FACULTÉ
DES SCIENCES

UNIVERSITÉ LIBRE DE BRUXELLES

Large scale spatio-temporal variation of carbon fluxes along the land-ocean continuum in three hotspot regions

Thesis submitted by Adam Hastie
in fulfilment of the requirements of the PhD Degree in Science
("Docteur en Science")
Academic year 2018-2019

Supervisor: Professor Pierre Regnier
Co-supervisor: Professor Lei Chou
Biogeochemistry and Earth System Modeling

Thesis jury:

Steeve BONNEVILLE (Université libre de Bruxelles, Chair)
Pierre SERVAIS (Université libre de Bruxelles, Secretary)
Philippe CIAIS (Université Paris-Saclay)
Bernhard WEHRLI (ETH Zurich)
Ronny LAUERWALD (Université libre de Bruxelles)

Financial support was received from the European Union's Horizon 2020 research and innovation program under the Marie Sklodowska- Curie grant agreement No. 643052 (C-CASCADES project)

Résumé

I. Résumé

L'état de l'art dans le domaine a montré qu'il y avait un lien étroit entre les cycles du carbone terrestre et aquatique : en effet, une partie du carbone fixé par photosynthèse (productivité primaire brute) est transférée vers les milieux aquatiques continentaux pour être ensuite transporté latéralement sous forme de carbone organique dissous (COD), de carbone organique particulaire (COP), de carbone inorganique dissous (CID). Durant ce transfert latéral, le carbone peut être minéralisé puis réémis vers l'atmosphère sous forme de CO_2 ou enfoui dans les sédiments. Cependant, nous sommes encore loin de bien comprendre et surtout de quantifier les variations temporelles et spatiales des flux de carbones à l'échelle régionale et globale, même si les études faites à l'échelle locale nous montrent qu'elles sont importantes. Au cours de cette thèse, nous nous sommes focalisés sur 3 grandes régions pour lesquelles la connaissance des flux de carbone le long du continuum aquatique reliant les écosystèmes terrestres aux océans étaient encore très parcellaire.

Pour la région boréale, un modèle empirique a été développé afin de produire les premières cartes à haute résolution de pCO_2 et d'émission de CO_2 pour les lacs boréaux. Les résultats du modèle nous ont permis de contraindre les émissions totales de CO_2 pour les lacs boréaux à 189 (74-347) Tg C an^{-1}, soit plus du double des estimations précédentes. Ce modèle a ensuite été couplé aux projections de production primaire brute terrestre et de précipitations afin de prédire les émissions de CO_2 pour ces lacs pour différents scénarios de changement climatique et d'occupation des sols. Les résultats montrent que même en prenant le scénario le plus conservatif, les émissions de CO_2 des lacs boréaux augmenteraient de 38% d'ici 2100.

Pour le bassin de l'Amazone, le modèle d'écosystème terrestre ORCHILEAK, paramétré par de nouvelles donnés de forçage des zones humides, a été utilisé pour démontrer que l'export de carbone terrestre vers les réseaux fluviaux ainsi que les émissions de CO_2 ont une très grande

variabilité interannuelle : émissions élevées lors des années à forte précipitation et basses lors des années sèches. Cependant, la productivité nette de l'écosystème (PNE) Amazone et la fixation nette de carbone à l'échelle du bassin sont plus élevées lors des années humides, en partie dû au taux de décomposition de carbone organique réduit lorsque les sols sont saturés en eau. De plus, les résultats montrent que les flux de carbone des systèmes aquatiques ont une plus grande variabilité que les flux terrestres, ce qui atténue considérablement la variabilité interannuelle de la PNE du bassin de l'Amazone.

Pour finir, nous avons appliqué ORCHILEAK au bassin du Congo afin d'étudier l'évolution intégrée des flux de carbone terrestres et aquatiques de 1861 à nos jours, ainsi que de projeter leur devenir au cours du 21eme siècle selon les scénarios de changement climatiques et de changement d'occupation des sols. Nous avons montré que les flux terrestres et aquatiques augmentent de façon significative durant la période historique et dans le futur, cette augmentation étant largement induite par l'augmentation du CO_2 atmosphérique et, dans une moindre mesure, par le changement climatique. En particulier, la proportion de la productivité primaire brute terrestre exportée vers le continuum aquatique passe de 3% en 1861 à 5% en 2099. Ce résultat contraste avec ceux obtenu pour la région boréale où cette proportion reste relativement constante et pour l'Amazone où c'est une baisse qui est en fait prédite. Ces différences s'expliquent par des trajectoires de changement climatique distinctes pour ces 3 régions.

Résumé

Previous research has shown a close relationship between the terrestrial and aquatic carbon (C) cycles, namely that part of the C fixed via terrestrial net primary production (NPP) is exported to inland waters. In turn, it has been demonstrated that once in the freshwater system C can not only be transported laterally as dissolved organic carbon (DOC), particulate organic carbon (POC) and dissolved inorganic carbon (DIC) but is also mineralized and evaded back to the atmosphere as CO_2, or buried in sediments. A number of hotspot areas of aquatic CO_2 evasion have been identified but there are considerable gaps in our knowledge, particularly associated with understanding and accounting for the temporal and spatial variation of aquatic C fluxes at regional to global scales, which we know from local scale studies, to be substantial. In this thesis, three important regional hotspots of LOAC activity were identified, where significant gaps in our understanding remain.

For the boreal region, an empirical model is developed to produce the first high resolution maps of boreal lake pCO_2 and CO_2 evasion, providing a new estimate for total evasion from boreal lakes of 189 (74–347) Tg C yr^{-1}, which is more than double the previous best estimate. The model is also used along with future projections of terrestrial NPP and precipitation, to predict future lake CO_2 evasion under future climate change and land-use scenarios, and it is found that even under the most conservative scenario CO_2 evasion from boreal lakes may increase 38% by 2100.

For the Amazon Basin, the ORCHILEAK land surface model driven by a newly developed wetland forcing file, is used to show that the export of C to and CO_2 evasion from inland waters is highly interannually variable; greatest during wet years and lowest during droughts. However, at the same time overall net ecosystem productivity (NEP) and C sequestration is highest during wet years, partly due to reduced decomposition rates in water-logged floodplain soils. Furthermore, it is shown that aquatic C fluxes display greater variation than terrestrial C

fluxes, and that this variation significantly dampens the interannual variability in NEP of the Amazon basin by moderating terrestrial variation.

Finally, ORCHILEAK is applied to the Congo Basin to investigate the evolution of the integrated aquatic and terrestrial C fluxes from 1861 to the present day, and in turn to 2099 under a future climate and land-use scenario. It is shown that terrestrial and aquatic fluxes increase substantially over time, both over the historical period and into the future, and that these increases are largely driven by atmospheric CO_2. The proportion of terrestrial NPP lost to the LOAC also rises from 3% in 1861 to 5% in 2099 and this trend is driven not only by atmospheric CO_2 but also by climate change. This is in contrast to the boreal region where the proportion of NPP exported to inland waters is predicted to remain relatively constant, and to the Amazon, where a decrease has been predicted, due to differences in projected climate change.

Table of acronyms	
Acronym	Meaning
C	Carbon
NPP	Terrestrial net primary production
DOC	Dissolved organic carbon
POC	Particulate organic carbon
SOC	Soil organic carbon
DIC	Dissolved inorganic carbon
LOAC	Land-ocean aquatic continuum
pCO_2	Partial pressure of CO_2
NEP	Net ecosystem productivity
GCP	Global carbon project
DGVM	Dynamic global vegetation model
SD	Standard deviation
k	Gas exchange velocity
k_{600}	k for CO_2 at 20°C in freshwater
T	Temperature
P	Precipitation
A_{lake}	Lake area
T_{water}	Water temperature
T_{air}	Air temperature
FCO_2	CO_2 evasion
N	Nitrogen
P	Phosphorus
TOC	Total organic carbon
DOM	Dissolved organic matter
MFF	Maximal fraction of floodplain
MFS	Maximal fraction of swamps
NCC	Climate forcing dataset (Ngo-Duc et al., 2005)
Princeton GPCC	Climate forcing dataset (Sheffield et al., 2006)
ISIMIP	Inter-Sectoral Impact Model Intercomparison Project
ENSO	El Niño-Southern Oscillation
RMSE	Root-mean square error
NSE	Nash sutcliffe-efficiency
LUC	Land use change
PFT	Plant functional type
SHR	Soil heterotrophic respiration
TF	Throughfall
NBP	Net biome productivity
$LE_{Aquatic}$	Export flux of carbon to the coast
CV	Coefficient of variation
MEI	Multivariate ENSO Index
RCP	Representative concentration pathway
CC	Climate change

0. Contents

Table of figures

Table of Tables

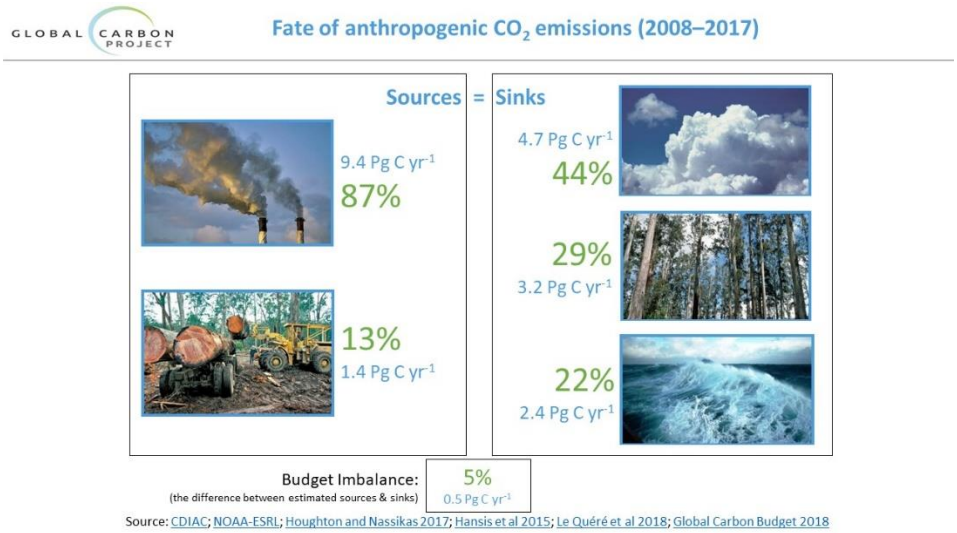
1.Introduction

1.1. The role of the land-ocean continuum in the global carbon cycle

In the last few decades, the global carbon (C) cycle has received an increasing amount of attention in the scientific research community, largely focused on its role in the climate-carbon cycle feedback loop. Approximately 50% of anthropogenic CO_2 emissions are absorbed by the oceans and terrestrial ecosystems of Earth (Le Quéré at al., 2018, Fig. 1), however research has long shown that this sink is highly sensitive to both atmospheric CO_2 levels and climate change (Cox et al., 2000; Friedlingstein et al., 2006). Both the ocean and terrestrial C sinks have increased over the 20th and 21st century (Le Quéré at al., 2018). Moreover, while the ocean sink shows limited fluctuation, the strength of the land sink exhibits considerable interannual to decadal variation (Fig. 2) and current estimates of the global land C sink include a substantial uncertainty range (Fig. 1). Until 2017, the Global Carbon Project (GCP) estimated the land sink indirectly as the residual of the other terms in the following equation:

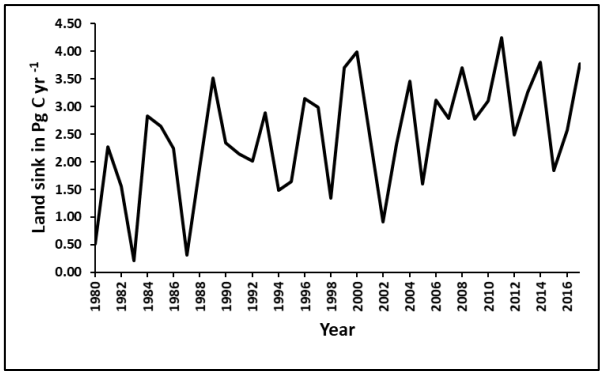
$$S_{LAND} = E_{FF} + E_{LUC} - (G_{ATM} + S_{OCEAN}) \quad (1)$$

Where S_{LAND} is the residual land sink, E_{FF} is CO_2 emissions from fossil fuels and industry, E_{LUC} is emissions from land-use change, G_{ATM} is the atmospheric CO_2 growth rate and S_{OCEAN} is the ocean sink.



Chapter 1 Figure 1: Breakdown of the main sources and sinks within the anthropogenic global C cycle, as estimated in the 2018 Global Carbon Budget (Le Quéré at al., 2018b). Units are Pg C yr^{-1}.

With this approach, the uncertainty in the land sink is by definition the sum of the uncertainty associated with the right-hand terms of equation 1 and is representative of the various biases and missing processes of the other terms. Following this, the GCP estimated the land sink at 3.1 ± 0.9 (1 σ) Pg C yr^{-1} for the decade 2006-2015 (Le Quéré at al., 2016). In 2017, the GCP began calculating the land sink directly from the multi-model mean output of 15 dynamic global vegetation models (DGVMs). Interestingly, the new mean annual land sink of 3.2 Pg C yr^{-1}, as well as the uncertainty range calculated as the mean standard deviation (SD) of the land sink (1959-2016) of 0.8 (1 σ) Pg C yr^{-1} are almost identical to those estimated using budget closure (Le Quéré at al., 2018[a]). For the latest budget, 2018 (Le Quéré at al., 2018[b]), the land sink remains unchanged at 3.2 (Fig. 1) ±0.8 (1 σ) Pg C yr^{-1} (uncertainty of 25%) and substantial interannual to decadal variation remains (Fig. 2). This is larger than the uncertainty associated with fossil fuel emissions of just 5% (9.4±0.5 Pg C yr^{-1}), and that associated with the ocean sink of 21% (2.4±0.5 Pg C yr^{-1}) (Le Quéré at al., 2018 [b]). Moreover, the overall budget is not closed with an imbalance (difference between the estimated C sources and C sinks) of 5% or 0.5 Pg C yr^{-1} globally (Fig. 1, Le Quéré at a`l., 2018 [b]).



Chapter 1 Figure 2: Interannual variation (1980-2017) in the global land C sink as estimated in the 2018 Global C Budget (Le Quéré at al., 2018 b). Units are Pg C yr^{-1}.

Moreover, bottom-up data driven approaches can give a wildly different picture of the global C balance. Zscheischler et al. (2017) used a data-driven approach to synthesize a range of estimates of the net C exchange (NCE, as defined in Zscheischler et al., 2017) between the surface (land, ocean, and coastal areas) and the atmosphere. Using this approach, they calculated a net present-day (2001-2010) transfer from the atmosphere to the surface (or surface sink) of -5.4 ± 2.0 Pg C yr^{-1}. Consider that over the same period (2001-2010), the GCP (Le Quéré et al., 2015) estimate the atmospheric growth rate (G_{ATM}, equation 1) to be 4.3 ± 0.1 Pg C yr^{-1}, a disagreement of almost 10 Pg C yr^{-1} and it becomes abundantly clear that the terrestrial C cycle is a research area that requires further attention.

The freshwater system or the land-ocean aquatic continuum (LOAC) is one aspect of the land carbon (C) cycle which has been neglected until relatively recently and is still not integrated into the DGVMs used to assess the S_{LAND} (equation 1) term of the global C budget (Le Quéré at al., 2018 [b]). Therefore, a major knowledge gap lies in our understanding of how the aquatic C cycle affects the wider land C budget and its interannual variation (Fig. 2). Historically, the LOAC was generally considered as a passive conduit through which carbon would pass from the terrestrial system to the ocean (Cole et al., 2007). Over the last decade a new paradigm has been established which places the LOAC as a much more active component where C can not only be transported laterally as dissolved organic carbon (DOC), particulate organic carbon

(POC) and dissolved inorganic carbon (DIC) but also mineralized and evaded back to the atmosphere as CO_2, or buried in lakes and floodplains (Cole et al., 2007; Battin et al., 2009; Regnier et al., 2013).

In 2013 Regnier et al. published a landmark paper on the LOAC and its perturbation, and this is still arguably the most comprehensive overview providing a global quantification of the various LOAC components, as well as an estimate of their uncertainties. They estimated that the input of C into inland waters is 2.8 Pg C yr^{-1} (Regnier et al., 2013 based on various sources such as Battin et al., 2009; Tranvik et al., 2009) and provided a breakdown of four sub-components of this flux. By far the largest is soil derived C in the form of DOC, POC or dissolved CO_2 which adds up to 1.9 Pg C yr^{-1} and was calculated by subtracting the proceeding three fluxes from the total input to inland waters (2.8 Pg C yr^{-1}). The second component is chemical weathering of carbonate and silicate rocks and is responsible for an input of around 0.5 Pg C yr^{-1} of largely dissolved inorganic C into inland waters (Regnier et al., 2013 based on various sources such as Gaillardet et al., 1999; Amiotte Suchet et al., 2003; Hartmann et al., 2009), with the majority of this total coming from the removal of atmospheric CO_2 in weathering reactions and the rest from the C contained in rocks. It is important to note that most of the C derived from weathering is indirect, namely that the majority of CO_2 removed from the atmosphere is soil derived, having firstly undergone photosynthetic fixation. The third component is C contained within sewage water, estimated at approximately 0.1 Pg C yr^{-1} Regnier et al., 2013 based on Ver et al., 1999 and Mackenzie et al., 2001). The fourth and final flux is photosynthetic fixation within inland waters and Regnier et al. (2013) estimate that around 20% of the C buried in and exported from inland waters is autochthonous, amounting to approximately 0.3 Pg C yr^{-1}.

Once in the LOAC, C can be transported further downstream as DOC, POC or DIC, mineralized and evaded back to the atmosphere as CO_2, or buried in lake and floodplain

sediments and there is consensus that CO_2 evasion from the water surface is the greatest flux. Regnier et al. (2013) put this flux conservatively at 1.1 Pg C yr^{-1}, but values range widely from around 1 to 3 Pg C yr^{-1} (Cole et al., 2007; Tranvik et al., 2009; Raymond et al., 2013; Drake et al., 2017) and estimates of its magnitude have gradually increased over time with new publications on the topic (Drake et al., 2017). Drake et al. (2017) argue that CO_2 evasion from inland waters could be as high as 3.9 Pg C yr^{-1}. Burial within lakes and reservoirs is also highly uncertain with estimates ranging from 0.2 to 0.6 Pg C yr^{-1} (Mendonça et al., 2017; Tranvik et al., 2009). Mendonça et al., (2017) combined observations with statistical modelling to produce their estimate of 0.15 Pg C yr^{-1} while the value of 0.6 from Tranvik et al. (2009) was based on a synthesis of the literature in combination with some simple assumptions to upscale to the entire globe. Of the 0.15 Pg C yr^{-1}, Mendonça et al., 2017) estimate that 0.06 is buried in reservoirs and this is close to the model-based estimate of 0.05 Pg C yr^{-1} calculated by Maavara et al. (2017) for the year 2030.

Moreover, a largely unaccounted for term is the burial of C on floodplains which could represent a non-negligible C sink (D'Elia et al., 2017). Export to the coast is arguably the best constrained LOAC flux and several estimates converge on a value of approximately 0.9 Pg C yr^{-1} (Cole et al., 2007; Tranvik et al., 2009) although estimates also depend on whether estuaries and coastal vegetation are accounted for or not (Cai, 2011; Bauer et al., 2013).

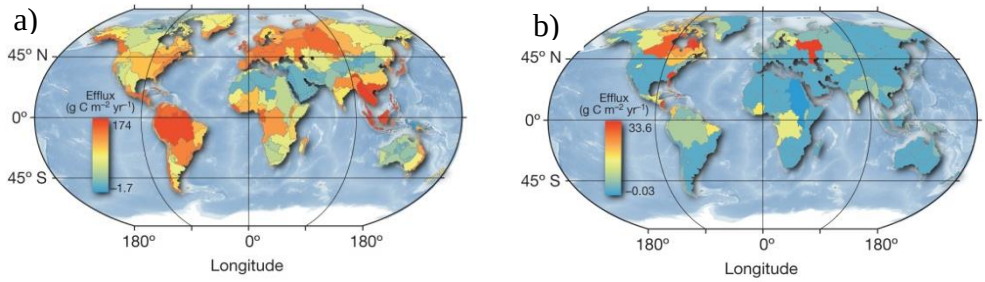
There are a number of reasons behind the large disagreements and uncertainties associated with inland water C fluxes, not least with CO_2 evasion as the largest component. Firstly, observed databases are very limited, both spatially and temporally. There are particular data gaps in regions such as the tropics and Siberia (Raymond et al., 2013; Lauerwald et al., 2015) while most fluxes are not sampled at sufficiently high frequencies to represent the true temporal variation; there are multiple examples of inland waters changing from a CO_2 source to a sink over the course of a single day (Reiman & Xu 2018; Xu et al., 2019). Secondly, the vast

majority of both pCO_2 and CO_2 flux "observations" are indirectly measured, resulting in substantial limitations. pCO_2 is typically calculated from alkalinity, pH, and water temperature, a method which is highly error prone in low-pH and low-alkalinity waters (Abril et al., 2015) and can lead to an overestimation of pCO_2 (Wallin et al., 2014). Moreover, calculating CO_2 evasion in turn relies on an empirical estimation of the gas exchange velocity *k* for which multiple different empirical parameterizations exist (Cole and Caraco, 1998; Read et al., 2012; Vachon & Prairie., 2013), and this is usually in addition to the uncertainty associated with pCO_2.

Another important research question, and addressed for the first time in the paper of Regnier et al. (2013), is how much these LOAC fluxes have been perturbed historically by anthropogenic divers, and it follows to ask how much they are likely to be perturbed under future climate change and land use change scenarios. Globally, Regnier et al. (2013) estimated that anthropogenic perturbation has increased the transport of C to the freshwater system by around 1 Pg C yr^{-1} and CO_2 evasion from inland waters by 0.5 Pg C yr^{-1} since pre-industrial times. Thus, an added layer of complexity is isolating the natural temporal and spatial variation from the anthropogenic perturbation along the LOAC, and indeed the budget of the GCP does not incorporate this perturbation as C transfer and transformation along the LOAC is not represented in the DGVMs that they used to estimate the S_{LAND} (equation 1) term. They acknowledge that it will affect the partitioning of C between the land and ocean sink but argue that the other terms would remain unaffected (Le Quéré at al., 2018[b]). In conclusion, the LOAC has been identified as a significant knowledge gap by the GCP (Le Quéré et al., 2015).

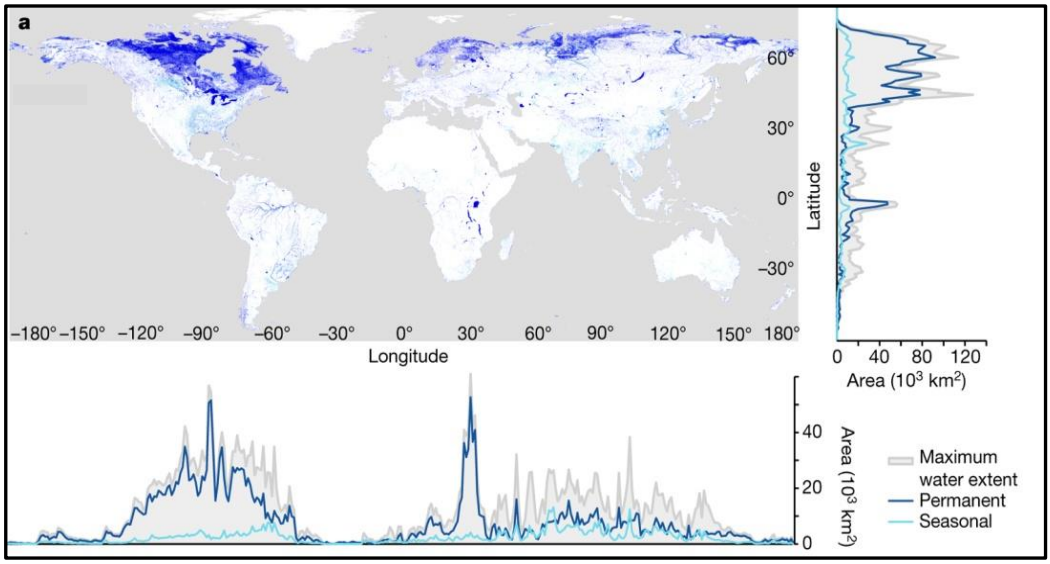
1.2. Estimating the spatial variation in LOAC fluxes using upscaling and empirical approaches

In order to better constrain the contribution of the LOAC to the global C cycle, we need to be able to understand and quantify the spatial and temporal variation of inland water C fluxes. In 2013, Raymond et al. (2013) published the first attempt to estimate the spatial variation in aquatic CO_2 evasion globally. They used a relatively simple upscaling approach combining a limited dataset of indirectly observed pCO_2 data with some empirical modelling, to estimate CO_2 evasion from rivers and lakes at the relatively course scale of COSCAT catchments (Meybeck et al., 2006, Fig. 3). Based on this method, they estimated that globally, rivers evade 1.8 (1.55-2.05) Pg C yr^{-1} (Fig. 3 a) to the atmosphere while lakes evade 0.32 (0.06-0.84) Pg C yr^{-1} (Fig. 3 b). Their results suggested some hotspot areas of CO_2 evasion from inland waters; tropical, sub-tropical and to a lesser extent, temperate regions dominated the flux from rivers (Fig. 3 a), while high latitudes dominated the flux from lakes (Fig. 3 b). Somewhat unsurprisingly, these hotspot areas coincide with water surface area (Fig. 4), rainfall and discharge being important drivers of the export of terrestrial NPP to the aquatic system, and in turn CO_2 evasion and the export of C to the coast. At high latitudes, the water surface area is dominated by lakes while most of the largest rivers of the world occur at tropical or sub-tropical latitudes (Fig. 4), and many of these rivers also have considerable floodplains.



Chapter 1 Figure 3: CO_2 efflux from a) rivers to the atmosphere and b) lakes to the atmosphere per COSCAT region, taken from Raymond et al. (2013). Units are g m^{-2} C yr^{-1}.

Note however, the particularly large uncertainty ranges and in particular for lake CO_2 evasion, largely caused by an inability to accurately estimate the number of small lakes (Raymond et al. 2013). Raymond et al. (2013) used simple empirical size- distribution relationships from the literature (Downing et al., 2012; McDonald et al., 2012) to predict the size and distribution of all lakes smaller than 3.16 km^2, which accounts for most of this uncertainty. Despite these large uncertainties, the large-scale spatial patterns in lake and river CO_2 evasion are generally supported by the findings of other global and regional studies. We discuss these studies in more detail in the following sections, as well as estimates for the export of C from the LOAC to the coast.



Chapter 1 Figure 4: Map showing global distribution of surface water and graphs of surface water area versus latitude and longitude, taken from Pekel et al. (2016).

1.2.1. CO_2 evasion from rivers

Raymond et al. (2013) could not find significant relationships between river pCO_2 and various environmental divers at the scale of COSCAT regions but a 2015 study (Lauerwald et al., 2015) utilized largely the same database to derive regression relationships at the river catchment scale, with the watersheds delineated based on the Hydrosheds (15 arcsec-Lehner et al., 2008) and Hydro1k (1km- U.S. Geological Survey) geodatabases. Lauerwald et al. (2015), estimated substantially smaller CO_2 evasion from rivers of 650 (483–846) Tg C yr^{-1}, with the uncertainty

range calculated using Monte Carlo simulations. They were able to explain almost 50% of the global spatial variation with only three drivers; terrestrial NPP, population density, and slope gradient with the first two showing positive relationships with pCO_2 and the last having an inverse relationship. Note that they also found significant positive relationships between pCO_2 and both air temperature and the flooded proportion of the catchment. Using this approach, they were also able to provide the spatial variation of CO_2 evasion at a much higher resolution of 0.5°. While the magnitude of the flux is considerably smaller than that estimated by Raymond et al. (2013), mainly due to lower estimates of mean pCO_2 in tropical rivers (an area of relatively sparse data), it largely confirmed the regional hotspots of Raymond et al. (2013).

There have also been various attempts to estimate CO_2 evasion within regional hotspot areas. Rather than using empirical modeling, these regional estimates have generally relied on even simpler approaches, namely aggregation of observed data and upscaling using an estimate of river surface area. For example, Rasera et al. (2013) upscaled observation to estimate CO_2 evasion from the Amazon Basin at 0.8 Pg C yr^{-1} while more recently, Sawakuchi et al. (2017) added observations from the basin area downstream of Obidos and concluded that CO_2 evasion from the entire Amazon Basin (down to mouth) could potentially be as high as 1.39 Pg C yr^{-1}. If accurate the latter estimate represents 77% of the 1.8 Pg C yr^{-1} global estimate by Raymond et al. (2013) and several times that estimated by Lauerwald et al. (2015). Incongruities such as these only seem to confirm how little we truly understand. The Congo Basin, the second largest river basin in the world, has also been confirmed as a hotspot area; Borges et al., (2015) used a simple upscaling approach to estimate present day CO_2 evasion from the rivers of the Congo basin at 133-177 Tg C yr^{-1}(Borges at al., 2015).

1.2.2 CO_2 evasion from lakes

Several studies have quantified the flux of CO_2 from lakes at the national scale, again largely relying on upscaling of observations and empirical modelling (Humborg et al., 2010;

McDonald et al., 2013; Butman et al., 2016). For the USA, Butman et al. (2016) delineated lake pCO_2 and FCO_2 for 19 sub-regions of the USA and calculated a total efflux of 16 (14.3–18.7) Tg C yr^{-1}, while Humborg et al. (2010), estimated pCO_2 and FCO_2 for 5 lake size classes and calculated a total FCO_2 from Swedish lakes of 1.74 Tg C yr^{-1}. No such national scale studies have been undertaken for Canada or Russia, the two most dominant countries in terms of total lake area (Verpoorter et al., 2014), and Canada in particular was identified as a hotspot are for lake CO_2 evasion in Raymond et al. (2013, Fig 3 b). Raymond et al. (2013), remains the only study to attempt to resolve the spatial variation of CO_2 evasion from lakes beyond the national scale, at the resolution of the COSCAT basins (Fig. 3), and a similar high-resolution study to what was undertaken by Lauerwald et al. (2015) for rivers, does not exist for lakes. Thus, an obvious knowledge gap in our attempts to constrain the LOAC budget and in turn the land sink, is a regional to global scale assessment of the spatial variation in lake CO_2 evasion in a similar manner to what has been undertaken for rivers. While substantial uncertainties in the spatial variation of river CO_2 evasion remain (Lauerwald et al., 2015), the current uncertainties associated with lake CO_2 evasion are considerably larger (0.06-0.84 Pg C yr^{-1}, Raymond et al., 2013). In summary, understanding and quantifying the spatial variation of CO_2 evasion in lakes is a clear research avenue, which requires further investigation. This discussion is continued in greater detail in chapter 2.

1.2.3. Export flux of C to the coast

Quantifying the export of C from land to the coast is important for the partitioning of C between the land and ocean sink, something that is not currently accounted for in the DGVMs used to estimate the Global Carbon Budget (GCB) (Le Quere et al., 2018[b]). Like the evasion fluxes, this has largely been estimated using empirical methods. Ludwig et al. (1996) found that they were able to explain the vast majority of the variation in DOC fluxes in 29 major rivers based on drainage, basin slope and soil C while they were also able to explain variation in POC as a

function of sediment flux, in turn related to drainage, rainfall and basin slope. Using regression models to upscale these relationships, they estimated a TOC flux from rivers to the coast of 0.38 Pg C yr^{-1}. Following on from this Meybeck et al. (2006) devised the COSCAT global database (Fig. 3); a set of 151 catchments characterized by area, latitudinal range, runoff and coastal limits, amongst other things, with the aim of providing a consistent framework within which to estimate the transfer of water, sediments, and nutrients from land to the ocean. The original study (Meybeck et al., 2006) used the COSCAT delineation to estimate average runoff and Nitrogen yields globally for the 151 catchments, but other studies have applied it in estimating C fluxes (Mayorga et al., 2010; Raymond et al., 2013). This approach is only designed to provide mean annual values of discharge and nutrient yields at the mouth of each COSCAT basin and therefore its main utility is in providing a picture of the spatial variation and in identifying hotspot catchments in land-ocean linkage, the Amazon Basin being one such region. The Global NEWS model is a hybrid model, incorporating both empirical and mechanistic aspects to estimate annual exports of C, N and P at river basin mouths, and builds on the work of both Ludwig et al. (1996) and Meybeck et al. (2006). Indeed, the POC component is modelled using the same empirical relationship of Ludwig et al. (1996), a function of sediment flux, while the basins are defined using a precursor of the methodology described in Meybeck et al. (2006), and at a similar scale. They estimate a TOC flux to the coast of 0.30 Pg C yr^{-1}, similar to the 0.38 Pg C yr^{-1} calculated by Ludwig et al. (1996), while an older upscaling study estimated that this flux could be as high as 0.53 Pg C yr^{-1}. More importantly, as in Meybeck et al. (2006) they are were able to identify hotspot areas of TOC export, namely the tropical latitudes. Additionally, several studies have put the DIC export flux from inland waters at around 0.3 Pg C yr^{-1} (Suchet & Probust., 1996; Stallard et al., 1998).

Data-driven models are useful in providing estimates of present-day fluxes of C along the LOAC and also allow estimates of the uncertainties associated with the data and statistical

relationships, as was for example performed using Monte-Carlo simulations in Lauerwald et al. (2015). However, data-driven models are reliant upon accurate measurements or at the very least a good understanding of the uncertainty associated with such observations. In reality, data accuracy is difficult to ensure given that measurements of pCO_2 are overwhelmingly indirect and associated with various errors (Wallin et al., 2014; Abril et al., 2015). CO_2 fluxes from inland waters are also usually indirectly calculated, using pCO_2 measurements along with an empirical parameterization of gas exchange velocity k, another considerable source of uncertainty. As such, if data-driven approaches are used, they need to be combined with robust and transparent estimations of all major sources of uncertainties. Additionally, in order to explore the effects of climate and land use change, and other perturbations on LOAC fluxes, data driven models need to be complemented by or indeed coupled to process based/ mechanistic models.

1.3. Temporal variation in aquatic C fluxes from seasonal to centennial timescales

1.3.1. Seasonal variations

Much like with spatial variation, in order to provide reliable estimates of present-day estimates of inland water C fluxes, we need to be able to account for the temporal variation. Moreover, if we are to model the seasonal and longer-term changes in these fluxes, it is essential that we can understand and represent the key drivers of this change. Existing global estimates of LOAC C fluxes such as Raymond et al. (2013) deal with mean annual fluxes, but in reality, aquatic C fluxes exhibit strong seasonal patterns. In the boreal region for example, research has found that CO_2 accumulates under ice covered lakes in winter, leading to very high emissions during ice melt (Striegl et al., 2001). A synthesis of data sampled during ice melt (Denfeld et al., 2018), concluded that the ice-melt period was on average responsible for 17% of annual CO_2 evasion

from high latitude lakes. Similarly, snowmelt has generally been found to drive an increase in pCO_2 and CO_2 evasion in lakes (Evans et al., 2013) and rivers (Lauerwald et al., 2015).

More generally, seasonal variation in hydrology has been identified as one of the most important drivers of the seasonal variation in both DOC and pCO_2 concentrations in rivers and lakes across different latitudes, though the manner of this relationship is variable. In tropical rivers for example, Richey et al. (2002) found that pCO_2 in the Amazon Basin is closely correlated with discharge, rising and falling with river flow as allochthonous C is transported to the LOAC from wetlands and terrestrial systems. Discharge has also been shown to drive higher pCO_2 in temperate rivers (Marescaux et al., 2018) again largely as a result of imported allochthonous C. Moreover, a positive relationship between seasonality in discharge and pCO_2 has been found in some boreal lakes (Rantakari & Kortelainen, 2005; Sobek et al., 2003). However, the relationship is different in areas with important groundwater contributions to aquatic CO_2 concentrations. Nydhal et al. (2017) found negative relationships between precipitation and pCO_2 in some boreal lakes and suggested that this may be a result of increased precipitation causing a dilution of CO_2 concentrations due to an altered balance between surface and CO_2 -rich groundwater flow.

1.3.2. Interannual variations and climate extremes

Significant interannual variation in river pCO_2 /CO_2 evasion has also been observed at various latitudes. A recent study by Almeida et al (2017) demonstrated that in addition to seasonal variation, large flood events also drive interannual variation in CO_2 evasion from the Madeira River (a tributary of the Amazon), namely that years with extreme flooding evade 20% more CO_2 to the atmosphere per unit area than years without. Similarly, infrequent but large flood events have also been shown to drive the transport and mineralization of terrestrial dissolved organic matter (DOM) in the USA (Raymond et al., 2016). Significant interannual variation in

pCO_2 has also been observed in boreal lakes, driven by interannual variation in precipitation (Rantakari & Kortelainen., 2005).

Furthermore, previous research has shown that the terrestrial C cycle varies both seasonally and interannually and is sensitive to the ENSO cycle (Zhao & Running., 2010; Wang et al., 2018) while the same has been demonstrated in regional studies of the Amazon (Gatti et al., 2014; Doughty et al., 2015; Feldpausch et al., 2016). This substantial interannual variation can be seen in the aggregated global land sink of the GCP (Fig. 2). However, as previously mentioned, the LOAC and its interannual variation is not integrated into the GCP estimates as it is not represented in the underlying models. As such, another knowledge gap is how the interannual variation of the terrestrial and aquatic cycles interact, particularly during years with extreme climatic events. This can only be achieved using a model which integrates the aquatic C cycle within the terrestrial.

1.3.3. Long-term trends associated with climate change and anthropogenic perturbation

Additionally, LOAC fluxes have also been shown to vary over decadal to centennial timescales and these changes are often associated with direct anthropogenic perturbation such as land-use change, hydrological management and climate change, though the effects are often not isolated. Globally, Regnier et al. (2013) estimated that anthropogenic perturbation has increased the transport of C to the freshwater system by around 1 Pg C yr^{-1} since pre-industrial times, largely as a result of increased export from soils. Local and regional scale studies have also found evidence of perturbation along the LOAC. A study in 2016 along the Mississippi (Ren et al., 2016) identified land-use change including land management practices (such as fertilization and irrigation), followed by change in atmospheric CO_2, as the biggest factors in a 40% historical increase in DOC export to the Gulf of Mexico (Ren et al., 2016). However, Jones et al. (2003) found a significant decrease in pCO_2 from 1973-1994 in a study of 417 streams and

rivers across the contiguous USA, and suggested that this was a consequence of a reduction in C import from the terrestrial environment. In the Seine Basin, land management practices have been shown to drive seasonal variation in pCO_2 while waste-water pollution has been found to influence longer term (1970-2015) changes in pCO_2 (Marescaux et al., 2018). Le et al. (2018) observed that between 1960 and 2015 both DOC and DIC had decreased significantly in the Red River in Vietnam, the former as a result of dam impoundment and the latter as a result of carbonate precipitation in irrigated, agricultural land in combination with reduced river flow. Damming has also been shown to have an impact on the LOAC globally. By the beginning of the 21st century, Maavara et al. (2017) showed that the combined effect of in reservoir C mineralization and burial had reduced OC export from inland waters to the coast by 13% (48±11 Tg C^{-1}) and predict that this value could increase to 19% by 2030 (83±18 Tg C^{-1}).

Increasing soil erosion of soil organic C (SOC) represents another long-term change in the C cycle and a recent study, Naipal et al. (2018) showed that this flux has been substantially perturbed since 1850, increasing the export of SOC to the LOAC. Finally, inputs to the LOAC from physical erosion of inorganic C (as well as highly resistant OC) represents a flux which operates at even longer timescales and in the context of the scope of this thesis (seasonal to centennial timescales) is refractory, i.e. is a passive flux. This discussion of the temporal variation in aquatic C fluxes is by no means exhaustive but aims to highlight the complexity of C dynamics along the LOAC and the challenges we face in trying to represent and quantify this variation across larger scales. This challenge is unlikely to be met using only the empirical approaches described in the previous section. Climatologies of pCO_2 have been produced for both the coastal and open ocean based on empirical methods (Landschützer et al., 2013; Laruelle et al., 2017) using the SOCAT database which contains several million measurements, but this requires a data network of sufficiently high temporal and spatial resolution. The current

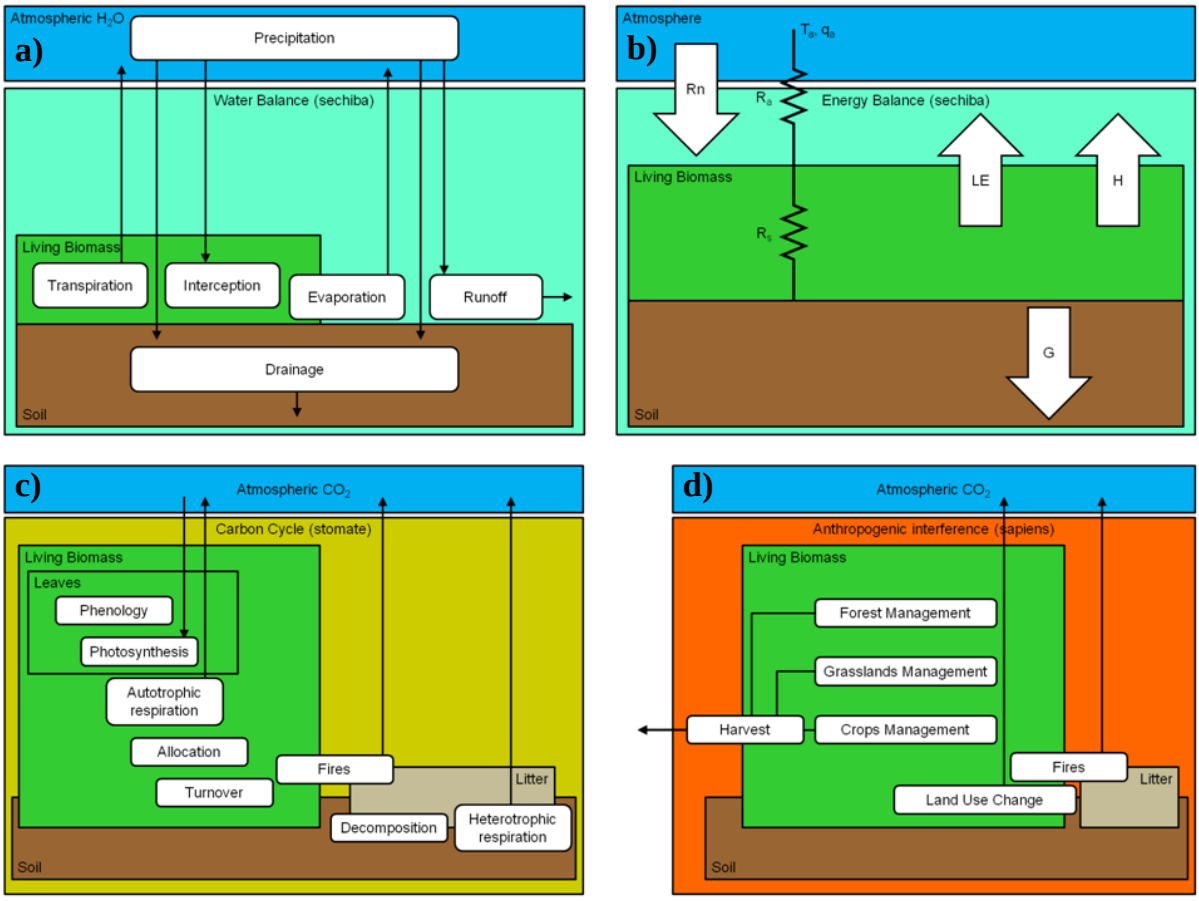
database of river and lake pCO_2 is simply insufficient to do the same for rivers and lakes, certainly at regional to global scales.

1.4. Process based modelling approaches to understanding the LOAC

Considering the current limitations in inland water C flux datasets, arguably the most efficient way to investigate the variation of these fluxes, from seasonal to centennial timescales, and particularly their response to future perturbations, is to use process based/ mechanistic models. With these models, one can apply different scenarios of atmospheric CO_2, climate and land-use change amongst other boundary conditions, to investigate how the C cycle will react to different perturbations, which can eventually be used to inform environmental management policies. As previously discussed, the LOAC is not currently integrated into any of the DGVMs used by the GCP. One of these is ORCHIDEE, the land surface model of the IPSL (Institut Pierre Simon Laplace) Earth System Model. Figure 5 illustrates the main processes simulated in the ORCHIDEE model (Krinner et al., 2005); while the water balance (a) and the C cycle (c) are both represented, the transfer and transformation of C to, and along the LOAC is neglected.

Attempts to develop mechanistic models of C fluxes along the LOAC began with the representation of C export fluxes from inland waters to the coast. In a 2013 paper, Kicklighter et al. (2013) further developed the existing Terrestrial Ecosystem Model (TEM) to represent the transfer of DOC along Arctic rivers and export to the Arctic Ocean. The existing TEM is a process-based model which uses high resolution input datasets on climate, elevation, soils and land-use amongst other things, to simulate plant growth, C and N exchange between vegetation and soil pools, as well as exchanges between terrestrial pools and the atmosphere via primary production, respiration and decomposition. The Kicklighter et al. (2013) study added two processes; DOC production from incomplete decomposition of litter and soil organic matter and DOC leaching into the river network, with leaching in turn depending on concentration of

DOC in combination with the runoff of rainwater and snowmelt. Applying this approach at a spatial resolution of 0.5° and a monthly temporal resolution, they estimated that an average of 32 Tg C yr^{-1} has been transported from the terrestrial to the aquatic system across the pan-artic over the 20th century. Moreover, they used the model to show that historical changes in environmental factors, mainly climate change, have increased cumulative DOC leaching by 258 Tg C between 1900 and 2006. Though it should be noted that the model performance was mixed; it was able to recreate present-day observed DOC loads relatively closely (± 25% of observed) in 2 of the 8 watersheds studied.



Chapter 1 Figure 5: The main processes simulated in ORCHIDEE (Krinner et al., 2005). The processes are grouped in terms of the water balance (a), the energy balance (b), the biogeochemical processes (c) and anthropogenic processes (d). Although the water and energy budget were separated in this presentation, both need to be run at the same time. These processes are coded in a group of modules called 'Sechiba' and are the backbone of ORCHIDEE. In addition to Sechiba, the biogeochemical code and the anthropogenic code, grouped in modules called 'Stomate', can be activated. Several individual processes can be

switched on or off, so supporting a wide range of model set-ups. Nevertheless, the model can be run without activating Stomate.

