, C. J., Kaufmann, R. K., Slayback, D., Shabanov, N. V., and Myneni, R. B.: Variations in northern vegetation activity inferred from satellite data of vegetation index during 1981 to 1999, Journal of Geophysical Research: Atmospheres, 106, 20069-20083, 10.1029/2000JD000115, 2001.

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 https://www.nature.com/articles/nclimate3004 - supplementary-information, 2016.

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.

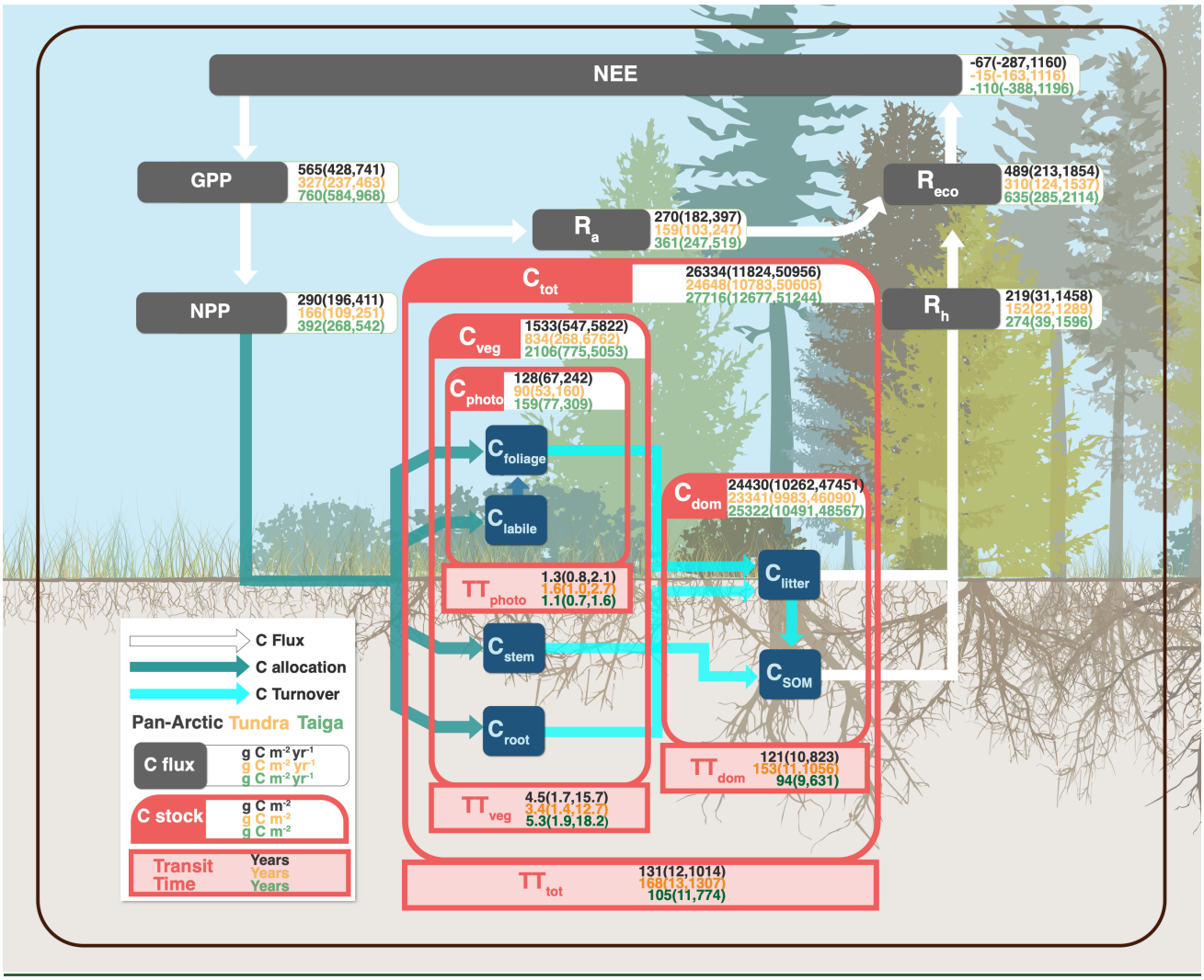


Figure 1. Schematic diagram of the terrestrial C processes modelled in CARDAMOM for the pan-Arctic (black values), tundra (yellow values) and taiga (green values) domains. The values characterize the median for the 2000-2015 period and the parentheses delimit the 90% confidence interval. C processes represented include flows for C fluxes in white [NEE, Net Ecosystem Exchange; GPP, Gross Primary Production; NPP, Net Primary Production; R_{eco}, ecosystem Respiration; R_a, autotrophic Respiration; R_h, heterotrophic Respiration], C allocation in blue [to labile, leaf, stem and root], and C turnover in cyan [from leaf, wood, roots and litter]. C stocks are represented in dark blue boxes [labile, leaf, stem, root, litter and SOM, Soil Organic Matter] and aggregated into photosynthetic (C_{photo} = leaf + labile), vegetation (C_{veg} = leaf + labile + wood + roots), soil (C_{dom} = litter + SOM) and total (C_{tot} = C_{photo} + C_{veg} + C_{dom}) C stocks in red boxes. Analogy, transit times (TT) are also aggregated into photosynthetic (TT_{photo} = leaf + labile), vegetation (TT_{veg} = leaf + labile + wood + roots), soil (TT_{dom} = litter + SOM) and total (TT_{tot} = TT_{photo} + TT_{veg} + TT_{dom}) C transit times.