A 2015 study employed the Dynamic Land Ecosystem Model 2.0 (DLEM 2.0) at a resolution of approximately 0.1° to simulate historic changes to not only DOC export but also POC and DIC export from North-Eastern USA to the Atlantic Ocean. DLEM is a full DGVM and meets the criteria for use in the GCP (Le Quere et al., 2018[b]). As well as the terrestrial C cycle, DLEM 2.0 includes DOC leaching to the LOAC, the export of POC to the LOAC via soil erosion and export of DIC to the LOAC as a result of dissolution of atmospheric CO_2, dissolution of soil CO_2, and carbonate rock weathering. Applying DLEM 2.0, Tian et al. (2015) found an increase in DIC export from eastern North America to the Atlantic Ocean from 1901-2008 but no significant trend in DOC or POC. They demonstrated that climate change and increasing atmospheric CO_2 increased long-term C export while land-use change decreased C export. In terms of model performance, they note that the R^2 of observed vs simulated DOC time series was over 0.85 for seven of the river basins in which the model was applied, though they do not provide values of root-mean square error (RMSE).

The ORCHILEAK model (Lauerwald et al., 2017) is a new model branch of ORCHIDEE (Organizing Carbon and Hydrology in Dynamic Ecosystems) (Krinner et al. 2005), the land surface component of the Institut Pierre-Simon Laplace (IPSL) earth system model (ESM) (Fig. 5). It stimulates the production of DOC in the canopy and soils, the leaching of DOC and CO_2 from soils to the river network, DOC mineralization and the subsequent CO_2 evasion from the water surface, as well as DOC and CO_2 export to the coast. It also simulates the exchange of C between litter, soils and water on floodplains and in swamps, and integrates the LOAC within a full representation of the terrestrial C cycling as simulated by ORCHIDEE. Prior to this thesis, ORCHILEAK had been applied to the Amazon basin where it was able to recreate seasonal patterns of observed DOC fluxes and CO_2 evasion across the Amazon basin (Lauerwald et al., 2017) It has also recently been used to estimate long term

perturbation of the integrated LOAC C cycle in the Amazon Basin (Lauerwald et al., in submission), see chapter 3 for further discussion).

The only process-based model to have been applied at the global scale to simulate the spatio-temporal variation in the transport of C along the river network is the recently published TRIPLEX-HYDRA (Li et al., 2019). The model performed moderately well against observations in terms of DOC yield (against annual averages for 26 major river, mean R^2 = 0.61) and was able to explain >50% of the interannual variation in all but three of the rivers. From 1951-2015, they simulated a decreasing trend in DOC flux in the mid-high latitude rivers (30-90°N) and an increasing trend at tropical latitudes (Li et al., 2019). They also highlighted spatial hotspots in DOC yield (g C m^{-2} yr^{-1}) for the present day, the Amazon and Congo Basin being two such regions, as well as some of the large high-latitude basins (Li et al., 2019). TRIPLEX-HYDRA is a simpler mechanistic model compared to ORCHILEAK and DLEM which means that it is arguably easier to apply at the global scale. However, it does not incorporate as many processes as the other two and moreover, it does not meet the criteria for the models used in the annual GCB.

These examples aim to highlight the power of process-based models to quantify the spatio-temporal variation in C fluxes along the LOAC and to better understand and attribute the most important processes driving their short-term variation and longer-term perturbation. However, there are disadvantages associated with process-based models. Firstly, it is more difficult to systematically assess the uncertainty associated with process-based estimates. With empirical models, one can for example, use the statistical parameters (b-estimates) associated with multiple linear regressions as a basis to provide an objective assessment of uncertainty. In process-based models, uncertainty can be assessed against point based or aggregated sets of observations using RMSE for example, but it is more difficult to do this in a systematic way across the entire mode domain because of data limitations. Another option

is to run sensitivity analyses by adjusting model parameters or to apply an uncertainty range to the model forcing files. In studies using multiple mechanistic models, one can also assess the range in multi-model outputs, as is done in for the Inter-Sectoral Impact Model Intercomparison Project (ISIMIP), though currently only two ISIMIP models integrate the LOAC.

More generally process-based models are always vulnerable to criticism for not incorporating all of the important processes; in its current form ORCHILEAK does not account for nutrient limitation for example (see chapters 4 and 5 for further discussion). These models are however, in constant evolution and even incomplete versions arguably represent the best options to understand the long-term dynamics of the LOAC at large scales, and crucially to integrate the aquatic and terrestrial C cycles.

1.5. Contribution of this thesis

Previous research has shown a close relationship between the terrestrial and aquatic C cycle, namely that terrestrial NPP is exported to the aquatic system, driven largely by the hydrological cycle. Moreover, as discussed, a number of hotspot areas of aquatic CO_2 evasion have been identified, and a substantial temporal variation has been demonstrated. However, there remains considerable gaps in our knowledge, particularly associated with understanding and accounting for the temporal and spatial variation of aquatic C fluxes at regional to global scales, which we know from local scale studies, are substantial. This is reflected in the large uncertainty bands associated with global estimates of these fluxes and in the fact that the terrestrial C cycle remains the sink/source with the largest uncertainty in the most recent Global Carbon Budget (Le Quéré at al., 2018[b]). In this thesis we identified three particularly important regional hotspots of the LOAC, with significant gaps in our understanding and designed research questions for each, to address some of these unknowns.

Chapter 2: CO_2 evasion from boreal lakes: Revised estimate, drivers of spatial variability, and future projections

Considerable past research has been undertaken to understand what drives pCO_2 variation in boreal lakes and to quantify the CO_2 evasion fluxes to the atmosphere from the scale of individual lakes to the national scale such as for Sweden (Roehm et al., 2009; Humborg et al., 2010). However, the study of Raymond et al. (2013., Fig. 3 b) remains the only attempt to estimate pCO_2 variation and CO_2 evasion from lakes at the regional to global scale, and this was only performed at the very course resolution of the COSCAT regions (Meybeck et al., 2006) using a relatively simple upscaling methodology, and is associated with large uncertainties. Additionally, no attempt has been made to predict how pCO_2 and CO_2 evasion are likely to change in the future across the boreal region and this is an important research question given that the region is predicted to undergo substantial climate change over the 21st century (Gauthier et al., 2015; Intergovernmental Panel on Climate Change (IPCC), Climate Change, 2013; Koven, 2013; Price et al., 2011).

With these considerations in mind, we address the following research questions in chapter 2:

- Can we develop an empirical model, using similar methods applied for rivers by Lauerwald et al. (2015), which can explain the spatial variation of existing boreal lake pCO_2 observations, using only universally available geodata as dependent driver variables?
- Can we in turn use estimates of gas exchange velocity *k* and lake area, and a state-of-the-art satellite-derived lake area database (Verpoorter et al., 2014), to calculate pan-boreal lake CO_2 evasion?

- Can we use this empirical model, along with future model projections of the drivers, to estimate the evolution of boreal lake pCO_2 and CO_2 evasion over the 21st century under different Representative Concentration Pathways (RCPs)?
- Finally, what are the uncertainties in our calculations and what is the principal source of uncertainty?

Chapter 3: Aquatic carbon fluxes dampen the overall variation of net ecosystem productivity in the Amazon basin: An analysis of the interannual variability in the boundless carbon cycle.

In the tropics, previous research has shown that the terrestrial C cycle varies both seasonally and interannually and is sensitive to the ENSO cycle (Zhao & Running., 2010; Wang et al., 2018) while the same has been demonstrated in regional studies of the Amazon (Gatti et al., 2014; Doughty et al., 2015; Feldpausch et al., 2016). However, with the exception of the Gatti et al. (2014) study which directly measured CO_2 fluxes in the lower troposphere, none of these studies accounted for the aquatic C cycle nor the interaction between it and the terrestrial C cycle, and the Gatti study is limited in that it only covered a two-year period (2010-2011). The LOAC is often dismissed as many argue that the majority of the terrestrial C cycle is effectively closed, i.e. it does not matter if the C is emitted back to the atmosphere as respiration from terra-firme or whether it is transported to the aquatic system before being evaded from the water surface. Either way the C is lost back to the atmosphere, though of course C can also be sedimented and buried in the LOAC or exported to the coast. Separately, research has shown that across the Amazon, aquatic CO_2 evasion is seasonally variable (Richey et al., 2002). More recently, Almeida et al (2017) demonstrated that in addition to seasonal variation, large flood events also drive interannual variation in aquatic CO_2 evasion from the Madeira River (a tributary of the Amazon). However, a major knowledge gap remains in understanding how the aquatic fluxes vary interannually across the basin, and in turn what effect they have on the

overall C cycle of the Amazon. These research questions are particularly pressing given that the region is increasingly vulnerable to extreme climatic events such as droughts and floods (Marengo et al., 2011; Chou et al., 2013; Gloor et al., 2013; Zulkafli et al., 2016).

With these knowledge gaps in mind, we set out to tackle the following research questions in chapter 3 using the ORCHILEAK model (Lauerwald et al., 2017):

- To what extent do the LOAC fluxes (aquatic CO_2 evasion and C export to the coast) vary inter-annually and seasonally throughout the entire Amazon Basin?
- How does interannual variation in discharge and flooding affect the LOAC fluxes, terrestrial NPP, soil heterotrophic respiration (SHR) and ultimately the NEP of the Amazon Basin, particularly in the context of increasing climatic extremes? More specifically, does the incorporation of LOAC fluxes amplify or dampen variation in NEP?

Chapter 4: Historic and future trends of aquatic carbon fluxes integrated within the Congo Basin carbon balance

The Congo has been shown to be a hotspot region for both the terrestrial and aquatic C cycle, as the world's second largest area of contiguous tropical rainforest and second largest river by discharge. In terms of the terrestrial C cycle, the largest intact tropical peatland in the world was recently discovered in the Congo, storing around 30 Pg C alone (Dargie et al., 2017). LOAC fluxes are also estimated to be substantial but again until now have only recently been calculated by upscaling from a spatially and temporally limited dataset using considerable assumptions. Research has shown that terrestrial net primary productivity (NPP) and in turn above ground C storage in living trees, have increased in recent decades in tropical Africa (Yin et al., 2017; Lewis et al., 2009) due in large part to a combination of increasing atmospheric CO_2 concentrations and climate change (Ciais et al., 2009; Pan et al., 2015). Moreover, these

trends are predicted to continue into the future. However, these studies did not incorporate inland water fluxes and we do not know how C export to the LOAC, CO_2 evasion or export to the coast have been perturbed over time, nor how they will change under future climate and land use change scenarios.

With these knowledge gaps in mind, in chapter 4 we aim to address the following research questions:

- To what extent have LOAC fluxes (CO_2 evasion and C export to the coast) in the Congo Basin changed from 1860 to the present day and what are the primary drivers of this change?
- How will these fluxes change under future climate and land use change scenarios (RCP 6.0)?
- What does the temporal evolution of LOAC fluxes mean for the wider C cycle of the Congo Basin?

2. CO_2 evasion from boreal lakes: revised estimate, drivers of spatial variability, and future projections

This chapter is published in Global Change Biology. Full reference: Hastie, A., Lauerwald, R., Weyhenmeyer, G., Sobek, S., Verpoorter, C., & Regnier, P. (2018). CO_2 evasion from boreal lakes: Revised estimate, drivers of spatial variability, and future projections. ***Global Change Biology*****, *24*(2), 711–728. http://doi.org/10.1111/gcb.13902**

Abstract

Lakes (including reservoirs) are an important component of the global carbon (C) cycle, as acknowledged by the 5th assessment report of the IPCC. In the context of lakes, the boreal region is disproportionately important contributing to 27% of the worldwide lake area, despite representing just 14% of global land surface area. In this study, we used a statistical approach to derive a prediction equation for the partial pressure of CO_2 (pCO_2) in lakes as a function of lake area, terrestrial net primary productivity (NPP) and precipitation ($r^2 = 0.56$), and to create the first high resolution, circumboreal map (0.5°) of lake pCO_2. The map of pCO_2 was combined with lake area from the recently published GLOWABO database and three different estimates of the gas transfer velocity k to produce a resulting map of CO_2 evasion (FCO_2). For the boreal region we estimate an average, lake area weighted, pCO_2 of 966 (678- 1325) µatm and a total FCO_2 of 189 (74-347) Tg C yr^{-1}, and evaluate the corresponding uncertainties based on Monte Carlo simulation. Our estimate of FCO_2 is approximately twofold greater than previous estimates, as a result of methodological and data source differences. We use our results along with published estimates of the other C fluxes through inland waters to derive a C budget for the boreal region, and find that FCO_2 from lakes is the most significant flux of the land-ocean aquatic continuum, and of a similar magnitude as emissions from forest fires. Using the model and applying it to spatially resolved projections of terrestrial *NPP* and precipitation while keeping everything else constant, we predict a 107% increase in boreal

lake FCO_2 under emission scenario RCP8.5 by 2100. Our projections are largely driven by increases in terrestrial NPP over the same period, showing the very close connection between the terrestrial and aquatic C cycle.

2.1. Introduction

Lakes (including reservoirs) are an important component of the global carbon (C) cycle, as acknowledged by the 5th assessment report of the IPCC (Ciais *et al*., 2013). Global, regional and local studies commonly report carbon dioxide (CO_2) supersaturation (e.g. Sobek *et al*., 2005; Lapierre & del Giorgio, 2012; Weyhenmeyer *et al*., 2012; Raymond *et al*., 2013), resulting in an evasive flux of CO_2 (FCO_2 in this paper), which on a global scale, equates to approximately 0.32 to 0.64 Pg C yr^{-1} (Cole *et al.*, 1994; Cole *et al*., 2007; Tranvik *et al*., 2009; Aufdenkampe *et al.*, 2011; Raymond *et al.*, 2013; Holgerson & Raymond, 2016). This CO_2 outgassing corresponds to roughly 12-25 % of the total carbon flux mobilised from soils and the bedrock into aquatic systems (Regnier *et al*., 2013).

In the context of lakes, the boreal forest region region (BF, as defined in Potapov *et al.*, 2008) is disproportionately important. According to the satellite-based GLOWABO product (Verpoorter *et al.*, 2014), 1.35 x 10^6 km^2 out of 5 x 10^6 km^2 lakes globally are located in this region, contributing to 27 % of the worldwide lake area, despite the BF representing just 14% of global land surface area. Boreal waters are also predicted to be very sensitive to future climate change; in particular with regard to increasing temperature and terrestrial net primary productivity (*NPP*), as well as mobilisation of C from thawing permafrost soils (Price *et al.*, 2011; IPCC, 2013; Koven, 2013; Gauthier *et al.*, 2015). It is thus important to understand the consequences of future climate change for lake C cycling and lake FCO_2. In order to project future boreal lake CO_2 partial pressure (pCO_2) and FCO_2, we first need robust and spatially resolved estimates for the present day, and we then need to identify the key environmental drivers of lake FCO_2.

The spatial heterogeneity in lake FCO_2 has so far only been assessed using an approach averaging observed local lake values of pCO_2 across a region of interest, from which an evasion rate was calculated and applied to the entire lake surface area of the region. At the national scale for the USA, McDonald *et al*. (2013) and Butman *et al.* (2016) resolved pCO_2 and FCO_2 using 16 and 19 regions, respectively. Similarly, Humborg *et al*. (2010) calculated average pCO_2 values and FCO_2 rates for 5 lake size classes to estimate the total FCO_2 from Swedish lakes, although only a single region was used here. Raymond *et al*. (2013) presented the first global map of lake pCO_2 and FCO_2 for different lake size classes at the resolution of the so-called COSCAT segmentation of Meybeck *et al*. (2006), which subdivides the world's catchments into 231 units. While such regionalization provides robust estimates where data availability is high, they are of less use where observations are scarce or missing. In the case of boreal lakes, large data sets are available for Sweden, Finland, and the southeastern part of Canada. In contrast, data availability is scarce across very large portions of NW America and the Asian continent. Thus, for the regions that lack empirical data, pCO_2 and FCO_2 need to be modelled.

In order to extrapolate pCO_2 to those locations where data is scarce or absent, globally available environmental drivers need to be identified. In previous studies, the spatial variability of lake pCO_2 in high latitude regions has been linked to a wide range of variables including dissolved or total organic carbon (DOC/TOC) concentration in lake water (*e.g.* Roehm *et al.,* 2009; Humborg *et al.,* 2010; Lapierre & del Giorgio, 2012; Weyhenmeyer *et al*., 2012 for most recent contributions), lake area and depth (Sobek *et al.*, 2003; Kortelainen *et al*., 2006; Roehm *et al.,* 2009; Humborg *et al.*, 2010; Raymond *et al*., 2013), dissolved inorganic carbon (DIC) input from the catchment (Maberly *et al.,* 2012; Weyhenmeyer *et al*., 2015; Perga *et al.,* 2016), chlorophyll-a concentration (Kortelainen *et al.,* 2006; Roehm *et al*., 2009; Kortelainen *et al.,* 2013; Maberly *et al*., 2012; Perga *et al.,* 2016), and precipitation (Sobek *et al.*, 2003; Rantakari

& Kortelainen, 2005). While hydrochemical and physical variables observed in the field give valuable insights into in-lake processes controlling pCO_2 and its short-term variability, they are of limited use for extrapolations, as lake pCO_2 can only be predicted where these variables have been locally observed. Considering the vast number of lakes in the boreal zone, the proportion that is covered by sampling programs is small. Therefore, geodata sets of potential environmental controls related to climate, terrain, geology, and vegetation, which cover the global landmass in a consistent way, are better alternatives for large-scale assessments.

In 2013 Lauerwald *et al.* showed that a significant proportion of the spatial variability in river pCO_2 in North America could be explained by catchment variables derived from geodatabases. Using a multiple regression analysis, they found that 43% (r^2 =0.43) of the river pCO_2 variability across North America is related to annual mean precipitation, annual mean air temperature and mean catchment slope gradient. A similar approach was later applied to derive global maps of river pCO_2 and FCO_2 (Lauerwald *et al*., 2015). A multiple regression using terrestrial NPP, population density, mean catchment slope gradient, as well as mean air temperature at the sampling location was able to explain 47% (r^2 =0.47) of the spatial variability in global river pCO_2. The map of predicted pCO_2 was then combined with estimates of stream surface area and gas exchange velocity *k* to derive a map of global FCO_2 from rivers at 0.5° resolution.

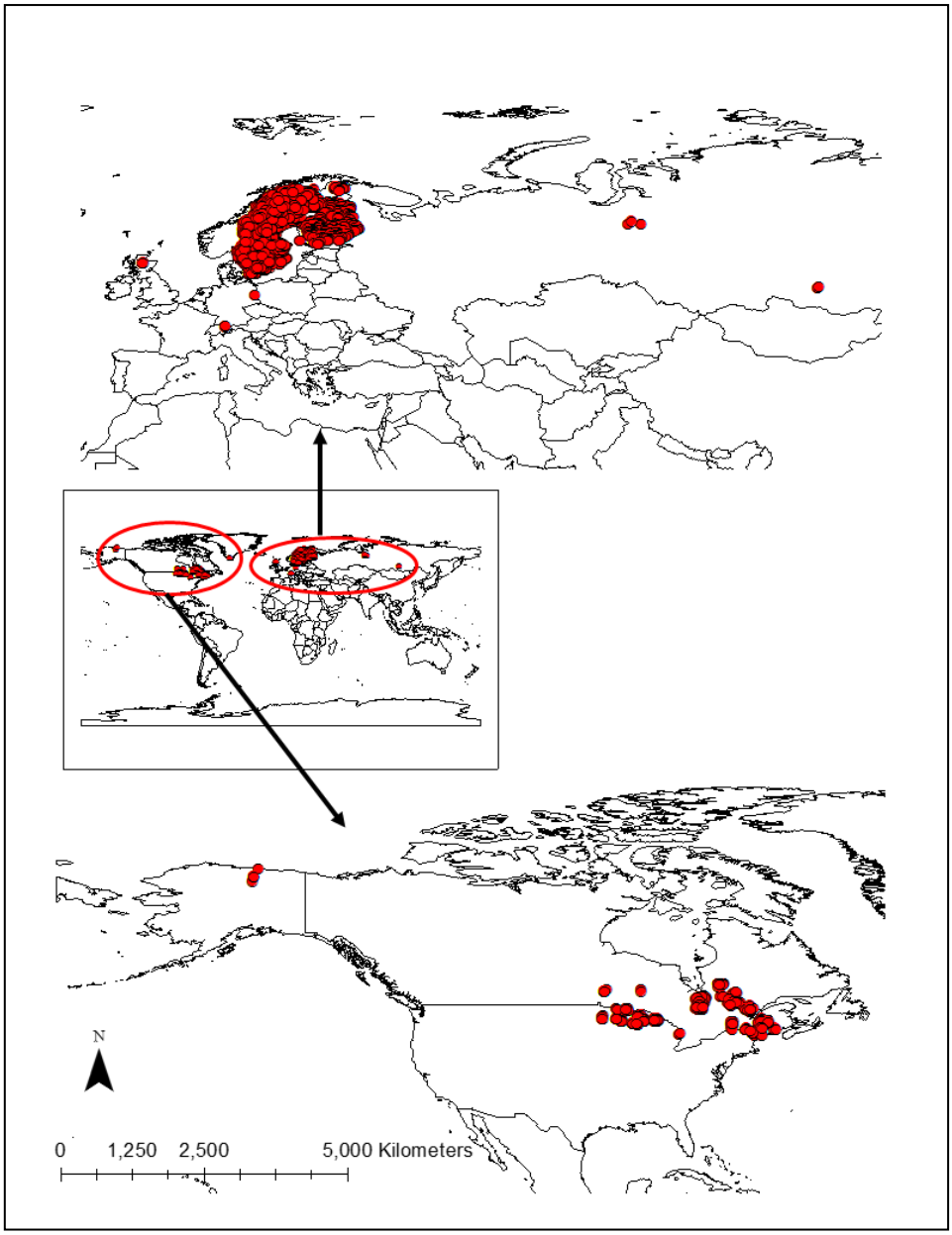
In this study, we used globally available environmental drivers derived from geodatabases to predict a 0.5° map of lake pCO_2 for the entire BF biome from a limited, spatially lumped set of sampling data. The map of pCO_2 was then combined with lake area from the GLOWABO database (Verpoorter *et al.*, 2014) and estimates of gas transfer velocity *k* to produce the first high-resolution map of boreal lake FCO_2. We then merged our new map with the spatially resolved river FCO_2 from Lauerwald *et al*. (2015), lateral land-ocean C exports (GlobalNews, Mayorga *et al.*, 2010) and lake C burial (Heathcote *et al.,* 2015), to derive a full aquatic carbon

budget for the circumboreal region. Finally, based on spatially resolved projections of terrestrial *NPP* and precipitation, we used the model to predict changes in lake pCO_2 and FCO_2 over the 21st century.

2.2. Materials and methods

2.2.1. Lake pCO_2 data

Lake chemistry data was collated from a number of different databases pertaining to distinct regions (Fig. 1): 1) Swedish lake pCO_2 data were taken from the Swedish lake chemistry database published in Weyhenmeyer *et al*. (2012), 2) pCO_2 data from a number of other boreal regions (in particular the whole of Scandinavia and North America) were taken from a global lake pCO_2 database published in Sobek *et al.* (2005), 3) Canadian lake pCO_2 data were taken from Lapierre & del Giorgio. (2012), and 4) Siberian data was taken from Shirokova *et al.* (2013). The Siberian data were reported in µmol L^{-1} and converted into partial pressure using Henry's constant, adjusted for water temperature, following Telmer & Veizer (1999).



Chapter 2 Figure 1:Geographical distribution of measured lake pCO_2 data (n= ~27,000 samples).

Approximately 99% of the pCO_2 values were calculated from measured alkalinity, pH, and water temperature. Samples with a pH ≤ 5.4 were discarded because calculating pCO_2 from alkalinity, pH, and water temperature is highly error prone in low-pH and low-alkalinity waters (Raymond *et al.*, 2013, Abril *et al.*, 2015). A recent study in Sweden (Wallin *et al.*, 2014) found a slight overestimation of the average pCO_2 in samples calculated with the alkalinity-based method. The majority of these samples had low alkalinity (< 0.07 meq L^{-1}). As such, we

compared the median and average pCO_2 values in our dataset above and below this alkalinity threshold. In fact, in our data the samples with an alkalinity > 0.07 meq L^{-1} had a marginally higher median/ average pCO_2 than those ≤ 0.07 meq L^{-1}. We further investigated the effect of alkalinity on the indirect pCO_2 data using simple linear regression. A significant but very weak correlation ($r = -0.0308$, $p < 0.05$, $N = 26530$) was found between the indirectly observed pCO_2 data and alkalinity (after discarding samples with pH ≤ 5.4). Therefore, we conclude that using a pH threshold of 5.4 is a sufficient criterion for filtering out unreliable data. Moreover, another study in Sweden and Finland (Denfeld *et al.*, 2015) did not find a statistically significant difference between directly and indirectly measured pCO_2 data (Wilcoxon each pair test: $p > 0.05$). The remaining 1% of the samples (which includes 25% of North American data) corresponded to direct pCO_2 measurements (see Sobek *et al*., 2005 and Lapierre & del Giorgio, 2012 for more details).

In order to remove pCO_2 data from lakes potentially covered by ice, where possible, only data measured at a water temperature greater than 4 °C and between the months of April to November were retained in the analysis. For a small number of samples (< 1%), water temperature was not reported and in these cases all data form April- November were retained. After the data selection procedure described above, approximately 27,000 pCO_2 samples were retained. See discussion for further deliberation on the issue of ice cover.

2.2.2 Data aggregation

Based on coordinates and lake name, median pCO_2 was calculated per sampling location. Then, using the median pCO_2 for all sampling locations within each 0.5° grid cell, a spatially-aggregated mean pCO_2 was calculated at this resolution. This produced a single 'open water' lake pCO_2 value for 584 grid cells located within the circumboreal region. In order to prevent a geographical bias and to have a similar number of pCO_2 grids representing Scandinavia and North America, only Scandinavian grids with a mean pCO_2 calculated from at least 22 median

pCO_2 values (in other words 22 sampled lakes) were retained in our training data. This rule did not apply to the grids containing the very largest lakes in the Scandinavian dataset, which were included regardless of the number of samples used in data aggregation to ensure that they are represented. Indeed, the largest lakes in the boreal region are larger than the area of a 0.5° grid cell, meaning that in some cases it is not possible to have more than one lake in a grid. Given the relative sparsity of the North American data, all of the North American grids were retained regardless of the number of lakes represented in each grid. This selective procedure left 168 grids evenly divided between Scandinavia and North America, as well as 4 grids in Siberia. Lake area data was also aggregated to a resolution of 0.5°, and $\log_{10}$ transformed prior to computing the mean $\log_{10}$ lake area per grid.

2.2.3 Predictors of pCO_2

A set of environmental and climatic variables was selected and sourced, largely from publicly available geodatabases (Table 1). The choice of retained variables was guided by two principles: firstly, we only chose data that were global in coverage with a resolution of at least 0.5°; secondly, we prioritized variables that have previously been shown to drive the variation of pCO_2 in inland water bodies. Among all variables, pCO_2 and average slope gradient of the catchment clearly showed a skewed distribution, and were thus log-transformed before regression analysis.

Firstly, the Pearson correlation coefficients between $\log_{10}(pCO_2)$ and all of the variables, as well as between variables themselves, were analyzed using the software package STATISTICA™. Secondly, we fitted multiple linear regression models with all possible combinations of three of the 17 predictors described in Table 1 using the software package R (R Core Team 2013). Initially, we placed no limit on the number of retained predictors, but found that adding a fourth predictor added little in the way of additional descriptive power. Therefore, we placed a limit of three predictors to ensure a parsimonious model.

Chapter 2 Table 1: Environmental geodata evaluated as potential predictors of log10(pCO_2) (training data): Basic statistics

Variable	Unit	Mean	Median	Min	Max	Source	Resolution
Lake area (A_{lake})	km²	10.95	0.52	0.01	82200.00	Sobek *et al*., *2005;* Weyhenmeyer *et al*., 2012; Lapierre & del Giorgio 2012	30"
Air temperature (*T*, Apr-Nov monthly mean)	°C	9.01	9.30	1.20	12.80	Hijmans *et al*., 2005	30"
Precipitation (*P*, Apr-Nov monthly mean)	mm $month^{-1}$	71.90	73.46	41.61	100.99	Hijmans *et al*., 2005	30"
Wind speed (Apr-Nov monthly mean)	[m s^{-1}]	3.94	3.83	2.09	5.89	Hijmans *et al*., 2005	30"
Soil pH index (top 5cm)	—	5.14	5.12	4.68	6.18	SoilGrids., 2014	30"
Soil carbon content (top 5cm)	g kg^{-1}	94.99	94.99	63.54	126.44	SoilGrids., 2014	30"
Terrestrial net primary productivity (*NPP*)	g C m^{-2} yr^{-1}	486.73	504.15	60.20	746.89	Zhao *et al*., 2005	30"
Population density	Inh. Km^{-2}	3.05	2.35	0.02	23.79	CIESIN & CIAT 2005	2.5"
Elevation	m	240.45	226.60	18.60	973.81	GLOBE, 1999	30"
Catchment slope gradient	Degrees	0.988	0.670	0.0945	8.744	Lauerwald *et al*., 2015 (GLOBE, 1999)	30"
Runoff	mm yr^{-1}	379.88	359.54	98.00	929.29	Fekete *et al*., 2002	30"
% cover evergreen trees	—	28	29	0	74	Global Land Cover., 2000	2"
% cover mixed trees	—	22	18	0	83	Global Land Cover., 2000	2"
% cover cultivated areas	—	10	02	0	88	Global Land Cover., 2000	2"
% silt of soil	—	34	34	0	43	Harmonized World Soil Database., 2000	30"
% sand of soil	—	52	53	0	62	Harmonized World Soil Database., 2000	30"
% clay of soil	—	13	13	0	21	Harmonized World Soil Database., 2000	30"

2.2.4 Upscaling of pCO_2 data

The fitted regression equation was applied in a geographical information system (GIS) to build a high-resolution map (0.5°) of pCO_2 from the drivers selected from the statistical treatment of the geodata. Terrestrial NPP was taken from the MODIS satellite derived dataset described in Zhao *et al*. (2005), while precipitation was taken from Hijmans *et al*. (2005), a high-resolution database based on interpolation of global weather station data. Log_{10} lake area (and in turn pCO_2) was adjusted to be representative for the total lake area within each grid, and hence the total gas exchange flux through the aquatic-atmosphere interface, using the following equation:

$$Adj.A_{lake} = \frac{\sum Log10(A_{lake}) * A_{lake}}{\sum A_{lake}} (per\ grid) \quad (1)$$

where A_{lake} denotes lake area

Lake area data from the literature (Sobek *et al.*, 2005; Lapierre & del Giorgio 2012; Weyhenmeyer *et al*., 2012; Shirokova *et al.*, 2013) was used for training the statistical model (i.e. lake area as a predictor of pCO_2) while lake area from the GLOWABO database (Verpoorter *et al.*, 2014) was used for extrapolation of pCO_2 and calculation of FCO_2. This is because each aggregated pCO_2 value is only representative of those lakes, which happen to have been sampled in that specific grid. When we extrapolate to the circumboreal scale, we use the GLOWABO database so that estimated pCO_2 is representative of the average size of all of the lakes in each grid. GLOWABO is a global inventory of lakes, which was developed by applying a lake extraction algorithm to high-resolution stationary satellite imagery (Verpoorter *et al*., 2012). As the database is stationary, it relies on an algorithm to filter the satellite images to minimize false detection due to mountain and cloud shadows.

The regression model was validated with 131 of the discarded grids from Scandinavia, selecting only those aggregated from a minimum of 10 samples per grid. All of the validation data were aggregated to the 0.5° grid scale before being compared to predicted pCO_2.

2.2.5 Calculation of FCO_2

FCO_2 was calculated using the equation:

$$FCO_2 = A_{lake} * k * \Delta CO_2 \quad (2)$$

where FCO_2 is in moles d^{-1}, A_{lake} (m^2) denotes lake area, k (m d^{-1}) the gas exchange coefficient, and ΔCO_2 (mol m^{-3}) difference between water and air pCO_2, assuming an atmospheric pCO_2 of 390 μatm. In equation 2, lake pCO_2 is computed according to the procedure described in

the preceding sections. pCO_2 was then converted to CO_2 concentrations using Henry's constant K_H, corrected for temperature.

Lake area was calculated from the GLOWABO database (Verpoorter *et al.*, 2014). The temperature adjusted k was derived from k_{600} (k for CO_2 at 20°C in freshwater- Schmidt number of 600) for each 0.5° grid using the Schmidt number, calculated from the mean water temperature over the April-November period. Water temperature was in turn calculated from mean air temperature over the same period using the equation reported by Lauerwald *et al.* (2015):

$$T_{water}[^\circ C] = 3.941 + 0.818\, T_{air}\,[^\circ C] \quad (R^2 = 0.88) \tag{3}$$

There are numerous methods for estimating k_{600} (m d^{-1}) in the literature. In this study, we have compared estimated boreal lake FCO_2 using three different methodologies for calculating k_{600}. The first method is based on lake area, where four different k_{600} values (0.54, 1.16, 1.32 and 1.90 m d^{-1}) were applied for four corresponding lake area bins (<0.1, 0.1-1, 1-10 and >10 km^2). This approach is taken from Raymond *et al.* (2013), following the relationship between k_{600} and lake area proposed by Read *et al.* (2012). The second method is based on the relationship between k_{600} and wind speed given by Cole & Caraco (1998):

$$k_{600} = 2.07 + 0.251 U_{10}^{1.7} \tag{4}$$

where U_{10} is the average wind speed in m s^{-1} at 10m (Hijmans *et al.*, 2005). This gives a k_{600} range between 0.58 m d^{-1} and 2.04 m d^{-1}.

The third is taken from Vachon & Praire (2013) and incorporates the effects of both wind speed and lake area:

$$k_{600} = 2.51 + 1.48 * U_{10} + 0.39 * U_{10} * log_{10} A_{lake} \tag{5}$$

where A_{lake} is in km^2 and U_{10} in m s^{-1}.

k_{600} has also been found to depend on buoyancy flux in addition to wind speed and lake area. MacIntyre *et al*. (2010) found that higher k_{600} values occur during convective cooling compared to heating, and that during overnight low wind conditions, k_{600} depends on buoyancy flux rather than wind speed. We were unable to take these additional considerations into account as we lacked the necessary data. Despite this, we are confident that with the use of three methods to calculate k_{600}, and by explicitly incorporating its variation in our Monte Carlo analysis, we adequately account for the uncertainty associated with gas exchange velocity.

FCO_2 was then converted to g C yr^{-1} by multiplying by the molar mass of carbon (12.01 g mol^{-1}) and the number of days per year, 365. FCO_2, was first calculated per grid and for the four lake size categories, <0.1, 0.1-1, 1-10 and >10 km^2 described previously (Read *et al.*, 2012), before being amalgamated.

Total FCO_2 was calculated by summing FCO_2 from each size class. After pCO_2 and FCO_2 had been extrapolated, various masks were applied in ArcGIS in order to estimate values for different regions and countries. The boreal forest land cover region (BF) was taken from Potapov *et al.* (2008) (Fig. S13).

2.2.6 Future projections of boreal pCO_2 and FCO_2

We acquired data from four Earth System Models used in the 5th Coupled Model Intercomparison Project (CMIP5) to predict pCO_2 and FCO_2 over the 21st century for the IPCC scenarios RCP2.6 and RCP8.5 (IPCC, 2013): the Canadian Centre for Climate Modeling and Analysis (CCCma) CanESM2 model, the Met Office (UKMO) HadGEM2-ES model, the Institute Pierre-Simon Laplace (IPSL) IPSL-CM5A-LR model and the Max Planck Institute for Meteorology (MPI) MPIESM-LR model. We selected these firstly, as they are all used in the CMIP5 project, and secondly because projected values of the environmental drivers of pCO_2 identified for the present day (terrestrial NPP and precipitation), as well as air

temperature for adjusting future K_H, and k via the Schmidt number (as described in preceding section), were easily accessible online. RCP2.6 an RCP8.5 will lead to radiative forcing levels of 2.6 and 8.5 Watts per square meter (Wm^{-2}) respectively by the year 2100. RCP8.5 is characterised by very high energy intensity as a result of high population growth and a comparatively low rate of technological development while RCP2.6 is characterised by very low energy intensity, declining oil consumption and the use of carbon capture and storage technologies (CCS) (van Vuuren et al., 2011). The data were downloaded from the Earth System Grid Federation (ESGF) node hosted by the IPSL, for RCP2.6 and RCP8.5 model runs, as well as historical runs (1850-2005). All data were taken from the r1i1p1 realization of the simulations. In order to account for inter-annual variation, the data were aggregated in time to obtain projections for 10 year-periods centered on the years 2000, 2030, 2050 and 2100. For example, total annual terrestrial NPP for the year 2030 was based on aggregating across the years 2025-2034. Some of the models only provide projections up to the year 2100 so for this specific period we aggregated the data across the preceding 10 years, 2090-2099. After aggregation, the projections of terrestrial NPP, air temperature and precipitation were aggregated in a GIS to ensure a uniform spatial resolution of 0.5°. Finally, for each 0.5° grid, the simulated terrestrial *NPP*, air temperature and precipitation was adjusted using the difference between modelled and observed data for the 2000 period. For example, the terrestrial *NPP* for the year 2030 was adjusted according to:

$$NPP_{2030,A} = NPP_{2000,O} + (NPP_{2030,M} - NPP_{2000,M}) \quad (6)$$

where $_A$ denotes adjusted, $_O$ denotes observed and $_M$ denotes modelled. $NPP_{2000,O}$ was taken from Zhao *et al*., (2005), which was also used for training the prediction equation.

This procedure is similar to the one applied in regional downscaling of future projections, to reflect the fact the Earth System models are mostly designed to predict future changes and not to capture spatial patterns at the (sub)-regional scale.

After calculating future pCO_2, we then added the projected increase in atmospheric pCO_2 (above the present-day value of 390 μatm used in this analysis) for the equivalent scenario and year. For example, under RCP8.5 atmospheric CO_2 is projected to be 936 μatm by 2100 (Meinshausen *et al.*, 2011), and thus we added 546 (936-390) to our projected estimate of boreal lake pCO_2 for the same year and scenario. With this approach, we account for the fact that we do not use present-day atmospheric pCO_2 in our prediction equation of lake pCO_2, but we assume it to be implicitly represented in the intercept of that equation, and we further assume that the increase in atmospheric pCO_2 will lead to a corresponding, total increase in lake pCO_2, preserving the total delta pCO_2. In our calculation of FCO_2 no such step is required as it is calculated from delta pCO_2, not absolute lake pCO_2.

2.2.7 Uncertainty estimates based on Monte Carlo simulation

We calculated the uncertainty associated with our pCO_2 and FCO_2 estimates (5^{th} and 95^{th} percentiles) using a Monte Carlo simulation comprising 10,000 runs. Here we calculated a probability density function for pCO_2 based on varying the b estimates for each of the three predictors (terrestrial NPP, precipitation and lake area) retained in the multiple linear regression, assuming a normal distribution constrained by the standard errors of each b estimate (Table 4). Uncertainty associated with our future projections of pCO_2 and FCO_2 were similarly calculated using a Monte Carlo simulation forced by future projections of terrestrial NPP and precipitation. However, to account for the variation of terrestrial NPP and precipitation projections resulting from using four different Earth System models, we performed 4x 2,500 runs (2,500 for each model). Uncertainty was propagated to the calculation of FCO_2 using equation (2), where k values were randomly selected between the minimum and maximum

values calculated for each grid by the three methods outlined in the previous section. 10,000 Monte-Carlo runs were again performed, assuming a uniform distribution of *k*. It was also assumed that pCO_2 and *k* vary independently of one another. The Monte Carlo analyses were performed using the statistical software R 3.2.2 (R Core Team, 2015).

2.3. Results

2.3.1 Controls of spatial variation in pCO_2

Table 2 presents the correlations between aggregated (0.5°) $\log_{10}(pCO_2)$ and 17 variables derived from environmental geodata (Table 2).

Highly significant ($p < 0.001$) positive correlations were found for wind speed (Apr-Nov monthly mean) ($r = 0.58$), annual terrestrial net primary productivity (terrestrial *NPP*) ($r = 0.35$), air temperature (*T*, Apr-Nov monthly mean) ($r = 0.34$), population density ($r = 0.46$), the percentage of the grid covered by needle- leaved evergreen trees (GLC land cover class 4) ($r = 0.36$) and by cultivated or managed land (GLC 16) ($r = 0.32$), while highly significant negative correlations were found with $\log_{10}(A_{lake})$ ($r = -0.51$), precipitation (*P*, April to November monthly mean) ($r = -0.42$), elevation ($r = -0.30$) and runoff ($r = -0.33$).

In order to ensure a parsimonious prediction model, we restricted the maximum number of drivers to three. All possible combinations of drivers were tested in the regression analysis. Based on this analysis, the model with both the highest r^2 and lowest Root-mean-square error (RMSE) included $\log_{10}(A_{lake})$, *P* and *T*, as predictors (Table 3). However, we chose to proceed with the second-best performing model, composed of $\log_{10}(A_{lake})$, terrestrial *NPP* and *P*. This choice is justified by the more direct mechanistic link between lake pCO_2 and terrestrial *NPP* compared to *T*. *T* is a proxy for many different variables and controls a variety of different processes in multiple ways, meaning that *T* related patterns at the spatial scale are not necessarily transferrable to the temporal scale (Weyhenmeyer *et al*., 2016). In addition, the

allochthonous C inputs to lakes, as related to terrestrial *NPP*, have been shown to be a stronger constraint on CO_2 evasion from boreal lakes than the direct temperature effect on decomposition rates (Kortelainen *et al*., 2006). Accordingly, terrestrial *NPP* has been identified as a main control of CO_2 emission from lakes (Maberly *et al*., 2012), underpinning its strong mechanistic linkage to lake FCO_2. It is likely that the effects of terrestrial *NPP* are statistically represented by *T* and *P* in the training data, and that is one explanation for terrestrial *NPP* not being retained in the equation with the lowest RMSE. The relationships between *P*, *T* and terrestrial *NPP* are, however, more complex than the statistics for the training data suggest, and the empirical relations between the three predictors in the training data are not necessarily representative of the extrapolation area or prediction periods. Thus, we assume the combination of terrestrial *NPP*, *P* and $\log_{10}(A_{lake})$ to be more robust for extrapolations in space and time. Together, these predictors explain 56% ($r^2 = 0.56$, Fig. 2) of the spatial variation in $\log_{10}(pCO_2)$. *P* has the strongest partial correlation coefficient (r = -0.535) followed by $\log_{10}(A_{lake})$ (r = -0.407) and terrestrial *NPP* (r = 0.360) (Table 4). The resulting prediction of boreal lake pCO_2 is given by the following equation:

$$log_{10}\,(pCO_2\,[\mu atm]) \;=\; 3.26 \pm 0.07 - 0.0844 \pm 0.0115 * log_{10}(A_{lake}\,[km^2]) \;+ 6.89 \pm 1.11 * NPP\,[10^4 g\ C\ m^{-2} yr^{-1}] - 8.30 \pm\ 0.84 *\ P\,[10^3 mm]\ (r^2 \;=\; 0.56).$$

where the ranges ($\pm$) represent the standard errors.

Fig. 2 shows the scatterplot of observed *versus* (vs) predicted $\log_{10}(pCO_2)$, categorized by region. Siberia is represented by only a handful of grids with pCO_2 data. Despite this geographical bias, Fig. 2 shows that there is no consistent pCO_2 prediction bias in the Siberian grids compared to those of Scandinavia and North America.