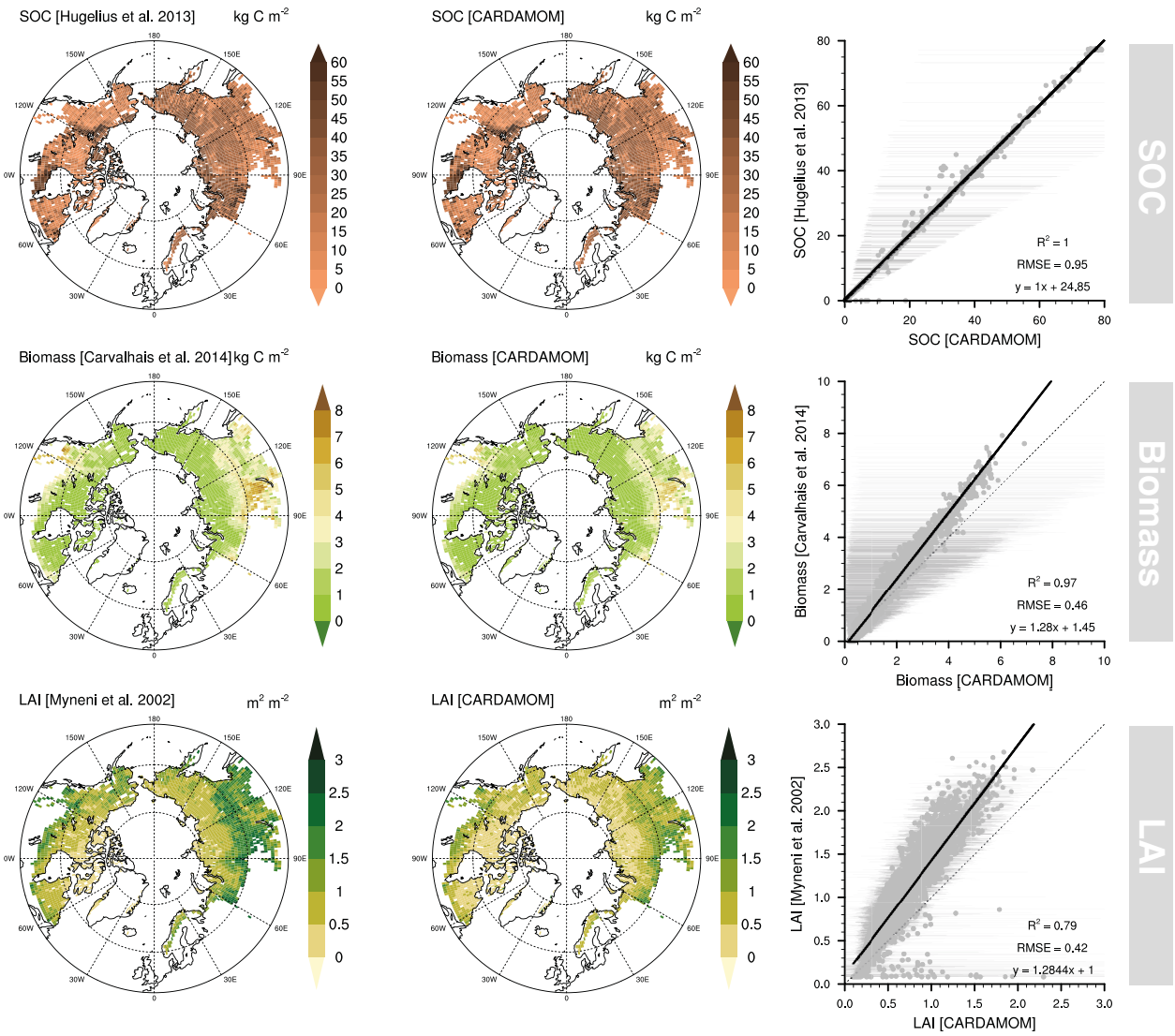


Figure 2~~1~~. Original soil organic carbon [SOC; Hugelius et al.~~,~~ ~~2014~~2013]~~,~~ ~~and~~ biomass [Carvalhais et al.~~,~~ 2014] and leaf area index [LAI; Myneni et al. 2002] datasets used in the data assimilation process within the CARDAMOM framework (left hand side), assimilated SOC, biomas and ~~biomass~~ LAI integrated in CARDAMOM (center), and their respective goodness-of-fit statistics between original and assimilated datasets (right hand side). The error bars represent the 90% confidence interval of the assimilated ~~biomass~~variable in CARDAMOM.

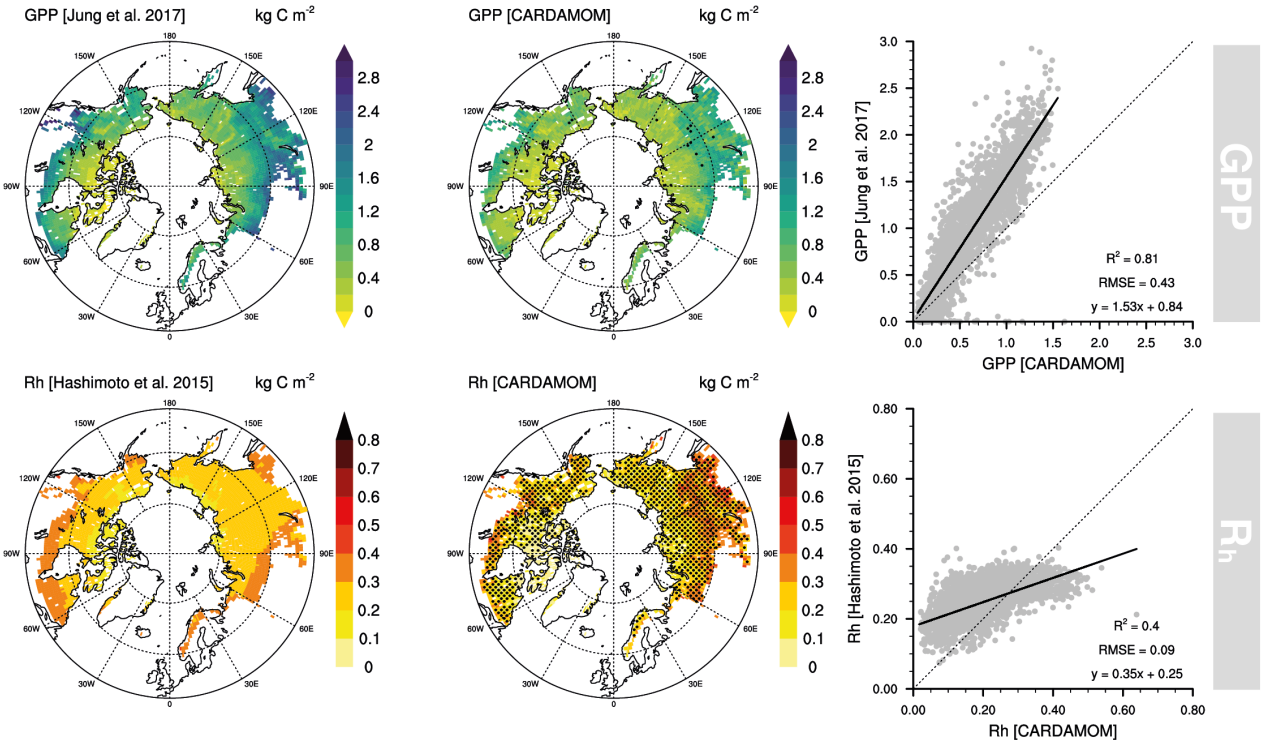


Figure 3~~2~~. Original gross primary productivity [GPP; Jung et al., 201~~76~~] and heterotropic respiration [R_h; Hashimoto et al., 2015] datasets used in the data validation process (left hand side), estimated GPP and R_h by CARDAMOM (center), and their respective goodness-of-fit statistics between original and assimilated datasets (right hand side). Stippling indicates locations where the independent datasets are within the CARDAMOM's 5th and 95th percentiles.

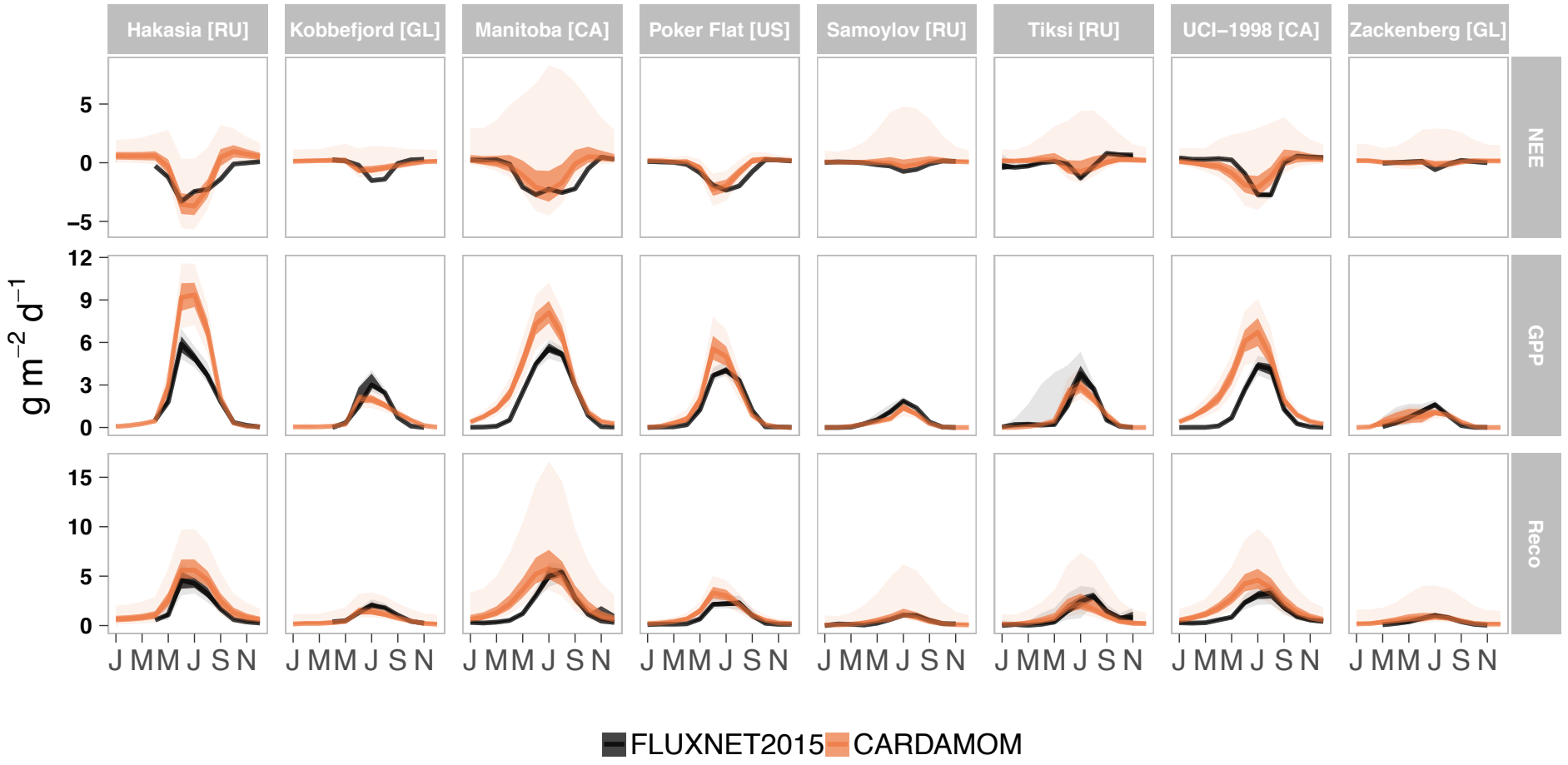


Figure 4~~3~~. Monthly-aggregated seasonal variability of observed [FLUXNET2015] and modelled [CARDAMOM] C fluxes [NEE, Net Ecosystem Exchange; GPP, Gross Primary Production; R_{eco}, ecosystem Respiration] across eight low- and high-Arctic sites [Hakasia, Kobefjord, Manitoba, Poker Flat, Samoylov, Tiksi, UCI-1998 and Zackenberg]. Each of these sites, located in different countries [RU-Russia, GL-Greenland, CA-Canada, US-Unite States,] feature different meteorological conditions and vegetation types (Table S4~~2~~). Uncertainties represent the 25th and 50th percentiles (darker shade) and the 5~~25~~th and ~~75th~~ 95th percentiles (lighter shade) of both field observations and the CARDAMOM framework.