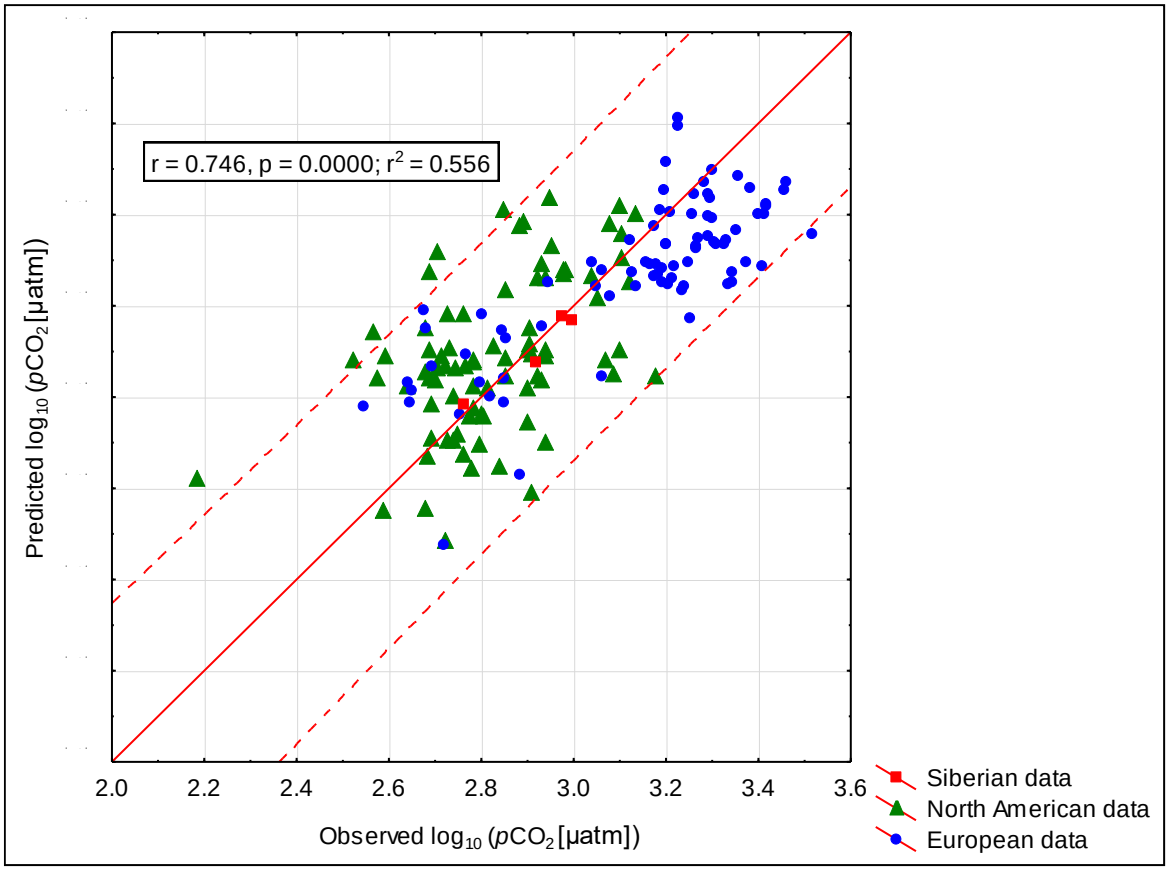
Chapter 2 Table 2: Pearson correlation (r) between the variables tested

	P	$\log_{10}(A_{lake})$	Wind speed	Terrestrial *NPP*	Soil pH	*T*	$\log_{10}$(slope gradient)	Elevation	Pop. density	Runoff	Soil org. C content	% cover evergreen trees	% cover mixed trees	% cover cultivated areas	% silt of soil	% sand of soil	% clay of soil
$\log_{10}(pCO_2)$	-0.42***	-0.51***	0.58***	0.35***	-0.17*	0.34***	0.11	-0.30***	0.46***	-0.33***	0.041	0.36***	0.029	0.32***	-0.022	0.14	-0.18*
P		-0.043	-0.25***	0.28***	-0.11	0.33***	0.057	0.30***	-0.25**	0.74***	-0.22**	-0.42***	0.30***	-0.17*	0.10	-0.093	-0.0018
$\log_{10}(A_{lake})$			-0.42***	-0.35***	0.092	-0.26***	-0.073	-0.13	-0.20**	0.10	0.17*	0.00	-0.15	-0.24**	-0.19*	-0.22**	-0.015
Wind speed				0.28***	0.08	0.29***	-0.31***	-0.20**	0.29**	-0.30***	-0.078	0.17*	-0.08	0.25***	-0.28***	0.25**	-0.057
Terrestrial *NPP*					-0.011	0.84***	-0.037	-0.24**	0.31***	0.018	-0.47***	-0.15	0.33***	0.22**	-0.11	0.10	-0.08
Soil pH						-0.07	-0.28***	-0.033	0.12	-0.14	-0.58***	-0.53***	-0.16*	0.26***	-0.25**	-0.45***	0.56***
T							-0.084	-0.41***	0.43***	0.10	-0.57***	-0.11	0.29***	0.38***	-0.15*	0.033	-0.10
$\log_{10}$(slope gradient)								0.33***	0.03	0.29***	0.042	-0.080	0.17*	-0.061	0.51***	-0.24**	-0.37***
Elevation									-0.48***	0.15*	-0.043	-0.23**	0.074	-0.30***	0.15	0.026	-0.19*
Pop. density										-0.084	-0.17*	0.08	-0.04	0.43***	-0.037	0.021	0.18*
Runoff											-0.055	-0.30***	0.12	-0.21**	0.27***	-0.13	-0.056
Soil org. C content												0.64***	-0.25**	-0.26***	0.011	0.43***	-0.19*
% cover evergreen trees													-0.42***	-0.10	-0.16*	0.43***	-0.34***
% cover mixed trees														0.06	0.34***	-0.16*	0.0092
% cover cultivated areas															-0.052	0.012	0.15
% silt of soil																-0.13	0.0034
% sand of soil																	-0.18*

*** p<0.001 ** p<0.01, *p<0.05, p>0.05

Chapter 2 Table 3: The top ten ranking multilinear regression equations composed of three drivers. Shown in descending order of ability (r^2) to predict the dependent variable $log_{10}(pCO_2)$

Predictors	r^2	Root-mean-square error (RMSE)
$log_{10}(A_{lake}\ [km^2])$, *P* [mm],*T* [°C]	0.59	0.16
$log_{10}(A_{lake}\ [km^2])$, *P* [mm], terrestriaL *NPP* [g C m^{-2} yr^{-1}]	0.56	0.17
$log_{10}(A_{lake}\ [km^2])$, *P* [mm], Wind speed [m s^{-1}]	0.53	0.18
$log_{10}(A_{lake}\ [km^2])$, Wind speed [m s^{-1}], *T* [°C]	0.53	0.18
$log_{10}(A_{lake}\ [km^2])$, *P* [mm], Pop. density [Inh. Km^{-2}]	0.52	0.18
P [mm], terrestrial *NPP* [g C m^{-2} yr^{-1}], Wind speed [m s^{-1}]	0.52	0.18
$log_{10}(A_{lake}\ [km^2])$, *P* [mm], Elevation [m]	0.52	0.18
Wind speed [m s^{-1}], Elevation [m], log_{10}(slope gradient [Degrees])	0.51	0.18
Wind speed [m s^{-1}], log_{10} (slope gradient [Degrees]), % cover evergreen trees	0.51	0.18
$log_{10}(A_{lake}\ [km^2])$, Wind speed [m s^{-1}],% cover evergreen trees	0.50	0.18



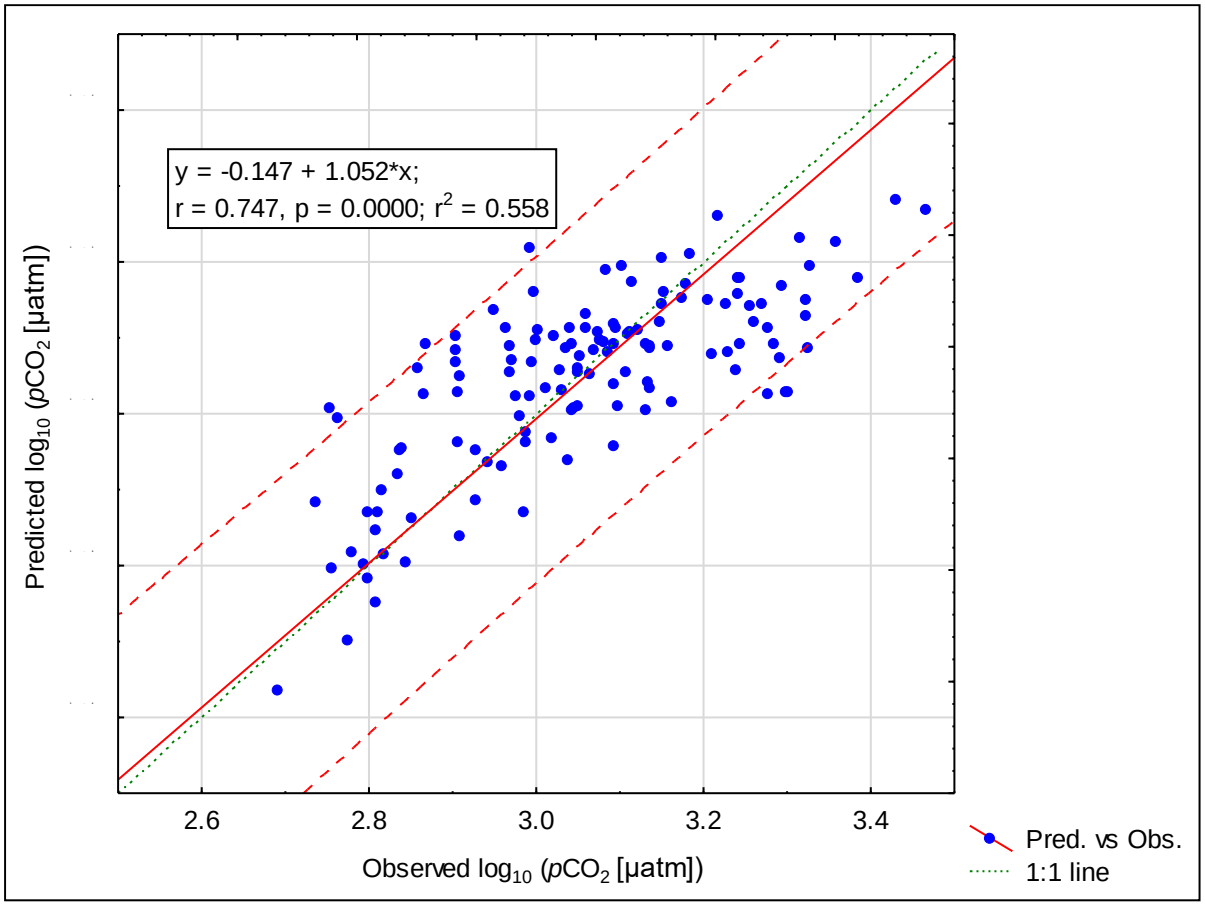
Chapter 2 Figure 2: Observed vs predicted $\log_{10}(pCO_2$ [µatm]) for model training (168 grids) categorised by region with 95% prediction interval (dashed lines)

Chapter 2 Table 4: Retained predictors with b-estimates, associated standard errors, and partial correlations to the dependent variable $\log_{10}(pCO_2)$

Predictor	b-Estimate	Standard Error	p-Value	Partial Correlation
Intercept	3.26	0.0696	<0.0001	
$log_{10}(A_{lake})$	-0.0844	0.0115	<0.0001	-0.407
Terrestrial *NPP* [10^4 g C m^{-2} yr^{-1}]	6.89	1.11	<0.0001	0.360
P [10^3 mm]	-8.30	0.841	<0.0001	-0.535

The regression model was validated with 131 of the discarded grids from Scandinavia. The resulting scatterplot (Fig. 3) has a slope of 1.052 and an intercept of -0.148, suggesting that

there is minimal prediction bias. Based on the multiple regression model described above, lake area weighted mean pCO_2 for the BF land cover region is estimated to be 966 µatm (Fig. 4 a) with an uncertainty range of 678- 1325 µatm corresponding to the 5th and 95th percentiles. The range of uncertainty was obtained from the Monte-Carlo analysis, outlined in the methods. Mean pCO_2 varies inversely with lakes size category (Table 5). Indeed, our estimate of mean pCO_2 for the smallest lake size category (<0.1km^2) of 1558 µatm (1110-2208) is approximately twice that of our estimate for the largest category (>10km^2) of 789 µatm (563-1120).

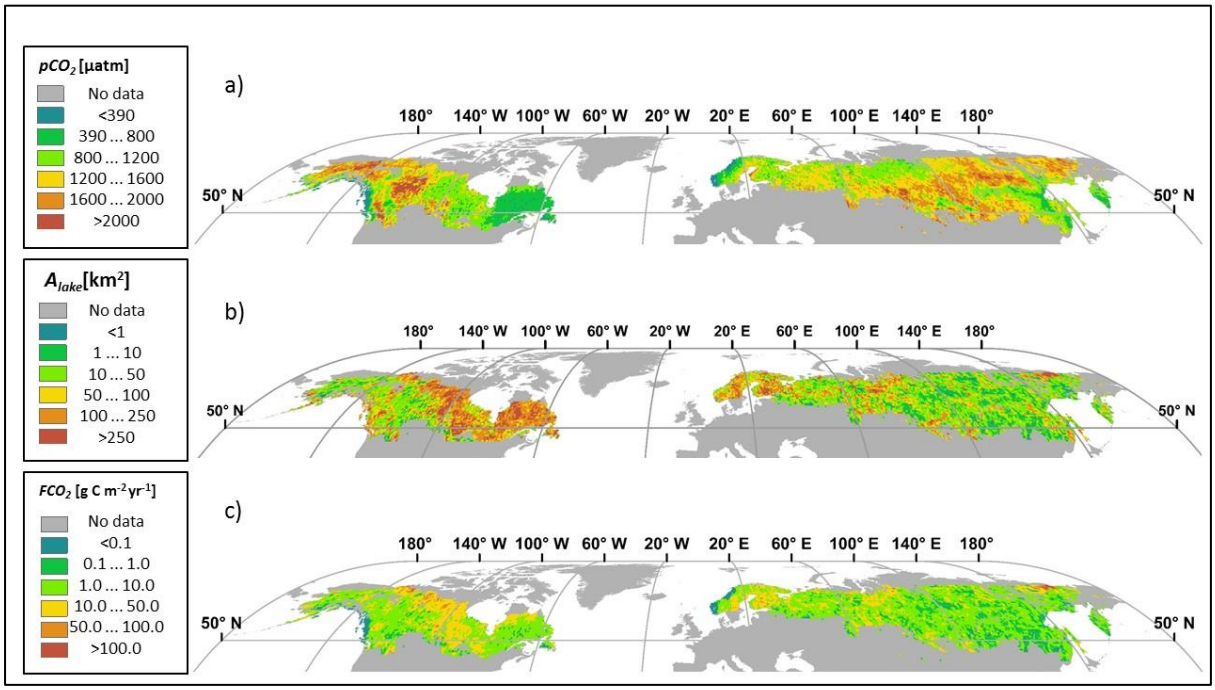


Chapter 2 Figure 3: Observed vs predicted log10(pCO_2 [µatm]) for model validation (131 grids) with 95% prediction interval (dashed lines)

2.3.2 Estimates of FCO_2 for present-day conditions

The map of FCO_2 (Fig. 4c) show a complex spatial pattern reflecting the high spatial variation of both pCO_2 (Fig. 4 a) and A_{lake} (Fig. 4 b). Integrated over the BF region, we estimated a total FCO_2 of 189 Tg C yr^{-1} (range of 74-347 Tg C yr^{-1}) while for the entire 50°- 70° N latitudinal

band (Fig. S12) we estimated a total evasion of 272 (115-487) Tg C yr^{-1}. Canada alone showed the highest FCO_2 with 137 (55-250) Tg C yr^{-1} (Table S2), mainly because Canada has the greatest total A_{lake}. We found a relatively even contribution for the four different lake size categories to total FCO_2, and this is a reflection of the contrasting relationships between lake area and pCO_2, and lake area and gas exchange velocity k. While the smallest lakes had the highest estimated pCO_2 values, they also had the lowest values of k, and this pattern is reversed in the largest lakes (Table 5).



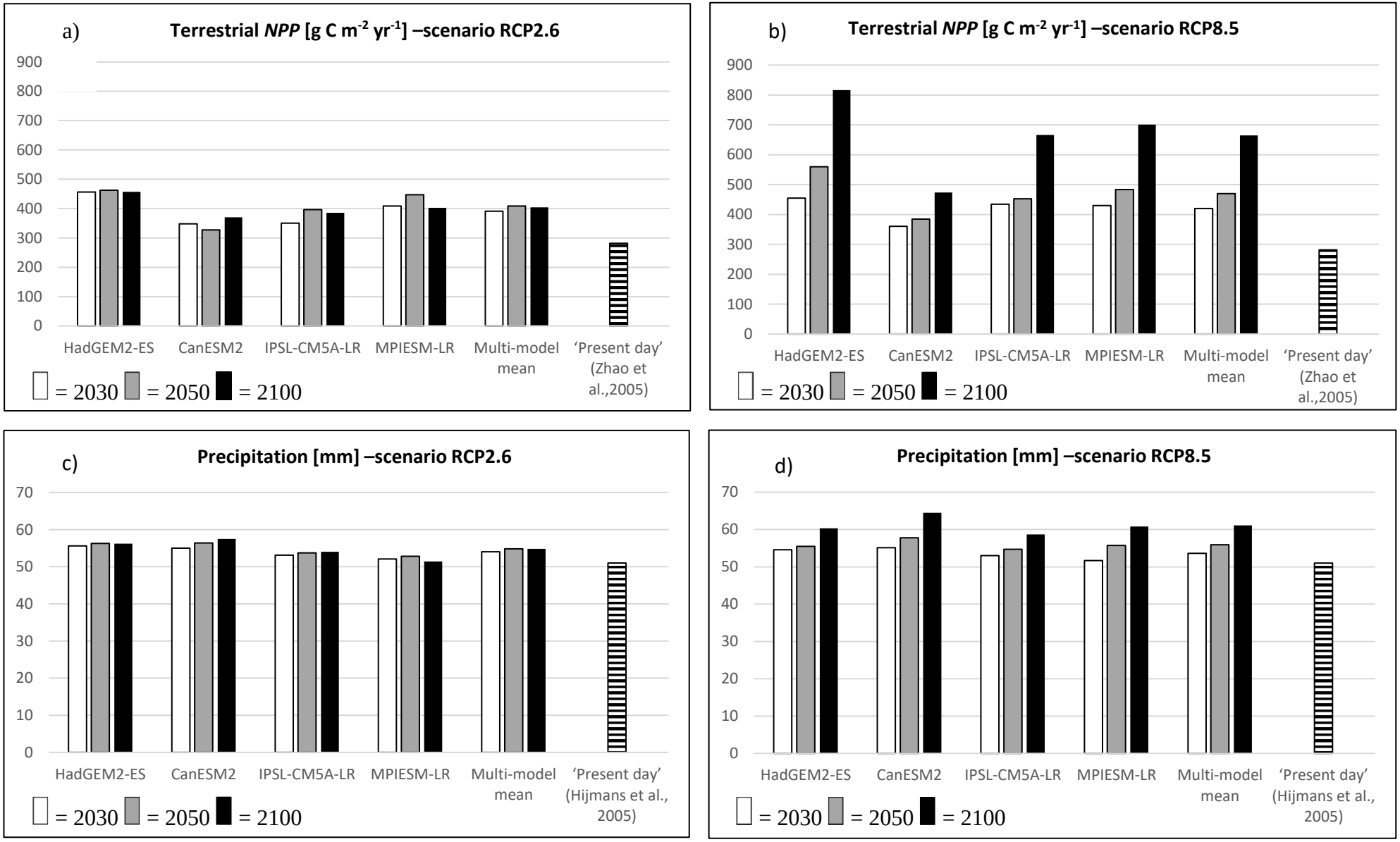
Chapter 2 Figure 4: Predicted maps of (a) pCO_2, (b) A_{lake}, and (c) FCO_2 for the BF land cover region

Chapter 2 Table 5; pCO_2, FCO_2, total A_{lake} and k values in relation to lake size classes for the BF land cover region

Lake size class	pCO_2 (µatm)	FCO_2 (Tg C yr^{-1})	Total A_{lake} (km^2)	Mean k (m d^{-1})
<0.1km^2	1558 (1110-2208)	38.17 (18.64-65.00)	208,008	0.63
0.1-1km^2	1237 (890-1739)	52.36 (9.38-87.1)	282,434	0.87
1-10km^2	1020 (734-1434)	44.80 (19.54-79.34)	286,624	0.99
>10km^2	789 (563-1120)	53.96 (9.91-116.5)	570,583	1.27

2.3.3 Projections of pCO_2 and FCO_2

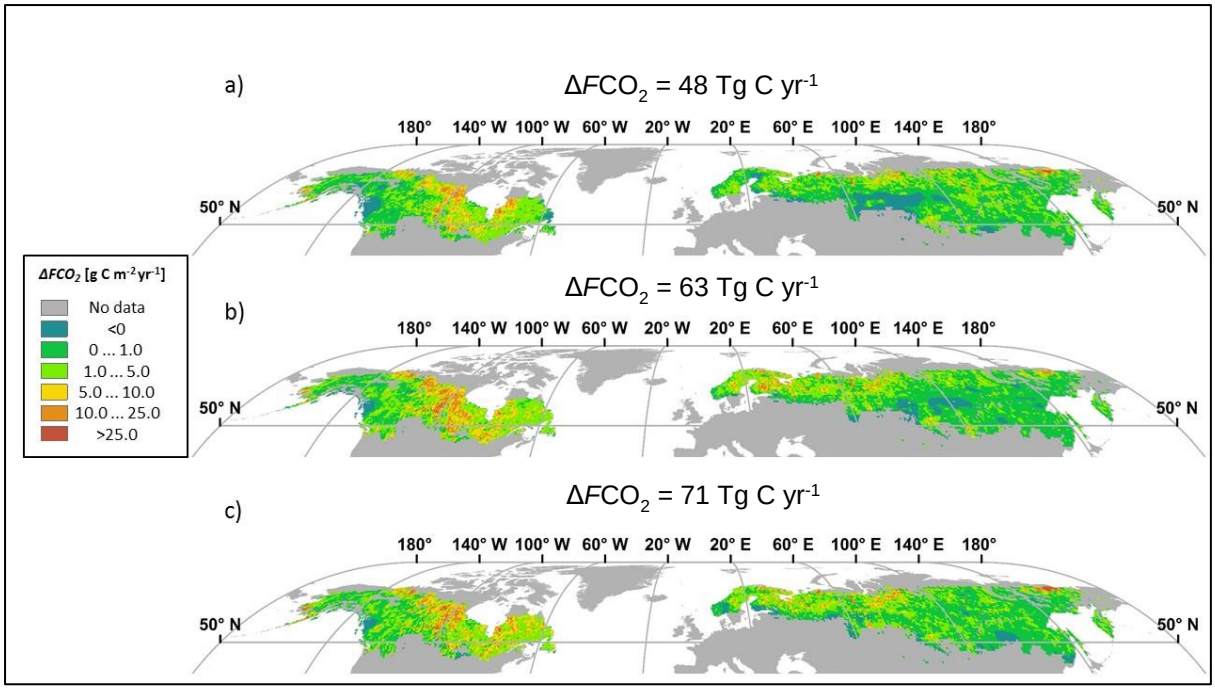
We used our empirical model and projections of terrestrial *NPP* and *P* to estimate the change in BF lake pCO_2 and FCO_2 over the 21st century, under two future GHG emission scenarios, namely the lowest (RCP2.6) and highest (RCP8.5) emission scenario of the IPCC. Additionally, we used future projections of *T*, to adjust K_H and *k*, and in turn FCO_2. Based on the multi-model mean, we estimated that annual terrestrial *NPP* will increase by 135% from 282 to 664 g C m^{-2} yr^{-1} (Fig. 5 b) by 2100 under RCP8.5, while we estimated that *P* (April-Nov monthly mean) will increase by 20% from 51 mm to 61 mm (Fig. 5 d). For the years 2030 and 2050 respectively, we estimated that terrestrial *NPP* will increase by 49% and 67% and that *P* will increase by 5% and 10%. Under RCP2.6 (Fig. 5 a, c), we estimated a 45% increase in terrestrial *NPP* by 2050 to 408.5 g C m^{-2} yr^{-1} before reducing slightly to 404 g C m^{-2} yr^{-1} by 2100. *P* is predicted to increase by just 7.5% to 54.8 mm for both the years 2050 and 2100. In both scenarios (RCP2.6 and 8.5), and across all three time periods (2030, 2050 and 2100), all four models (CanESM2, HadGEM2-ES, IPSL-CM5A-LR & MPIESM-LR) project an increase in both terrestrial *NPP* and *P* across the BF region, relative to the year 2000 (Fig. 5 a-d). The HadGEM2-ES model consistently projects the largest increase in terrestrial *NPP* while the CanESM2 model consistently projects the smallest increase. In contrast, the CanESM2 model projects the largest increase in *P* across all scenarios and time periods, with the exception of RCP2.6 for the year 2030 (for which HadGEM2-ES projects the highest *P*). In all our projections, we have assumed that A_{lake} will remain constant.



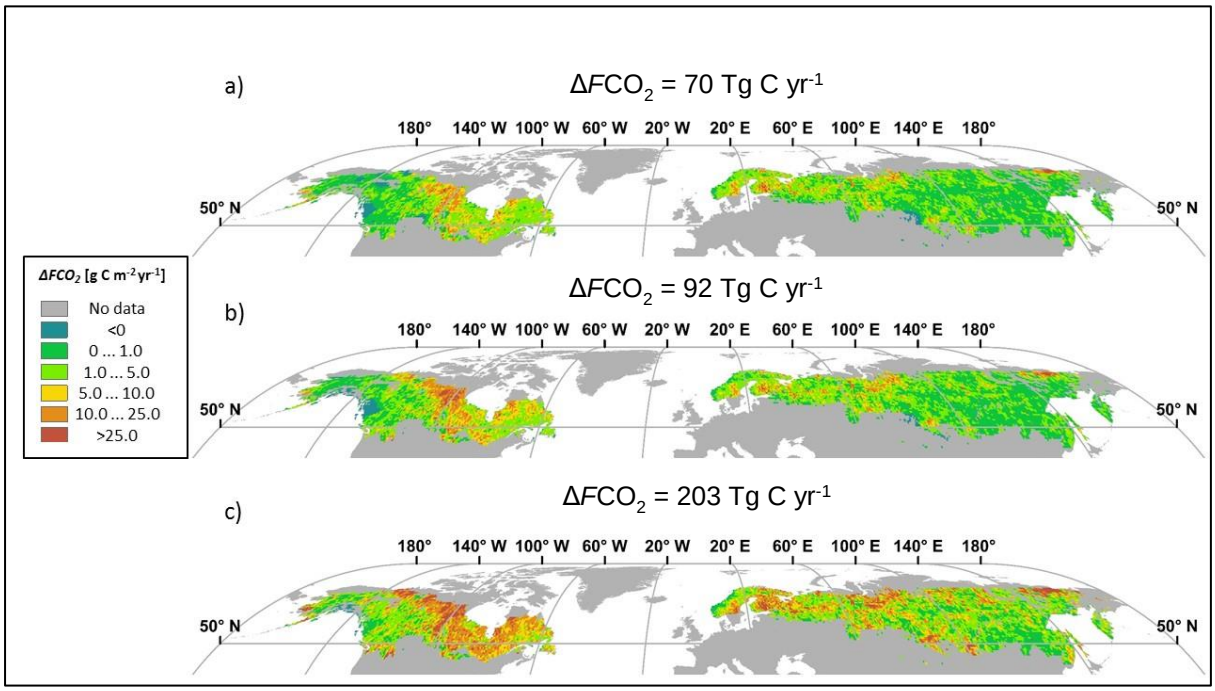
Chapter 2 Figure 5:Projected BF terrestrial NPP under scenario (a) RCP2.6 (b) RCP8.5 and P under scenario (c) RCP2.6 (d)RCP8

Under scenario RCP8.5, both pCO_2 (Table S3) and FCO_2 (Fig. 7) are predicted to increase substantially across the BF region, peaking at 2198 (1303-3761) µatm and 392 (96.5-922) Tg C yr^{-1} respectively for the year 2100. This represents a 127.5% increase in pCO_2 and a 107% increase in FCO_2. For the years 2030 and 2050 respectively, we estimate that pCO_2 will increase by 31% and 48% and that FCO_2 will increase by 37% and 49%. For RCP2.6 pCO_2 and FCO_2 (Table S3, Fig. 6) are estimated at 1230 (739-1981) µatm and 260 (82.9-537) Tg C yr^{-1} respectively for the year 2100. This represents a 27% increase in pCO_2 and a 38% increase in FCO_2. Note that the increases in pCO_2 and FCO_2 are not equivalent, largely due to the fact that the former is a reflection of absolute pCO_2 while the latter is calculated from delta pCO_2.

The projected increase in lake CO_2 evasion is mainly driven by the substantial projected increase in the positive driver terrestrial *NPP* over the 21st century. In contrast, the projected increase in *P*, a negative driver of pCO_2 in our model, is very small. The smaller increase in both pCO_2 and FCO_2 predicted under the RCP2.6 scenario is reflected in the smaller increase in terrestrial *NPP* (Fig. 5) under this scenario. By applying a linear function between our FCO_2 values for 2000, 2030, 2050 and 2100, we estimate a cumulative CO_2 evasion flux of around 29 Pg C from boreal lakes to the atmosphere over the course of the 21st century, under RCP8.5. Under RCP2.6, we estimate a smaller cumulative flux of 24 Pg C. See Table S1 in Supporting information for more detailed results of future FCO_2 projections relating to different options for the adjustment of k and K_H.



Chapter 2 Figure 6:Spatially resolved (0.5°) predicted change in CO_2 evasion, ΔFCO_2 (from a year 2000 baseline) under scenario RCP2.6 for the year (a) 2030, (b) 2050 and (c) 2100 for the BF land cover region.



Chapter 2 Figure 7:Spatially resolved (0.5°) predicted change in CO_2 evasion, ΔFCO_2 (from a year 2000 baseline) under scenario RCP8.5 for the year (a) 2030, (b) 2050 and (c) 2100 for the BF land cover region.

2.4. Discussion

2.4.1 Drivers of pCO_2 and FCO_2 spatial variability

We found significant relationships between lake pCO_2 and a variety of environmental drivers (Table 2). The majority of these relationships concur with existing literature, i.e. the observed positive correlation between terrestrial *NPP* and pCO_2 is in line with previous studies, which report on CO_2 supersaturation in boreal and temperate lakes as a result of allochthonous inputs of organic carbon (OC) and inorganic carbon (IC) from the catchment (Maberly *et al.,* 2013; Weyhenmeyer *et al*., 2015; Wilkinson *et al*., 2016). Similarly, the presence of boreal coniferous forest has previously been shown to exert a strong positive control on pCO_2 by inducing a net heterotrophic state from elevated DOC concentrations (Hanson *et al*., 2003; Sobek *et al*., 2007; Chmiel *et al*., 2016). There is less of a clear consensus regarding the effects of temperature and wind. Several previous studies have found pCO_2 and organic C mineralization in lakes to be strongly positively linked to temperature (Marotta *et al,*. 2009; Gudasz *et al.*, 2010; Kosten *et al*., 2010), while other studies have found only a weak relationship (Sobek *et al.*, 2005; Sobek *et al*., 2007; Lapierre *et al.*, 2015). In our study, temperature had a moderately positive effect on lake pCO_2, similar to that of terrestrial *NPP*, and indeed the two are highly intercorrelated ($r = 0.84$). At present, it is not clear in how far our observed positive relationship between *T* and pCO_2 is related to increased aquatic respiration or other in-lake processes or to increased terrestrial *NPP*. Since terrestrial *NPP* was almost equally powerful in explaining pCO_2 in lakes as *T*, we built our predictive model on terrestrial *NPP* to get a better mechanistic understanding.

Wind speed has previously been shown to exert a strong negative control on pCO_2 in reservoirs (Morales-Pineda *et al*., 2014), the likely mechanism being higher wind speeds leading to higher FCO_2 (Cole & Caraco, 1998; Reed *et al*., 2012; Vachod & Praire, 2013) and in turn a decrease in pCO_2. In contrast, our study revealed a strong positive correlation between wind speed and lake pCO_2. This could be explained by higher wind speeds promoting the vertical mixing of

waters, especially during summer in lakes where a thermocline has formed. Nevertheless, wind speed was not retained for the empirical model, because the b-estimate was associated with a very high standard error (>20% of the b-estimate).

In terms of the negative controls of pCO_2, it is well established in the literature that smaller lakes generally have higher pCO_2 values, due to their proportionately greater surface area in contact with the catchment and greater allocthonous C inputs per unit volume (Sobek *et al.*, 2003; Kortelainen *et al.*, 2006; Humborg *et al.*, 2010; Catalan *et al*., 2016). Perhaps the most interesting result is the strong negative control of *P* on pCO_2. Although, previous large-scale studies have found a positive relationship between *P* and open water pCO_2 (Sobek *et al*., 2003; Rantakari & Kortelainen, 2005), a recent temporal study over a 17-year study period found a negative or no relation at all between precipitation and pCO_2 in boreal inland waters (Nydahl *et al*., 2017). Nydahl *et al.* (2017) suggested that increased precipitation results in a dilution of CO_2 concentrations in inland waters due to an altered balance between surface and CO_2-rich groundwater flow. In addition, *P* induced increased surface water runoff can cause a faster water flushing through the landscape giving less time for *in situ* CO_2 production in inland waters. It is however important to keep in mind that *P* is highly intercorrelated with a number of the other variables tested, most notably elevation, % of coniferous tree cover per grid, population density and wind speed. As such, it may be that the relationship with *P* is also representing the effects of these environmental and physical drivers on pCO_2.

Our maps show a high degree of spatial variation and a complex pattern of pCO_2, reflecting the fact that no single driver is dominant. One region where a clear divergence in pCO_2 can be observed is Scandinavia. For instance, we estimated an average area weighted pCO_2 of 949 (637-1345) µatm for Sweden and a substantially smaller value of 552 (372-779) µatm for Norway (Table S2), as a result of the differing topography and climate found in the two countries. Due to its close proximity to the sea and relatively high mean elevation, Norway

receives a substantially greater amount of monthly precipitation (87 mm as a monthly mean over the April-November period) compared to Sweden (58 mm as a monthly mean over the April-November period), and a lower annual terrestrial *NPP* of 208 g C m^{-2} yr^{-1} compared to Sweden's total of 373 g C m^{-2} yr^{-1} (Fig. S14 a, c). Another region where a relatively strong pattern can be seen is in Quebec, where low pCO_2 (Fig. 4 a) coincides with low terrestrial *NPP* (Fig. S14 a) and relatively high precipitation (Fig. S14 c). The spatial pattern in FCO_2 is even more complex, because the hotspots of pCO_2 generally do not coincide with those of A_{lake}.

2.4.2 Comparison to previous global studies

In Table 6, we compare our results of FCO_2 and pCO_2 to values found in the literature, averaged across the boreal region. For an extended table with additional regional breakdowns of results, please refer to Table S2 of Supporting information. Our estimate of total FCO_2 of 189 (74-347) Tg C yr^{-1} from lakes in the BF region is substantially higher (by a factor of nearly 2.5) than the estimate of 79 Tg C yr^{-1} proposed by Raymond *et al.* (2013) for the same region. Our estimate of FCO_2 over the 50°-70°N latitudinal band is also higher than the two previous estimates of Raymond *et al*. (103 Tg C yr^{-1}) and Aufdenkampe *et al*. (110 Tg C yr^{-1}) by a comparable factor.

There are several explanations for our relatively high estimates of FCO_2. One substantial difference in our study is the incorporation of the new GLOWABO lake database. Across the BF region, the GLOWABO database contains a total A_{lake} of 1,350,353 km^2, compared to the total of 931,619 km^2 estimated by Raymond *et al.* (2013). We calculated an area specific FCO_2 of 140 g C m^{-2} yr^{-1}, which is still 64% larger than that of Raymond *et al*. (2013). Indeed using Raymond's value of total A_{lake}, we would calculate a total FCO_2 of 130 Tg C yr^{-1}. Therefore, total A_{lake} is not the only reason for the substantial difference in FCO_2 between the two studies. The greater number of the smallest, high pCO_2 lakes in GLOWABO compared to previous methods (see Verpoorter *et al.*, 2014) is another plausible explanation for our high estimate.

In comparison to Raymond *et al.* (2013), we also used a substantially different methodology, as well as different data to train our model. We used additional boreal pCO_2 in the training of our model from Canada (Lapierre & del Giorgio, 2012), Sweden (Weyhenmeyer *et al*., 2012) and Siberia (Shirokova *et al*., 2013). Our methodology for estimating *k* also differed compared to previous studies. We used the same two methodologies for deriving *k* as Raymond *et al.* (2013) but added an additional method outlined in Vachon and Prairie (2013), which led to slightly higher *k* values.

Using only the two methods for calculating *k* (Cole & Caraco, 1998; Read *et al.*, 2012)

used in Raymond *et al.* (2013) we obtain a total BF evasion of 150 (67-258) Tg C yr^{-1}, which gives an area specific CO_2 evasion rate of 111 g C m^{-2} yr^{-1}. If we multiply this flux density by the total A_{lake} from Raymond *et al*. (2013) we reach a total evasion of 104 Tg C yr^{-1}. Thus we conclude that the remaining discrepancy of 25 Tg C yr^{-1} between our results and those of Raymond *et al*. is due to methodological and pCO_2 data differences.

Chapter 2 Table 6: pCO_2, FCO_2, total A_{lake} and *k* values compared to previous studies

Region	pCO_2 (µatm)	FCO_2 (Tg C yr^{-1})	Total A_{lake} (km^2)	Mean *k* (m d^{-1})	Source
BF	1278 Area weighted 966 (678-1325)	189 (74-347)	1,350,353	0.86	This study
BF	790	79	931,619	0.82	Raymond *et al*., 2013 (INTERPOLATED)
50-70°N	1305 Area weighted 1006 (715-1366)	272 (115-487)	1,751,985	0.88	This study

50-70°N	812	103	1,194,701	0.84	Raymond *et al.*, 2013 (INTERPOLATED)
50-90°N*	1100	110	80,000-1,650,000	0.96	Aufdenkampe *et al.*, 2011

2.4.3 Sources of uncertainty

2.4.3.1. Upscaling

Using a statistical model to extrapolate pCO_2 in regions of minimal data coverage is suitable if the variation in the environmental parameters in the predictor equation is similar in the training areas and in the extrapolated areas. This condition is fulfilled in our study (see Fig. S6-S10 in Supporting information) where 99.6% of the variation in extrapolated terrestrial *NPP* and 98.8% of the variation in extrapolated *P* lies within the range recorded in the grids used for training. We are thus confident that we are not extrapolating too far beyond the statistical model boundaries. However, it is important to note that the mean values of both terrestrial *NPP* and *P* are substantially higher over the grids used in training the data compared to the mean values over the entire extrapolated region. For the training data mean terrestrial *NPP* and *P* are 477 g C m^{-2} yr^{-1} and 71 mm respectively, compared to 282 g C m^{-2} yr^{-1} and 51 mm across the BF land cover region.

Our mean estimated pCO_2 across the extrapolated BF region of 1278 µatm is higher than the value of 1133 µatm observed in our training data but more importantly, the vast majority of the variation in our extrapolated pCO_2 lies within the range of observed pCO_2. While the minimum pCO_2 of 25.6 µatm over our extrapolated grids is lower than the minimum of 152 µatm in the observed grids, the maximum value over our extrapolated grids is also lower (Fig. S3 and Fig. S4 of Supporting information), resulting in a smaller pCO_2 range over the extrapolated grids. Moreover, reducing the number of grids in our analysis from 584 to 168 could have resulted in certain geographical/ climatic areas being underrepresented but grids from the vast majority of

boreal latitudes remained after this edit and therefore we are confident that most of the variation in the original data is retained. In addition, the variation in pCO_2, as well as terrestrial *NPP* and *P*, is similar across the 584 grids and the 168 grids (Fig. S3-S11 of Supporting information).

In calculating annual FCO_2 across the boreal region, we multiplied our daily evasion estimates by number of days per year irrespective of location and associated ice cover duration. This choice is guided by the fact that significant CO_2 accumulation has been previously reported under ice covered lakes and very high emissions during ice melt (Striegl *et al.,* 2001). Such findings concur with our own preliminary analysis of the seasonality of pCO_2 at individual sampling locations, where peak pCO_2 values were often measured during spring before April and at temperatures below 4 °C. These conditions were excluded from our analysis as we restricted our dataset to samples measured at a water temperature greater than 4 °C and between the months of April to November. Moreover, a disproportionate percentage (45%) of our raw data was sampled during the summer (July-Sept) and this data had a median value of 997 µatm. We can compare this to the spring (April-June) data with a median value of 1416 µatm, which comprised just 20% of our data, or the annual median value of 1478 µatm (all data samples including winter and < 4 °C), and conclude that the data used in our final analysis likely leads to a conservative estimate of pCO_2. This choice compensates somewhat for the lack of accounting for variable ice cover duration in our estimation of FCO_2. As discussed in the methods, we only included pCO_2 data with a pH of > 5.4 in order to filter out unreliable data. However, this could also lead to underestimation of pCO_2. Note that Raymond *et al.,* (2013) used the same approach, meaning that the results from both studies can be compared.

2.4.3.2 Lake area

There are a number of limitations associated with GLOWABO. Despite the use of a number of filters to minimize errors, some false detection of lakes due to cloud and mountain shadow is

unavoidable. Other sources of errors include the elucidation of lakes from large rivers and wetlands.

While these limitations are significant, validation of GLOWABO against a high resolution map of Sweden (Verpoorter *et al*., 2012), an area which encounters all of the aforementioned problems, achieved a performance index of 91% for lake area, while lake number differed by less than 3% (see Verpoorter *et al*., 2012 for further discussion).

2.4.3.3 Gas exchange velocity k

Gas exchange velocity *k* represents one of the largest sources of uncertainty. We assessed this uncertainty by using three different methods to calculate *k* and we reported a best estimate as the average of these three *k* quantification methods. We further accounted for this uncertainty by incorporating *k* into the Monte Carlo analysis. Additionally, in order to assess the uncertainty associated with the variation in *k* alone, we undertook an extra Monte Carlo analysis in which we only accounted for variation in *k* (that is, uncertainty associated with pCO_2 calculation was excluded). Based on this analysis, the mean FCO_2 is 185 Tg C yr^{-1} for the BF region, very close to the 189 Tg C yr^{-1} estimated in the original Monte Carlo analysis. The range of uncertainty is only moderately smaller; we estimate 5th and 95th percentile FCO_2 at 98 Tg C yr^{-1} and 297 Tg C yr^{-1} respectively, compared to the original range of 74-347 Tg C yr^{-1}. Thus, we conclude that *k* is indeed the largest source of uncertainty in our calculation of FCO_2.

2.4.3.4 Future changes in lake CO_2 evasion

Our study does not account for future changes in the extent of the boreal forest, predicted as a result of increasing temperature (Koven, 2013; Gauthier *et al.*, 2015). However, recent research estimating future changes in the boreal C stock under scenario RCP4.5, suggests that any C gained from northern expansion of the boreal forest is likely to be offset by loss from southern boreal retreat, and thus little net change is predicted (Gauthier *et al.*, 2015). Finally, our study

does not account for the future impact of permafrost thaw, which will become an increasingly important source of C. Drake *et al*. (2015) report that by 2100, between 5 and 10 Tg C will be released annually from Yedoma soils alone.

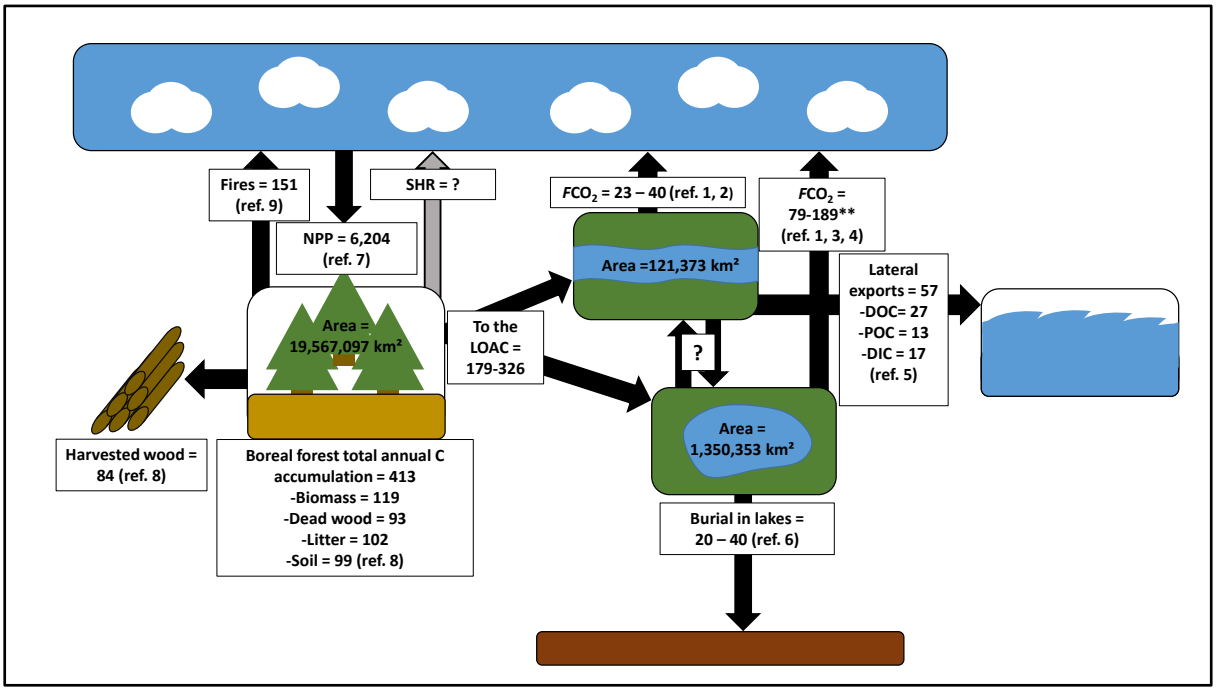
2.4.4 Present and future carbon budget for the boreal region

This study is the first to spatially resolve lake pCO_2 and FCO_2 across the boreal region, moreover using only environmental drivers derived from freely available geodata. The resolution of our maps is compatible with most global land surface and inversion models (Ciais *et al*., 2013), and thus could potentially be used for validation purposes. High resolution estimates of C fluxes such as those reported here are crucial in deriving more reliable regional C budgets, particularly along the land-ocean aquatic continuum (LOAC) where large uncertainties remain (Regnier *et al.,* 2013). Figure 8 integrates our results within a C budget for the boreal region using previous spatially resolved estimates of terrestrial *NPP* (Zhao *et al.,* 2005), FCO_2 in rivers (Lauerwald *et al*., 2015), C burial in lake sediments (Heathcote *et al*., 2015), lateral C exports to the ocean (Mayorga *et al.,* 2010), C accumulation in forests (Pan *et al*., 2011), and emissions from fires (van der Werf *et al*., 2017, in review). Our updated budget suggests that lakes are the most significant contributor to the LOAC budget. This is largely due to their substantially greater surface area; we estimate that it is 11 times that of rivers in the BF region. Moreover, we estimate that in the order of 3-5 % of the C fixed by terrestrial vegetation (terrestrial *NPP*) is leaking each year into inland water bodies. This value is comparable to the global estimate of 3.2% proposed by Regnier *et al.,* (2013) and the 5% recently calculated for the Amazon basin by Lauerwald *et al*. (2017), which ignores the lateral mobilization of POC. Interestingly, the magnitude of the LOAC C flux is of the same order as the mean C storage in the boreal forest biomass and soils combined. It is also greater than the vertical flux as a result of boreal forest fires (van der Werf *et al*., 2017, in review) and the lateral C flux from harvested wood (Pan *et al*., 2011). Our findings imply that the leakage through the LOAC considerably

reduces the C accumulation in boreal forests. This could particularly be true for Canada where our estimate of lake FCO_2 alone of 137 Tg C yr^{-1} is substantially larger than the mean (1990-2007) C storage of 20 Tg C yr^{-1} for the Canadian boreal forested proposed by Pan *et al*., (2011). Our budget is also likely to be conservative given that we do not account for methane (CH_4) fluxes. A recent study (Rasilo *et al*., 2015) of 224 lakes in Quebec found that as much as 25% of the emissions from lakes, in terms of atmospheric warming potential, are in the form of CH_4. Moreover, there are a small number of additional C fluxes contributing to the Net Ecosystem Exchange budget, such as emissions associated with the consumption of crop products, which we do not include but are of relatively minor importance (Ciais *et al*., in review).