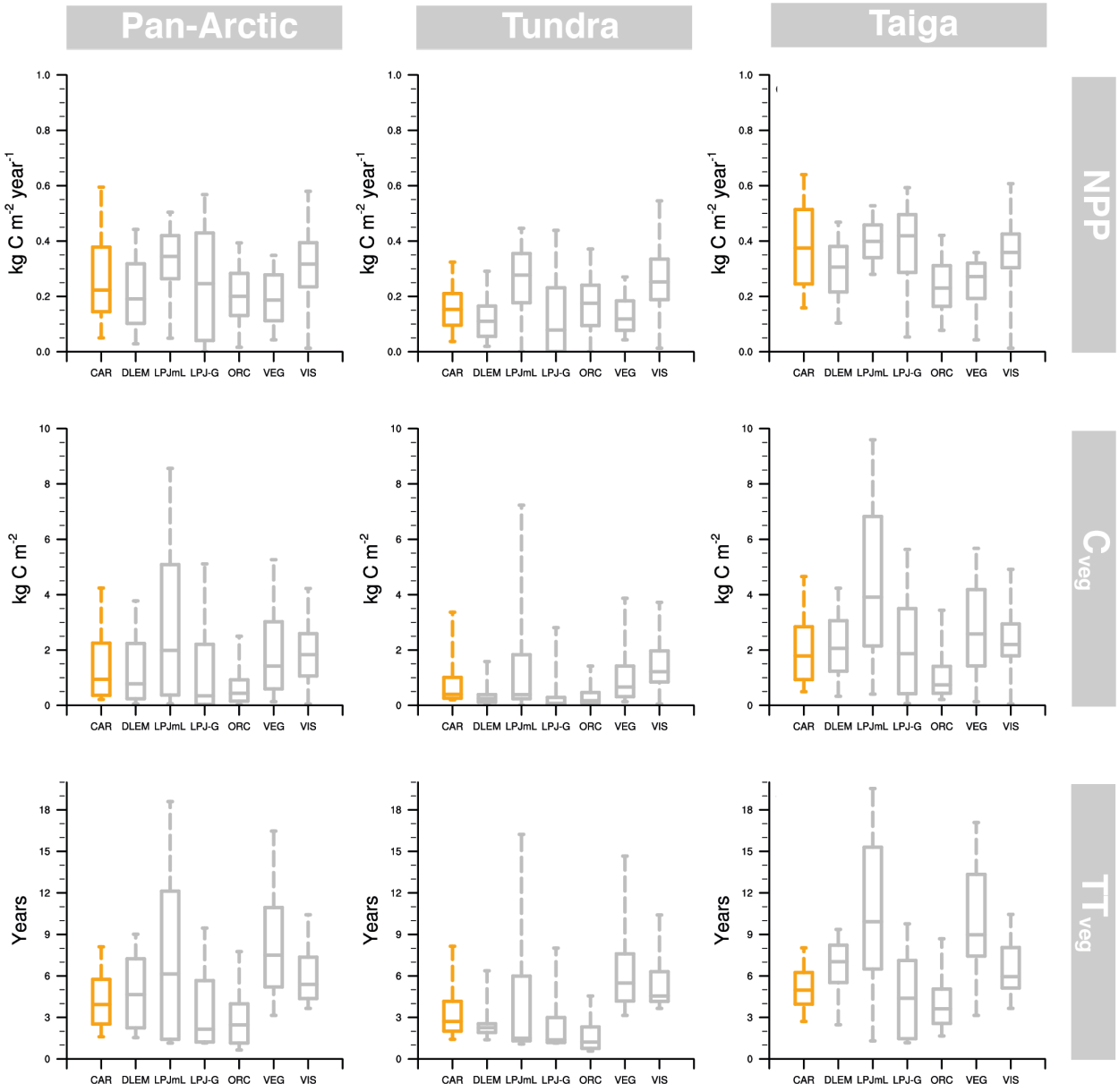


Figure 54. Central tendency and variability of NPP [Net Primary Production], C_{veg} [Vegetation C stock~~pool~~], TT_{veg} [Vegetation transit time] estimated by CARDAMOM (orange) and ISI-MIP2a models (grey) in the Pan-Arctic, tundra and taiga regions. The box whisker plots comprise~~s~~ the estimations between the 5th and 95th percentiles, and the box encompasses the 25th to 75th percentiles. The line in each box mark the median of studied variables in each region.

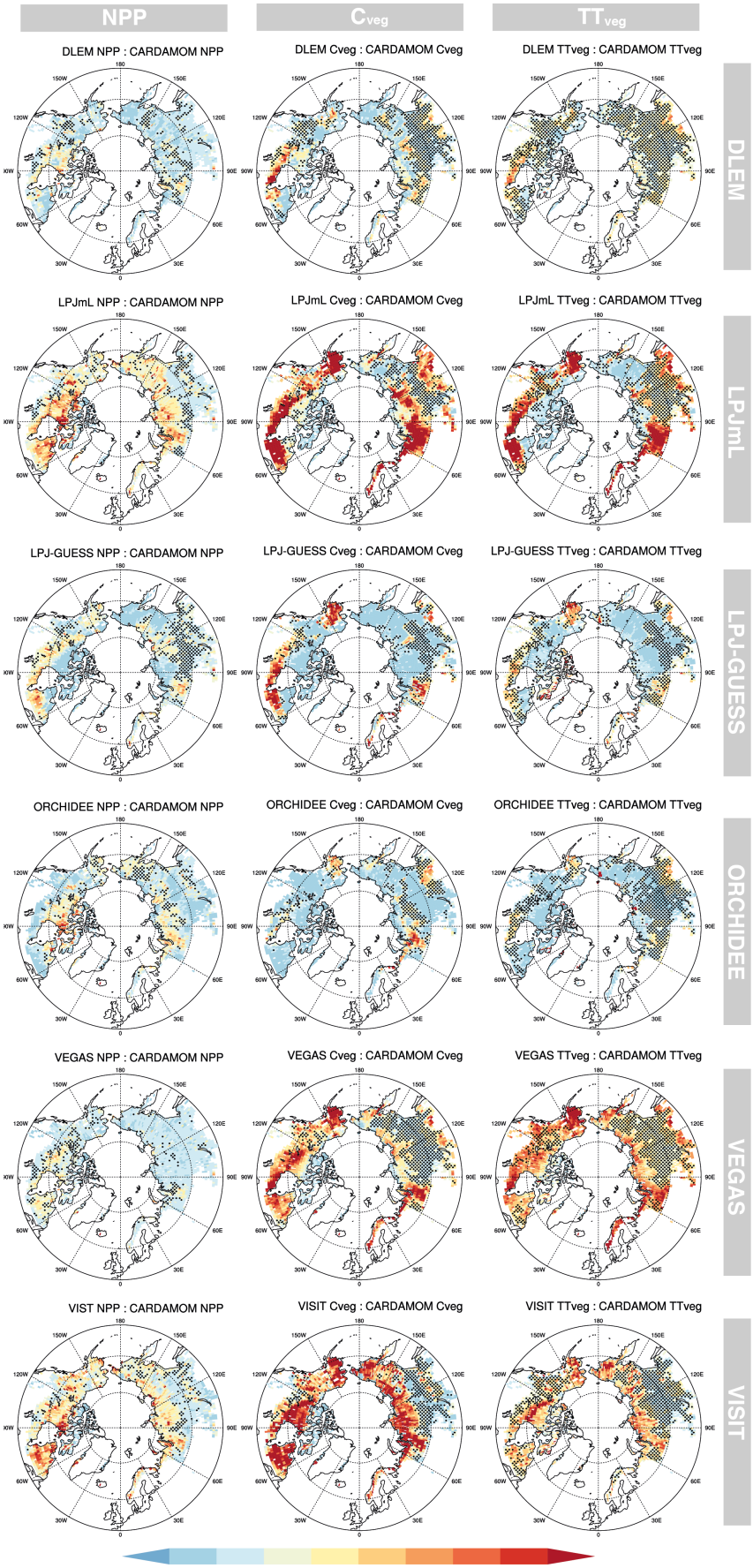
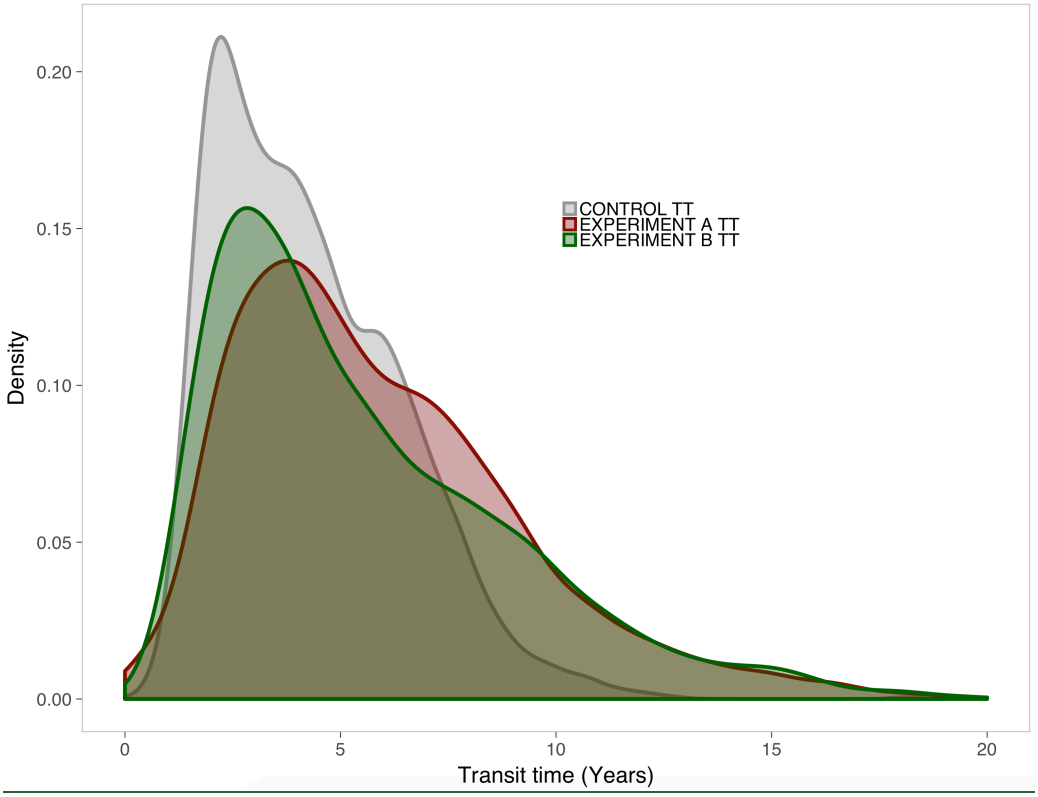