We estimate that lake pCO_2 and FCO_2 will increase substantially over the 21st century relative to our present-day estimates. Under RCP8.5, we predict a 37%, 49% and 107% increase in boreal lake FCO_2 by 2030, 2050 and 2100 respectively, amounting to a cumulative perturbation of the lake to atmosphere CO_2 flux of about 9 Pg C over the 21st century. This is a significant perturbation, of a similar magnitude to predicted future changes in boreal soil organic C stocks in some land C models (Nishina *et al*., 2014). Our projections are largely driven by increases in terrestrial *NPP* of 46%, 67% and 135% over the same period. Interestingly, even under the GHG scenario RCP2.6, we predict a 25%, 33% and 38% increase in boreal lake FCO_2 by 2030, 2050 and 2100 respectively. This suggests that a substantial strengthening of the CO_2 evasion flux from boreal lakes is expected irrespective of the emission scenario. Our results concur with those of Larsen *et al*. (2011), which projected a 65% increase in TOC concentration in Norwegian lakes by 2100 under the superseded IPCC B2 scenario, an intermediate GHG emission scenario. In our study, NPP increases at an equivalent rate while the increase in precipitation is much smaller, meaning that the proportion of *NPP* lost from lakes to the atmosphere remains relatively constant at approximately 3% under both scenarios. Finally, our estimates of future FCO_2 are likely to be conservative, due to our lack of accounting for the

impact of permafrost thaw on remobilizing, old labile C. Accounting for this substantial source of future C should be prioritized in future studies of boreal and high latitude regions, and may require the explicit representation of these processes in mechanistic Earth System models.



Chapter 2 Figure 8 Updated carbon budget along the land-ocean aquatic continuum (LOAC) for the boreal region. Units are Tg C yr^{-1}. Ref. 1- Aufdenkampe et al. (2011), ref. 2- Lauerwald et al. (2015), ref. 3- Raymond et al. (2013), ref. 4 (this study), ref. 5- Mayorga et al. (2010), ref. 6- Heathcote et al. (2015), ref. 7- Zhao et al. (2005), ref. 8- Pan et al. (2011), ref. 9- van der Werf et al. (2017, in review). This scheme does not include estuarine C fluxes which are relatively minor in this region (Laruelle et al., 2013), or the C fluxes between lakes and rivers for which no estimate could be found for the boreal region.**

2.5 Supporting information

2.5.1. Comparison to previous regional studies

As with the boreal region as a whole, our estimates of FCO_2 for individual countries are higher than previous values in the literature. For Sweden, we estimated a total FCO_2 of 7.9 (2.8-15.3) Tg C yr^{-1}, approximately 4.5 times greater than the value of 1.74 Tg C yr^{-1} estimated in Humborg *et al.* (2010) and roughly 3 times greater than the value of 2.55 Tg C yr^{-1} that is obtained by interpolating the results of Raymond *et al.* (2013) over the same area. In Humborg *et al.* (2010), a low and uniform *k* value for lakes of 0.21 m d^{-1} was used, as well as a lower

total lake area (Table S2). Despite our smaller average estimate for pCO_2, the combination of higher and spatially variable *k* values and a larger lake surface area results in a substantially higher FCO_2 for Sweden. Interestingly, our estimate of total FCO_2 for Norway of 0.84 (0.53-2.61) Tg C yr^{-1} is very close to the downscaled value of 0.81 Tg C yr^{-1} by Raymond et al. (2013). The substantially lower average pCO_2 predicted in our study is here balanced by our larger total lake surface area. In fact, Raymond *et al*. (2013) estimated a very similar average pCO_2 for Norway and Sweden, 830 and 812 µatm respectively, while our spatially resolved estimate for Sweden is almost twice that of Norway (Table S2).

Supporting tables and figures

Chapter 2 Table S 1: Projected total annual FCO_2 for the boreal forest region by model configuration

Model configuration	**RCP 2.6**			**RCP 8.5**		
	2030	**2050**	**2100**	**2030**	**2050**	**2100**
			FCO_2 (Tg C yr^{-1})			
K_H and gas exchange velocity *k* adjusted for future temp.	237 (78.6-478)	252 (83.3-504)	260 (82.9-537)	259 (89.1-516)	281 (91.1-578)	392 (96.5-922)
No adjustment of K_H or gas exchange velocity *k* (present day values)	246 (81-509)	262 (85-542)	270 (85-572)	269 (91-553)	293 (94-619)	416 (101-989)
Only K_H adjusted for future temp.	230 (76.5-462)	242 (80.3-482)	245 (78.5-504)	251 (86.5-497)	262 (85.1-536)	329 (81.0-768)
Only gas exchange velocity *k* adjusted for future temp.	237 (79.3-477)	255 (84.9-509)	268 (86.4-553)	261 (90.1-517)	295 (96.5-604)	466 (116-1,108)

Chapter 2 Table S 2: pCO_2, FCO_2, total A_{lake} and *k* values compared to previous studies

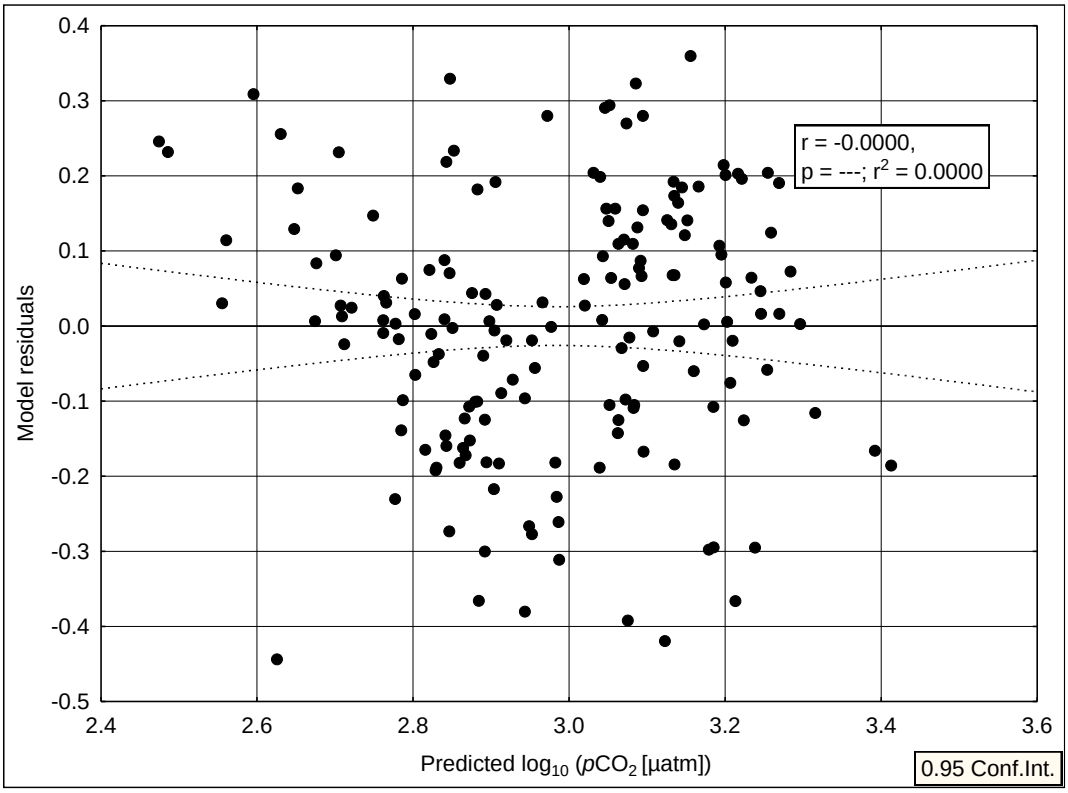
Region	pCO_2 (µatm)	FCO_2 (Tg C yr^{-1})	Total A_{lake} (km^2)	Mean *k* (m d^{-1})	Source
BF	1278 Area weighted 966 (678-1325)	189 (74-347)	1,350,353	0.86	This study
BF	790	79	931,619	0.82	Raymond *et al*., 2013 (INTERPOLATED)
50-70°N	1305 Area weighted 1006 (715-1366)	272 (115-487)	1,751,985	0.88	This study
50-70°N	812	103	1,194,701	0.84	Raymond *et al*., 2013 (INTERPOLATED)
50-90°N*	1100	110	80,000-1,650,000	0.96	Aufdenkampe *et al*. 2011
Sweden	1148 Area weighted 949 (637-1345)	7.9 (2.8-15.3)	50,608	1.05	This study
Sweden	1288	1.7	33,301	0.21	Humborg *et al*., 2010
Sweden	812	2.6	32,212	0.91	Raymond *et al*., 2013
Norway	Area weighted 552 (372-779)	0.84 (0.53-2.61)	23,565	1.04	This study
Norway	830	0.81	9,378	0.88	Raymond *et al*., 2013
Finland	Area weighted 973 (669-1355)	6.56 (2.59-12.2)	46,175	0.95	This study

Finland	753	1.98	31,861	0.92	Raymond *et al*., 2013
Russia	Area weighted 1024 (724-1396)	104 (45-186)	695,333	0.81	This study
Russia	832	42.4	378,535	0.7	Raymond *et al*., 2013
Canada	Area weighted 952 (677-1293)	137 (55-250)	879,741	0.93	This study
Canada	816	80	920,920	0.81	Raymond *et al*., 2013

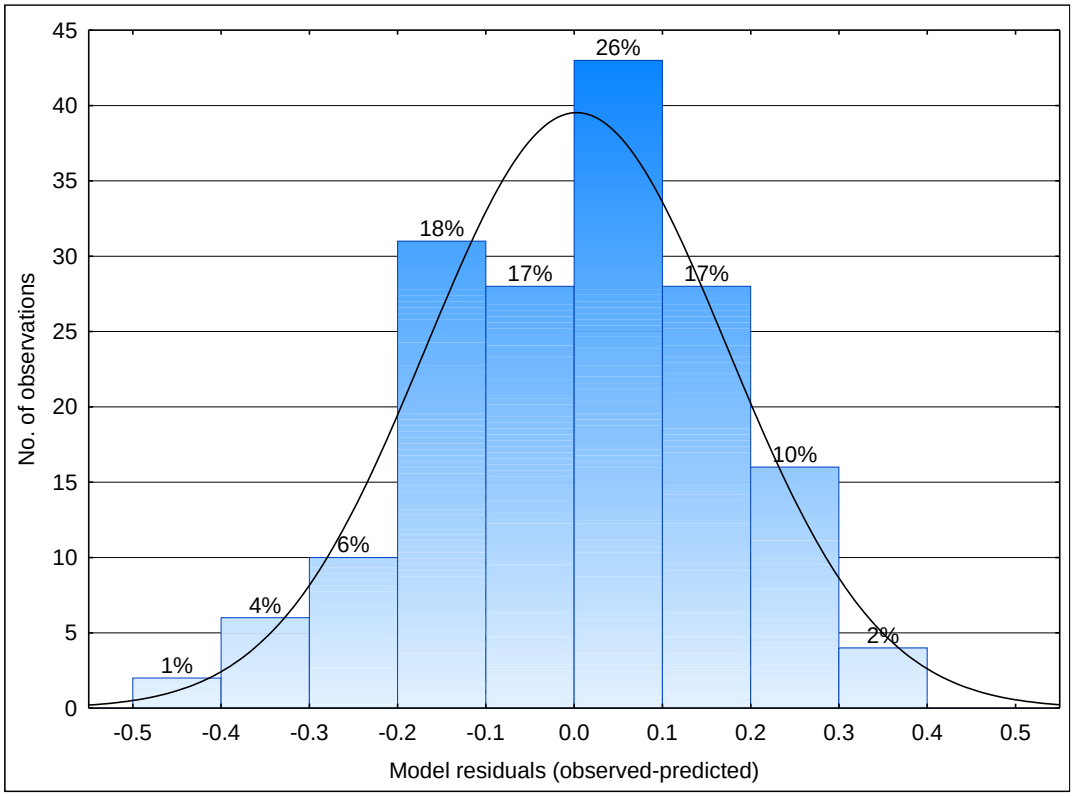
Chapter 2 Table S 3: Projected area weighted mean pCO_2 and total annual FCO_2 by region

Region	RCP 2.6			RCP 8.5		
	2030	2050	2100	2030	2050	2100
			pCO_2 (µatm)			
BF	1183 (744-1849)	1238 (762-1938)	1230 (739-1981)	1270 (792-1982)	1428 (890-2260)	2198 (1303-3761)
50-70°N	1198 (766-1849)	1236 (773-1925)	1249 (757-2034)	1277 (806-1987)	1425 (904-2245)	2219 (1313-3938)
Sweden	1044 (659-1601)	1105 (606-1697)	1064 (629-1695)	1232 (713-2023)	1283 815-1950)	1950 (1302-2960)
Norway	631 (407-952)	659 (426-1036)	629 (370-984)	724 (440-1121)	808 (533-1188)	1269 (885-1842)
Finland	1228 (842-2085)	1414 (711-2530)	1268 (624-2317)	1502 (797-2626)	1534 (895-2576)	2321 (1405-3921)
Russia	1254 (796-1944)	1258 (777-1943)	1298 (779-2087)	1389 (797-2625)	1486 (895-2576)	2398 (1405-3921)
Canada	1131 (733-1726)	1194 (765-1844)	1183 (729-1918)	1165 (759-1778)	1356 (868-2128)	2038 (1278-3509)
			FCO_2 (Tg C yr^{-1})			
BF	237 (78.6-478)	252 (83.3-504)	260 (82.9-537)	259 (89.1-516)	281 (91.1-578)	392 (96.5-922)
50-70°N	328 (118-646)	340 (118-675)	360 (120-750)	354 (128-698)	377 (129-769)	541 (135-1,337)
Sweden	8.54 (2.52-17.3)	9.14 (2.36-19.2)	9.02 (2.34-19.2)	11.0 (3.11-23.3)	10.4 (3.22-21.0)	14.3 (4.50-29.6)
Norway	1.12 (-0.52-3.37)	1.36 (-0.47-3.85)	1.15 (-0.68-3.64)	1.63 (-0.41-4.45)	1.56 (-0.40-4.24)	1.94 (-0.67-5.80)
Finland	7.04 (2.32-14.1)	8.70	7.49 (1.75-16.6)	9.45 3.21-18.9)	8.98 (3.05-17.8)	12.6 (4.08-26.6)

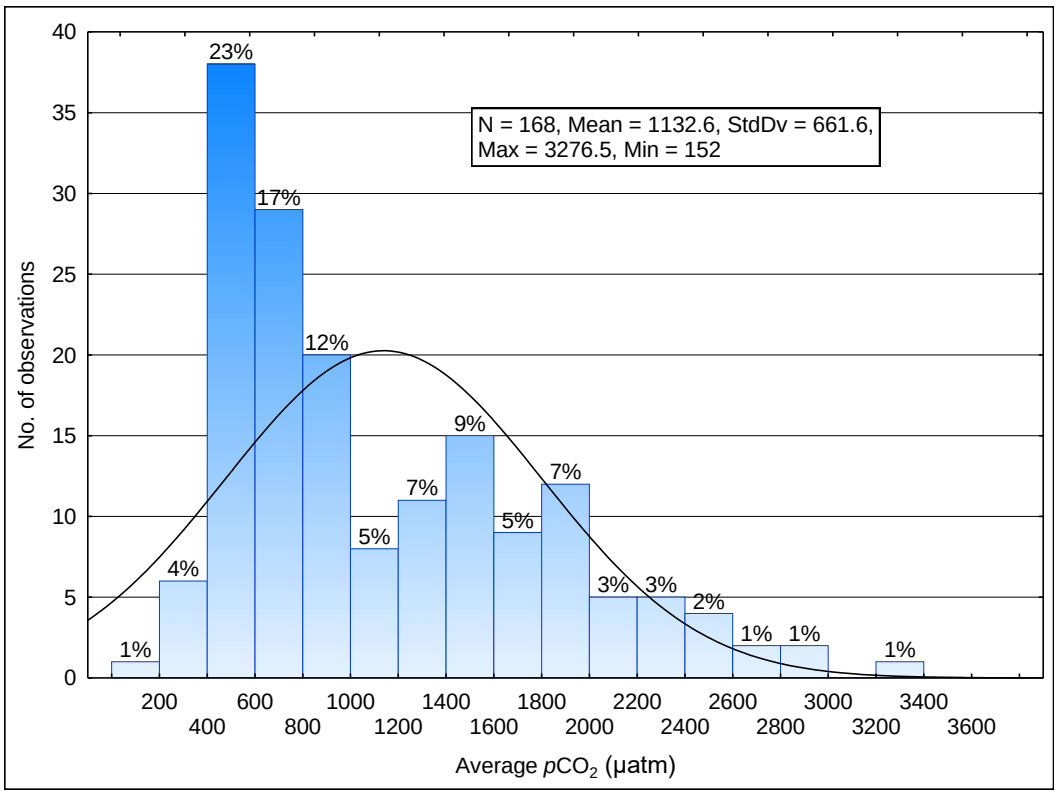
		(2.44-18.3)				
Russia	130 (47.5-257)	129 (43.6-254)	141 (47.1-289)	144 (53.7-281)	148 (52.3-298)	229 (53.3-568)
Canada	168 (57.8-332)	180 (63.0-358)	185 (60.6-390)	172 (60.5-341)	195 (64.1-404)	261 (67.3-647)



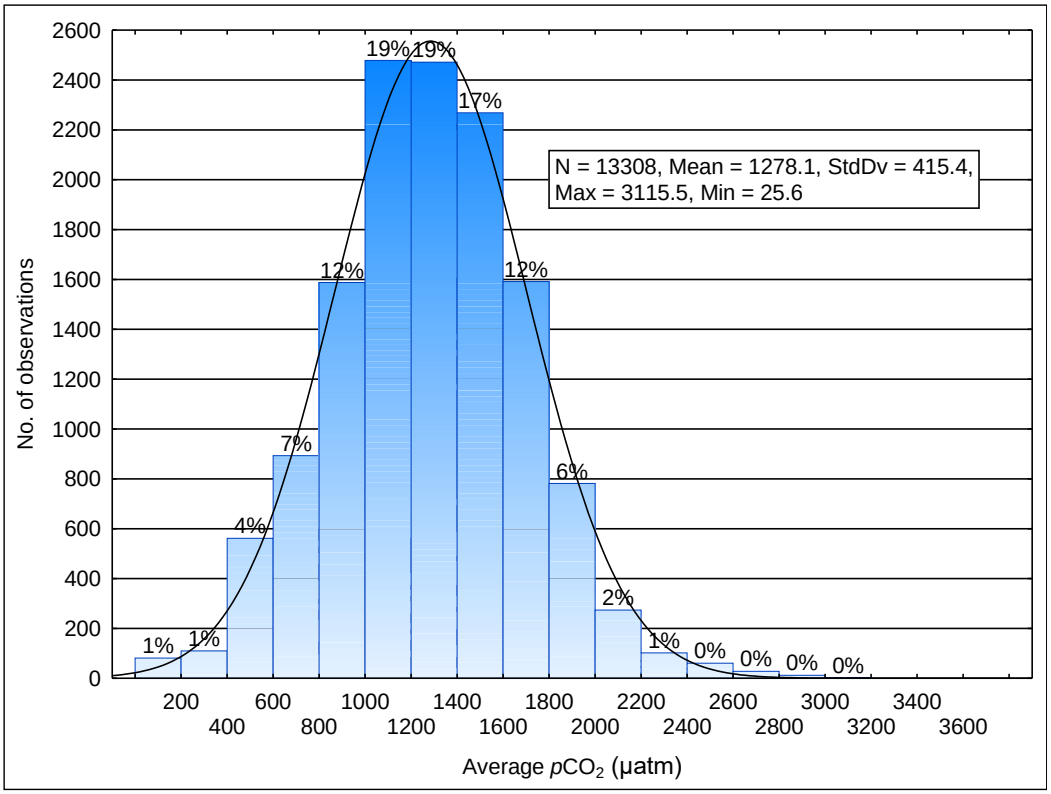
Chapter 2 Figure S 1: Scatterplot of predicted $\log_{10}$ (pCO_2 [µatm]) vs model residuals



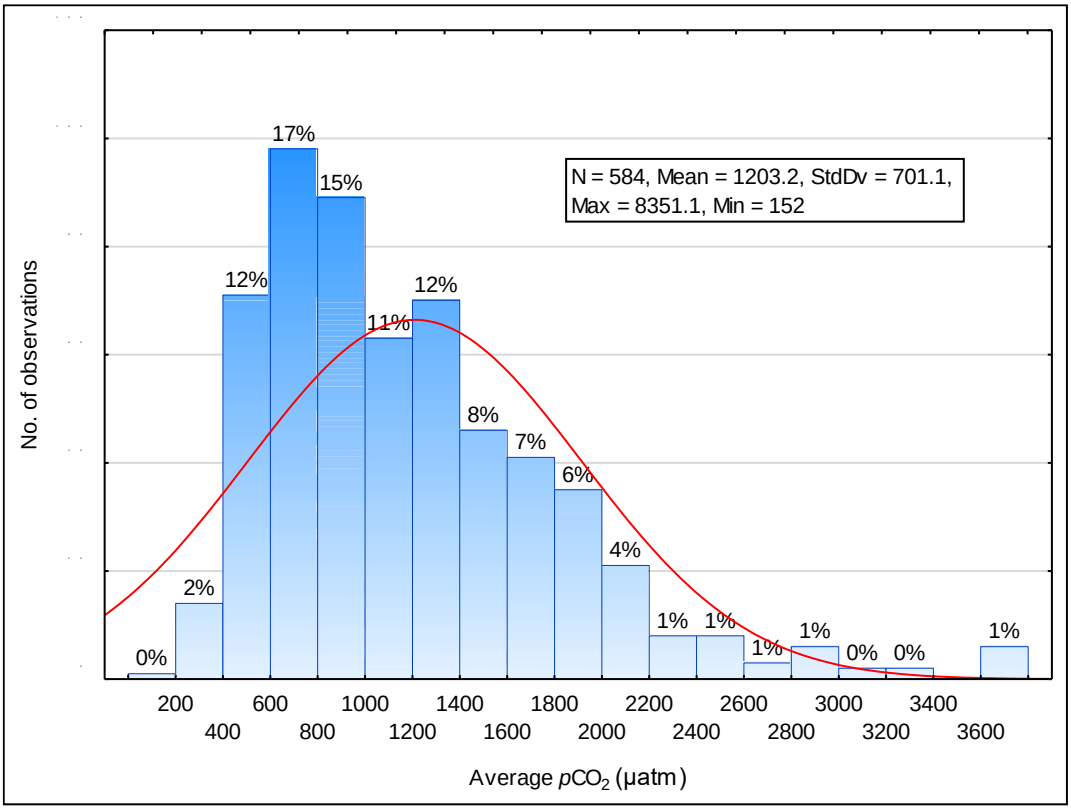
Chapter 2 Figure S 2:Histogram of regression model residuals (observed-predicted $log_{10}(pCO_2[uatm])$)



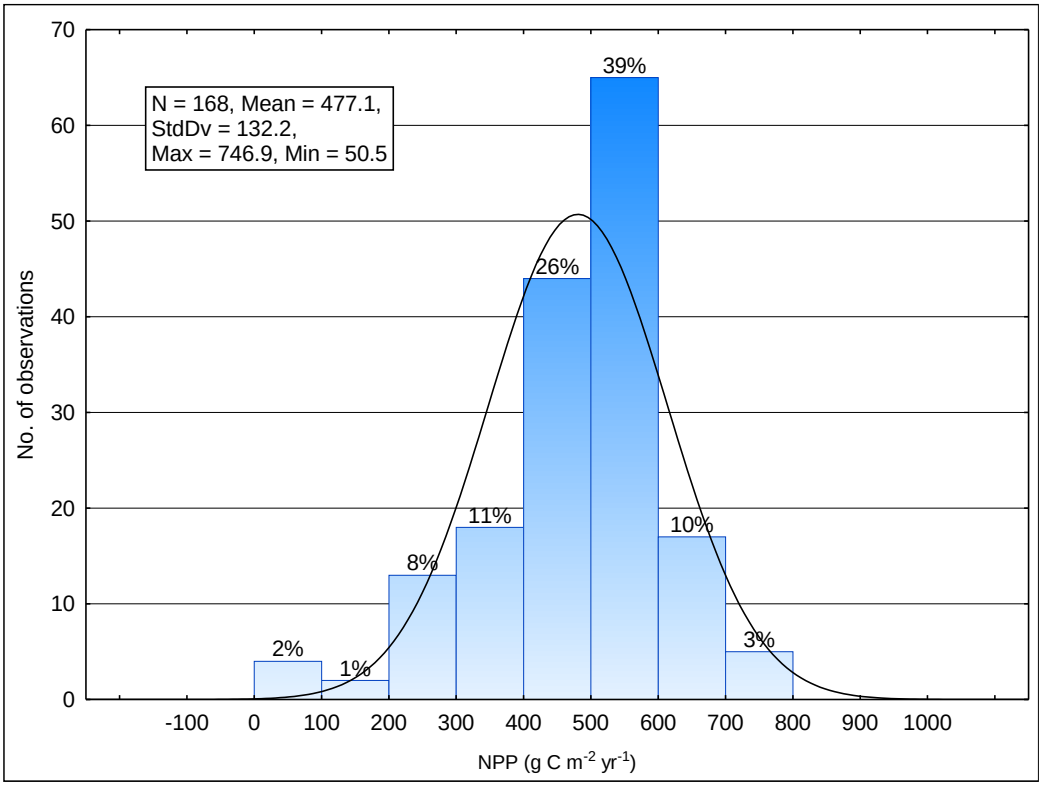
Chapter 2 Figure S 3:Histogram showing frequency of observed average pCO_2 (µatm) values (training data)



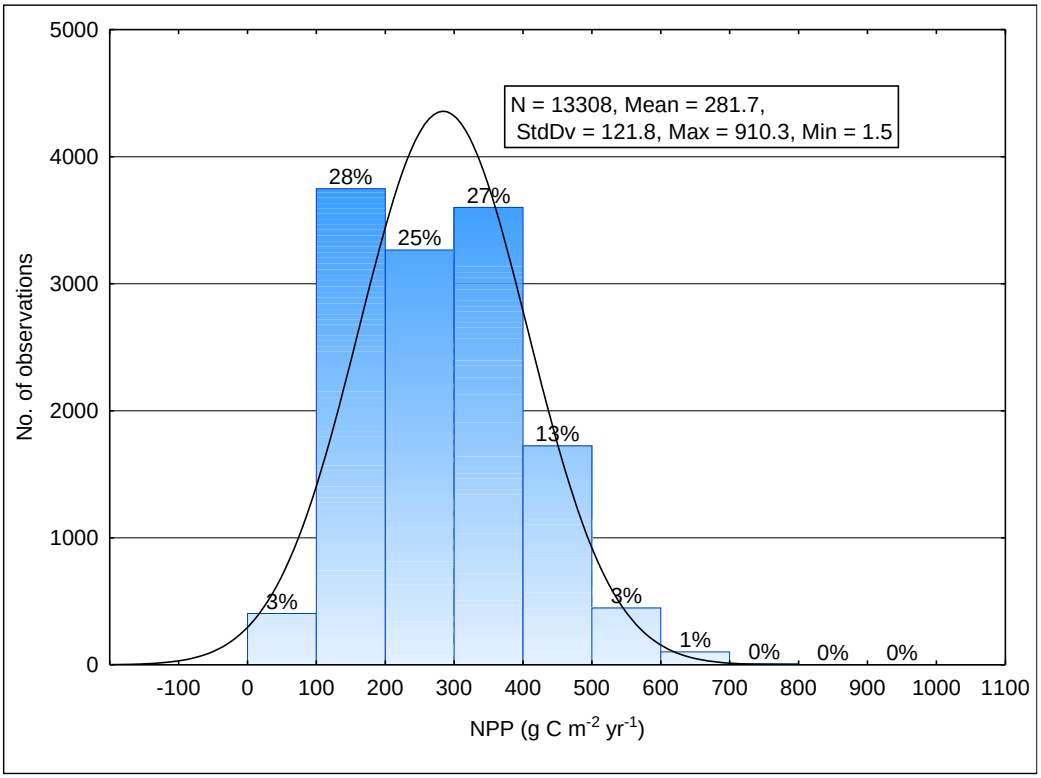
Chapter 2 Figure S 4:Histogram showing frequency of extrapolated average pCO_2 (µatm) values for BF region



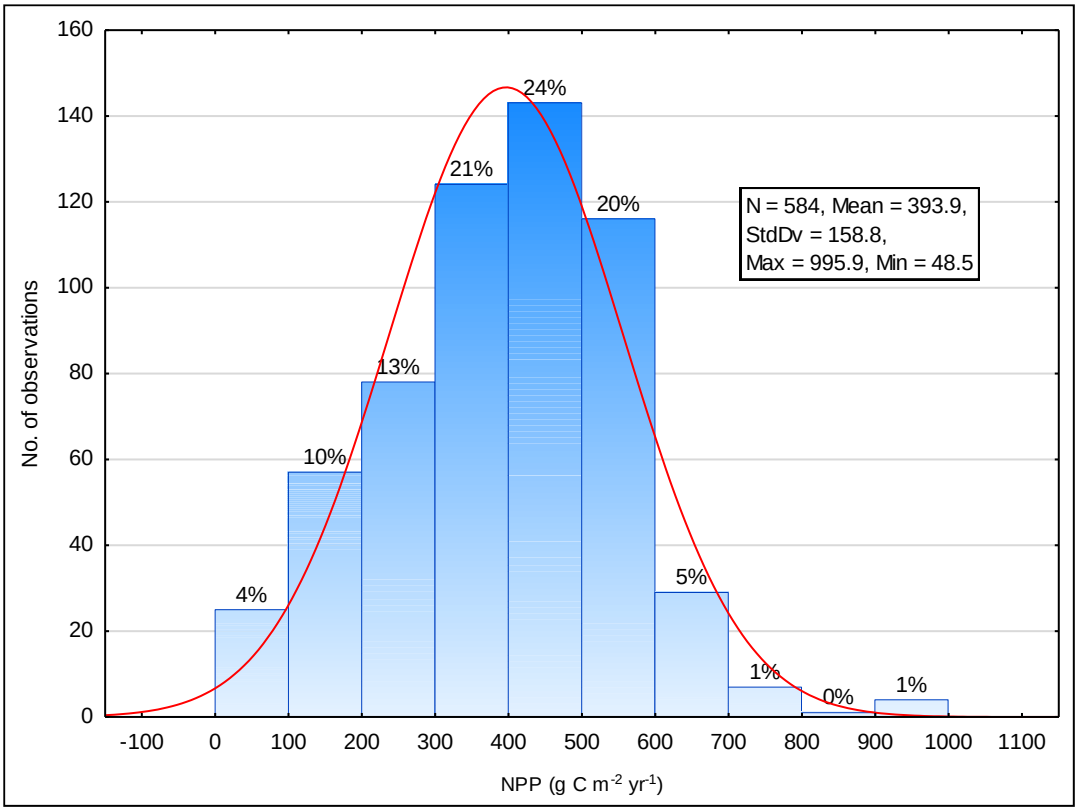
Chapter 2 Figure S 5:Histogram showing frequency of observed average pCO_2 (µatm) values (training data- original 584 grids)



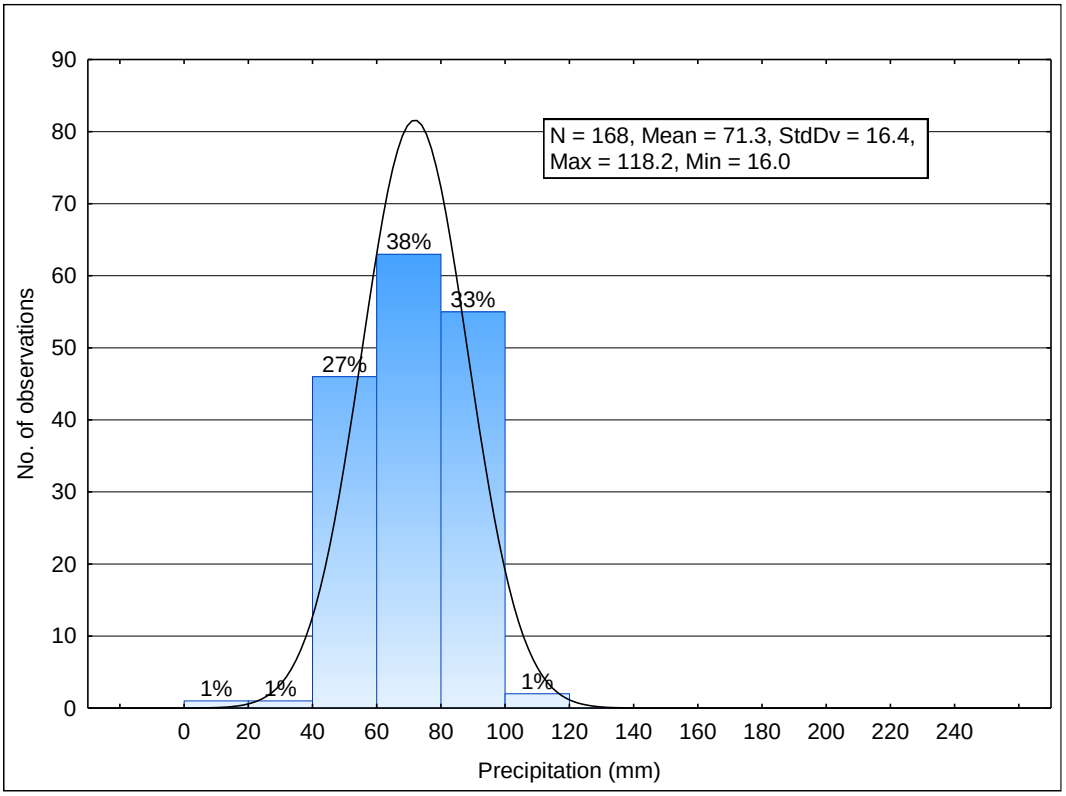
Chapter 2 Figure S 6:Histogram of observed terrestrial net primary productivity (g C m^{-2} yr^{-1}) (training data)



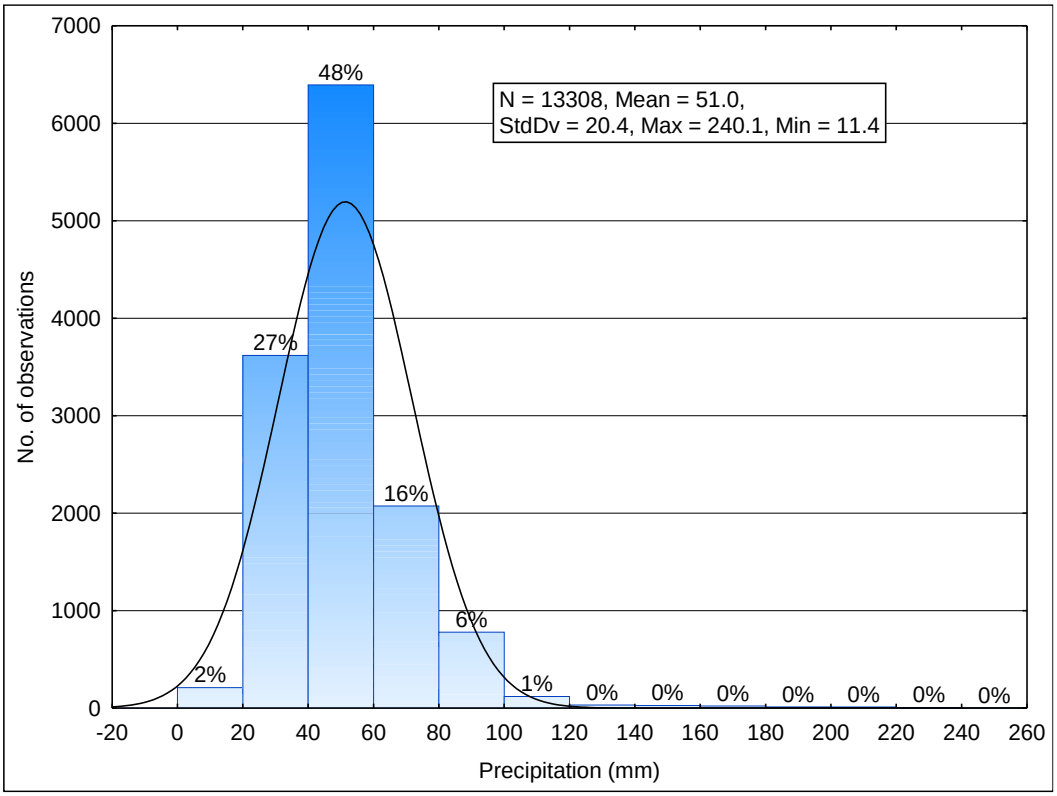
Chapter 2 Figure S 7:Histogram of extrapolated terrestrial net primary productivity (g C m^{-2} yr^{-1}) for BF region



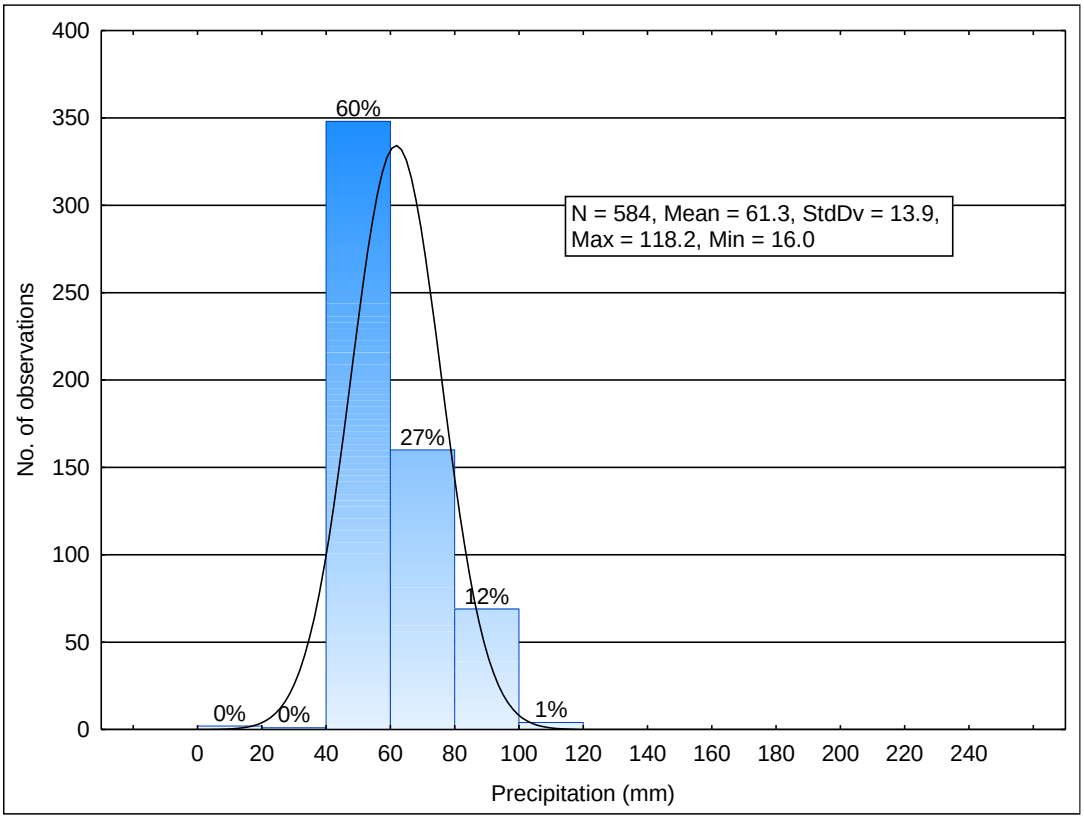
Chapter 2 Figure S 8:Histogram of observed terrestrial net primary productivity (g C m^{-2} yr^{-1}) (training data- original 584 grids)



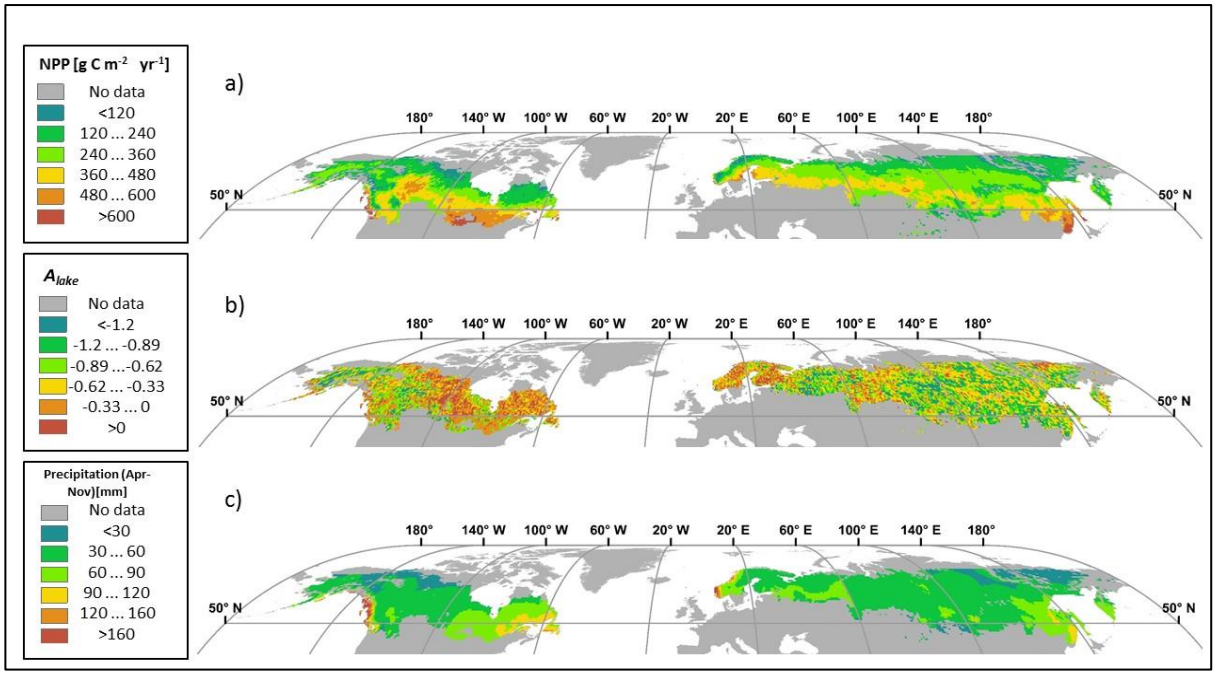
Chapter 2 Figure S 9:Histogram of observed precipitation (April-Nov. monthly mean [mm]) (training data)



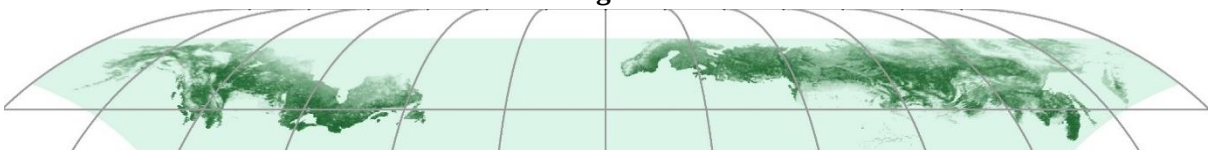
Chapter 2 Figure S 10:Histogram of extrapolated precipitation (April-Nov. monthly mean [mm]) for BF region



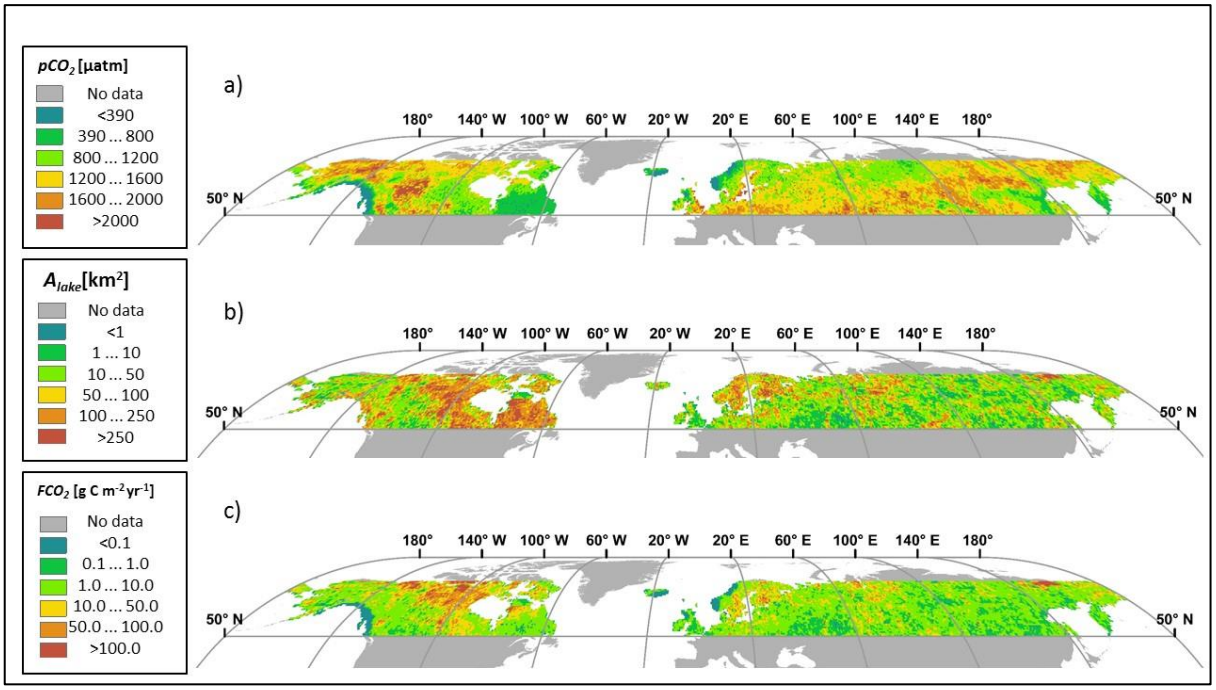
Chapter 2 Figure S 11:Histogram of observed precipitation (April-Nov. monthly mean [mm]) (training data- original 584 grids*)*



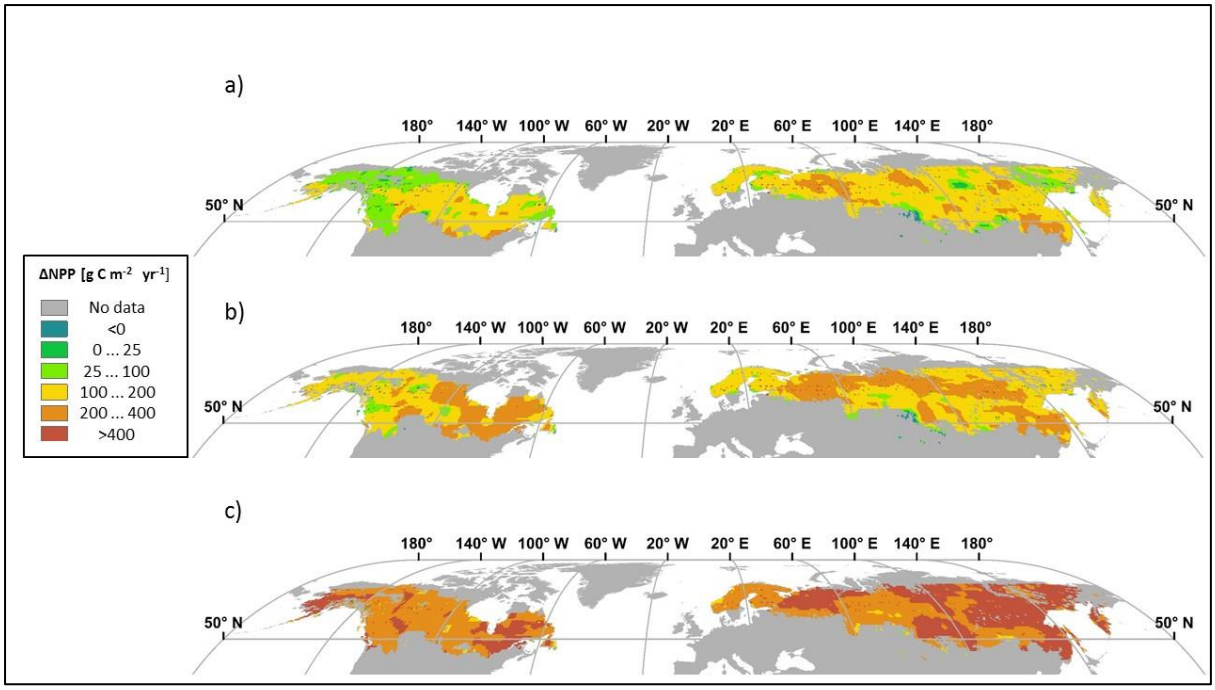
Chapter 2 Figure S 13:Maps of retained drivers (a) terrestrial NPP, (b) adjusted mean log_{10} lake area (km2), (A_{lake}), and (c) precipitation (monthly mean of April-Nov) for the BF land cover region



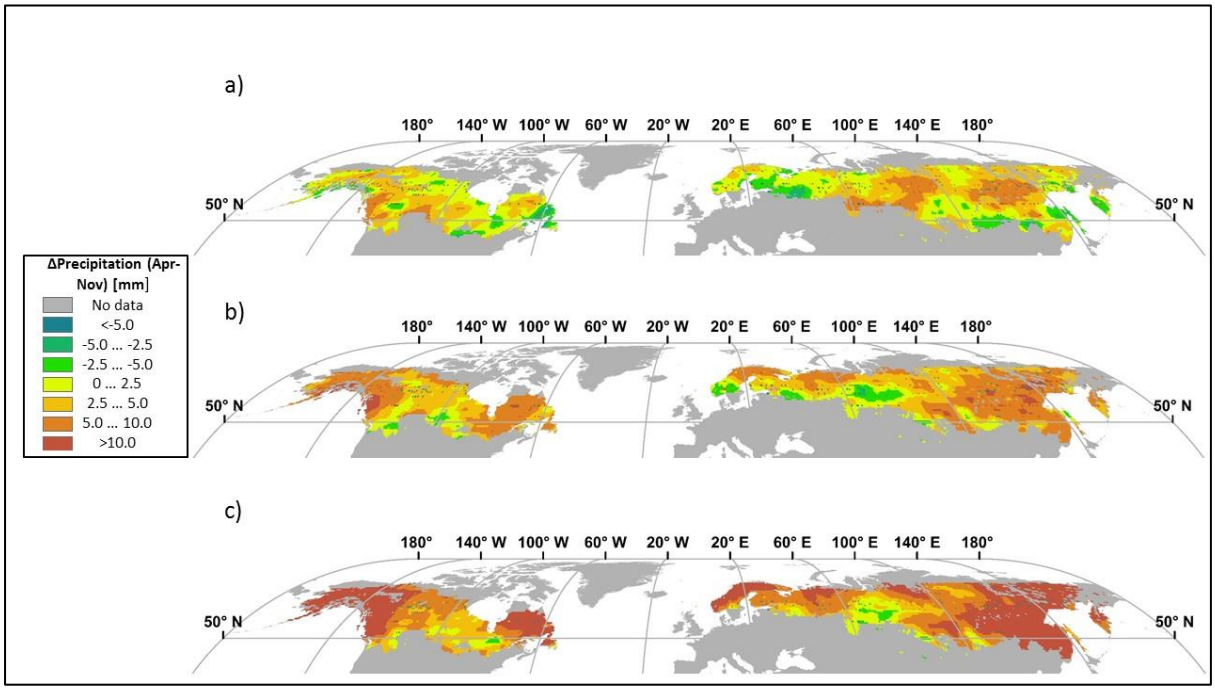
Chapter 2 Figure S 14:Map of boreal forest land cover taken from Potapov et al. (2008)



Chapter 2 Figure S 12:Predicted maps of (a) pCO_2, (b) A_{lake}, and (c) FCO_2 for the region of 50-70° latitude



Chapter 2 Figure S 15:Spatially resolved (0.5°) multi-model mean predicted change in terrestrial NPP, ΔNPP (from a year 2000 baseline) under scenario RCP8.5 for the year (a) 2030, (b) 2050 and (c) 2100 for the BF land cover region



Chapter 2 Figure S 16:Spatially resolved (0.5°) multi-model mean predicted change in precipitation, Δprecipitation (from a year 2000 baseline) under scenario RCP8.5 for the year (a) 2030, (b) 2050 and (c) 2100 for the BF land cover region

Acknowledgements

Financial support was received from the European Union's Horizon 2020 research and innovation program under the Marie Sklodowska-Curie grant agreement No. 643052 (C-CASCADES project). RL acknowledges funding from the European Union's Horizon 2020 research and innovation program under grant agreement no.703813 for the Marie Sklodowska-Curie European Individual Fellowship "C-Leak". GW acknowledges funding from the Swedish Research Council (Grant No. 2016-04153*)* and the Knut and Alice Wallenberg Foundation (KAW project). The research leading to these results has received additional funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013) / ERC grant agreement n° 336642 to SS. Many thanks to Jean-François Lapierre & Paul del Giorgio for providing us with Canadian lake pCO_2 data. We are also grateful to the very constructive comments of two reviewers.