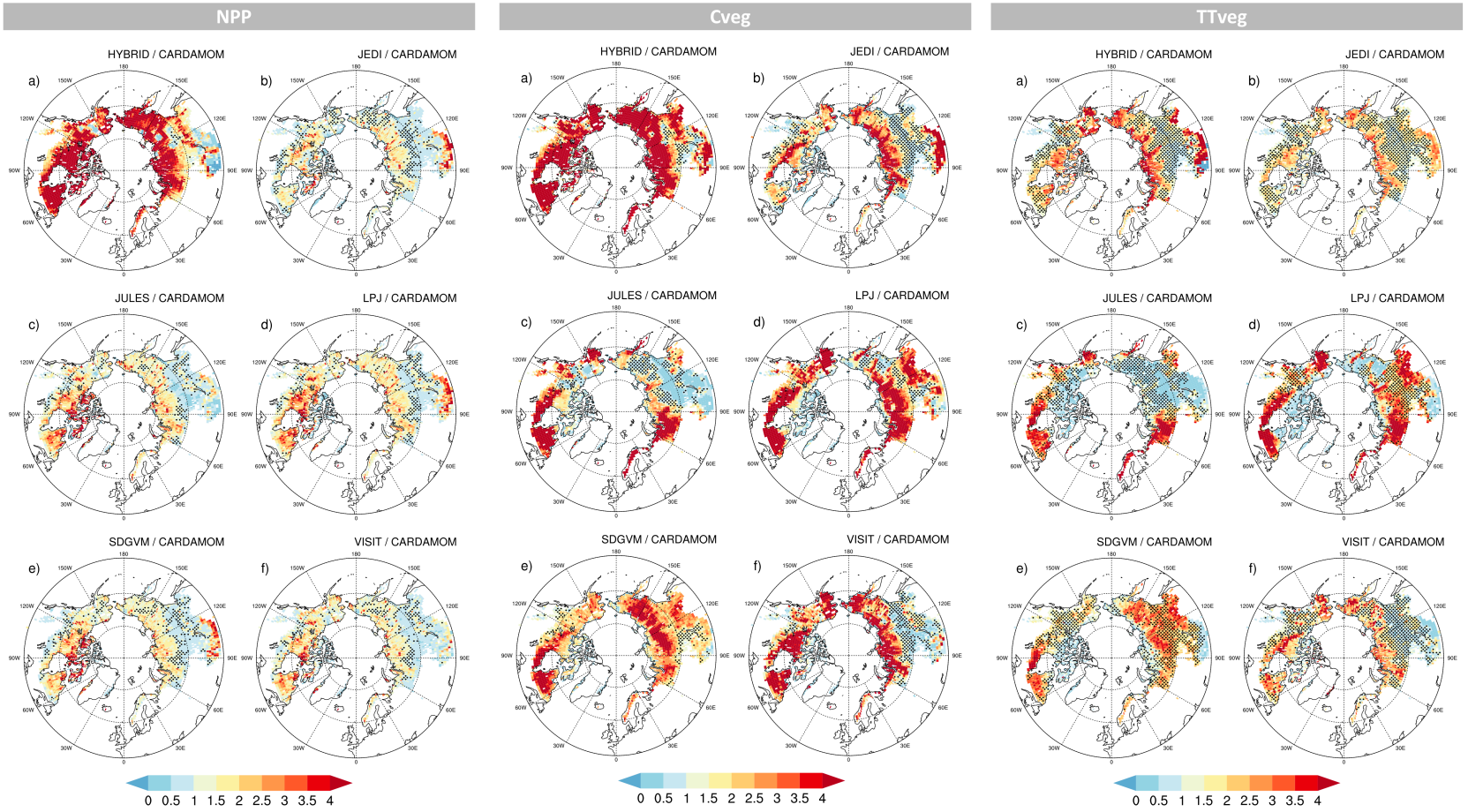