3. Aquatic carbon fluxes dampen the overall variation of net ecosystem productivity in the Amazon basin: An analysis of the interannual variability in the boundless carbon cycle

This chapter has been accepted for publication in Global Change Biology. Full reference: Hastie, A., Lauerwald, R., Ciais, P., Regnier, P (accepted). Aquatic carbon fluxes dampen the overall variation of net ecosystem productivity in the Amazon basin: An analysis of the interannual variability in the boundless carbon cycle. *Global Change Biology*.

Abstract

The river-floodplain network plays an important role in the carbon (C) budget of the Amazon basin, as it transports and processes a significant fraction of the C fixed by terrestrial vegetation, most of which evades as CO_2 from rivers and floodplains back to the atmosphere. There is empirical evidence that exceptionally dry or wet years have an impact on the net C balance in the Amazon. While seasonal and interannual variations in hydrology have a direct impact on the amounts of C transferred through the river-floodplain system, it is not known how far the variation of these fluxes affects the overall Amazon C budget.

Here, we introduce a new wetland forcing file for the ORCHILEAK model, which improves the representation of floodplain dynamics and allows us to closely reproduce data-driven estimates of net C exports through the river-floodplain network. Based on this new wetland forcing and two climate forcing datasets, we show that across the Amazon, the percentage of NPP lost to the river-floodplain system is highly variable at the interannual timescale and wet years fuel aquatic CO_2 evasion. However, at the same time overall net ecosystem productivity (NEP) and C sequestration is highest during wet years, partly due to reduced decomposition rates in water-logged floodplain soils. It is years with the lowest discharge and floodplain inundation, often associated with El Nino events, that have the lowest NEP and the highest

total (terrestrial plus aquatic) CO_2 emissions back to atmosphere. Furthermore, we find that aquatic C fluxes display greater variation than terrestrial C fluxes, and that this variation significantly dampens the interannual variability in NEP of the Amazon basin. These results call for a more integrative view of the C fluxes through the vegetation-soil-river-floodplain continuum, which directly places aquatic C fluxes into the overall C budget of the Amazon basin.

3.1. Introduction

The land-ocean aquatic continuum (LOAC) is now well established as an important component of the global carbon (C) cycle (Ciais et al., 2013). Atmospheric C fixed in terrestrial ecosystems and wetlands can be lost through respiration, and stored in biomass and soil, but can also be transferred laterally to the LOAC as dissolved organic carbon (DOC), particulate organic carbon (POC) and dissolved CO_2. Along the LOAC this C can in turn undergo biogeochemical transformations, be lost back to the atmosphere via CO_2 evasion, transferred further downstream to estuaries and the coast, or undergo sedimentation in wetlands (incl. lakes and reservoirs). It has been demonstrated at the catchment (Cole & Caraco, 2001) to global scale (Battin et al., 2009; Regnier et al., 2013; Ciais et al. in review), that these fluxes are important and should not be neglected in land C budgets.

Globally, there remains a high degree of uncertainty associated with the amounts of C being transferred through and processed within the LOAC. Estimates of the total amount of terrestrial C inputs to inland waters range widely from 1.1 to 5.1 Pg C yr^{-1} (Cole at al., 2007; Aufdenkampe et al., 2011; Regnier et al., 2013; Drake et al., 2017), reflecting the fact that this flux is indirectly derived by summing estimates of aquatic CO_2 evasion, C exports to the coast and burial in the LOAC. Of the three constituent fluxes, CO_2 evasion is the largest (Drake et al., 2017) and thus uncertainties in CO_2 evasion dominate the subsequent uncertainty in the export of terrestrial C to inland waters. Moreover, aquatic CO_2 evasion is highly spatially

variable and hotspot regions have been identified; the boreal and tropical regions contributing disproportionately to global CO_2 evasion from lakes (Hastie et al., 2018) and rivers (Lauerwald et al., 2015), respectively.

In the Tropics, high terrestrial net primary productivity (NPP) and high rainfall drive a large export of C to inland waters and in turn high aquatic CO_2 evasion. In 2002, Richey et al. extrapolated observed pCO_2 measurements to estimate a total CO_2 evasion flux of 0.47 Pg C yr^{-1} from the inland waters of the Amazon Basin (upstream of Obidos, see Fig. S1), 13 times greater than their 36 Tg C yr^{-1} estimate of the total organic C (TOC) export to the coast. In 2013, Rasera et al. calculated a substantially higher CO_2 evasion of 0.80 Pg C yr^{-1} over the same basin area, largely as a result of higher values of gas exchange velocity (K_{600}). More recently, Sawakuchi et al. (2017) added observations from the basin area downstream of Obidos and concluded that CO_2 evasion from the entire Amazon Basin (down to mouth) could potentially be as high as 1.39 Pg C yr^{-1}.

Previous studies have shown that there is considerable seasonal variation in aquatic CO_2 evasion. Richey et al. (2002), found that the partial pressure of CO_2 (pCO_2) and in turn CO_2 evasion was tightly coupled to discharge, increasing and decreasing with rising and falling water respectively. Moreover, they measured exceptionally high pCO_2 values (>44,000 µatm) on the floodplain of the mainstem of the Amazon, and speculated that the source of the C is likely to be organic matter exported from flooded forests.

This was later confirmed by Abril et al. (2014) who demonstrated that Amazonian wetlands export around 50% of their GPP to inland waters in contrast to the typical values of <2% exported from terrestrial landscapes. They went on to conclude that the lateral C flux from wetlands is enough to account for around 0.21 Pg C yr^{-1} of the total CO_2 evasion flux from the inland waters of the Amazon river-floodplain network. A recent study by Almeida et al (2017)

demonstrated that in addition to seasonal variation, large flood events also drive interannual variation in CO_2 evasion from the Madeira River (a tributary of the Amazon), namely that years with extreme flooding evade 20% more CO_2 to the atmosphere per unit area than years without. Another flux linked to flood events is C burial and a recent study estimated the POC burial flux in Amazon floodplain lakes at 16 Tg C yr^{-1} (Sanders et al., 2017), at least an order of magnitude lower than estimates of CO_2 evasion.

These observed seasonal and interannual signals in C fluxes are particularly important given that the region is increasingly vulnerable to extreme climatic events such as droughts and floods (Marengo et al., 2011; Chou et al., 2013; Gloor et al., 2013; Zulkafli et al., 2016). Indeed, recent studies have shown substantial decreases in terrestrial net primary productivity (NPP), and in turn C uptake from the atmosphere as a result of the 2005 and 2010 droughts (Zhao & Running, 2010; Potter et al, 2011; Gatti et al., 2014; Doughty et al., 2015 and Feldpausch et al., 2016). However, most of these studies do not account for LOAC fluxes. For these reasons, it is important that we understand the interannual variation in LOAC fluxes and how they influence the overall net ecosystem production (NEP) of the entire Amazon Basin.

With this in mind, we aim to tackle the following research questions:

- To what extent do the LOAC fluxes (aquatic CO_2 evasion and C export to the coast) vary inter-annually and seasonally throughout the entire Amazon Basin?
- How does interannual variation in discharge and flooding affect the LOAC fluxes, terrestrial NPP, soil heterotrophic respiration (SHR) and ultimately the NEP of the Amazon Basin, particularly in the context of increasing climatic extremes? More specifically, does the incorporation of LOAC fluxes amplify or dampen variation in NEP?

Upscaling studies and empirical models are useful in providing estimates of individual components of the LOAC fluxes for the present day. However, these methods cannot represent the interaction between the different aspects of the Amazon Basin C cycle. A more complex and integrated modelling approach is required to understand and, ultimately, predict the longer-term variation in LOAC fluxes and how this variation affects the net C balance of these ecosystems.

In 2017, Lauerwald et al. developed the first full Land Surface Model (ORCHILEAK model) approach to represent the lateral C fluxes along the LOAC in the Amazon Basin and similarly demonstrated the significance of wetlands, concluding that 51% of total CO_2 evasion comes from the floodplains. The study estimated a total CO_2 evasion of 0.38 Pg C yr^{-1}, close to the value produced by Richey et al. (2002) from up-scaling of measurements. In addition, they substantiated the idea that wetlands are a disproportionately important source of C to rivers, calculating that the CO_2 inputs from root and heterotrophic respiration in flooded soils are almost twice that from non-flooded soils.

The land surface model approach undertaken by Lauerwald et al. (2017) provides a valuable tool for further research, in particular the capability to make future projections of the LOAC C fluxes. However, while they were able to reproduce the seasonality in discharge on the main stem of the Amazon, the total flooded area was substantially underestimated when compared to the observed data of Richey et al., 2013 (after Hess et al., 2003). This is because Lauerwald et al. (2017) relied on the coarse (0.25°) global inundation dataset of Prigent et al. (2007), which tends to underestimate the total floodable area (Lauerwald et al., 2017). Given previous estimates of the magnitude of the CO_2 evasion flux from the Amazon floodplain, the importance of wetlands, and the region's increasing vulnerability to climatic extremes; it is vital that we can accurately model its floodplain dynamics.

In this study, an improved representation of floodplain and wetland dynamics is achieved through the production of a new floodplain forcing file for the ORCHILEAK model, from the high resolution (100m or 0.0008°) synthetic aperture radar (SAR) dataset of Hess et al. (2015). We use this new forcing file to improve the simulation of the interannual variation of LOAC fluxes. In turn, we are able to address the research questions previously outlined, and more specifically to evaluate the impact of flood extent on the dynamics of LOAC fluxes, and ultimately how interannual variation in these aquatic C fluxes influences the overall variation in NEP in the Amazon.

3.2. Methods

3.2.1 A brief description of the ORCHILEAK land surface model

ORCHILEAK (Lauerwald et al., 2017) is a new model branch of ORCHIDEE (Organizing Carbon and Hydrology in Dynamic Ecosystems) (Krinner et al. 2005), the land surface component of the Institut Pierre-Simon Laplace (IPSL) earth system model (ESM). It simulates the production of DOC in the canopy and soils, the leaching of DOC and CO_2 from soils to the river network, DOC mineralization and the subsequent CO_2 evasion from the water surface. Crucially, it also simulates the exchange of C between litter, soils and water on floodplains and in swamps. The representation of these fluxes is in turn closely coupled to the hydrology scheme, namely the representation of precipitation, throughfall, surface runoff, drainage, and the routing of discharge along the river-floodplain network. At the same time, ORCHILEAK also simulates vegetation dynamics of 12 plant functional types, 5 of which are present in the Amazon, as well as the C balance of biomass, litter and soils. In short, ORCHILEAK integrates LOAC fluxes within a full representation of the terrestrial C cycling as simulated by ORCHIDEE. However, in its current form ORCHILEAK does not account for the burial of POC in fluvial and floodplain sediments or the evasion of C to the atmosphere as CH_4. These fluxes are further discussed later. While the model does not simulate the lateral transport of

POC, it does account for the decomposition of submerged litter as a substantial source of DOC and dissolved CO_2 to the water column; in other words, POC from submerged litter is assumed to decompose locally in ORCHILEAK. The model is described in more detail in the proceeding sections.

3.2.2 Overview of the hydrology, soil C scheme, and the transport and transformation of aquatic C fluxes in ORCHILEAK

Precipitation and other meteorological input parameters are prescribed by a forcing file (Fig. S2). The hydrology module of ORCHILEAK, just like that of the standard version of ORCHIDEE, partitions the precipitation between interception loss in the vegetation canopy and throughfall to the ground. The throughfall is further partitioned into infiltration and surface runoff. The soil water storage is refilled by infiltration and depleted by evapotranspiration and drainage. The soil hydrology is represented using a 2 m soil column vertically discretized into 11 layers of geometrically increasing thickness from top to bottom. These processes are all represented at a 30 min time step (see d'Orgeval et al., 2008, Rosnay et al., 2002 for details).

ORCHILEAK incorporates a soil C module largely based on ORCHIDEE-SOM (Camino-Serrano, 2018). The soil module uses the hydrological module outputs to simulate microbial production and consumption of DOC, sorption and desorption of DOC on soil organic matter, the advection and diffusion of DOC and dissolved CO_2 within the soil column and their subsequent lateral export via runoff and drainage as well as the throughfall of DOC onto the soil or water surface. There are 3 pools of DOC in the soil which are defined by their source material and residence times (τ_{carbon}); the active, slow and passive pool. ORCHILEAK distinguishes between flooded and non-flooded soils; decomposition rates of litter, SOC and DOC being 3 times lower in flooded soils. Furthermore, it simulates the input of C to the water column from flooded soils; DOC from litter and SOC decomposition from the top 4.5 cm of the soil column feeds directly to the DOC pool of the overlying waterbody.

The river routing module of ORCHILEAK routes the runoff and drainage from the hydrology module and the corresponding dissolved C fluxes from the soil C module as river flow at a daily time-step along a gridded river routing scheme at 0.5° resolution (Vorosmarty et al., 2000). The river network (Fig. S2) is connected to two sorts of wetland, floodplains and swamps. Where a swamp is present, a constant fraction of the river flow is feeding into the bottom of the soil column. Where a floodplain is present, a temporary water body of time-variant surface and volume may be formed beside the river channel and it is fed by a fraction of river flow when bank-full discharge is surpassed. In the case of the Amazon basin, the bankfull discharge threshold was defined as the median discharge simulated over the period 1980-2000 (see Lauerwald et al., 2017). From the inundated floodplain, water and dissolved C may infiltrate back into the soil or flow back into the river channel, while water may also evaporate. The maximal floodable area (MFF) and the areal fraction of swamps (MFS) per simulation grid is prescribed by a forcing file. The water that infiltrates back into the soil is returned to the hydrology module. The dissolved C contained in that water is returned to the soil C module.

ORCHILEAK simulates the transport and decomposition of terrestrial C inputs within the routing scheme, with the assumption that the lateral transport of DOC and CO_2 are proportional to discharge. Within the water column, DOC is separated into a labile and refractory pool, with half-life times of 2 and 80 days, respectively. The labile pool corresponds to the active pool of the soil C scheme, while the refractory pool is derived from the slow and passive soil solution DOC pools combined. In order to ensure numerical precision, CO_2 production and evasion from the water column, as well CO_2 inputs from flooded litter and SOC are simulated at the high temporal resolution of 1/240 day (6 min). pCO_2 is calculated at the same 6 min time-step based on the dissolved CO_2 concentration, and the temperature-dependent solubility of CO_2. pCO_2 is then used along with a gas exchange velocity and a diurnally variable water surface area, to

calculate CO_2 evasion. Fixed gas exchange values of 3.5 m d^{-1} and 0.65 m d^{-1} are used for rivers (and open floodplains) and forested floodplains, respectively. Flooded forests are given a lower gas exchange velocity due to the reduced impact of wind (i.e. lower wind speeds). For a comprehensive description of the ORCHILEAK model, including the underlying equations, please refer to Appendix 8.1 "ORCHILEAK model developments" taken directly from Lauerwald et al. (2017).

3.2.3 New wetland forcing files

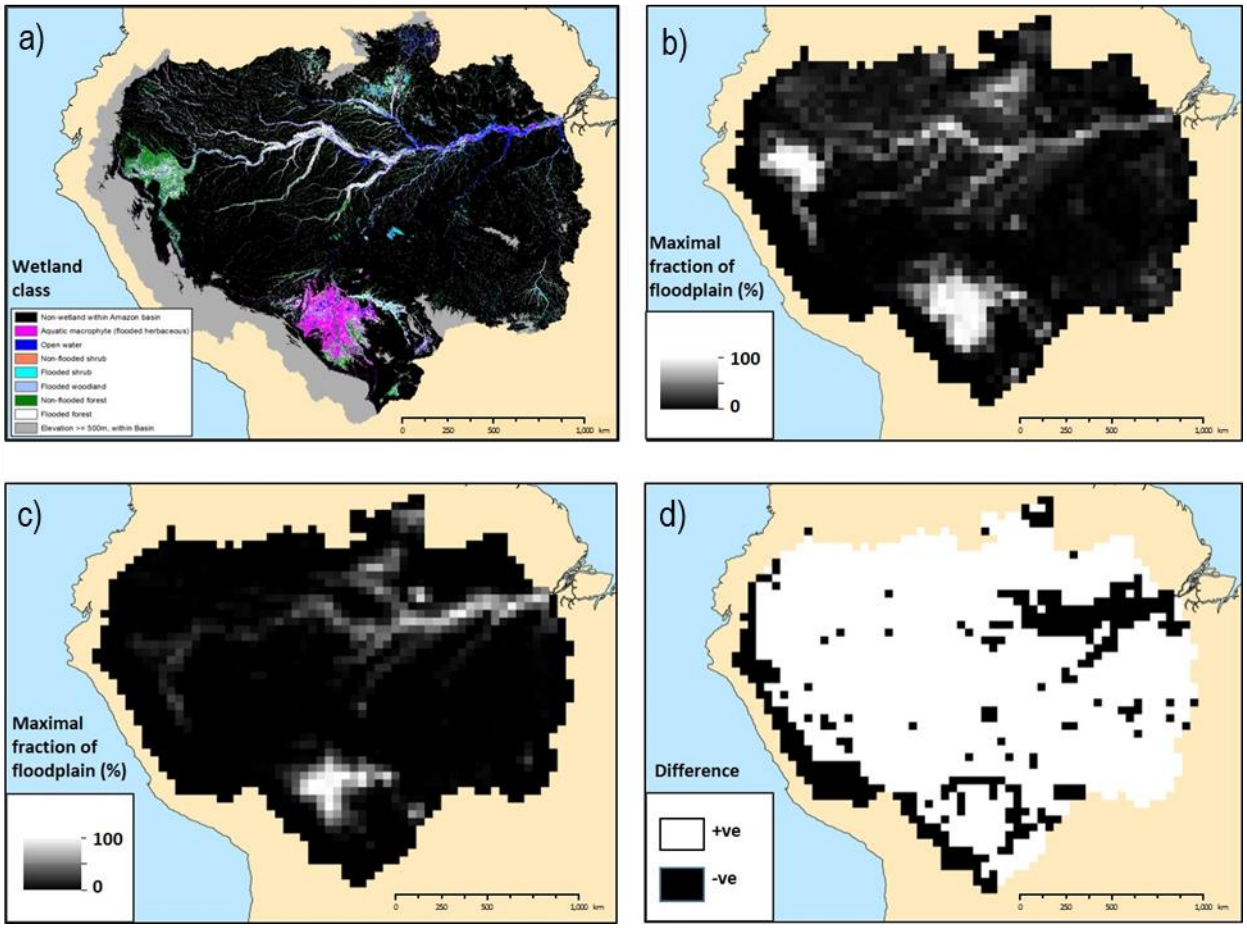
The original routing scheme of ORCHIDEE used universal MFF and MFS derived from the Global Lakes and Wetlands Database (GLWD, Lehner and Doll, 2004) that were shown to considerably underestimate inundated areas in the Amazon (Guimberteau et al., 2012). As a result, Guimberteau et al. developed new MFF and MFS based on the 0.25° datasets of Prigent et al. (2007) and Martinez and Le Toan (2007), respectively. This led to some improvement but inundation was still substantially underestimated in the Amazon Basin.

In the Guimberteau datasets, "swamps", defined as the vegetated part of maximum floodplain, were subtracted from the MFF and used to create the separate swamp (MFS) forcing file. In ORCHILEAK (Lauerwald et al., 2017), swamps were reincorporated into the MFF forcing file, creating a larger, more realistic MFF, and representing the total flooded area from which inland water CO_2 is evading. While these modifications again led to some improvement in the representation of floodplains and swamps in ORCHILEAK, it fundamentally still relied on a low resolution (0.25°) dataset, missing smaller areas of inundation, and meaning that the overall maximum floodplain extent was too small (Lauerwald et al., 2017).

With these limitations in mind, we created a new maximal fraction of floodplain (MFF, Fig.1) forcing file for the ORCHILEAK model based on the 100m Synthetic Aperture Radar (SAR) data described in Hess et al. (2015, see Fig.1 a). This dataset represents different

wetland types during the 1996 May-June flood season. Firstly, we merged all of the wetland categories in Fig. 1 a) into one class, with the exception of the 'non-wetland within the Amazon Basin', 'Open water' and 'Elevation >= 500 m, in Basin' categories. We then aggregated the merged dataset to a resolution of 0.5° (Fig.1 b). Note that in the MFF we included three classes of land cover that were not flooded during the 1996 flood season, namely 'non-flooded shrubs', 'non-flooded woodlands' and 'non-flooded forest' (classes 44, 66 and 88). This decision is based on the justification provided in Hess et al. that these "areas not flooded on either date, but adjacent to flooded areas and displaying landforms consistent with wetland geomorphology". In other words, while these areas were not flooded in 1996, they are likely prone to inundation in other years with greater precipitation and thus should be included in maximum flood extent. Across the Amazon basin, the new forcing file prescribes an average MFF of 13.6%, approximately twofold greater than the 6.3% produced with the original ORCHILEAK forcing derived from Prigent et al. (2007) (Fig. 1 c, d). The addition of the 44, 66 and 88 land cover classes makes a moderate difference; we produce an average MFF of 10% without these 3 classes. For comparison, we also aggregated the 232m resolution wetland dataset of Gumbricht et al. (2017). Assuming that all of the wetland categories in Gumbricht et al. (2017) contribute to the maximum flood extent, we produce an average MFF of 14.9%. However, we chose to use the MFF derived from Hess et al. (2015) as it is measured at a higher resolution and considers wetlands as synonymous with floodplains, while Gumbricht et al. (2017) has a wider definition. In order to account for the uncertainty associated with the MFF forcing file we created two new versions of it; one in which the MFF of each grid was systematically increased by 7% (excluding "highland" areas >=500m identified in Hess et al., 2015) (MFF+7), and another where the MFF was decreased by 7% (MFF-7). We chose a value of 7% as this is the inferred error of the original dataset, described in Hess et al. (2015). Across the Amazon basin, the MFF-7 forcing gives an

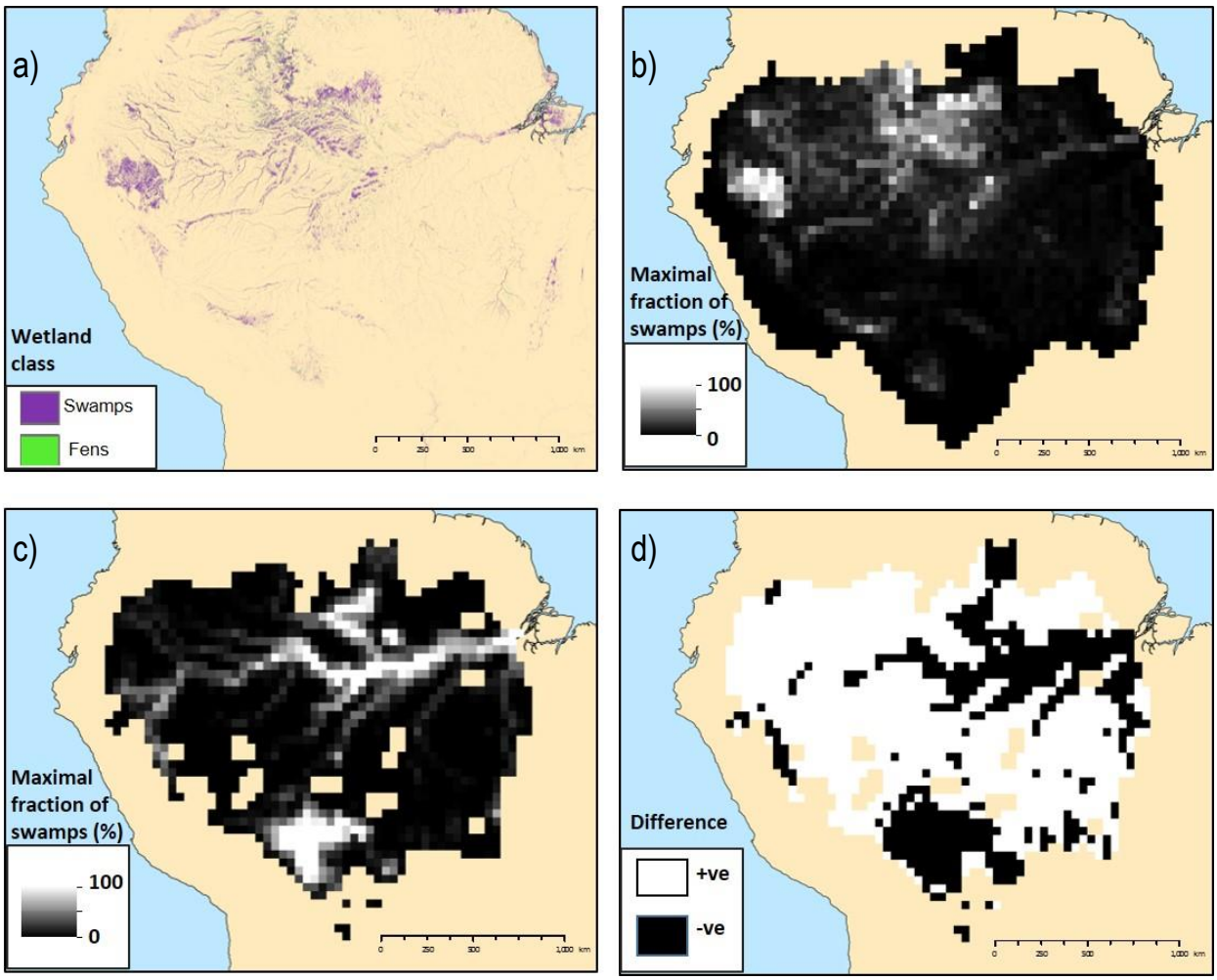
average MFF of 9.3% while the MFF+7 gives an average of 18.3%. This range also envelops the uncertainty associated with the inclusion or exclusion of classes 44, 66 and 88, as well as that associated with the difference between the Hess and Gumbricht datasets.



Chapter 3 Figure 1: a) Wetland classification within the Amazon Basin (Hess et al., 2015), b) the new maximal fraction of floodplain (MFF) forcing file derived from Hess et al. (2015) data, c) the previous MFF forcing file (Guimberteau et al., 2012), and d) the difference between the new and old MFF. In pane d), "+ve" refers to an increase in MFF with the new MF forcing, while "-ve" refers to a decrease. Maps in panels b-d are at a resolution of 0.5°.

We also created a new ORCHILEAK maximal fraction of swamps (MFS, Fig. 2) forcing file based on the 232 m resolution tropical wetland dataset of Gumbricht et al (2017), as the Hess et al. (2015) dataset does not define an explicit "swamp" category. We extracted class 30 (Swamps incl. bogs) and 40 (Fens), before merging these classes and aggregating them to the 0.5° resolution. Across the Amazon basin, the new forcing file prescribes an average MFS of

5.4% (Fig.2, b) which is comparable to the 6% produced with the previous approach (Guimberteau et al., 2012) (Fig. 2, c, d).



Chapter 3 Figure 2: a) Swamps and ferns classification within Amazon Basin from Gumbricht et al (2017), b) the new maximal fraction of swamps (MFS) forcing file derived from Gumbricht et al. (2017) data, c) the previous MFS forcing file (Lauerwald et al., 2017), and d) the difference between the new and old MFS. In pane d), "+ve" refers to an increase in MFF with the new MF forcing, while "-ve" refers to a decrease. Maps in panels b-d are at a resolution of 0.5°.

3.2.4 Simulation Set-up

3.2.4.1. Model configuration

The model was initially run from 1980 until 2000 using two different climate forcing datasets, namely Princeton GPCC (Sheffield et al., 2006), and a regionally updated version of NCC (Ngo-Duc et al., 2005) which was introduced by Guimberteau et al. (2012). This was done in order to test which dataset is able to better recreate observed discharge and the associated

seasonal and interannual variability in floodplain inundation, as well as to account for the uncertainty associated with choice of climate forcing. With the combination of the two climate forcing files and the three MFF forcing files, we ran six different model configurations. Model parameterisation can also cause uncertainty such as the setting of decomposition rate constants for labile and refractory DOC within ORCHILEAK. However, the impact of these parameters was already investigated via a sensitivity analyses in the paper describing the development of the ORCHILEAK model (Lauerwald et al., 2017). As such, we chose to focus on climate forcing and floodplain area as sources of uncertainty in combination with substantial validation against observations and model outputs from the literature.

The original ORCHILEAK simulation (Lauerwald et al., 2017) used only the updated NCC climate forcing. Here, we ran four simulations with the NCC climate forcing dataset; one with the new versions of MFF (hereafter referred to as "standard MFF") and MFS, two more to account for the uncertainty in MFF (MFF+7 & MFF-7), and another with the old MFF (Lauerwald et al., 2017), in order to determine the impact of the new wetland forcing files. We ran the Princeton GPCC simulations with the new versions of MFF and MFS only (three runs). Model parameterisation follows Lauerwald et al. (2017).

3.2.4.2. Hydrology statistics

Following Lauerwald et al. (2017), we calculated a series of statistical parameters in order to calibrate the flood dynamics of the model in a robust and consistent manner. After an initial run, we calculated bank-full discharge and the median water storage for each grid cell (1980-2000), for each model configuration. Any discharge in excess of the median water storage will overtop and begin to inundate the floodplains. After updating bank-full discharge and re-running each model configuration, we calculated the 95^{th} percentile of all simulated water level heights (1980-2000) for each grid cell. This represents the maximum water level, at which the

maximum floodable area is inundated. Once this was updated, each model configuration was re-run once more.

3.2.4.3. Soil carbon spin up

In order to reach a steady state soil carbon pool, we spun-up the model for a total of approximately 7,000 years, looping over 10 years of climate forcing data (1948-1957). To reach steady state more quickly, we first ran the model for 2000 years with the default soil carbon residence time (τ_{carbon}) values halved and a constant atmospheric CO_2 concentration of 350 µatm. Land-cover, representative of the first year of climate forcing data (1948), remained constant over these spin-up runs. After this procedure, all of the soil C pools were approximately at steady state (<0.01% change over the last century of the spin up). Note that it is assumed that soil C pools were in quasi steady state before significant human impact.

3.2.4.4. Transient simulations

We then performed a transient (industrial) run from 1860, until the year that the particular climate forcing dataset starts from (for example to 1948 for Princeton GPCC), again looping over 10 years of climate data but with transient land-cover (LUH-CMIP5) and atmospheric CO_2. Finally, we performed a fully transient simulation (land-cover, atmospheric CO_2 and climate) to the final year of each climate forcing dataset. Note that the NCC climate forcing data is only available until 2000 while the Princeton GPCC data runs until 2010.

3.2.5. Model evaluation and analysis of simulation results

We started by evaluating the hydrology, concentrating on flooded area as this was underestimated in the original ORCHILEAK model set up (Lauerwald et al., 2017). The new MFF and MFS forcing files meant that we had to re-evaluate both discharge and floodplain inundation dynamics. We firstly focused on recreating observed discharge at Obidos (Fig. S1), the most downstream gauging station for which an observed time-series is available (Cochonneau et al., 2006). Total flooded area of the central quadrant of the Amazon basin (Fig.

S1) was tested against remote sensing data (Melack et al., 2011). Note that the Melack et al. dataset uses the same wetland mask as we use here, but the seasonality and area of inundation is completely independent. We then performed a model validation for the DOC and aquatic CO_2 evasion fluxes using the same validation data and methodology as described in Lauerwald et al. (2017), as well as an in-depth comparison of our results to those of previous studies. In addition, we examined the interannual variation of both the terrestrial (meaning NPP and SHR) and aquatic C fluxes (also referred to as LOAC fluxes, and meaning CO_2 evasion from the water surface and the export flux of C to the coast) of the Amazon, and assessed how this variation relates to rainfall and temperature variation through linear regression analysis. As we found long-term (decadal) trends in several of the fluxes, most notably NPP (Tables S2 & S3), we detrended the annual times series using the Detrend function within the "SpecsVerification" package in R (R Core Team 2013), before performing the regression analyses using STATISTICA^{TM}. Finally, we sum the various C fluxes to calculate the net C balance of the Amazon Basin (see 2.6) and examine the importance of the LOAC fluxes to the overall C balance.

3.2.6. Calculating the net carbon balance of the Amazon

In order to estimate the net C balance of the Amazon basin, we summed the terrestrial and aquatic C fluxes to estimate Net Ecosystem Production (NEP) and Net Biome Production (NBP). Positive values of NEP and NBP correspond to a net sink.

We define NEP as follows:

$$NEP = NPP + TF - SHR - FCO_2 - LE_{\text{Aquatic}} \quad (1)$$

Where *NPP* is terrestrial net primary production, *TF* is the throughfall flux of DOC, *SHR* is soil heterotrophic respiration (only the part evading from the soil surface); FCO_2 is CO_2 evasion from the water surface and $LE_{Aquatic}$ is the export flux of C to the coast. *NBP* is the

same as *NEP* but with the addition of the C lost (or gained) through land use change (*LUC*, including fires and the export of woody biomass) and crop harvest (*Harvest*):

$$NBP = NEP - (LUC + Harvest) \quad (2)$$

3.3. Results

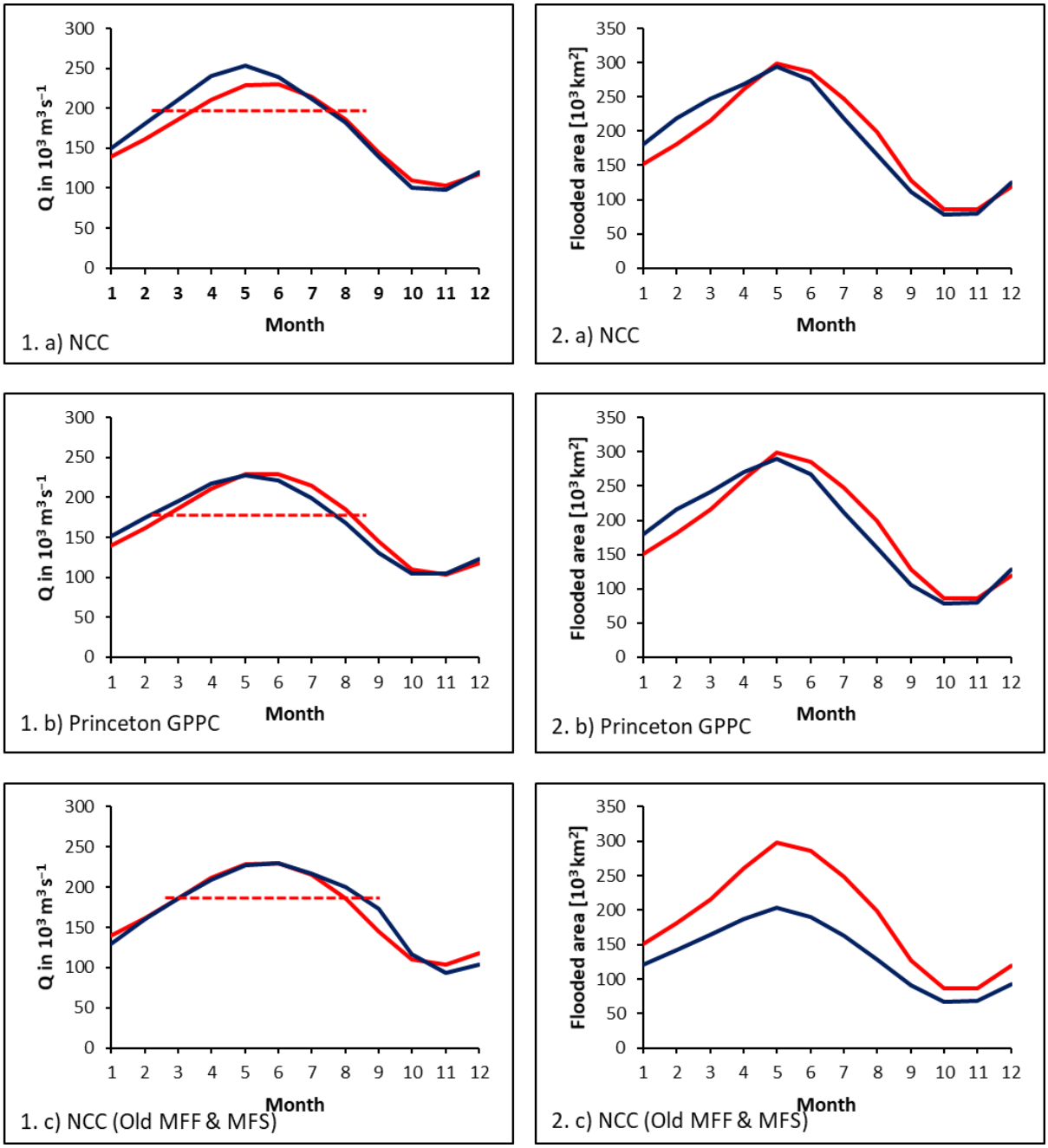
3.3.1. Representation of Hydrology

The model is able to reproduce river discharge at Obidos (1980-2000), the farthest downstream river gauge (Fig. S1), both in terms of total magnitude and seasonal variability. Simulation with the old floodplain/swamp forcing used by Lauerwald et al. (2017) and simulations based on the new floodplain/swamp forcing file showed a similarly good performance (Fig.3-1a-c, Table 1). There was no substantial difference in the simulated discharge from the Amazon basin after the implementation of the new floodplain. However, the new floodplain forcing substantially improved the ability of the model to reproduce the seasonality in flooded area (Fig. 3-2a-c); Nash Sutcliffe-Efficiency (NSE) and Root Mean Square Error (RMSE) were 0.91 and 12% respectively with the new floodplain forcing, compared to -0.75 and 32% with the old (Table 1).

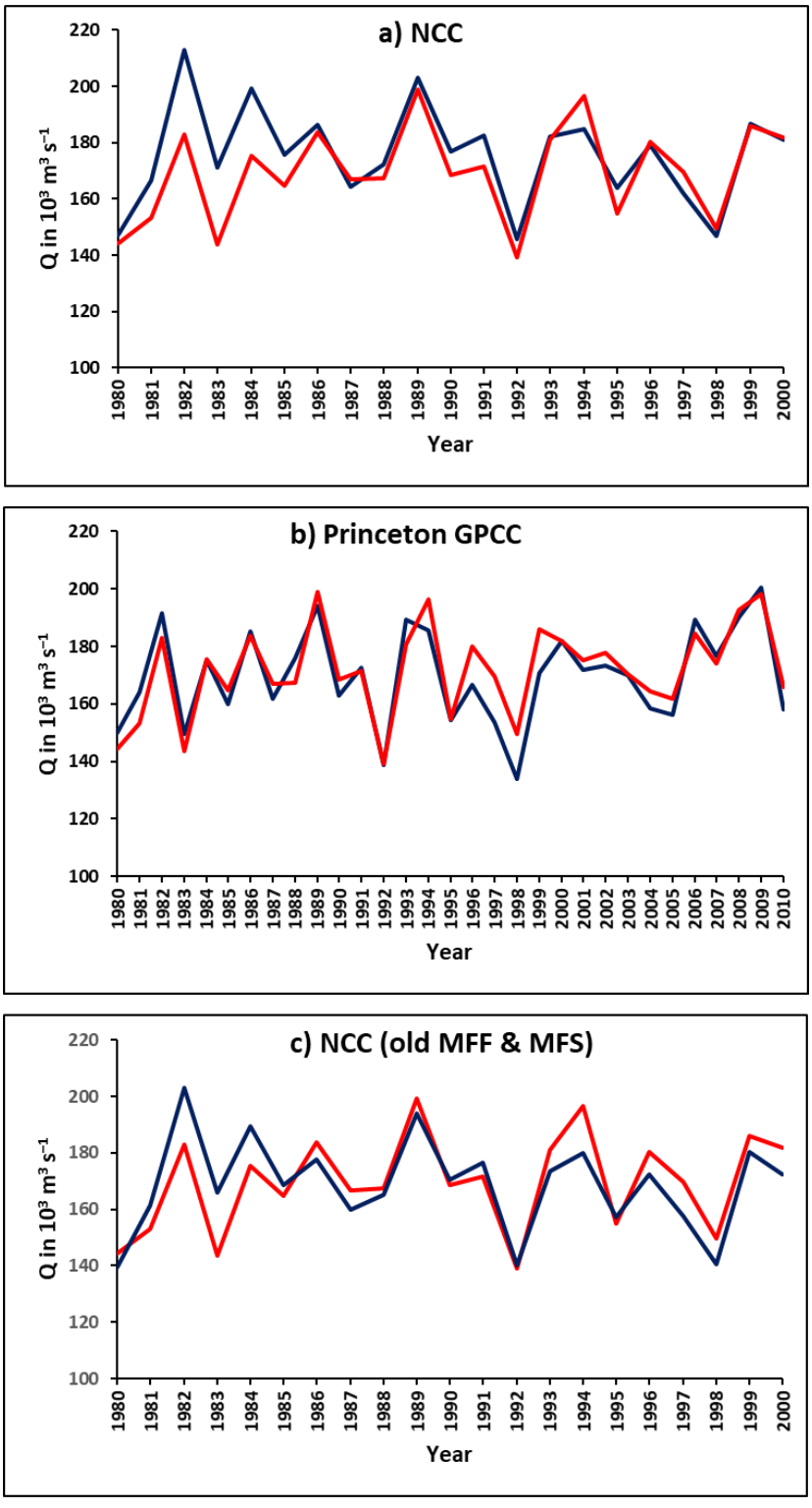
Comparing model runs driven by the two different climate forcing, NCC and Princeton GPCC climate data, we find a similarly good performance as well. With both forcing data sets, we were able to recreate the observed mean magnitude and seasonality in discharge at Obidos (1980-2000) (Fig. 3-1 a, b) and flooded area in the central (Fig. S1) Amazon (1981-1996) (Fig. 3-2 a, b).

While the model was mostly able to reproduce the observed interannual variation in discharge, there was some minor difference in performance related to the choice of climate forcing (Figure 4, Table S1). The simulation driven by the Princeton GPPC data had an NSE of 0.79 and a RMSE of 4% against observations, compared to 0.50 and 7% for the NCC run (Figure 4, Table

S1). The year with the highest observed discharge was 1989 with a mean of 199 10^3 m^3s^{-1}. The Princeton GPCC run correctly simulated 1989 as the year with the highest discharge, with a mean of 194 10^3 m^3s^{-1}. The NCC run ranks 1989 as the year with the second highest discharge, and actually predicts a higher 1989 mean discharge of 203 10^3 m^3s^{-1}. With NCC, the year with highest discharge is 1982, which is the 5th highest discharge in the observed time series. Conversely, the NCC simulation correctly modelled 1992 as the year with the lowest discharge (146 10^3 m^3s^{-1}) while the run driven with Princeton ranked 1992 second lowest (Figure 4, Table S1). It is important to note that the differences in observed discharge between both the highest (1989) and second highest (1994), and lowest (1992) and second lowest (1983) are minor (Figure 4, Table S1).



Chapter 3 Figure 3-1: Seasonality of simulated *versus* observed discharge (Cochonneau et al., 2006) at Obidos (1980-2000 monthly mean), with a) NCC climate forcing with standard MFF b) Princeton GPCC climate forcing with standard MFF and c) NCC with old MFF & MFS. 2: Seasonality of simulated versus observed flooded area (Melack et al., 2011) in the central Amazon basin (1981-1996 monthly mean) with a) NCC climate forcing with standard MFF b) Princeton GPCC climate forcing with standard MFF and c) NCC with old MFF & MFS.



Chapter 3 Figure 4 Annual variation of simulated *versus* observed discharge (Cochonneau et al., 2006) at Obidos (1980-2000) for a); run with NCC climate forcing with standard MFF, b) Princeton GPCC climate forcing with standard MFF and c) NCC climate forcing with old MFF & MFS

:

Chapter 3 Table 1: Performance statistics for modelled versus observed discharge Q at Obidos and flooded area in the central Amazon basin for different climate forcing configurations

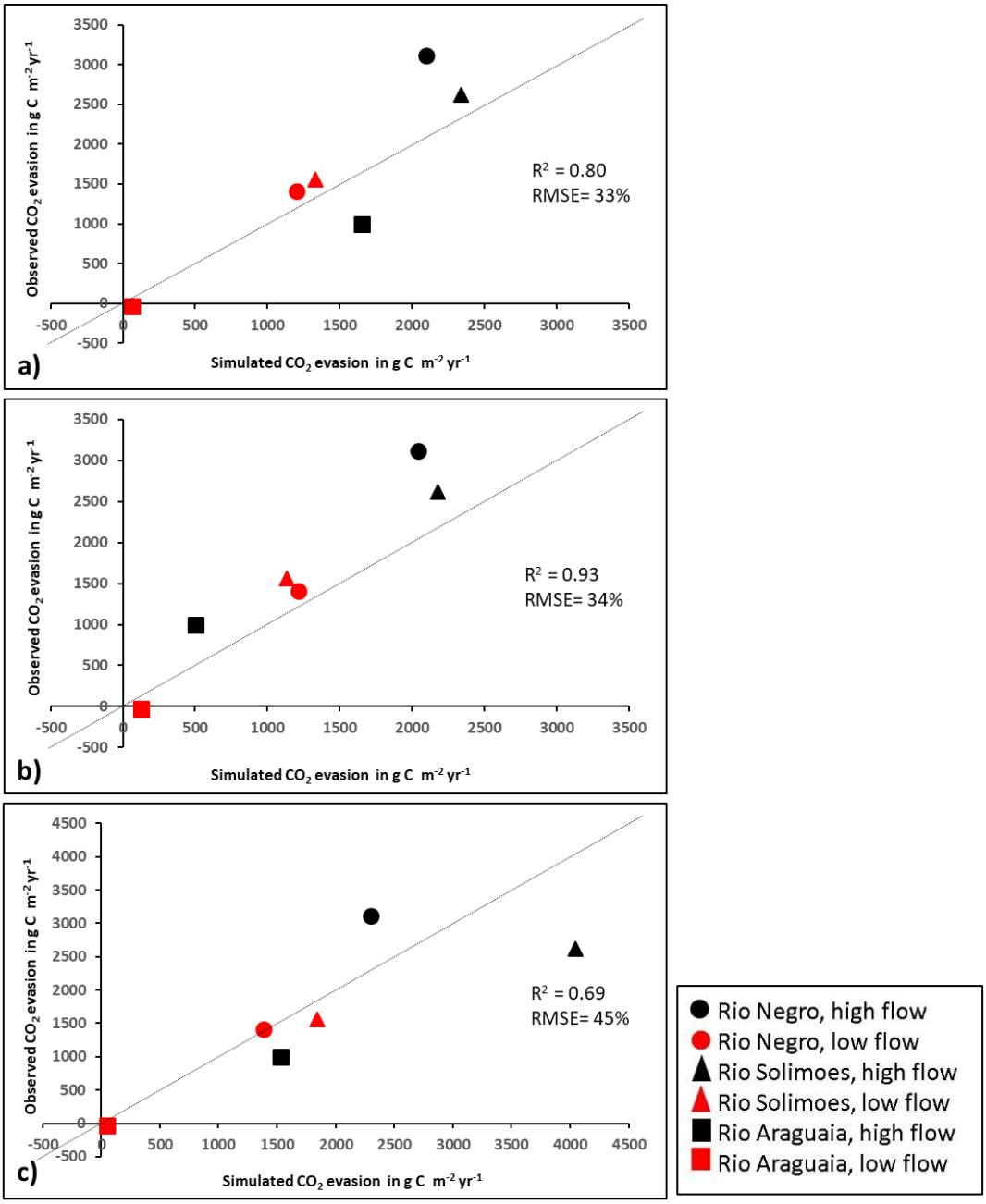
	Seasonality in Q at Obidos (1980-2000)			Flooded area in central Amazon (1981-1996)			Interannual variation in Q at Obidos (1980-2000)		
Climate forcing	RSME	NSE	R^2	RSME	NSE	R^2	RSME	NSE	R^2
NCC	9%	0.91	0.95	12%	0.91	0.91	7%	0.50	0.66
Princeton GPCC	6%	0.94	0.95	13%	0.89	0.90	4%	0.79	0.81
NCC (old MFF & MFS)	6%	0.95	0.95	32%	-0.75	0.97	6%	0.62	0.67

3.3.2. Carbon fluxes along the Amazon Basin

We estimate a long-term mean (1980-2000 across six model runs) NPP rate of 1,214 (1,204-1,223) g C m^{-2} yr^{-1} (range represents the variation caused by the combination of the two climate forcing and the three MFF forcing files; standard, MFF +7 and MFF-7), amounting to a total NPP of 6.81 (6.75-6.86) Pg C yr^{-1} for the entire Amazon Basin (5.6 x 10^6 km^2). If we only consider the uncertainty associated with climate forcing alone, the range is reduced to 6.77-6.85 Pg C yr^{-1}. The effect of the new MFF and MFS on NPP was negligible; mean annual NPP being 1,220 g C m^{-2} yr^{-1} (total of 6.84 Pg C yr^{-1}) and 1,222 g C m^{-2} yr^{-1} (total of 6.85 Pg C yr^{-1}) with the original (Lauerwald et al., 2017) and new forcing files, respectively, both driven by NCC. We estimate a mean annual soil heterotrophic respiration (SHR) of 5.87 (5.62-6.16) Pg C yr^{-1}. The new forcing file had a significantly greater effect on SHR than on NPP; the original forcing file (with NCC) produces a higher mean annual SHR of 6.30 Pg C yr^{-1}, compared to 5.94 Pg C yr^{-1} (with NCC) this difference due to the greater suppression of organic matter decomposition with the new MFF (Rueda-Delgado et al., 2006). We estimate a mean annual throughfall DOC flux (TF) of 79 (78-79) Tg C yr^{-1}.

We simulate a mean annual (1980-2000) CO_2 evasion of 746 (526-998) Tg C yr^{-1} from the water surfaces of the Amazon basin, a 97% increase from the 379 Tg C yr^{-1} produced with the

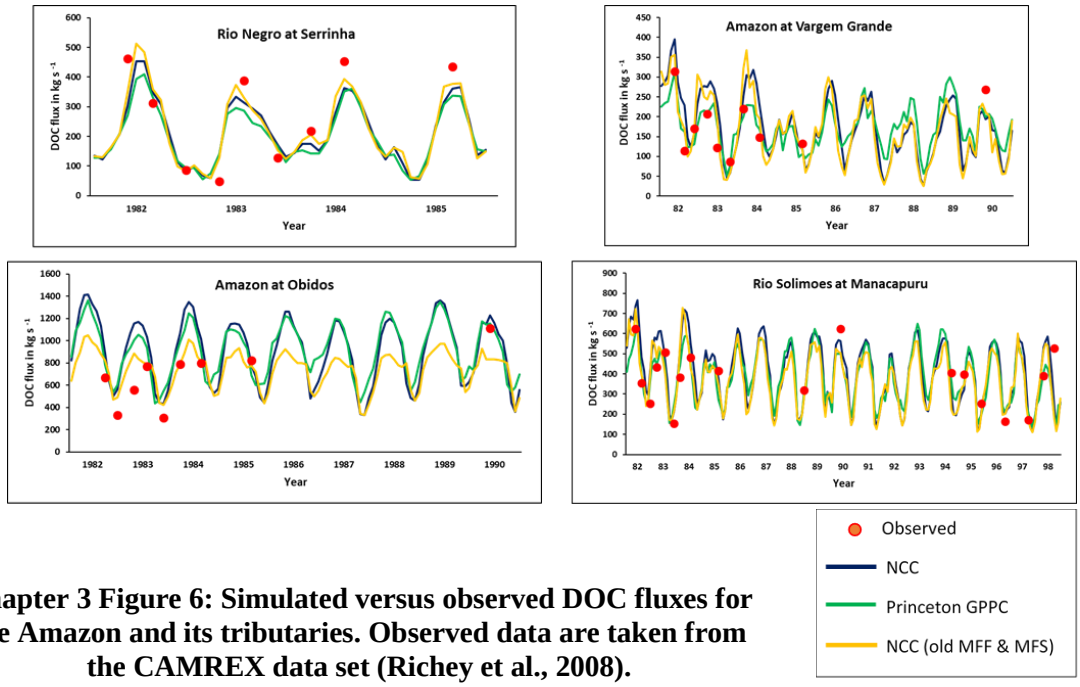
original ORCHILEAK configuration (Lauerwald et al., 2017). If we only include the uncertainty associated with climate forcing, we produce a mean of 729 Tg C yr^{-1} and the range is substantially reduced to 700-758 Tg C yr^{-1}, meaning that the majority of the uncertainty in the evasion flux comes from the MFF forcing. We attribute approximately 75% of the CO_2 evasion flux to the floodplain compared to 51% in the original study (Lauerwald et al., 2017). With the new MFF forcing, we moderately improved the reproduction of observed CO_2 evasion fluxes during low (monthly avg. discharge < yearly avg. discharge) and high flow (monthly avg. discharge > yearly avg. discharge) periods at three sites in the Amazon (Rasera et al., 2013, Fig. 5) (R^2 =0.80, RMSE = 1.4 µmol CO_2 m^{-2} s^{-1} vs R^2 =0.69, RMSE = 1.9 µmol CO_2 m^{-2} $s^{-}1$, with new (a) and old MFF (c) respectively, both driven by NCC). The performance was further improved with the Princeton GPCC climate data; R^2 =0.93, RMSE = 1.4 µmol CO_2 m^{-2} s^{-1} (Fig. 5, b).



Chapter 3 Figure 5: Observed versus simulated CO_2 evasion rates per water surface area for a) run with NCC climate forcing (standard MFF), b) Princeton GPCC climate forcing (standard MFF), and c) NCC climate forcing with old MFF & MFS. Observed data are from Rasera et al. (2013). Reported are means of the observed values, 2006 -2010. The simulated values refer to the mean evasion rate during low (monthly avg. discharge < yearly avg. discharge) and high flow periods (monthly avg. discharge > yearly avg. discharge) (1981–2000), see Figure 3. Note that the scale of the axes c) is slightly different to a) and b).

We simulate a mean annual (1980-2000) DOC export to the coast (downstream of Obidos) of 38 (33-44) Tg C yr^{-1}. In Figure 6, we compare simulated DOC flux against the observations at several sites (see Fig. S1 for locations) and find that the model can recreate the temporal variation in DOC relatively well (Table S5). The effect of the new forcing files is mixed, with

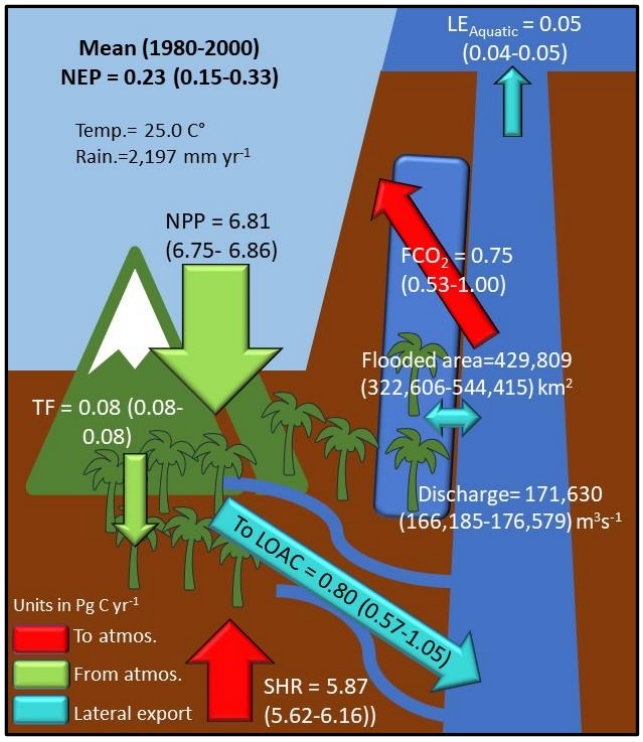
the performance improving at some sites but worsening at others (Fig. 6). The largest impact can be seen at Obidos where the new forcing files result in a substantially larger DOC flux during high flow. The model run using the old MFF and MFS appears to perform better at moderate discharge, while the new set up appears to perform better during periods when observed DOC is very high (i.e. 1990). Both appear to overestimate DOC flux at Obidos during low flow. We simulate a mean annual flux (to the coast) of dissolved CO_2 of 7.1 (6.8-7.7) Tg C yr^{-1}.



Chapter 3 Figure 6: Simulated versus observed DOC fluxes for the Amazon and its tributaries. Observed data are taken from the CAMREX data set (Richey et al., 2008).

3.3.3. The net carbon balance of the Amazon Basin

The long-term mean (1980-2000) C balance; that is the components of the Net Ecosystem Production (NEP, equation 1), is presented in Fig. 7. We estimate a mean (1980-2000) NEP of 0.23 (0.15-0.33) Pg C yr^{-1} and a mean Net Biome Production (NBP, equation 2) of 0.04 (-0.04-0.14) Pg C yr^{-1}. Using the original floodplain and swamp forcing files (with NCC), we estimate a mean annual NEP of 0.17 Pg C yr^{-1}. Using the same set up (with NCC) but with the new MFF and MFS forcing files we produce a higher sink of 0.21 Pg C yr^{-1}.

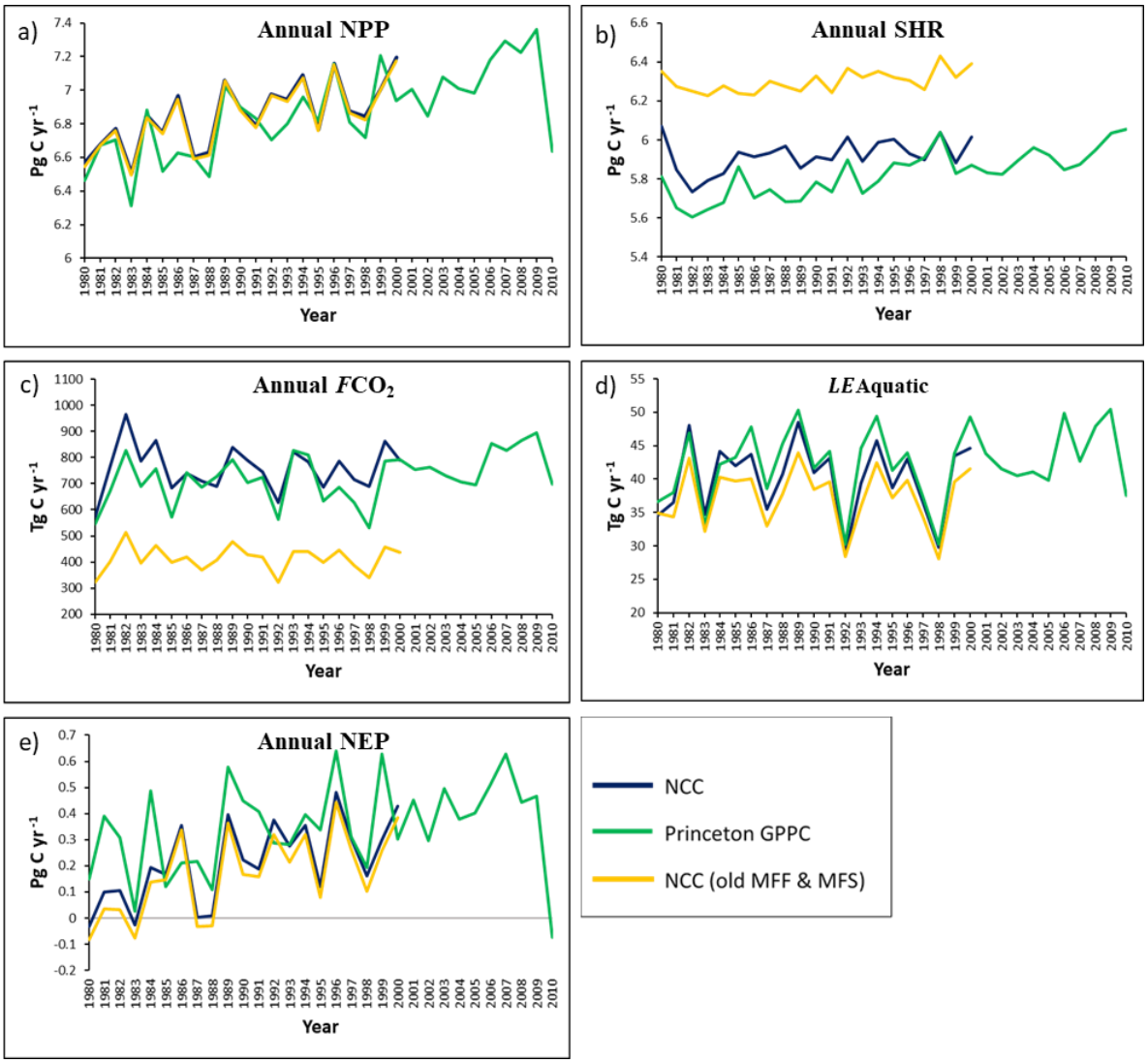


Chapter 3 Figure 7: Simulated annual C budget (NEP) for the Amazon basin annual mean (1980-2000), where NEP is net ecosystem production, NPP is terrestrial net primary productivity, TF is throughfall, SHR is soil heterotrophic respiration, FCO_2 is aquatic CO_2 evasion, LOAC is C leakage to the land-ocean aquatic continuum (FCO_2 + to $LE_{Aquatic}$), and $LE_{Aquatic}$ is the export C flux to the coast. Numbers refer to mean across the six simulations while numbers in parentheses refer to range.

3.3.4. Interannual variation of the carbon fluxes within the Amazon Basin

Our results show considerable interannual variation in NPP, from a mean low of 6.41 (6.29-6.52) Pg C yr^{-1} in 1983, to a high of 7.16 (7.14- 7.16) Pg C yr^{-1} in 1996 (Fig.8-a), though the Princeton GPCC simulation, which runs until 2010, has several years (2006-2009 inclusive) with slightly higher NPP. This variation has a strong positive correlation with precipitation (detrended R^2 =0.48, p<0.001 with NCC; detrended R^2 =0.43, p<0.0001 with Princeton GPCC, Table S6 & S7, Fig. 9-a) and a strong negative correlation with temperature (detrended R^2 =0.56, p<0.0001 with NCC; detrended R^2 =0.43, p<0.0001 with Princeton GPCC, Table S6 & S7, Fig. 9-b). In addition, NPP is inversely correlated with the multivariate ENSO index (MEI, sum of monthly MEI from July of preceding year to June of concurrent year, detrended R^2 =0.40, p<0.01 with NCC; detrended R^2 =0.35, p<0.001 with Princeton GPCC, Table S6 & S7, Fig. 9-c) (Wolter et al., 2011). We also find substantial interannual variation in SHR from a mean (across the two runs with new floodplain forcing) low of 5.69 (5.41- 6.03) Pg C yr^{-1} in 1982 to a high of 6.06 (5.91- 6.24) Pg C yr^{-1} in 1998 (Fig. 8-b). Conversely to NPP, SHR is positively correlated with temperature, and negatively correlated with rainfall, though these

relationships are relatively weak (relationship with temperature not significant with NCC, detrended temperature R^2= 0.13 with Princeton GPCC, $p<0.05$; detrended rainfall R^2 =0.19, $p<0.05$ with NCC, detrended rainfall R^2 = 0.24, $p<0.01$ with Princeton GPCC).

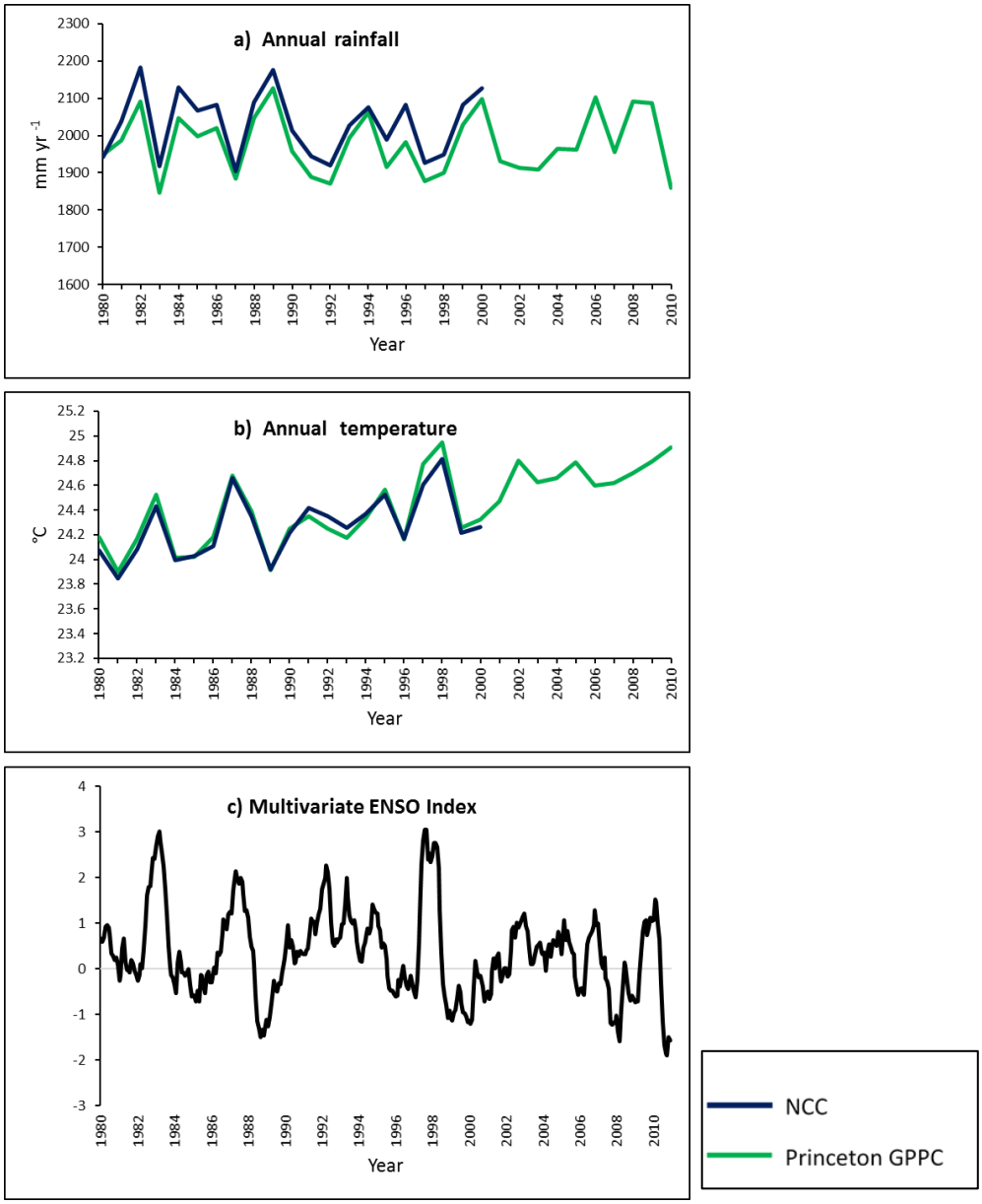


Chapter 3 Figure 8: Simulated annual variation in NEP and its components over the Amazon Basin from 1980-2000 (to 2010 in case of Princeton GPCC).

Our results also show considerable inter-annual (1980-2000) variation in inland water CO_2 evasion from a mean low of 571 (402- 759) Tg C yr^{-1} in 1980 to a high of 920 (633- 1,267) Tg C yr^{-1} in 1982 (Fig.8-c), strongly correlated with precipitation (detrended R^2 = 0.55, $p<0.001$ with NCC; detrended R^2 = 0.64, $p<0.0001$ with Princeton GPCC, Table S6 & S7) and inversely correlated with temperature (detrended R^2 = 0.21, $p<0.05$ with NCC; detrended R^2 = 0.18, $p<0.05$ with Princeton GPCC, Table S6 & S7). While both model runs rank 1982 as having the

highest CO_2 evasion over the simulation period (1980-2000), there is some divergence in regards to the lowest ranking year. The NCC run ranks 1980 lowest with 584 (422-759) Tg C yr^{-1} whereas the Princeton GPCC run ranks 1998 lowest with a total of 538 (399-685) Tg C yr^{-1}. In 1980 the Amazon rainy season was exceptionally dry (Andreoli et al., 2012), while 1998 coincides with a strong El Nino event (Fig.9-c) and associated anomalously low precipitation and high temperatures (Wenhong et al., 2011; Gloor et al., 2013, 2015). Conversely, 1982 experienced an exceptionally wet rainy season (Andreoli et al., 2012). These temporal patterns are also exhibited in the rainfall and temperature parameters from both of the climate forcings used in this study (Fig.9).

At the interannual timescale, aquatic CO_2 evasion is only weakly to moderately correlated with NPP (detrended $R^2 = 0.19$, $p<0.05$ NCC run; detrended $R^2 = 0.28$, $p<0.01$ with Princeton GPCC run, Table S6 & S7) and therefore the proportion of NPP lost through the LOAC is variable, ranging from 9% to 13%. In contrast, inland water CO_2 evasion is strongly inversely correlated with SHR (detrended $R^2 = 0.76$, $p<0.0001$ NCC run; detrended $R^2 = 0.66$, $p<0.0001$ with Princeton GPCC run, Table S6 & S7), indicating that years with less SHR have more evasion, and vice versa. Again, we find considerable interannual variation in C flux to the coast (Fig. 8 d) displaying a similar pattern to aquatic CO_2 evasion (aquatic CO_2 evasion *versus* C flux to coast $R^2 = 0.48$ for NCC, $p<0.001$; $R^2 = 0.64$ $p<0.0001$ for Princeton GPCC, Table S6 & S7). In relative terms, the LOAC fluxes show far greater interannual variation than the terrestrial C fluxes. For example, aquatic CO_2 evasion (NCC, 1980-2000) has a coefficient of variation (CV) of 11.7%, while the lateral flux of C to the coast has a CV of 13.6%. In contrast, NPP and SHR have a CV of only 2.9% and 1.5%, respectively.



Chapter 3 Figure 9:Interannual variation in a) rainfall and b) temperature. c) Monthly multivariate ENSO Index from 1980-2010 (Wolter et al., 2011)

As with its constituent components, simulated NEP shows considerable interannual variation (Figure 8-e) from a low of -0.05 (-0.11 – 0.03) Pg C yr^{-1} in 1983 to a high of 0.52 (0.41- 0.64) Pg C yr^{-1} in 1996. NEP is positively correlated with rainfall (detrended $R^2 = 0.27$, $p<0.05$ NCC run; detrended $R^2 = 0.25$, $p<0.01$ with Princeton GPCC run, Table S6 & S7) and negatively correlated with temperature (detrended $R^2 = 0.45$, $p<0.001$ NCC run; detrended $R^2 = 0.41$, $p<0.001$ with Princeton GPCC run, Table S6 & S7). The association with ENSO (detrended R^2

= 0.35, p<0.01 NCC run; detrended R^2 = 0.26, p<0.01 with Princeton GPCC run, Table S6 & S7) can be clearly seen in the simulated time series of NEP. Of the top six years with the lowest NEP (largest source of C to the atmosphere), four coincide with strong El Nino events, namely 1983, 1988, 1987 and 1998. Conversely, several of the years with the highest NEP (largest sink of atmospheric CO_2) take place during La Nina events, notably the strong La Nina event of 1988-1989, which results in the second highest simulated NEP; note that 2011 was one of the strongest La Nina on record but is not included in our forcing period. Taking the Princeton GPCC run alone, 2010 has the lowest NEP being a net CO_2 source to the atmosphere of -0.12 Pg C yr^{-1} (-0.14- -0.07) and coincides with another El Nino event combined with anomalously high Atlantic sea surface temperatures (SSTs) (Lewis et al., 2011).

We diagnosed the covariance between aquatic CO_2 evasion and the terrestrial C balance (defined as NPP-SHR) to determine how the variance in aquatic CO_2 evasion contributes to the overall variance in NEP across the simulation period. We find a negative covariance between aquatic CO_2 evasion and the terrestrial C balance of -0.024 and -0.022 for NCC and Princeton GPCC, respectively. Moreover, the terrestrial C balance is substantially more sensitive to changes in both precipitation and temperature than NEP (Tables S8-S11). For example (NCC run, Table S8), across the Amazon basin we find that the terrestrial C balance increases by 120 Tg C yr^{-1} for every 100mm increase in rainfall, while NEP only increases by 57 Tg C yr^{-1}. Note that these values are based on simple linear regression and thus the sensitivity to rainfall may be exaggerated but this is the case for both values.

As a consequence of this change in sensitivity, the variation of the budget is less pronounced once the aquatic components are incorporated; the terrestrial C balance has a SD of 0.20 Pg C yr^{-1} and 0.24 Pg C yr^{-1} with NCC and Princeton GPCC respectively, while NEP has a SD of 0.15 Pg C yr^{-1} and 0.17 Pg C yr^{-1}. These results concur with the idea of CO_2 evasion having a

moderating effect on overall heterotrophic respiration and suggest that accounting for CO_2 evasion from the river-floodplain network dampens the interannual variation in NEP.

3.4. Discussion

Our value of mean (across two models) NPP rate of 1,214 g C m^{-2} yr^{-1} matches closely to previous estimates in the Amazon. Rodig et al. (2018) estimated a mean annual NPP of 1,130 g C m^{-2} yr^{-1} using the forest gap FORMIND model, while a value of 1,030 g C m^{-2} yr^{-1} was derived from MODIS remote-sensing data (Zhao & Running, 2010).

Our estimate of mean total annual aquatic CO_2 evasion of 746 (526-998) Tg C yr^{-1} is relatively close to the 800 Tg C yr^{-1} proposed by Rasera et al. (2013) from upscaling of observations, over a larger basin area of 6×10^6 km^2. If we adjust our estimate (calculated across a smaller basin area of 5.6×10^6 km^2) to the same area, then we get a closer estimate of 799 Tg C yr^{-1}. Moreover, if we only base our mean CO_2 evasion estimate on the same years as Rasera et al. (i.e. 2006- 2010), we actually produce a larger value of 887 Tg C yr^{-1} (based on Princeton GPCC run only). We also estimate a similar distribution of CO_2 evasion between low and high flow periods (Table S4). Like those of Rasera et al. (2013), our results exhibit a strong seasonal cycle in CO_2 evasion, with the high flow season (monthly avg. discharge > yearly avg. discharge) contributing approximately 75% of the annual total. In contrast, our results are considerably higher than those of Richey et al. (Table S4). It is encouraging that our results are similar to those of Rasera et al. (2013) as their upscaling was based on an extensive 5-year field campaign where the flux of CO_2 was directly measured while those of Richey et al. (2002) were derived indirectly from pCO_2 measurements. In terms of flood extent, the Rasera et al. study used the same assumptions for water surface area as Richey et al (2002), who in turn used an older version (Hess et al., 2002) of the Hess et al. (2015) floodplain product use in this study.

For the central quadrant of the Amazon basin alone (area = 1.77×10^6 km^2), we simulate a mean annual aquatic CO_2 evasion (1980-2000) of 341 and 318 Tg C yr^{-1} with NCC and Princeton GPPC, respectively, close to the 360 Tg C yr^{-1} estimated by Rasera et al. (2013), but considerably higher than the 210 Tg C yr^{-1} of Richey et al. (2002) and the 229 Tg C yr^{-1} of Lauerwald et al. (2017). Our results concur with both previous upscaling studies that the central Amazon basin contributes approximately 45% of the basin wide aquatic CO_2 evasion (Table S4). The differences between our CO_2 evasion estimates and those of Richey et al. (2002) are largely due to gas exchange velocity; we applied a fixed k_{600} rate of 3.5 m day^{-1} for rivers, while they used very conservative gas exchange velocities of 1.2 to 2.3 m day^{-1}. Conversely, the differences between our results and those of Lauerwald et al. (2017) are largely a result of the increase in maximal fraction of floodplain (MFF) across the basin, and the resultant increase in direct C inputs to inundated areas from canopy through-fall, submerged litter and soils. Our estimated DOC export to the coast (downstream of Obidos) of 34 (34-44) Tg C yr^{-1} is relatively high; Lauerwald et al. (2017), Richey et al. (1990) and Moreira-Turcq et al. (2003) estimated this flux at 23.4 Tg C yr^{-1}, 24.4 Tg C yr^{-1} and 27 Tg C yr^{-1}, respectively.

Our results for the mean NEP of 0.23 (0.15-0.33) generally concur with previous estimates. Tian et al. (1998) used the Terrestrial Ecosystem Model to estimate a mean annual NEP, without considering the LOAC loop of the carbon cycle (undisturbed ecosystems, 1980-1994), of 0.2 ±0.9 Pg C yr^{-1}. Another modelling study (S. Sitch, B. Smith and J. Kaplan, unpublished but cited in Prentice and Lloyd, 1998, page 620) also settled on a mean annual NEP of around 0.2 ±1.2 Pg C yr^{-1} over the same 15-year period. A 2016 review (Grace, 2016), compiled all of the existing literature to produce two estimates of the net C balance of the Amazon Basin; one 'bottom-up' approach using "plot data and remote sensing" and one 'top-down approach' using "aircraft-based measurements in the planetary boundary layer", the latter based on Gatti et al. (2014). These two approaches include perturbation fluxes such as deforestation and harvesting

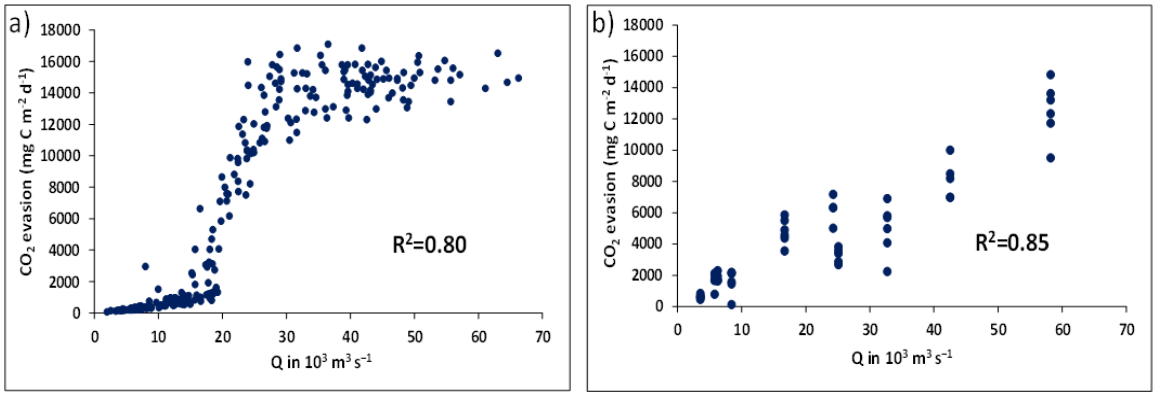
and evasion emissions in the atmospheric inversion estimate of Gatti et al. and are thus equivalent to our estimate of NBP. The bottom-up approach concludes that the Amazon Basin is a net C source to the atmosphere of 0.11 Pg C yr^{-1} when including land use change emissions but with an uncertainty of ± 0.16, in other words not markedly different from zero. The top-down approach came to a similar conclusion; that the Amazon is a net source to the atmosphere of only 0.06 Pg C yr^{-1} in a 'normal year' but only two years (2010 and 2011) were analyzed in Gatti et al. Again, the near neutral balance of Gatti et al. (2014) intrinsically includes aquatic CO_2 evasion (though not the lateral fluxes of C to the coast). They argue that the impact of riverine CO_2 evasion on the Amazon C balance is minimal as the "riverine organic carbon loop is very nearly closed". In other words, the vast majority of LOAC export to aquatic systems return to the atmosphere before leaving the Amazon Basin. In summary, the results of Gatti et al. (2014) are arguably the most comparable to our own and it is therefore encouraging that we produce a relatively similar NBP of 0.04 (-0.04-0.14) Pg C yr^{-1} (a difference of 100 Tg C^{-1} but with overlapping uncertainty ranges). It is important to note that ORCHILEAK does not incorporate methane fluxes. Indeed, if we include the recent estimate of the annual methane flux of approximately 40 Tg C^{-1} (Pangala et al., 2017) measured from the lower troposphere via aircraft; our NBP reduces to a neutral C balance.

While the new maximal fraction of floodplain (MFF) forcing leads to a dramatic increase in aquatic CO_2 evasion, it actually causes an overall decrease in the flux of CO_2 from the entire Amazon basin to the atmosphere. The greater inundation leads to a reduction in decomposition rates of litter, and soil organic matter. This suppression of organic matter decomposition has been observed in further field experiments (Dos Santos & Nelson, 2013), in addition to the study that informed the model configuration (Rueda-Delgado et al., 2006). This means that there is an additional net land C sink of approximately 40 Tg C yr^{-1} per year with the new floodplain compared to the old floodplain. While in a single year these differences are not so

substantial, over long time periods they could lead to significant differences in the long-term net C balance of the Amazon.

We found that the interannual variation in NPP is positively correlated with rainfall and negatively correlated with temperature and our results concur with previous research showing that drought years have significantly lower NPP. In our outputs, two of the years with the lowest NPP are 1983 and 1988, coinciding with two strong El Nino events (1982-1983 and 1987-1988, Figure 9), and corroborating the findings of Asner & Townsend (2000) based on analysis of remote sensing data from 1982-1993. Previous modelling studies such as Botta et al (2002) have also found 1983 and 1988 to be years with anomalously low NPP in the Amazon. Moreover, a 2011 study that combined remote sensing and modelling (Potter et al., 2011) estimated that the 2010 drought caused a reduction in NPP in the Amazon of 7% relative to the La Nina year 2008, and we produce a similar value of 8% (0.58 Pg C). However, a more recent study (Doughty et al., 2015) contradicts these findings. Doughty et al. (2015) measured NPP, autotrophic respiration and heterotrophic respiration at thirteen 1ha plots across South America from 2009-2011 and found that NPP remained relatively constant throughout the period. They observed a reduction in CO_2 uptake via photosynthesis by 0.38 Pg C yr^{-1} during the 2010 drought, but this was offset by a concurrent reduction in autotrophic respiration. They observed that the trees prioritised investment in growth (canopy tissue), while they reduced autotrophic respiration investment in tissue maintenance and defence, which ultimately may have caused an increase in tree mortality post drought (Doughty et al., 2015). The inability of dynamic global vegetation models (DGVMs), as well as remote sensing driven algorithms (Zhao & Running, 2010; Medlyn, 2011; Wang et al., 2013) to represent these complex biological interactions is a major limitation in current efforts to estimate NPP at the regional to global scale.

Our results show that both the seasonality and interannual variation in aquatic CO_2 evasion, are closely correlated with discharge. In Figure 10 a) we show the relationship between simulated monthly discharge and CO_2 evasion on the Madeira River at Porto Velho (R^2=0.81) (see Fig. S1 for location). The Madeira basin contains approximately one fourth of Amazonian wetlands (Melack and Hess 2010), including the extensive Llanos de Moxos and was the subject of a recent CO_2 evasion field campaign (Almeida et al., 2017). Our relationship follows a sigmoid curve where aquatic CO_2 evasion increases slowly at first while discharge remains in bank. Once the river over-tops its banks, CO_2 evasion increases rapidly before levelling out once the full area of the floodplain is saturated. Thus, at the basin scale, aquatic CO_2 evasion not only increases because of larger floodplain surface area, but also because of higher areal rates. This highlights the disproportionate importance of floodplains as a source of C and supports the findings of Almeida et al. (2017, Fig. 10, b). While they found a similarly strong relationship between observed discharge and aquatic CO_2 evasion at Porto Velho (R^2=0.85), as well as a similar range of values, the relationship does not follow precisely the same shape as ours. Their increase in evasion rate is more gradual and they do not observe a plateauing of CO_2 evasion above a certain discharge. This perhaps suggests that we underestimate the maximum extent of the floodplain in this specific model grid, and indeed, the location of Porto Velho, is in the minority of model grids where the maximum inundation actually decreases with the implementation of the new MFF forcing file.



Chapter 3 Figure 10:a); Monthly (1980-2000) simulated (NCC) aquatic CO_2 evasion versus simulated discharge on the Madeira River at Porto Velho and b); Observed aquatic CO_2 evasion versus observed on the Madeira River at Porto Velho, measured between 2009 and 2011.

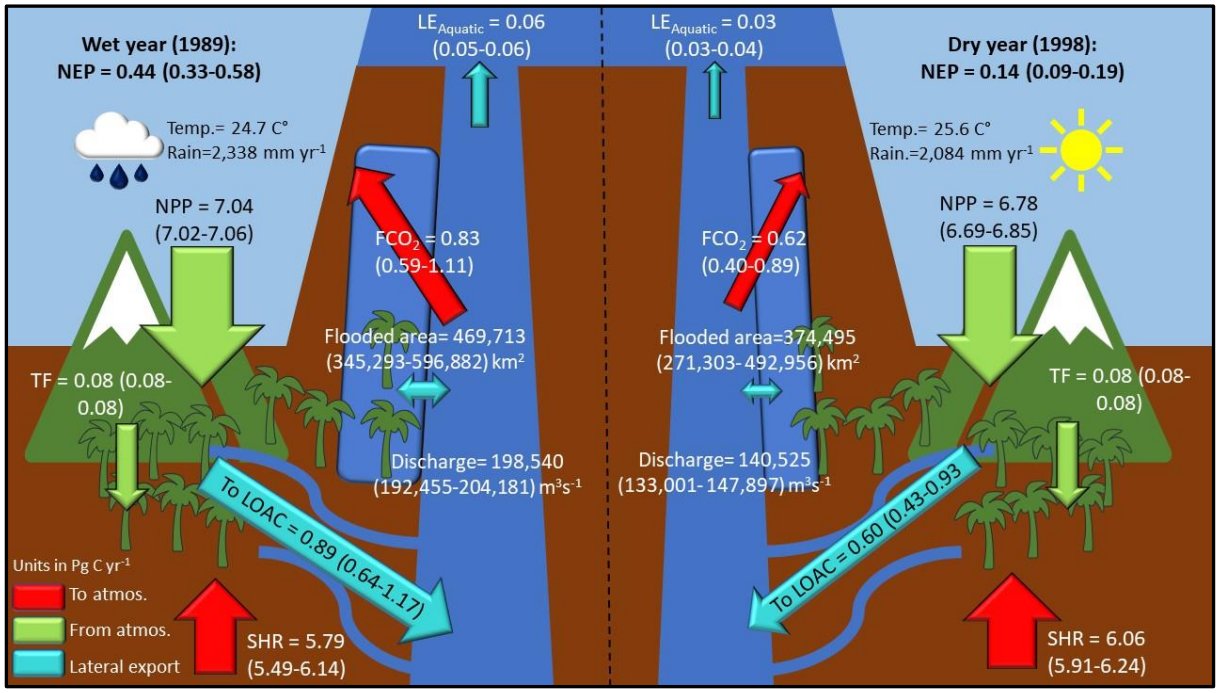
The pattern of interannual variation in NEP over the 1980s and 1990s in our results is consistent with that found in previous modelling studies over the same period (Prentice and Lloyd, 1998; Tian et al., 1998). Interestingly we find smaller interannual variation than these previous modelling studies that did not include inland water fluxes, further supporting the idea that incorporating aquatic fluxes dampens the interannual variation in NEP. Indeed, a 2013 study (Wang et al., 2013) found results to suggest that some DGVMs overestimate the sensitivity of net ecosystem exchange (NEE) to precipitation. The relationship between our simulated NEP and precipitation is generally weaker than that found in previous models across the tropical region (Wang et al., 2013, in this case NEE), and the addition of the aquatic C fluxes appears to be at least partly responsible for this; the sum of terrestrial fluxes (NPP-SHR) is more strongly correlated with precipitation (detrended $R^2 = 0.58$, $p<0.0001$ NCC run; detrended $R^2 = 0.51$, $p<0.0001$ with Princeton GPCC run) than NEP (detrended $R^2 = 0.27$, $p<0.05$ NCC run; detrended $R^2 = 0.25$, $p<0.01$ with Princeton GPCC run), which includes aquatic components.