Figure 6. NPP [Net Primary Production], C_{veg} [Vegetation C ~~pool~~stock] and TT_{veg} [Vegetation transit time] ratios between ISI-MIP2a model ensembles [DLEM, LPJmL, LPJ-GUESS, ORCHIDEE, VEGAS and VISIT] and CARDAMOM. Stippling indicates locations where the ISI-MIP2a model mean is within the CARDAMOM's 5th and 95th percentiles.


0.20
0.15
0.10
0.05
0.00
Density
CONTROL TT
EXPERIMENT A TT
EXPERIMENT B TT
0
5
10
15
20
Transit time (Years)



~~Figure 65. NPP [Net Primary Production], Cveg [Vegetation C pool] and TTveg [Vegetation transit time] ratios between ISI-MIP model ensembles [HYBRID, JEDI, JULES, LPJ, SDGVM and VISIT] and CARDAMOM. Stippling indicates locations where the ISI-MIP model mean is within the CARDAMOM's 5th and 95th percentiles.~~

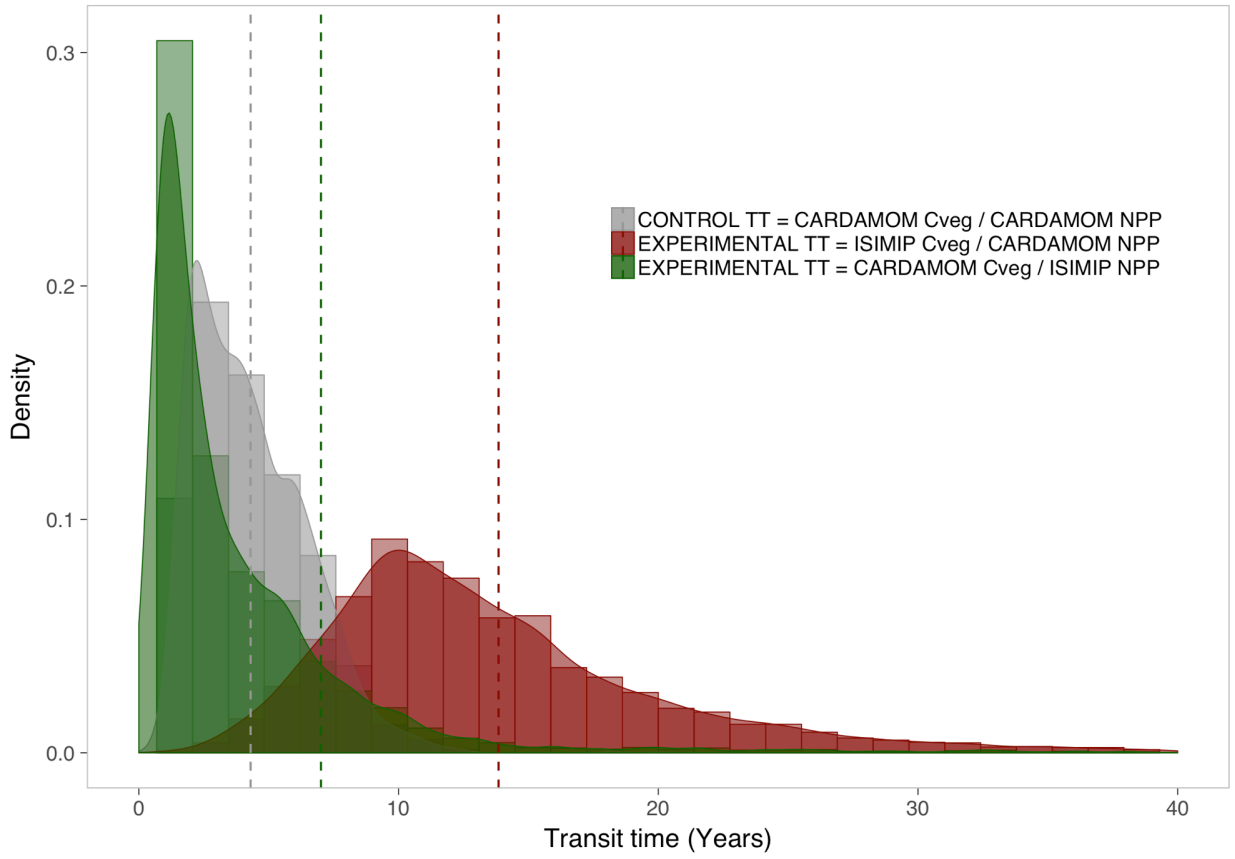


Figure 7~~6~~. Distribution functions derived from the attribution analysis used to estimate the origin of vegetation transit time (TT_{veg}) bias from ISI-MIP2a models. The CONTROL TT (grey) includes both biomass (C_{veg}) and net primary production (NPP) estimated by CARDAMOM. EXPERIMENT A TT (dark red) incorporates C_{veg} from ISI-MIP2a and NPP from CARDAMOM while EXPERIMENT B TT (dark green) includes NPP from ISI-MIP2a and C_{veg} from CARDAMOM. The lower the overlapped area is between control and experimental TT, the larger the contribution for TT biases is. For readability purposes, the scale in X-axis is delimited to 20 years.

~~Distribution functions derived from the attribution analysis used to estimate the origin of vegetation transit time (TT_{veg}) bias from ISIMIP models. The control TT includes both biomass (C_{veg}) and net primary production (NPP) estimated by CARDAMOM (grey), while each of the two experimental TTs include C_{veg} (yellow) and NPP (blue) from ISIMIP models. The lower the overlapped area is between control and experimental TT, the larger the contribution for TT biases is. Dashed lines represent the average TT value for each population. For readibility purposes, the scale in X-axis is delimited to 40 years.~~

Table 1. Multi-year (2000-2015) annual average of main ecosystem C fluxes [NEE, GPP, NPP, R_{eco}, R_a, R_h; g C m^{-2} yr^{-1}], C ~~pools~~ stocks [C_{photo}, C_{veg}, C_{dom}, C_{tot}; kg C m^{-2}] and transit times [TT_{photo}, TT_{veg}, TT_{dom}, TT_{tot}; years] for the pan-Arctic, tundra (non-forested) and taiga (forested), region. The averages contain the the median in bold (50^{th} percentile) and the uncertainty estimations across the 90% confidence range between the 5th and 95th percentiles assuming no spatial correlation between uncertainties in all pixels. We assume spatial correlation between ~~uncertainties in all~~ pixels: P50, P05 and P95 represents the area-weighted aggregate of all pixels' media, P05 and P95.~~, and the median in bold (50^{th} percentile).~~

		Pan-Arctic					Tundra					Taiga				
		P05	P25	**P50**	P75	P95	P05	P25	**P50**	P75	P95	P05	P25	**P50**	P75	P95
C fluxes	**NEE**	-286.7	-170.5	**-67.4**	149.9	1159.9	-163.4	-84.5	**-14.9**	176.2	1116.1	-387.7	-241.0	**-110.4**	128.2	1195.8
	GPP	427.8	504.1	**565.0**	633.5	740.7	236.8	285.0	**327.2**	379.4	463.3	584.1	683.5	**759.8**	841.6	967.9
	NPP	196.4	248.3	**290.3**	337.1	410.7	109.1	139.3	**165.9**	198.3	250.7	268.0	337.6	**392.1**	450.8	541.8
	R_{eco}	212.8	345.8	**488.8**	764.0	1854.1	124.3	211.9	**310.0**	540.3	1536.8	285.3	455.5	**635.3**	947.3	2114.0
	R_a	181.8	229.2	**269.8**	317.4	396.5	102.8	132.0	**158.5**	191.5	247.4	246.6	308.8	**361.0**	420.6	518.6
	R_h	31.0	116.6	**219.0**	446.6	1457.6	21.6	79.9	**151.5**	348.8	1289.4	38.7	146.7	**274.3**	526.7	1595.4
C pools	**C_{photo}**	0.1	0.1	**0.1**	0.2	0.2	0.1	0.1	**0.1**	0.1	0.2	0.1	0.1	**0.2**	0.2	0.3
	C_{veg}	0.5	1.0	**1.5**	2.6	5.8	0.3	0.5	**0.8**	2.0	6.8	0.8	1.4	**2.1**	3.1	5.1
	C_{dom}	10.3	18.3	**24.4**	32.2	47.5	10.0	17.4	**23.3**	31.0	46.1	10.5	19.0	**25.3**	33.2	48.6
	C_{tot}	11.8	20.0	**26.3**	34.5	51.0	10.8	18.4	**24.6**	33.0	50.6	12.7	21.3	**27.7**	35.7	51.2
Transit times	**TT_{photo}**	0.8	1.0	**1.3**	1.6	2.1	1.0	1.2	**1.6**	2.0	2.7	0.7	0.9	**1.1**	1.2	1.6
	TT_{veg}	1.7	2.8	**4.5**	7.5	15.7	1.4	2.2	**3.4**	5.9	12.7	1.9	3.2	**5.3**	8.8	18.2
	TT_{dom}	9.8	51.5	**120.5**	245.9	822.6	11.0	61.6	**152.8**	318.7	1055.9	8.7	43.3	**94.1**	186.3	631.4
	TT_{tot}	11.5	56.9	**133.1**	276.0	1013.6	12.5	67.0	**167.7**	357.6	1306.8	10.7	48.5	**104.7**	209.4	774.3

Table 2. Parameter uncertainty reduction in percentage ranked from least (red) to most (blue) constrained in the pan-Arctic, tundra and taiga domains. The reduction percentage is calculated based on the difference between the 90% CI prior range and the 90% CI posterior range. [Efren1]~~Statistics of linear fit between the CARDAMOM framework (independent) and the FLUXNET2015 field observations (dependent) per individual site and per C flux [NEE, Net Ecosystem Exchange; GPP, Gross Primary Production; R_{eco}, ecosystem Respiration]. The units for RMSE and bias are g C m^{-2} $month^{-1}$ in NEE, GPP and R_{eco}.~~