Despite some of the limitations of DGVMs discussed, namely their inability to fully capture the complex effects of droughts on NPP, the response of our model to drought events concurs with observational based studies, and most significantly to those based on the measurement of

atmospheric CO_2 fluxes. The 2010 Amazon drought was one of the most severe ever recorded and related to another El Nino event, as well as anomalous SSTs (Lewis et al., 2011). Gatti et al. (2014) used small aircraft to measure CO_2 fluxes just above the Amazon rainforest (lower-troposphere) and found that in 2010, the Amazon basin was a net source to the atmosphere of 0.48± 0.18 Pg C yr^{-1}. A 2015 study (van der Laan-Luijkx et al., 2015), further constrained the results of Gatti et al. using remote sensing data and estimated a smaller atmospheric CO_2 source between 0.07 and 0.31 Pg C yr^{-1} for 2010. Based on our Princeton GPCC run, we similarly estimate that in 2010, the Amazon was an overall CO_2 source for the atmosphere of 0.33 (0.35 - 0.29) Pg C yr^{-1} (based on NBP). Additionally, using a combined remote sensing and modelling approach, Potter et al. (2011) estimated that the 2010 drought caused a loss of biomass in the Amazon of 0.50 Pg C yr^{-1} relative to the strong La Nina year of 2008, and we produce a similar NEP deficit 0.51 Pg C yr^{-1}.

In Figure 11, we show our simulated C budget for a drought year, 1998, and an anomalously wet year, 1989, to illustrate how both terrestrial and aquatic C fluxes react to climatic extremes. In 1989, high aquatic CO_2 evasion to the atmosphere driven by high rainfall and large floodplain inundation, partly offsets a relatively large terrestrial sink, caused by high terrestrial NPP and low SHR. In 1998 the opposite occurs; low rainfall results in a low flux of CO_2 from inland waters to the atmosphere, which moderates a relatively high SHR flux and low terrestrial NPP. As previously noted, aquatic CO_2 evasion is highly sensitive to rainfall and in turn both discharge and inundation, and displays greater interannual variation than the terrestrial C fluxes. Aquatic CO_2 evasion is positively correlated with NPP but the two fluxes represent opposite signals in terms of C exchange with the atmosphere, while aquatic CO_2 evasion is inversely correlated to SHR, both fluxes being C sources for the atmosphere. For these two reasons, the aquatic fluxes generally act to compensate the difference between terrestrial NPP and SHR and thus dampen overall interannual variation in the net C balance.

Another process not accounted for in our model is C sequestration on floodplains. Interestingly, a 2003 study (Aalto et al., 2003) showed that, sediment accumulation on Amazon floodplains is closely linked to the ENSO cycle. Like our findings for aquatic CO_2 evasion, sediment accumulation was found to be higher during La Nina years, and most notably in 1988. Despite not accounting for this C sink term in our model, the comparison of our net C balance for the Amazon (NBP) against observations (Grace et al., 2016) suggests that if anything we are still underestimating the net flux of C from the Amazon basin to the atmosphere.



Chapter 3 Figure 11:Simulated annual C budget for left; the Amazon basin for the year 1989, and right; the Amazon basin for the year 1998, where NEP is net ecosystem production, NPP is terrestrial net primary productivity, TF is throughfall, SHR is soil heterotrophic respiration, FCO_2 is aquatic CO_2 evasion, LOAC is C leakage to the land-ocean aquatic continuum (FCO_2 + LEAquatic), and LEAquatic is the export C flux to the coast. Numbers refer to mean across the six simulations while numbers in parentheses refer to range.

3.4.1. The importance of integrating the LOAC within the land carbon cycle

The Amazon is facing a number of threats including climate change, land use change and dam construction (Nobre et al., 2016). Climatic events such as droughts and floods are becoming more frequent (Marengo et al., 2011; Gloor et al., 2013; Zulkafli et al., 2016), while southern Amazonia has experienced a general lengthening of the dry season (Fu et al., 2013). The region

is also undergoing a boom in dam construction with 140 dams under construction or already in operation, and a further 288 planned (Latrubesse et al., 2017) with direct impact on the C retention efficiency within the LOAC (Maavara et al., 2017). In addition, a recent study demonstrated that the lowland floodplain forests of the Amazon are less resilient to fires than terra firme forests (Flores et al., 2017).

For these reasons, it is vital that the flood dynamics of the Amazon can be correctly represented in biogeochemical models. The implementation of a new floodplain forcing file based on high resolution SAR data substantially improves our ability to accurately simulate the seasonality in observed flooding. Moreover, it leads to a 97% increase in our estimate of mean annual CO_2 evasion from the river-floodplain aquatic continuum and supports some larger previous estimates based on simple upscaling approaches (Table S4). Our results show that the LOAC fluxes, highly sensitive to hydrological variation, display greater interannual variation than the terrestrial C fluxes (NPP – SHR), and are thus disproportionately important to the overall variation of the net C balance, relative to their magnitude. We also find that the percentage of NPP lost to the LOAC is variable at the interannual timescale (Fig. 11).

Our results suggest that the linkage between the terrestrial and aquatic environment may be larger than previously thought and our estimate of aquatic CO_2 evasion from the Amazon is of a globally significant magnitude in terms of aquatic C fluxes. However, these results must be placed within the context of their overall impact on the net C balance of the Amazon Basin. While greater inundation increases aquatic CO_2 evasion, it simultaneously decreases the decomposition of organic matter in litter and soils and we show that the net impact of greater flooding is in fact a reduction in the flux of CO_2 from the Amazon basin to the atmosphere. It is during years with the lowest precipitation, often associated with El Nino events that highest net flux of CO_2 to the atmosphere are simulated. Indeed, we find that aquatic C fluxes partly compensate terrestrial C fluxes, and therefore moderate the overall interannual variation in

NEP. Thus, DGVMs that do not account for aquatic fluxes may overestimate the magnitude of interannual variation in NEP. This calls for a fully integrated view of the land carbon cycle, which cannot be achieved with empirical studies alone and highlights the value of a model that can integrate the terrestrial and aquatic C cycles.

3.5. Supporting Information

Chapter 3 Table S 1: Observed and simulated interannual variation in discharge at Obidos				
Year	Discharge Observed	Discharge-NCC	Discharge-Princeton GPCC	Discharge-NCC Old MFF & MFS
1980	144,370	147,285	149,910	139,553
1981	153,225	166,707	164,040	161,241
1982	182,842	212,865	191,576	202,947
1983	143,687	171,066	149,387	165,846
1984	175,450	199,124	175,196	189,411
1985	164,633	175,744	160,053	168,478
1986	183,717	186,200	185,116	177,521
1987	166,811	164,406	161,826	159,829
1988	167,403	172,462	175,977	165,191
1989	199,050	203,105	193,973	193,937
1990	168,582	176,768	162,801	170,335
1991	171,414	182,388	172,742	176,686
1992	139,094	145,868	138,766	140,125
1993	180,933	182,199	189,133	173,582
1994	196,483	184,887	185,723	179,884
1995	154,871	163,971	154,392	157,061
1996	180,167	179,224	166,462	172,236
1997	169,687	162,210	153,735	157,735
1998	149,533	146,927	134,056	140,562
1999	185,917	186,924	170,880	180,278
2000	181,942	181,013	182,018	172,380
RMSE %		7	4	6
NSE		0.50	0.79	0.62
R2		0.66	0.81	0.67

Chapter 3 Table S 2: Pearson correlation coefficient (r) between the various carbon fluxes and climate variables (NCC climate data). NPP is terrestrial net primary productivity, SHR is soil heterotrophic respiration, *F*CO_2 is aquatic CO_2 evasion, NEP is net ecosystem production, MEI is multivariate ENSO Index and $LE_{Aquatic}$ is the export C flux to the coast.

	SHR	*F*CO_2	$LE_{Aquatic}$	NEP	Rain.	Temp.	Year	MEI
NPP	0.14	0.33	0.43	0.97	0.43	-0.19	0.68	-0.51
SHR	1.00	-0.79	-0.44	0.10	-0.44	0.38	0.40	0.19
Aquatic CO_2 evasion		1.00	0.69	0.25	0.74	-0.41	0.02	-0.36
Lateral C			1.00	0.36	0.79	-0.55	-0.01	-0.75
NEP				1.00	0.34	-0.19	0.62	-0.51
Rain					1.00	-0.68	-0.10	-0.56
Temp.						1.00	0.44	0.68
Year								-0.07

Chapter 3 Table S 3: Pearson correlation coefficient (r) between the various carbon fluxes and climate variables (Princeton GPCC climate data). NPP is terrestrial net primary productivity, SHR is soil heterotrophic respiration, FCO_2 is aquatic CO_2 evasion, NEP is net ecosystem production, MEI is multivariate ENSO Index and $LE_{Aquatic}$ is the export C flux to the coast.

	SHR	FCO_2	$LE_{Aquatic}$	NEP	Rain.	Temp.	Year	MEI
NPP	0.41	0.62	0.51	0.86	0.48	0.19	0.71	-0.53
SHR		-0.15	-0.18	0.01	-0.28	0.72	0.79	0.03
Aquatic CO_2 evasion			0.80	0.47	0.75	0.01	0.39	-0.48
Lateral C				0.44	0.86	-0.25	0.18	-0.74
NEP					0.48	-0.21	0.30	-0.53
Rain						-0.40	0.03	-0.64
Temp.							0.72	0.36
Year								-0.16

Chapter 3 Table S 4: Mean annual CO_2 evasion from the river-floodplain network of the Amazon Basin (Tg C yr^{-1})

	NCC (standard MFF, 1980-2000)	Princeton GPCC (standard MFF, 1980-2010)	Lauerwald et al., 2017 (1980-2000)	Richey et al., 2002 (1995-1996)	Rasera et al., 2013 (2006-2010)
Amazon Basin	758	725	379	470	800
Central Amazon Basin	341 (45%)	318 (44%)	229 (60%)	210 (45%)	360 (45%)
High flow (central)	252 (74%)	235 (74%)	/	/	290 (81%)
Low flow (central)	89 (26%)	83 (26%)	/	/	70 (19%)

Chapter 3 Table S 5: Performance statistics for modelled vs observed DOC flux for different model configurations

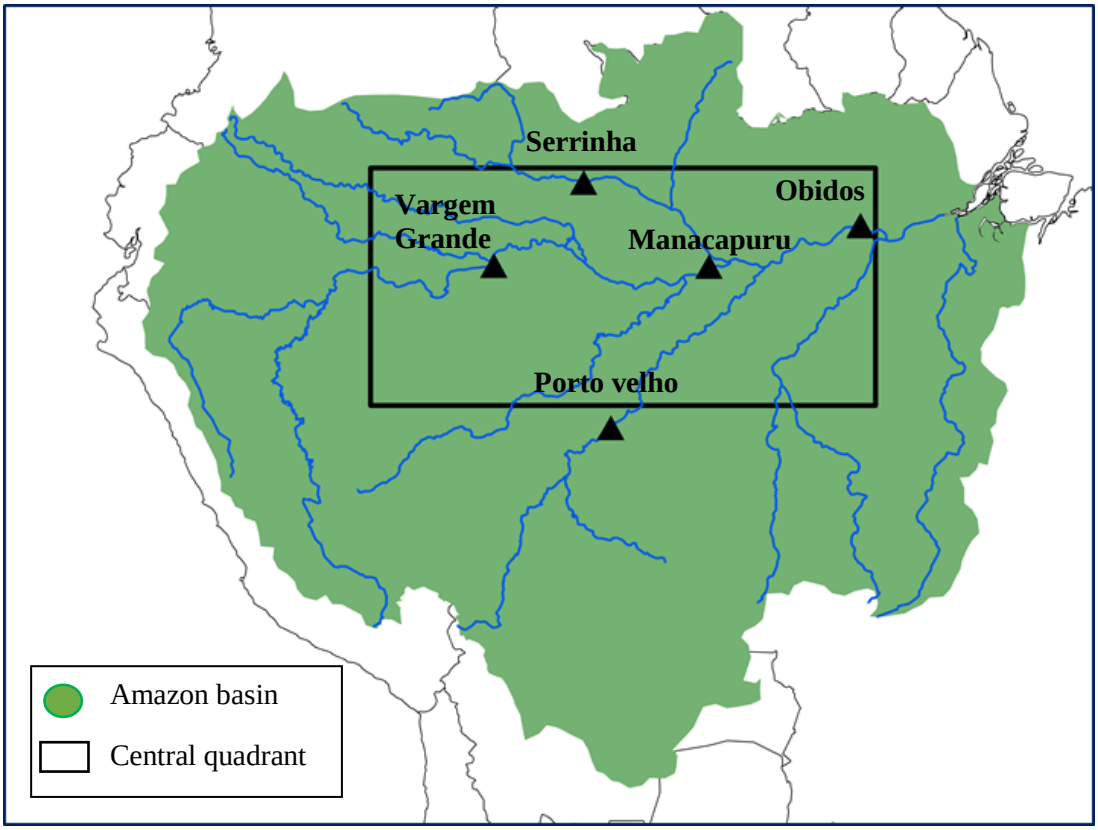
	NCC τ_{flood} 1.4			Princeton GPCC τ_{flood} 1.4			Old floodplain NCC τ_{flood} 1.4		
Location	RSME	NSE	R^2	RSME	NSE	R^2	RSME	NSE	R^2
Rio Negro at Serrinha	28%	0.74	0.84	34%	0.62	0.78	21%	0.85	0.91
Amazon at Obidos	40%	-0.32	0.62	33%	0.14	0.59	24%	0.52	0.58
Rio Solimoes at Manacapuru	30%	0.34	0.52	27%	0.44	0.49	33%	0.20	0.41
Amazon at Vargem Grande	40%	-0.08	0.50	21%	0.71	0.72	30%	0.41	0.68

Chapter 3 Table S 6: Pearson correlation coefficient (r) between detrended carbon fluxes and detrended climate variables (NCC climate data). NPP is terrestrial net primary productivity, SHR is soil heterotrophic respiration, FCO_2 is aquatic CO_2 evasion, NEP is net ecosystem production, MEI is multivariate ENSO Index and $LE_{Aquatic}$ is the export C flux to the coast.

	SHR	FCO_2	$LE_{Aquatic}$	NEP	Rain.	Temp.	MEI
NPP	-0.19	0.43	0.60	0.95	0.69	-0.75	-0.63
SHR		-0.87	-0.47	-0.21	-0.44	0.25	0.23
Aquatic CO_2 evasion			0.69	0.31	0.74	-0.46	-0.36
Lateral C				0.47	0.80	-0.61	-0.75
NEP					0.52	-0.67	-0.60
Rain						-0.71	-0.58
Temp.							0.80

Chapter 3 Table S 7: Pearson correlation coefficient (r) between detrended carbon fluxes and detrended climate variables (Princeton GPCC climate data). NPP is terrestrial net primary productivity, SHR is soil heterotrophic respiration, FCO_2 is aquatic CO_2 evasion, NEP is net ecosystem production, MEI is multivariate ENSO Index and $LE_{Aquatic}$ is the export C flux to the coast.

	SHR	FCO_2	$LE_{Aquatic}$	NEP	Rain.	Temp.	MEI
NPP	-0.34	0.53	0.55	0.96	0.66	-0.66	-0.59
SHR		-0.81	-0.54	-0.38	-0.49	0.36	0.25
Aquatic CO_2 evasion			0.80	0.40	0.80	-0.43	-0.45
Lateral C				0.41	0.87	-0.56	-0.72
NEP					0.50	-0.64	-0.51
Rain						-0.61	-0.63
Temp.							0.69



Chapter 3 Figure S 1:the geographical extent of the Amazon basin, and its central quadrant, along with major monitoring stations on the Amazon, Rio Madeira (Porto Velho), and Rio Negro (Serrinha).

Chapter 3 Table S 8: Climate sensitivities of C fluxes in Tg C yr^{-1} per 100mm increase in rainfall and 1 °C increase in temperature, for the NCC run (standard MFF) based on simple linear regression. NS = not significant (p>0.05). NEP is net ecosystem production, NPP is terrestrial net primary productivity, SHR is soil heterotrophic respiration, *F*CO_2 is aquatic CO_2 evasion, and $LE_{Aquatic}$ is the export C flux to the coast.

	NPP	*F*CO_2	SHR	$LE_{Aquatic}$	NEP	NPP-SHR
Change per 100mm ↑ in rain.	89	60	-31	4	57	120
Change per 1 °C ↑ in temp.	-462	-179	NS	-14	-354	-546

Chapter 3 Table S 9: Climate sensitivities of C fluxes in Tg C yr^{-1} per 100mm increase in rainfall and 1 °C increase in temperature, for the Princeton GPCC run (standard MFF) based on simple linear regression. NS = not significant (p>0.05). NEP is net ecosystem production, NPP is terrestrial net primary productivity, SHR is soil heterotrophic respiration, *F*CO_2 is aquatic CO_2 evasion, and $LE_{Aquatic}$ is the export C flux to the coast.

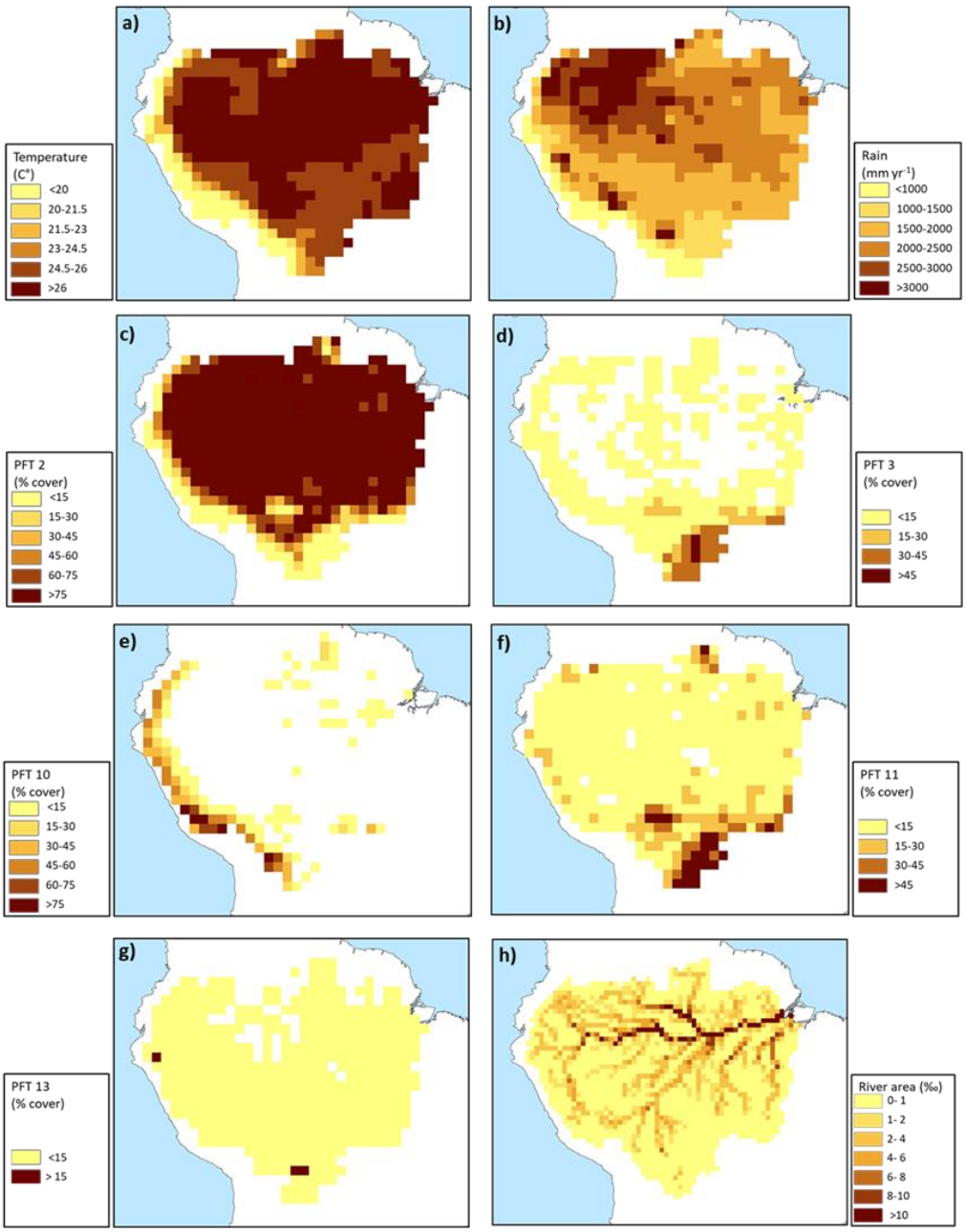
	NPP	*F*CO_2	SHR	$LE_{Aquatic}$	NEP	NPP-SHR
Change per 100mm ↑ in rain.	113	65	-34	4.4	78	147
Change per 1 °C ↑ in temp.	-566	-174	123	-14	-503	-689

Chapter 3 Table S 10: Climate sensitivities of C fluxes in Tg C yr^{-1} per 100mm increase in rainfall and 1 °C increase in temperature, for the NCC run (standard MFF) based on multiple linear regression (with both rainfall and temperature as independent variables). NS = not significant (p>0.05). NEP is net ecosystem production, NPP is terrestrial net primary productivity, SHR is soil heterotrophic respiration, *F*CO_2 is aquatic CO_2 evasion, and $LE_{Aquatic}$ is the export C flux to the coast.

	NPP	*F*CO_2	SHR	$LE_{Aquatic}$	NEP	NPP-SHR
Change per 100mm ↑ in rain.	NS	68	NS	3.6	NS	79
Change per 1 °C ↑ in temp.	-320	NS	NS	NS	-323	NS

Chapter 3 Table S 11: Climate sensitivities of C fluxes in Tg C yr^{-1} per 100mm increase in rainfall and 1 °C increase in temperature, for the Princeton GPCC run (standard MFF) based on multiple linear regression (with both rainfall and temperature as independent variables). NS = not significant (p>0.05). NEP is net ecosystem production, NPP is terrestrial net primary productivity, SHR is soil heterotrophic respiration, *F*CO_2 is aquatic CO_2 evasion, and $LE_{Aquatic}$ is the export C flux to the coast.

	NPP	*F*CO_2	SHR	$LE_{Aquatic}$	NEP	NPP-SHR
Change per 100mm ↑ in rain.	70	70	-30	4.3	NS	100
Change per 1 °C ↑ in temp.	-352	NS	NS	NS	-422	-385



Chapter 3 Figure S 2: Present day (1981-2010) spatial distribution of the principal climate and land-use drivers used in ORCHILEAK, across the Amazon Basin; a) mean annual temperature in °C, b) mean annual rainfall in mm yr^{-1}, c)-g) mean annual maximum vegetated fraction for PFTs 2,3, 10,11, and 13, h) river area. All at a resolution of 1° except for river area (0.5°).

Acknowledgments

Financial support was received from the European Union's Horizon 2020 research and innovation program under the Marie Sklodowska- Curie grant agreement No. 643052 (C-CASCADES project). RL acknowledges funding from the European Union's Horizon 2020 research and innovation program under grant agreement no. 703813 for the Marie Sklodowska-Curie European Individual Fellowship "C-Leak." PR acknowledges funding from the Belgian Federal Science Policy Office (BELSPO), project "Global impacts of hydrological and climatic extremes on vegetation" (SAT-EX) – Belgian research programme for Earth Observation Stereo III.

4. Historic and future trends of aquatic carbon fluxes integrated within the Congo Basin carbon balance

Abstract

As the world's second largest area of contiguous tropical rainforest and second largest river basin, the Congo basin plays an important role in the global carbon (C) cycle. Research has shown that terrestrial net primary productivity (NPP) and C storage in tree biomass has increased in recent decades in tropical Africa, due in large part to a combination of increasing atmospheric CO_2 concentrations and climate change. For the present day, it has been shown that a significant proportion of global terrestrial NPP is transferred laterally to the land-ocean aquatic continuum (LOAC) as dissolved organic carbon (DOC), particulate organic carbon (POC) and dissolved CO_2. Whilst the importance of LOAC fluxes in the Congo basin has been demonstrated for the present day, it is not known to what extent these fluxes have been perturbed historically, how they are likely to change under future climate change and land use scenarios, and in turn what impact these changes might have on the overall C cycle of the Congo. Here we apply the ORCHILEAK model to the Congo Basin and show that aquatic C fluxes have undergone considerable perturbation since 1861 to the present day, with CO_2 evasion and the export of C to the coast increasing by 28% and 29% respectively, largely because of rising atmospheric CO_2 concentrations. Moreover, under RCP 6.0 we predict that this perturbation will continue; over the full simulation period (1861-2099), we estimate that aquatic CO_2 evasion and the export of C to the coast will increase by 79% and 67% respectively. Finally, we show that the proportion of NPP lost to the LOAC also increases from approximately 3% to 5% from 1861-2099 as a result of both atmospheric CO_2 concentrations and climate change.

4.1. Introduction

As the world's second largest area of contiguous tropical rainforest and second largest river basin, the Congo basin plays an important role in the global carbon (C) cycle. Around 50 Pg C is stored in its above ground biomass (Verhegghen et al., 2012) while its peatlands contain another 30 Pg C (Dargie at al., 2017). Research has shown that terrestrial net primary productivity (NPP) has increased by an average of 10 g C m^{-2} yr^{-1} per year between 2001 and 2013 in tropical Africa (Yin et al., 2017), while storage in tropical tree biomass has increased by 0.34 Pg C yr^{-1} from 1968-2007 (Lewis et al., 2009) due in large part to a combination of increasing atmospheric CO_2 concentrations and climate change (Ciais et al., 2009; Pan et al., 2015). Moreover, these trends are predicted to continue into the future. It has also been evidenced that terrestrial NPP across the African continent varies by around 1.6 Pg C yr^{-1} between wet and dry years, though tropical evergreen forest NPP was shown to be the least variable with a coefficient of variation of 3.9% (Pan et al., 2015).

For the present day, it has been shown that a significant proportion of global terrestrial NPP is transferred laterally to the land-ocean aquatic continuum (LOAC) as dissolved organic carbon (DOC), particulate organic carbon (POC) and dissolved CO_2 (Cole at al., 2007; Battin et al., 2009; Regnier et al., 2013; Ciais et al. in review). This C can subsequently be evaded back to the atmosphere as CO_2, undergo sedimentation in wetlands and inland waters, or be transported to estuaries or the coast. The tropical region is a hotspot area for inland water C cycling (Lauerwald et al., 2015) due to high terrestrial NPP and precipitation, and a recent study used an upscaling approach based on limited observations to estimate present day CO_2 evasion from the rivers of the Congo basin at 133-177 Tg C yr^{-1} and the lateral C (TOC +DIC) export to the coast at 15.5 (13-18) Tg C yr^{-1} (Borges at al., 2015[a]). To put this into context, their estimate of aquatic CO_2 evasion represents 20- 27% of the global value estimated by Lauerwald et al.

(2015, 650 Tg C yr^{-1}) or 7-10% of the global estimate of Raymond et al. (2013, 1,800 Tg C yr^{-1}).

Whilst the importance of LOAC fluxes in the Congo basin has been demonstrated for the present day, it is not known to what extent these fluxes have been perturbed historically, how they are likely to change under future climate change and land use scenarios, and in turn what impact these changes might have on the overall C cycle of the Congo. With these knowledge gaps in mind, we aim to tackle the following research questions:

- To what extent have LOAC fluxes (CO_2 evasion and C export to the coast) changed from 1860 to the present day and what are the primary drivers of this change?
- How will these fluxes change under future climate and land use change scenarios (RCP 6.0 which represents the "no mitigation scenario")?
- What does the temporal evolution of LOAC fluxes mean for the wider C balance of the Congo Basin?

Understanding and quantifying these long-term changes requires a complex and integrated modelling approach. The ORCHILEAK model (Lauerwald et al., 2017), a new version of the land surface model ORCHIDEE (Krinner et al., 2005), is capable of representing both terrestrial and aquatic C fluxes for the present day in the Amazon basin (Lauerwald et al., 2017). Moreover, it was recently demonstrated that it can also recreate observed seasonal and interannual variation in Amazon C fluxes (Hastie et al., accepted).

In order to accurately simulate aquatic C fluxes, it is crucial that we can provide a realistic representation of the hydrological dynamics of the Congo River, including wetlands. Here, we develop new wetland forcing files for the ORCHILEAK model from the high-resolution dataset of Gumbricht et al. (2017) and apply the model to the Congo basin. After validating the model against observations of discharge, flooded area and DOC concentrations for the present day,

we then use the model to understand and quantify the long- term (1860-2100) temporal trends in both the terrestrial and aquatic C fluxes of the Congo Basin.

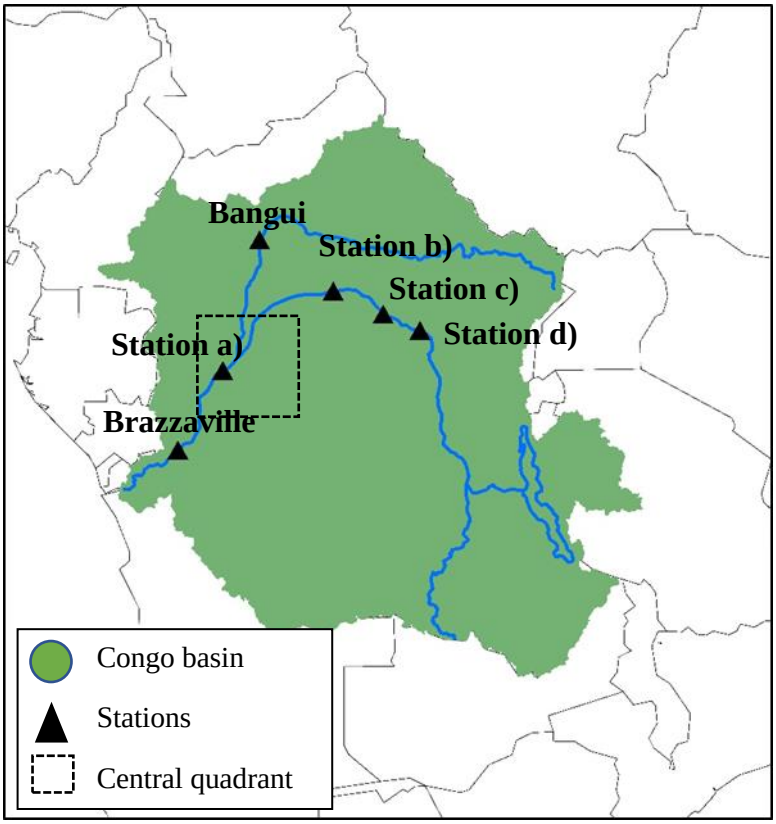
4.2. Methods

ORCHILEAK is a new branch of the ORCHIDEE land surface model (LSM), building on past model developments such as ORCHIDEE-SOM (Camino Serrano, 2015), and represents the first LSM-based approach which fully integrates the aquatic C cycle within the terrestrial. In this study, as in previous (Lauerwald et al., 2017, Hastie et al. accepted), we run the model at a spatial resolution of 1° and use the default time step of 30 min for all vertical exchanges of C, water and energy between soil, vegetation and the atmosphere, as well as the 1-day times-step for the lateral routing of water. Until now, ORCHILEAK has been parameterized and calibrated only for the Amazon Basin (Lauerwald et al., 2017, Hastie et al. accepted). In order to adapt and apply ORCHILEAK to the specific characteristics of the Congo River basin (2.1), we had to establish new forcing files representing the maximal fraction of floodplains (MFF) and the maximal fraction of swamps (MFS) (2.2) and to recalibrate the river routing module of ORCHILEAK (2.3). All of the processes represented in ORCHILEAK remain identical to those previously represented for the Amazon ORCHILEAK (Lauerwald et al., 2017). In the following methodology sections, we describe; 2.1- Site description, 2.2- Development of wetland forcing files, 2.3- Calibration of hydrology, 2.4- Simulation set-up, 2.5- Evaluation and analysis of simulated fluvial C fluxes, and 2.6- Calculating the net carbon balance of the Congo Basin. For a full description of the ORCHILEAK model please refer to sections Chapter 3 and Appendix 8.1.

4.2.1. Site description

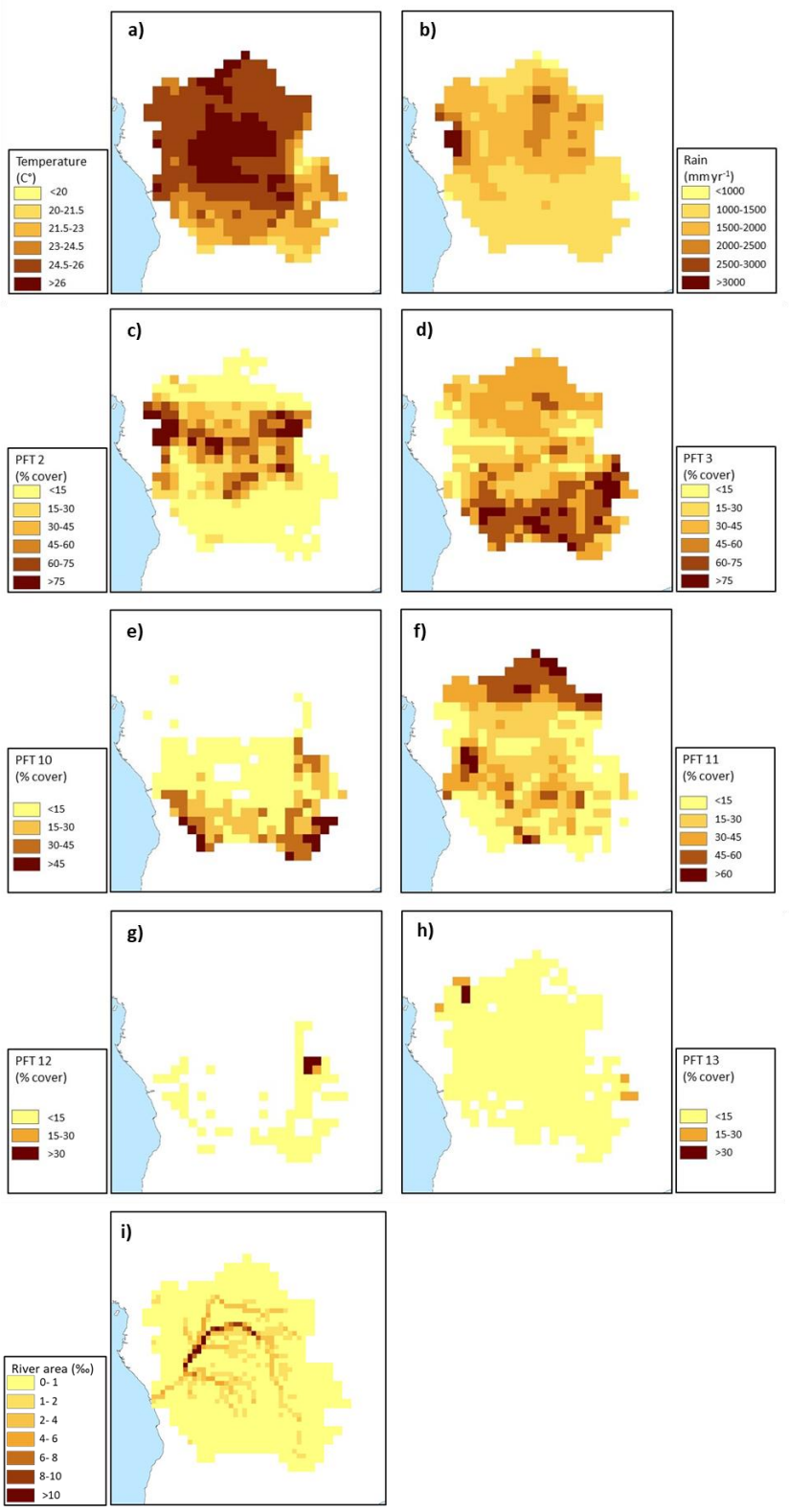
The Congo Basin is the world's second largest area of contiguous tropical rainforest and second largest river basin (Fig. 1), covering an area of 3.7 $x10^6$ km^2, with a mean discharge of around

42,000 m^{-3} s^{-1} (O'Loughlin et al., 2013) and a monthly variation of between 24,700–75,500 m^{-3} s^{-1} (Coynel et al., 2005).



Chapter 4 Figure 1:Extent of the Congo Basin, central quadrant of the "Cuvette Centrale" and sampling stations along the Congo and Ubangi Rivers.

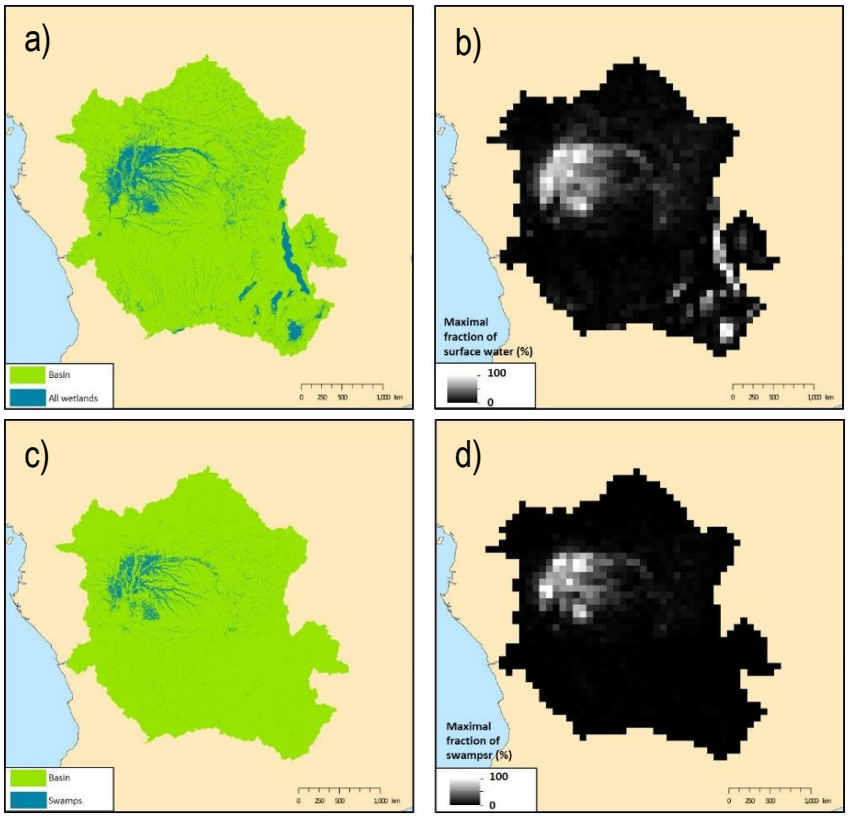
The major climate (ISMSIP2b) and land-cover (LUH-CMIP5) characteristics of the Congo Basin for the present day (1981-2010) are shown in Figure 2. The mean annual temperature is 25.2 °C but with considerable spatial variation from a low of 18.4°C to a high of 27.2°C (Fig. 2 a), while mean annual rainfall is 1520mm, varying from 733 mm to 4087 mm (Fig. 2 b). Land-use is mixed with tropical broad-leaved evergreen (PFT2, Fig. 1 c), tropical broad-leaved rain green (PFT3, Fig. 1 d), C3 grassland (PFT10, Fig. 2 e) and C4 grassland (PFT11, Fig. 2 f) covering a maximum of 25%, 35%, 15% and 25% of the basin area respectively. Agriculture covers only a small proportion of the basin, with C3 (Fig. 2 g) and C4 (Fig. 2 h) agriculture making up a maximum basin area of 4 and 3% respectively. Note that the fractions add up to a little over 100% as they represent the maximum vegetated fraction over the course of a year.



Chapter 4 Figure 2: Present day (1981-2010) spatial distribution of the principal climate and land-use drivers used in ORCHILEAK, across the Congo Basin; a) mean annual temperature in °C, b) mean annual rainfall in mm yr^{-1}, c)-h) mean annual maximum vegetated fraction for PFTs 2,3, 10,11,12 and 13, i) river area. All at a resolution of 1° except for river area (0.5°).

4.2.2. Development of wetland forcing files

We created an MFF forcing file for the Congo basin, derived from the 232m resolution dataset of Gumbricht et al. (2017) (Fig. 3 a and b). We firstly merged all of the wetland categories of the Gumbricht dataset before aggregating them to a resolution of 0.5° (the resolution of the sub grid basins in ORCHILEAK), to represent the maximum extent of inundation in the basin. This results in a mean MFF of 10.3%, i.e. a maximum of 10.3% of the surface area of the Congo basin can be inundated with water. This is very similar to the mean MFF value of 10% produced with the Global Lakes and Wetlands Database, GLWD (Lehner, & Döll, P.,2004; Borges et al., 2015[b]). We also created an MFS forcing file from the same dataset (Fig. 3 c) and d), merging the 'swamps' and 'fens' wetland categories and again aggregating them to a 0.5° resolution.



Chapter 4 Figure 3: a) Wetland extent (from Gumbricht et al., 2017). b) The new maximal fraction of floodplain (MFF) forcing file derived from from a). c) Swamps (including ferns) classification within Congo Basin from Gumbricht et al (2017). d) the new maximal fraction of swamps (MFS) forcing file derived from c). Panels a) and b) are at the same resolution as the Gumbricht dataset (232m) while b) and d) are at a resolution of 0.5°. Note that 0.5° is the resolution of the sub unit basins in ORCHILEAK (Lauerwald et all., 2015), with each 1° grid containing four sub basins.

4.2.3 Calibration of hydrology

As the main driver of the export of C from the terrestrial to aquatic system, it is crucial that the model can represent present day hydrological dynamics, at the very least on the main stem of the Congo. As this study is primarily concerned with decadal- centennial timescales our priority was to ensure that it can accurately recreate observed mean annual discharge at the most downstream gauging station Brazzaville, however we also tested its ability to simulate observed seasonality, as well as flood dynamics. Moreover, no data is available with which to directly evaluate the simulation of DOC and CO_2 exports from the soil to the river network, and thus we tested the model's ability to recreate the spatial variation of observed riverine DOC concentrations (Borges at al., 2015[b]), which can be regarded as an integrator of the C transport at the terrestrial- aquatic interface.

We first ran the model for the present day (in this case defined as 1990-2005/2010 depending on which climate forcing data was applied) using four climate forcing datasets; namely Princeton GPCC (Sheffield et al., 2006), ISIMIP2b (Frieler et al., 2017), GSWP3 (Kim., 2017) and CRUNCEP (Viovy., 2018), to assess which configuration is best able to recreate observed discharge on the Congo River at Brazzaville (Fig. 1), the most downstream river gauging station. After deciding to proceed with ISIMIP2b (see proceeding paragraphs), we calibrated a number of hydrological model parameters, namely the constants which dictate the water residence time of the slow, fast and floodplain reservoirs. In order to improve the simulation of observed discharge at Brazzaville, we reduced the residence time of all three reservoirs (compared to those values used in the original ORCHILEAK calibration for the Amazon, Lauerwald at al., 2017).

In order to evaluate the simulated discharge against observations, we first calibrated the flood dynamics of ORCHILEAK in the Congo Basin for the present day by calculating a number of

statistical parameters (Lauerwald et al., 2017). For each configuration, we ran the model once before calculating the median water storage for each grid cell (see Lauerwald et al., 2017). This represents the bank-full discharge (Table 1) for each grid cell; any water in excess of this value will start to inundate the floodplains. After re-running each model configuration with the new median water storage values, we calculated the 95th percentile of water level heights over the simulation period for each grid cell (Table 1). This value represents water level over the river banks at which the maximum horizontal extent of inundation (MFF) is reached. Each model configuration model was then re-run for a final time and the outputs were then validated against discharge data at Brazzaville (Cochonneau et al., 2006, Fig. 1). This procedure was repeated iteratively with the ISIMIP2b climate forcing, modifying the water residence times of each reservoir in order to find the best performing parametrization, as discussed in the previous section.

Limited observed discharge data is available for the Congo basin, with the majority concentrated on the main stem of the Congo, at Brazzaville station. After identifying ISIMIP2b the best performing climate dataset, we used the data of Bouillon et al. (2014) to further validate discharge at Bangui (Fig. 1) on the River Ubangi and additionally compared the simulated seasonality of flooded area against the satellite derived dataset GIEMS (Prigent et al., 2007; Becker et al., 2018), within the Cuvette Centrale wetlands (Fig. 1).

4.2.3. Simulation set-up

A list of the forcing files used, along with data sources, is presented in Table 1. The derivation of the wetland and swamp (MFF & MFS) is described in section 2.2 while the calculation of "bankfull discharge" and "95th percentile of water table height over flood plain" (Table 1) is described in the preceding section (2.3). The "poor soils" forcing file (Fig. S3 a) is derived from the Harmonized World Soil Database (FAO/IIASA/ISRIC/ISS-CAS/JRC, 2009), which

prescribes reduced decomposition rates in low pH and nutrient soils such as Podzols and Arenosols.

4.2.3.1. Soil carbon spin up

In order for the soil C pools to reach approximately steady state, we spun-up the model for approximately 9,000 years, with fixed land-use representative of 1861, and looping over the first 30 years of the ISMSIP2b climate forcing data (1861-1890). During the first 2,000 years of spin-up, we ran the model with a higher atmospheric CO_2 concentration of 350 µatm and default soil carbon residence time (τ_{carbon}) values halved, which allowed us to reach a steady-state more quickly. Following this, we ran the model for a further 7,000 years reverting to the default τ_{carbon} values. At the end of this process, the soil C pools had reached approximately steady state; <0.02% change in each pool over the last century of the spin up. Prior to 1861, we assume that the soil C was in quasi steady state.