Parameter	Name	Process	Pan-Arctic	Tundra	Taiga
MR_{litter}	Litter mineralization	Turnover	3.3	3.6	2.9
TOR_{roots}	Root turnover	Turnover	4.8	7.2	2.2
TOR_{wood}	Wood turnover	Turnover	9.0	8.5	9.7
C_{litter}	Litter C stock	Stocks	13.9	13.7	14.1
D_{rate}	Decomposition rate	Turnover	18.2	18.6	17.8
f_{rau}	Fraction of GPP respired (Autotropic respiration)	Allocation	30.9	31.7	30.2
L_f	Leaf fall duration	Phenology	37.3	25.0	51.1
LMA	Leaf mass per area	Phenology	42.8	46.3	38.9
C_{roots}	Fine root C stock	Stocks	52.4	72.1	30.3
R_l	Labile C release duration	Phenology	53.1	52.0	54.4
f_{wood}	Fraction of NPP to wood C pool	Allocation	65.8	68.1	63.3
F_{day}	Leaf fall day	Phenology	67.0	51.1	84.8
MR_{som}	Soil organic matter mineralization	Turnover	69.1	69.6	68.6
f_{labile}	Fraction of NPP to labile C pool	Allocation	74.2	75.5	72.8
C_{eff}	Canopy efficency	Phenology	74.7	75.5	73.7
f_{roots}	Fraction of NPP to roots C pool	Allocation	75.7	74.7	76.8
B_{day}	Leaf onset day	Phenology	76.2	67.4	86.1
C_{SOM}	Soil organic matter C stock	Stocks	80.7	81.4	80.0
L	Lifespan	Turnover	83.4	76.4	91.4
f_{foliar}	Fraction of NPP to foliage C pool	Allocation	88.0	88.6	87.4
C_{labile}	Labile C stock	Stocks	92.2	95.3	88.8
C_{wood}	Woody C stock	Stocks	92.6	90.1	95.5
C_{foliar}	Foliar C stock	Stocks	95.2	96.0	94.3

Table 3. Statistics of linear fit between the CARDAMOM framework (independent) and the ISI-MIP2a models (dependent) per individual model and per NPP [Net Primary Production; kg C m^{-2} yr^{-1}], C_{veg} [Vegetation C stock~~pool~~; kg C m^{-2}] and TT_{veg} [Vegetation transit time; years]. The units for RMSE and bias are kg C m^{-2} yr^{-1} in NPP, kg C m^{-2} yr^{-1} in C_{veg} and years in TT_{veg}.

		Panarctic					Tundra					Taiga				
		Intercept	Slope	R^2	RMSE	Bias	Intercept	Slope	R^2	RMSE	Bias	Intercept	Slope	R^2	RMSE	Bias
NPP (kg C m^{-2} y^{-1})	DLEM	0.04	0.61	0.58	0.09	-0.07	0.04	0.48	0.23	0.08	-0.05	0.12	0.47	0.44	0.08	-0.09
	LPJmL	0.19	0.51	0.43	0.10	0.06	0.12	0.88	0.38	0.10	0.10	0.31	0.23	0.21	0.07	0.02
	LPJ-GUESS	0.01	0.93	0.61	0.12	-0.01	-0.03	1.00	0.38	0.12	-0.03	0.13	0.67	0.45	0.12	0.00
	ORCHIDEE	0.14	0.27	0.17	0.10	-0.06	0.07	0.64	0.31	0.09	0.01	0.20	0.12	0.03	0.10	-0.14
	VEGAS	0.07	0.46	0.60	0.06	-0.07	0.05	0.55	0.36	0.07	-0.02	0.12	0.36	0.52	0.05	-0.13
	VISIT	0.18	0.47	0.26	0.13	0.04	0.10	0.95	0.30	0.13	0.09	0.30	0.18	0.06	0.12	-0.01
C_{veg} (kg C m^{-2})	DLEM	0.44	0.61	0.40	1.00	-0.13	0.38	0.10	0.03	0.65	-0.37	0.92	0.61	0.43	0.91	0.11
	LPJmL	1.70	0.88	0.15	2.80	1.48	1.40	0.16	0.01	2.30	0.65	2.80	0.79	0.12	2.80	2.42
	LPJ-GUESS	0.30	0.69	0.30	1.40	-0.15	0.37	0.13	0.02	0.95	-0.41	0.51	0.81	0.33	1.50	0.13
	ORCHIDEE	0.40	0.23	0.12	0.82	-0.71	0.33	0.04	0.01	0.46	-0.50	0.71	0.20	0.06	1.00	-0.94
	VEGAS	1.10	0.64	0.27	1.40	0.58	1.20	0.10	0.01	1.30	0.37	1.30	0.76	0.38	1.30	0.80
	VISIT	1.60	0.23	0.06	1.30	0.49	1.40	0.03	0.00	1.10	0.53	2.30	0.11	0.01	1.30	0.44
TT_{veg} (yr)	DLEM	1.90	0.69	0.29	2.30	0.56	2.30	0.18	0.05	1.80	-0.42	3.40	0.63	0.29	1.70	1.56
	LPJmL	4.00	0.75	0.07	6.10	2.91	4.10	0.08	0.00	5.10	0.82	7.30	0.60	0.03	5.80	5.27
	LPJ-GUESS	1.30	0.54	0.14	2.90	-0.68	1.70	0.28	0.04	2.90	-0.81	0.95	0.71	0.16	2.80	-0.53
	ORCHIDEE	1.40	0.34	0.10	2.20	-1.42	1.60	0.04	0.00	1.70	-1.78	2.30	0.35	0.07	2.10	-1.03
	VEGAS	5.90	0.62	0.11	3.90	4.23	6.60	0.10	0.00	3.80	3.42	5.50	0.93	0.17	3.40	5.12
	VISIT	5.40	0.12	0.01	2.30	1.65	5.20	0.06	0.00	2.30	1.92	6.70	-0.04	0.00	2.10	1.36