4.2.3.2. Transient simulations

After the spin-up, we ran a historical simulation from 1861 until the present day, 2005 in the case of the ISMSIP2b climate forcing data. We then ran a future simulation with the IPSL-CM5A-LR model outputs for RCP 6.0 (Frieler et al., 2017) until 2099, using the final year of the historical simulation as a restart file. In both of these simulations, climate, atmospheric CO_2 and land-use change were fully transient. As our aim is to investigate long-term trends, we calculated the 30-year running means of simulated C flux outputs in order to suppress interannual variation. RCP 6.0 is an emissions pathway that leads to a "stabilization of radiative forcing at 6.0 Watts per square meter (Wm^{-2}) in the year 2100 without exceeding that value in prior years" (Masui et al., 2011). It is characterised by intermediate energy intensity, substantial population growth, mid-high C emissions, increasing cropland area to 2100 and decreasing natural grassland area (van Vuuren et al., 2011). In the paper which describes the development of the future land use change scenarios under RCP 6.0 (Hurtt et al., 2011), it is shown that land

use change is highly sensitive to model set up, such as the choice of historical start date of, and whether or not shifting cultivation is included. Moreover, Africa is one the regions with the largest uncertainty range, and thus, there is considerable uncertainty associated with the effect of future land-use change (Hurtt et al., 2011). We chose RCP 6.0 as it represents a no mitigation (mid-high emissions) scenario and because it was the scenario applied in the recent paper of Lauerwald et al. (submitted) to examine the long-term LOAC fluxes in the Amazon basin. Therefore, we can directly compare our results for the Congo to those for the Amazon. Moreover, the ISIMIP2b data only follows two RCPs; RCP 2.6 (low emission) and RCP 6.0.

With the purpose of evaluating the effect of climate change, land-use change and rising atmospheric CO_2, we ran a series of control simulations. In each control simulation, one of these factors was fixed at its 1861 level (the first year of the simulation), or in the case of climate change looped over the years 1861-1890. The outputs of these simulations (also 30-year running means) were then subtracted from the outputs of the fully transient (original run) so that we could determine the contribution of each driver (Fig. 10, Table 1).

Chapter 4 Table 1: Forcing files used for simulations

Variable	**Spatial resolution**	**Temporal resolution**	**Data source**
Rainfall, Snowfall, Incoming shortwave and longwave radiation, Air Temperature, Relative humidity and Air pressure (close to surface), Wind speed (10 m above surface)	1°	1 day	ISIMIP2b, IPSL-CM5A-LR model outputs for RCP6.0 (Frieler et al., 2017)
Land cover (and change)	0.5°	annual	LUH-CMIP5
Atmospheric CO_2	1°	annual	ISIMIP2b, IPSL-CM5A-LR model outputs for RCP6.0 (Frieler et al., 2017)
Soil texture class	0.5°	annual	Reynolds et al. (1999)
Soil pH, bulk density	0.5°	annual	after HWSD v 1.1 (FAO/IIASA/ISRIC/ISS-CAS/JRC, 2009)
Poor soils	0.5°	annual	after HWSD v 1.1 (FAO/IIASA/ISRIC/ISS-CAS/JRC, 2009)
Stream flow directions	0.5°	annual	STN-30p (Vörösmarty et al., 2000)

Floodplains and swamps (MFF & MFS)	0.5°	annual	derived from Gumbricht et al. (2017)
River surface areas	0.5°	annual	Lauerwald et al. (2015)
Bankfull discharge	1°	annual	derived from pre-runs with ORCHILEAK (see text)
95th percentile of water table height over flood plain	1°	annual	derived from pre-runs with ORCHILEAK (see text)

4.2.4. Evaluation and analysis of simulated fluvial C fluxes

We first evaluated simulated DOC concentrations at several locations along the Congo mainstem, and on the Ubangi river against the data of Borges at al. (2015[b]). We also compared the various simulated components of the net C balance (e.g., NPP) of the Congo against values described in the literature. In addition, we assessed the relationship between the interannual variation in present day C fluxes of the Congo basin and variation in temperature and rainfall. This was done through linear regression using STATISTICA™. We found decadal trends in several of the fluxes and thus detrended the time series using the Detrend function within the "SpecsVerification" package in R (R Core Team 2013) before undertaking the statistical analysis.

4.2.5. Calculating the net carbon balance of the Congo Basin

We calculated the Net Ecosystem Production (NEP) by summing the terrestrial and aquatic C fluxes of the Congo basin (Eq. 1), while we also incorporated disturbance fluxes (Land-use change flux and harvest flux) to calculate Net Biome Production (NBP) (Eq. 2). Positive values of NEP and NBP equate to a net terrestrial C sink.

We define NEP as follows:

$$NEP = NPP + TF - SHR - FCO_2 - LE_{\text{Aquatic}} \quad (1)$$

Where *NPP* is terrestrial net primary production, *TF* is the throughfall flux of DOC, *SHR* is soil heterotrophic respiration (only the part evading from the soil surface); FCO_2 is CO_2 evasion from the water surface and $LE_{Aquatic}$ is the export flux of C to the coast. NBP is the

same as NEP but with the addition of the C lost (or gained) through land use change (*LUC*) and crop harvest (*Harvest*). *LUC* includes emissions from fires and the lateral export of woody biomass:

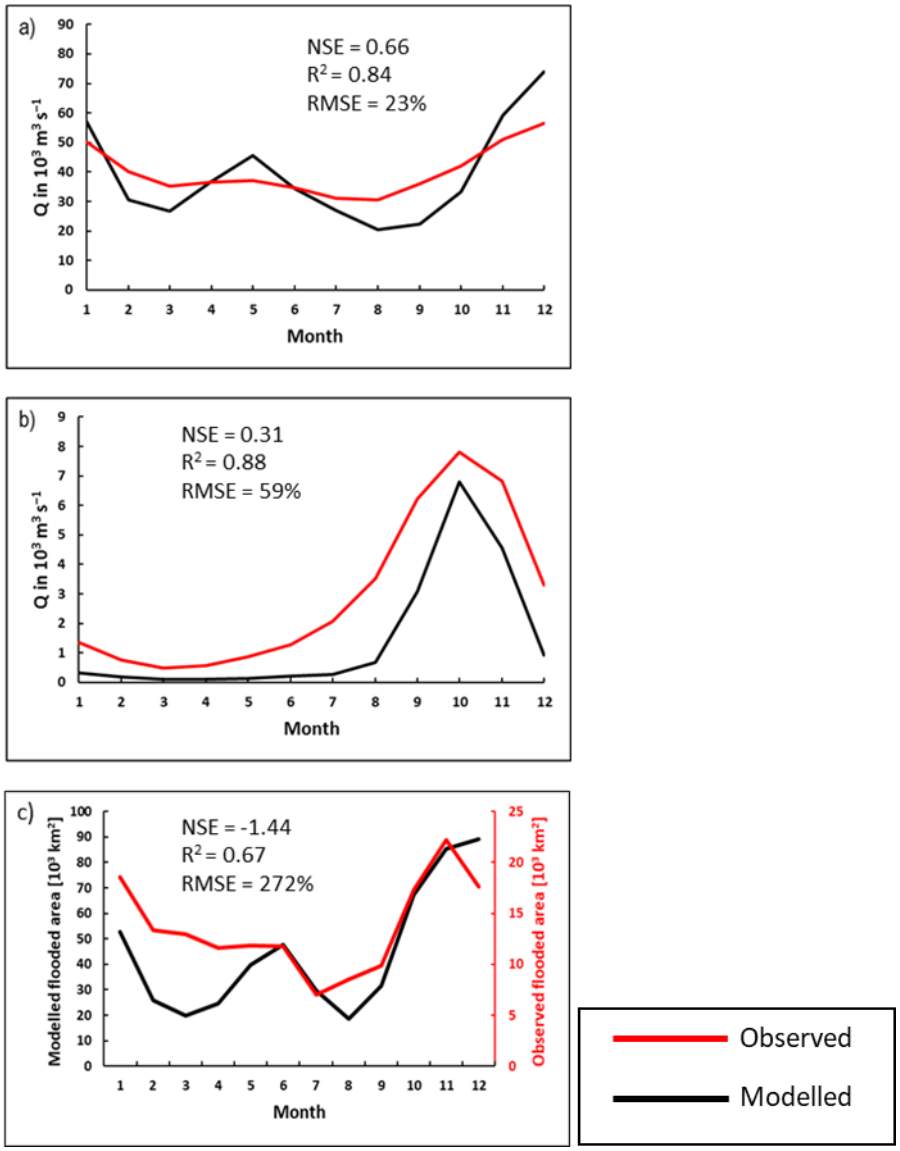
$$NBP = NEP - (LUC + Harvest) \quad (2)$$

4.3. Results

4.3.1. Representation of Hydrology

Without calibration, the majority of the different climate forcing model runs performed poorly, unable to accurately represent the seasonality nor the mean monthly discharge at Brazzaville (Table S1. Fig. 1). The best performing climate forcing dataset was ISIMIP2b followed by Princeton GPCC with root mean square errors (RMSE) of 29% and 40% and Nash Sutcliffe efficiencies (NSE) of 0.20 and -0.25, respectively. NSE is a statistical coefficient specifically used to test the predictive skill of hydrological models (Nash & Sutcliffe, 1970). As the clear best performing climate dataset, we chose to proceed with ISIMIP2b. After calibrating various model parameters, namely the residence times of the fast, slow and floodplain hydrological reservoirs (Lauerwald et al., 2017), we were able to further improve the simulation of discharge. The final model configuration is able to closely simulate the mean monthly discharge at Brazzaville (Fig. 4 a), Table 2) and is also able to represent the seasonality moderately well (Fig. 4 a), Table 2, RMSE =23%, R^2 =0.84 versus RMSE= 29% and R^2 =0.23 without calibration, Table S1). At Bangui on the Ubangi River (Fig. 1), the model is able to closely recreate observed seasonality (Fig. 4 b), RMSE =59%, R^2 =0.88) but substantially underestimates the mean monthly discharge, our value being only 50% of the observed. We produce reasonable NSE values of 0.66 and 0.31 for Brazzaville and Bangui respectively, indicating that the model is relatively accurate in its simulation of seasonality.

We also evaluated the simulated seasonal change in flooded area in the central (approx. 200,000 km^2, Fig. 1) part of the Cuvette Centrale wetlands against the GIEMS inundation dataset (1993-2007, maximum inundation minus minimum or permanent water bodies, Prigent et al., 2007; Becker et al., 2018). While our model is able to represent the seasonality in flooded area relatively well (R^2 =0.75 Fig. 4 c), it considerably overestimates the magnitude of flooded area relative to GIEMS (Fig. 4 c, Table 2). However, the dataset that we used to define the MFF and MFS forcing files (Gumbricht et al., 2017) is produced at a higher resolution than GIEMS and will capture smaller wetlands than the GIEMS dataset, and thus the greater flooded area is to be expected. GIEMS is also known to underestimate maximum inundation as it based on space-borne remote sensing data that is unable to capture the smallest inundated areas (Lauerwald et al., 2015; Hastie et al., accepted). Indeed, with the GIEMS data we produce an overall MFF for the Congo Basin of just 3%, less than one-third of that produced with the Gumbricht dataset (Gumbricht et al., 2017). As such, it is to be expected that there is a large RMSE (272%, Table 2) between simulated flooded area and GIEMS; more importantly, the seasonality of the two is highly correlated (R^2 = 0.67, Table 2). Overall the hydrological performance of the model against several published datasets is satisfactory as the main purpose of this study is to estimate the long-term changes of aquatic C fluxes. The most important result is that it can closely recreate the mean monthly/ annual discharge at Brazzaville (Table 2), the most downstream gauging station on the Congo (Fig. 1). As such, we consider the hydrological performance to be sufficiently good for our aims.



Chapter 4 Figure 4: Seasonality of simulated versus observed discharge at a) Brazzaville on the Congo (Cochonneau et al., 2006), b) Bangui on the Ubangi (Bouillon et al., 2014) 1990-2005 monthly mean and c) flooded area in the the central (approx. 200,000 km^2) area of the Cuvette Centrale wetlands versus GIEMS (1993-2007, Becker et al., 2018). The observed flooded area data represents the maximum minus minimum (permanent water bodies such as rivers) GIEMS inundation. See Figure S1 for locations

Chapter 4 Table 2: Performance statistics for modelled versus observed seasonality of discharge with calibrated ISIMIP climate forcing

Station	RSME	NSE	R^2	Simulated mean monthly discharge (m^3 s^{-1})	Observed mean monthly discharge (m^3 s^{-1})
Brazzaville	23%	0.66	0.84	38,944	40,080
Bangui	59%	0.31	0.88	1,448	2,923

Flooded area (Cuvette Centrale)	272%	-1.44	0.67	Simulated mean monthly flooded area (10^3 km^2)	Observed mean monthly flooded area (10^3 km^2)
				44	14

4.3.2. Carbon fluxes along the Congo basin for the present day

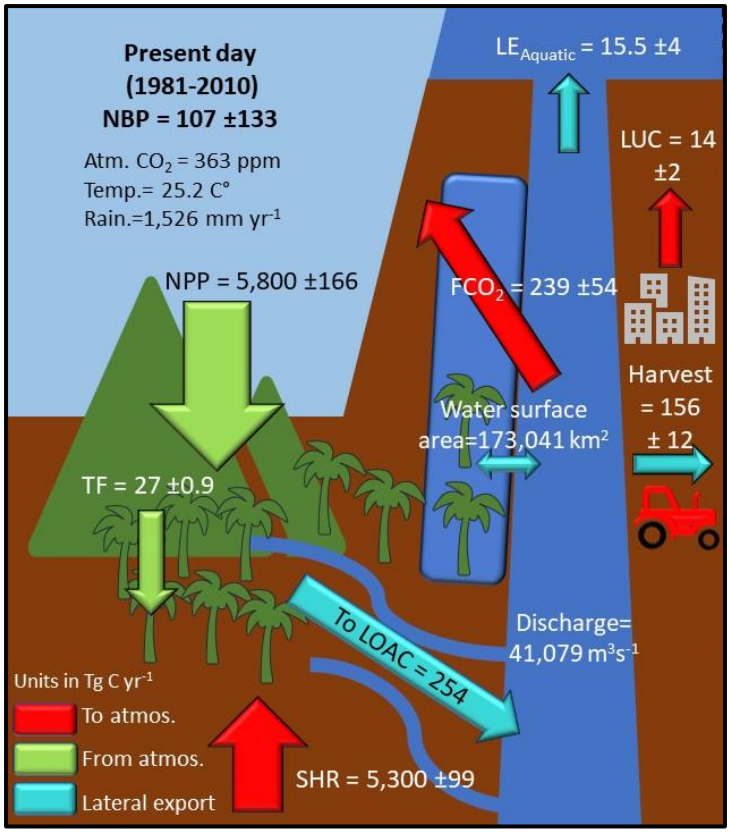
For the present day (1981-2010) we estimate a mean annual terrestrial net primary production (NPP) of 5,800 ±166 (standard deviation, SD) Tg C yr^{-1} (Fig. 5), corresponding to a mean areal C fixation rate of approximately 1,500 g C m-2 yr-1 (Fig. 6 a). We find a strong positive correlation between the interannual variation of NPP and rainfall (detrended R^2= 0.41, p<0.001, Table S2) and a moderate negative correlation between annual NPP and temperature (detrended R^2= 0.32, p<0.01, Table S2). We also see considerable spatial variation in NPP across the Congo Basin (Fig.6 a).

We simulate a mean soil heterotrophic respiration (SHR) of 5,300 ±99 Tg C yr^{-1} across the Congo basin (Fig. 5). Contrary to NPP, interannual variation in annual SHR is positively correlated with temperature (detrended R^2= 0.57, p<0.0001, Table S2) and inversely correlated with rainfall (detrended R^2= 0.10), though the latter relationship is not significant (p>0.05). We estimate a mean annual aquatic CO_2 evasion of rate of 1,363 ±83 g C m^{-2} yr^{-1}, amounting to a total of 235±54 Tg C yr^{-1} across the total water surfaces of the Congo basin (Fig. 5) and attribute 85% of this flux to flooded areas, meaning that only 32 Tg C yr^{-1} is evaded directly from the river surface. Interannual variation in aquatic CO_2 evasion (1981-2010) shows a strong positive correlation with rainfall (detrended R^2= 0.75, p<0.0001, Table S2) and a weak negative correlation with temperature (detrended R^2=0.09, not significant, p>0.05). Aquatic CO_2 evasion also exhibits substantial spatial variation (Fig.6, d), displaying a similar pattern to both DOC leaching (DOC_{inp}) (R^2= 0.81, p<0.0001, Fig.6, b) as well as CO_2 leaching (CO_{2inp})

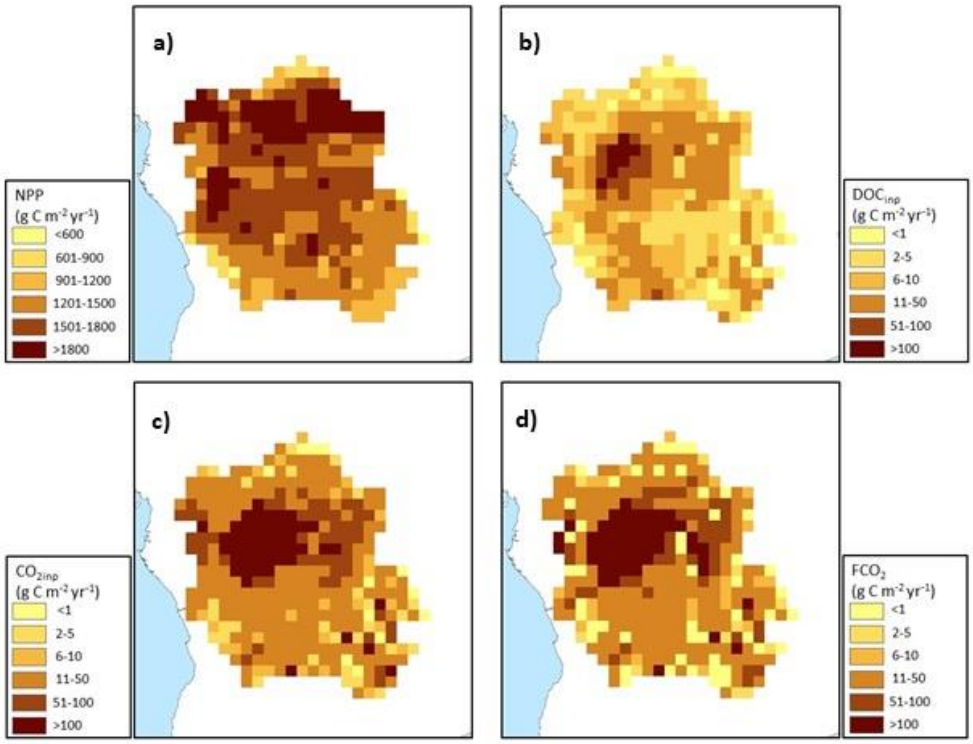
(R^2= 0.96, p<0.0001, Fig.6, c) into the aquatic system, but not terrestrial NPP (R^2= 0.01, p<0.05, Fig.6, a).

We simulate a mean annual C export to the coast of 15.5±4 Tg C yr^{-1} (Fig. 5). In Figure 7, we compare simulated DOC concentrations at six locations (Fig. 1) along the Congo River and Bangui tributary, against the observations of Borges at al. (2015[b]). The simulated DOC concentrations represent the average values across the particular sampling period at each site detailed in Borges et al. (2015[b]). We show that the model can recreate the spatial variation in DOC concentration within the Congo basin relatively closely with an R^2 of 0.82 and an RMSE of 19% (Fig. 7)

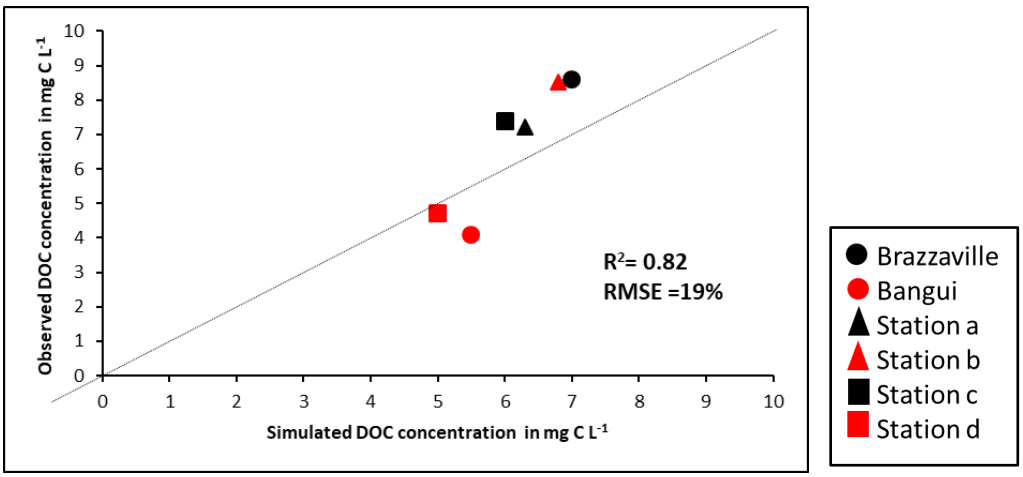
For the present day (1981-2010) we estimate a mean annual net ecosystem production (NEP) of 275 ±137 Tg C yr^{-1} and a net biome production (NBP) of 107 ±133 Tg C yr^{-1} (Fig. 5). Interannually, both NEP and NBP exhibit a strong inverse correlation with temperature (detrended NEP R^2=0.55, p<0.0001, detrended NBP R^2=0.54, p<0.0001) and weak positive relationship with rainfall (detrended NEP R^2=0.16, p<0.05, detrended NBP R^2=0.14, p<0.05). Furthermore, we simulate a present day (1981-2010) living biomass of 41 ±0.8 Pg C and a total soil C stock of 109 ±1.1 Pg C.



Chapter 4 Figure 5: Simulated annual C budget (NBP) for the Congo basin for the present day (1981-2010), where NPP is terrestrial net primary productivity, TF is throughfall, SHR is soil heterotrophic respiration, FCO_2 is aquatic CO_2 evasion, LOAC is C leakage to the land-ocean aquatic continuum (FCO_2 + $LE_{Aquatic}$), LUC is flux from Land-use change, and $LE_{Aquatic}$ is the export C flux to the coast. Range represents the standard deviation (SD).



Chapter 4 Figure 6:Present day (1981-2010) spatial distribution of a) terrestrial NPP, b) DOC leaching into the aquatic system, c) CO_2 leaching into the aquatic system and d) aquatic CO_2 evasion. All at a resolution of 1°



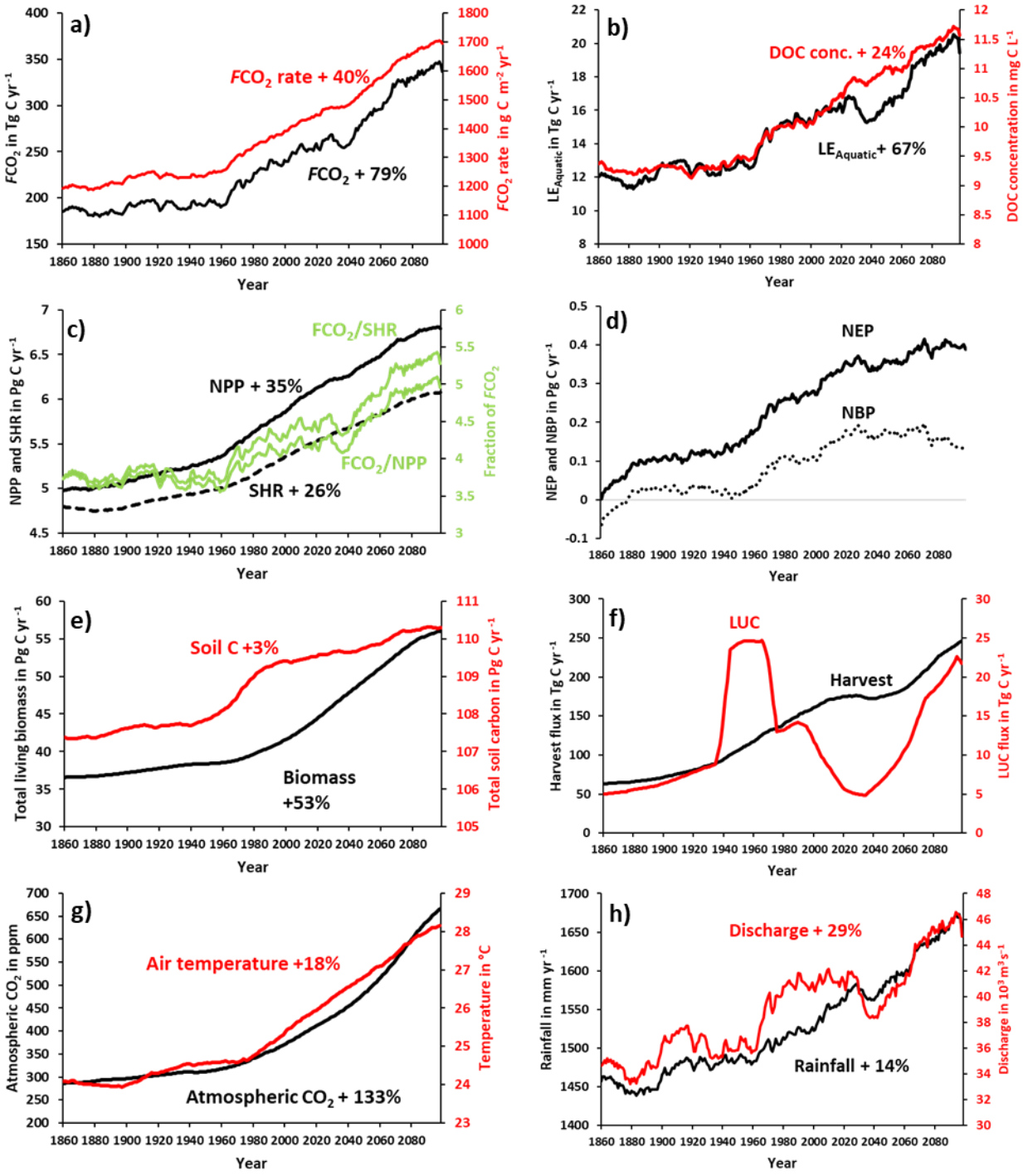
Chapter 4 Figure 7: Observed (Borges et al., 2015[a]) versus simulated DOC concentrations at several sites along the Congo and Bangui rivers. See Fig. S1 for locations. The simulated DOC concentrations represent the mean values across the particular sampling period at each site detailed in Borges et al. (2015[a]).

4.3.3. Long-term temporal trends in carbon fluxes

We find an increasing trend in aquatic CO_2 evasion (Fig. 8 a) throughout the simulation period, rising slowly at first until the 1960s when the rate of increase accelerates. In total CO_2 evasion rises by 79% from 186 Tg C yr^{-1} at the start of the simulation (1861-1890 mean) (Fig. 9) to 337 Tg C yr^{-1} (2070-2099 mean, Fig. 9), while the trend until the present day (1981-2010 mean) represents an increase of 28 % (to 239 Tg C yr^{-1}), though these trends are not uniform across the basin (Fig S1). The lateral flux of C to the coast ($LE_{Aquatic}$) follows a similar pattern (Fig. 8 b), rising by 67% in total, from 12 Tg C yr^{-1} (Fig. 9) to 15.5 Tg C yr^{-1} for the present day, and finally to 20 Tg C yr^{-1} (2070-2099 mean, Fig. 9). This is greater than the equivalent increase in DOC concentration (24%, Fig. 8 b) due to the concurrent rise in rainfall (by 14%, Fig 8. h) and in turn discharge (by 29%, Fig. 8 h). Interestingly, the proportion of NPP lost to the LOAC also increases from approximately 3% to 5%.

NPP and SHR also exhibit substantial increases of 35% and 26% respectively across the simulation period and similarly rise rapidly after 1960 (Fig. 8 c). NEP, NBP (Fig. 8 d) and

living biomass (Fig. 8 e) follow roughly the same trend as NPP but NEP and NBP slow down around 2030 and in the case of NBP, we actually simulate a decreasing trend over approximately the final 50 years. We also find that living biomass mass increases by a total of 53% from 1861 to 2099. Total soil C also increases over the simulation but only by 3% from 107 to 110 Pg C yr^{-1} (Fig. 8 e). The flux from land-use change (LUC) shows considerable decadal fluctuation increasing rapidly in the second half of the 20th century and decreasing in the mid-21st century before rising again towards the end of the simulation (Fig. 8 f). The harvest flux (Fig.8 f) rises throughout the simulation with the exception of a period during in the mid-21st century in which it stalls for several decades. This is reflected in the change in land-use from 1861- 2099 (Fig. S2 Table S3) during which the natural forest and grassland PFTs marginally decrease while both C3 and C4 agricultural grassland PFTs increase.



Chapter 4 Figure 8: Simulation results for various C fluxes and stocks from 1861-2099, using IPSL-CM5A-LR model outputs for RCP 6.0 (Frieler et al., 2017). All panels except for atmospheric CO_2, biomass and soil C correspond to 30-year running means of simulation outputs. This was done in order to suppress interannual variation, as we are interested in longer-term trends.

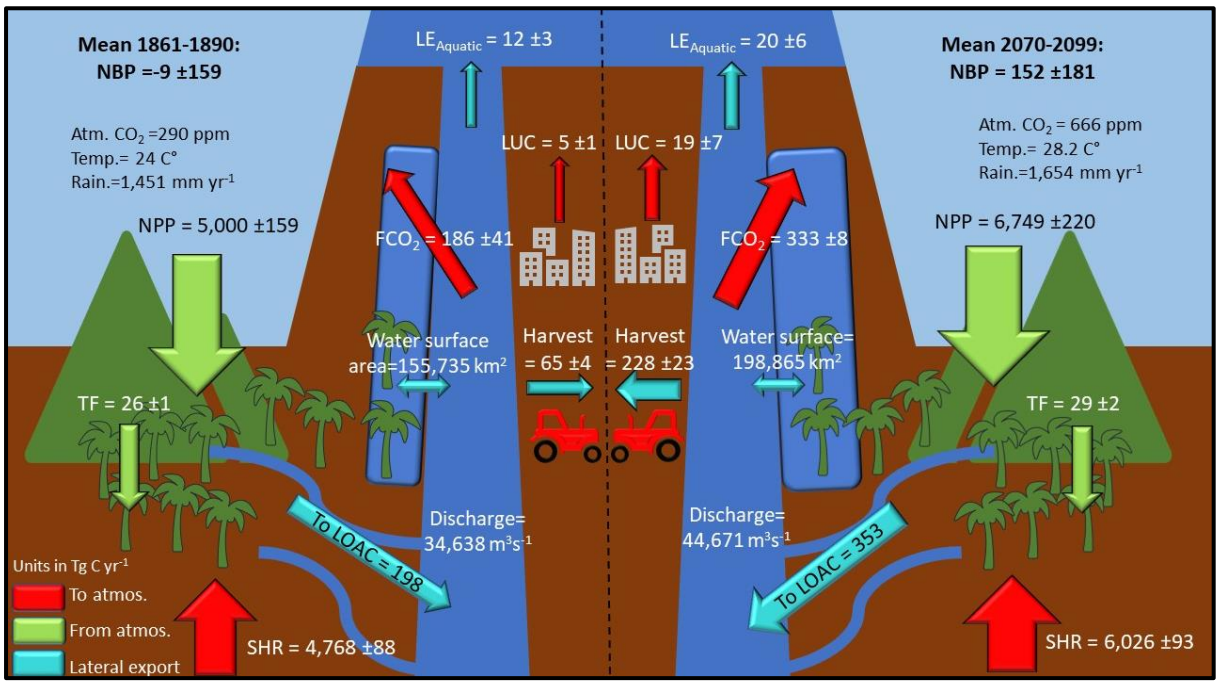
4.3.4. Drivers of simulated trends in carbon fluxes

The dramatic increase in the concentration of atmospheric CO_2 (Fig. 8 g) and subsequent fertilization effect on terrestrial NPP has the greatest overall impact on all of the fluxes across the simulation period (Fig. 10). It is responsible for the vast majority of the growth in NPP, SHR, aquatic CO_2 evasion and flux of C to the coast (Fig. 10 a, b, c & d). The effect of LUC on these four fluxes is more or less neutral, while the impact of climate change is more varied. The aquatic fluxes (Fig. 10 c, d) respond positively to an increase in the increase of both rainfall (and in turn discharge, Fig. 8 h) and temperature (Fig. 8 g) starting around 1970. From around 2020, the impact of climate change on the lateral flux of C to the coast (Fig 10 d) reverts to being effectively neutral, likely a response to a slowdown in the rise of rainfall and indeed a decrease in discharge (Fig 8 h), as well as perhaps the effect of temperature crossing a threshold. The response of the overall loss of terrestrial C to the LOAC (i.e. the ratio of LOAC/NPP, Fig. 10 e) is relatively similar to the response of the individual aquatic fluxes but crucially, climate change exerts a much greater impact, contributing substantially to an increase in the loss of terrestrial NPP to the LOAC in the 1960s, and again in the second half of the 21st century. These changes closely coincide with the pattern of rainfall and in particular with changes in discharge (Fig. 8 h).

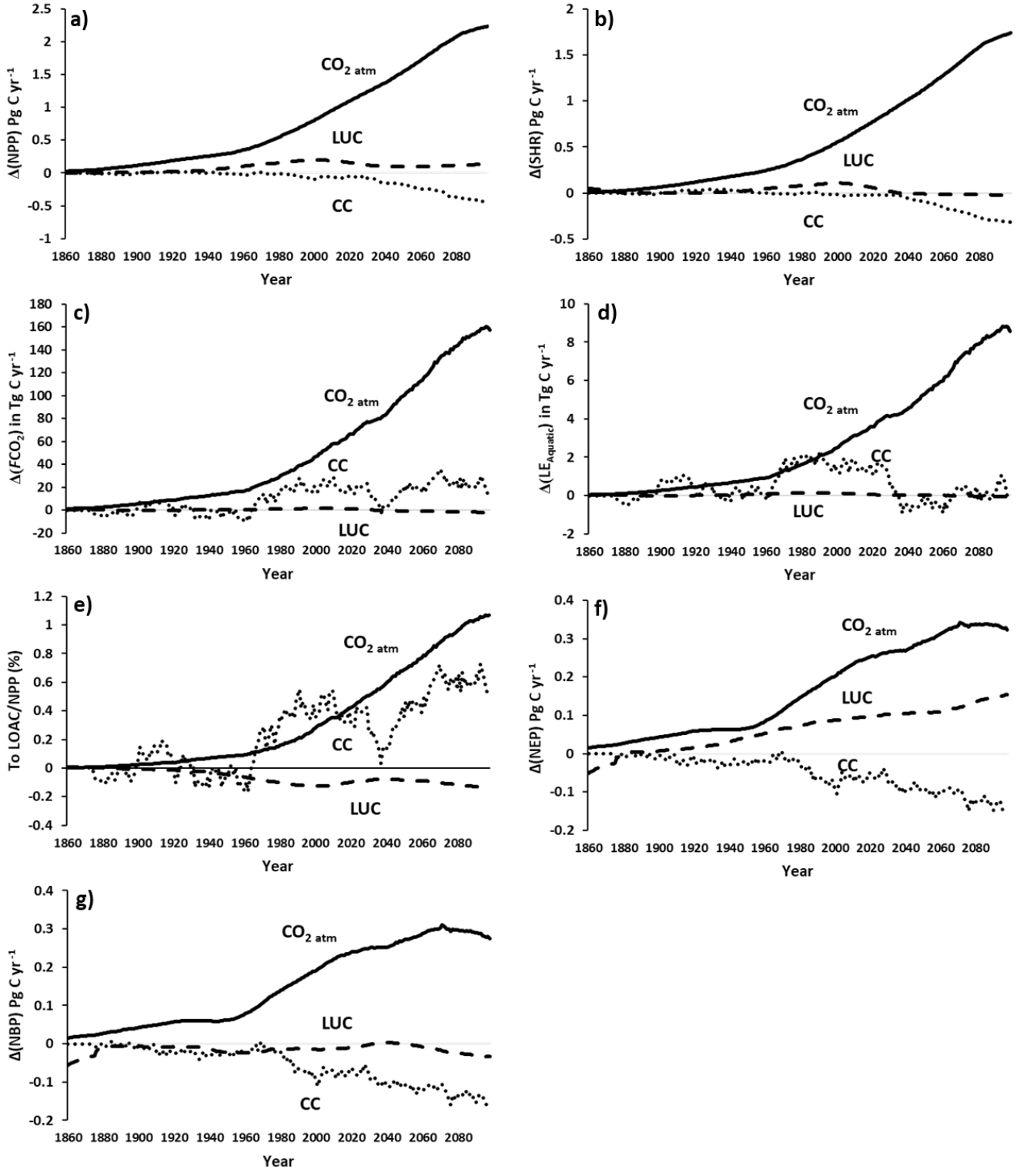
Overall temperature and rainfall increase by 18% and 14% respectively but in Fig. S2 one can see that this increase is non-uniform across the basin. Generally speaking, the greatest increase in temperature occurs in the south of the basin while it is the east that sees the largest rise in rainfall (Fig. S2). Land-use changes are similarly non-uniform (Fig. S2).

The response of NBP and in NEP (Fig.10 f, g) to the anthropogenic drivers is more complex. The simulated decrease in NBP towards the end of the run is influenced by a variety of factors; LUC and climate begin to have a negative effect on NBP (contributing to a decrease in NBP) at a similar time while the positive impact (contributing to an increase in NBP) of atmospheric

CO_2 begins to slow down and eventually level-off (Fig.10 g). LUC continues to have a positive effect on NEP (Fig.10 f) which prevents NEP from reducing in the same manner as NBP at the end of the simulation.



Chapter 4 Figure 9: Simulated annual C budget (NBP) for the Congo basin for the left, the Year 1861 and right, the Year 2099. Where NPP is terrestrial net primary productivity, TF is throughfall, SHR is soil heterotrophic respiration, FCO_2 is aquatic CO2 evasion, LOAC is C leakage to the land-ocean aquatic continuum (FCO_2 + LEAquatic), LUC is flux from Land-use change, and LEAquatic is the export C flux to the coast. Range represents the standard deviation (SD).



Chapter 4 Figure 10: Contribution of anthropogenic drivers; atmospheric CO_2 concentration ($CO_{2\ atm}$), climate change (CC) and land use change (LUC) to changes in the various carbon fluxes along the Congo Basin, under IPSL-CM5A-LR model outputs for RCP 6.0 (Frieler et al., 2017).

4.4. Discussion

Our estimate of present-day aquatic CO_2 evasion from the river surface of the Congo basin (32 Tg C yr^{-1}) is the same as that estimated by Raymond et al. (2015) (also 32 Tg C yr^{-1}), downscaled over the same basin area, but smaller than the 59.7 Tg C yr^{-1} calculated by Lauerwald et al. (2015) and far smaller than that of Borges et al. (2015[a]), 133-177 Tg C yr^{-1}. As previously discussed, we simulate the spatial variation in DOC concentrations measured by Borges et al. (2015[a,b], Fig. 7) relatively closely and moreover, our mean riverine gas exchange velocity *k* of 3.5 m d^{-1} is similar to the 2.9 m d^{-1} used by Borges et al. (2015[a]). It is therefore surprising that our estimate of riverine CO_2 evasion is so different, and likely to be related to methodological differences. While they based their upscaling on a relatively large database of observations compared to previous estimates (Raymond et al., 2013; Lauerwald et al., 2015) they still relied on some assumptions. They explain their approach in Borges et al. (2015[a]) as follows "The F data were aggregated to derive one value per tributary and per river mainstem, before averaging for a given river system. The global F values were computed as averages weighted by waterbody surface area for each river catchment derived from the percentage of river/stream effective surface area per catchment given in ref.3." F values refer to CO_2 evasion, while global F values refers to that across the entire African continent and ref. 3 refers to Raymond et al. (2013). We interpret this as meaning that while they accounted for the relative contribution of the surface area of each major African river basin by weighting, they did not do the same for the contribution of individual tributaries within the Congo Basin. Therefore, one explanation for their higher estimate could be the over-representation of CO_2 evasion smaller tributaries in the Congo. Another reason for the difference could be that the resolution of ORCHILEAK (1° for C fluxes) is not sufficient to fully capture the dynamics of the smallest streams of the Congo Basin which have been shown to have the highest DOC concentrations (Borges et al., 2015b).

Note also that in our simulations, the evasion flux from rivers only contributes 15% of total aquatic CO_2 evasion, and including the flux from wetlands, we produce a total of 235 Tg C yr^{-1}.

Our simulated export of C to the coast of 15.5 Tg C yr^{-1} is identical to the TOC+DIC export estimated by Borges et al. (2015[a]) of 15.5 Tg C yr^{-1}, which is perhaps unsurprising given that our validation against their data demonstrate that we simulate a similar spatial variation of DOC concentrations (Fig. 7 and Fig. 1 for locations). It is also relatively similar to the 19 Tg C yr^{-1} (DOC + DIC) calculated by Valentini et al. (2014). Valentini et al. (2014) used the largely empirical based Global Nutrient Export from WaterSheds (NEWS) model framework and they point out that Africa was underrepresented in the training data used to develop the regression relationships which underpin the model, and thus this could explain the small disagreement.

We simulate a mean present-day terrestrial NPP of approximately 1,500 g C m^{-2} yr^{-1} (Fig. 6), substantially larger than the MODIS derived value of around 1,000 g C m^{-2} yr^{-1} from Yin et al. (2017) across central Africa. Though, our stock of the present-day living biomass of 41.1 Pg C is relatively close to the total Congo vegetation biomass of 49.3 Pg C estimated by Verhegghen et al. (2012) based on the analysis of MERIS satellite data. Moreover, our simulated Congo Basin soil C stock of 109 ±1.1 Pg C is consistent with the approximately 120-130 Pg C across Africa between the latitudes 10°S to 10°N in the review of Williams et al. (2007), between which the Congo represents roughly 70% of the land area. (10°S to 10°N). Therefore, their estimate of soil C stocks across the Congo only would likely be marginally smaller than ours. It is also important to note that neither estimate of soil C stocks explicitly take into account the newly discovered peat store of 30 Pg C (Dargie et al., 2017) and therefore both are likely to represent conservative values. In addition, Williams et al. (2007) estimate the combined fluxes from conversion to agriculture and cultivation to be around 100 Tg C yr^{-1} in

tropical Africa (largely synonymous with the Cong Basin), which is relatively close to our present day estimate of harvesting + land-use change flux of 170 Tg C yr^{-1}.

There is sparse observed data available on the long-term trends of terrestrial C fluxes in the Congo. Yin et al. (2017) used MODIS data to estimate NPP between 2001 and 2013 across central Africa. They found that NPP increased on average by 10 g C m^{-2} per year, while we simulate an average annual increase of 4 g C m^{-2} yr^{-1} over the same period across the Congo Basin. The two values are not directly comparable as they do not cover precisely the same geographic area but it is encouraging that our simulations exhibit a similar trend to remote sensing data.

Our results of the historic trend in NEP (not including LUC and harvest fluxes) also generally concur with other modelling studies of tropical Africa (Fisher et al., 2013). Fisher et al. (2013) used nine different land surface models to show that the African tropical biome already represented a natural (i.e. no disturbance, but also neglecting LOAC fluxes) net uptake of around 50 Tg C yr^{-1} in 1901 and that this more than doubled by 2010. We find a similar trend though we simulate higher absolute NEP. Indeed, one of the models used in Fisher was ORCHIDEE and using this model alone, they calculate a virtually identical estimate of 277 Tg C yr^{-1} for the present day, though this estimate neglects the transfer of C along the LOAC and would therefore be significantly reduced with their inclusion. Our results also generally concur with estimates based on the upscaling of biomass observations (Lewis et al., 2009). Lewis et al., up-scaled forest plot measurements to calculate that tropical African forests represented a net uptake of approximately 300 Tg C yr^{-1} between 1968 and 2007 and this is consistent with our NEP estimate 275 Tg C yr^{-1} over the same period.

For the present day, we show that aquatic C fluxes, and in particular CO_2 evasion, are important components of the Congo Basin C balance, larger than for example the combined fluxes from

LUC and harvesting, with around 4% of terrestrial NPP being lost to the aquatic system each year. However, this value is only one third of that found in the Amazon, where around 12% of NPP is lost to the aquatic system each year (Hastie et al., accepted). There are a number of differences between the drivers in the two basins which could explain this. Mean annual rainfall is 44% greater in the Amazon, while mean annual discharge is 4 times higher. Moreover, 7.7% of the surface of the Amazon Basin is water compared to only 4.7% of the Congo.

We find that these fluxes have undergone considerable perturbation since 1861 to the present day, with CO_2 evasion and the export of C to the coast increasing by 28% and 29% respectively, largely because of rising atmospheric CO_2 concentrations. Moreover, under RCP 6.0 we predict that this perturbation will continue; over the entire simulation period (1861-2099), we estimate that aquatic CO_2 evasion and the export of C to the coast will increase by 79% and 67% respectively. Interestingly, this increase is considerably higher than 25% and 30% rise predicted for the Amazon basin (Lauerwald et al., in submission), over the same period and under the same scenario. This is largely due to the fact climate change is predicted to have a substantial negative impact on the aquatic C fluxes in the Amazon, something that we do not find for the Congo where rainfall is projected to substantially increase over the 21st century. In the Amazon, Lauerwald et al. (in submission) show that while there are decadal fluctuations in precipitation and discharge, total values across the basin remain unchanged in 2099 compared to 1861. However, changes in the spatial distribution of precipitation mean that the total water surface area actually decreases in the Amazon. Indeed, while we find an increase in the ratio of C exports to the LOAC/NPP from 3 to 5%, Lauerwald et al. (in submission) find a comparative decrease. The increase in the proportion of NPP lost to the aquatic system (Fig. 8, 9), as well as in the concentration of DOC at Brazzaville (by 23%), could also have important secondary effects, not least the potential for greater DOC concentrations to lower pH levels (Laudon & Buffam, 2008).

Conversely, our simulated increase in DOC export to the coast up to the present day is smaller than findings recently published for the Mississippi River using the Dynamic Land Ecosystem Model (DLEM, Ren at al., 2016). In addition, the Mississippi study identified LUC including land management practices (such as fertilization and irrigation), followed by change in atmospheric CO_2, as the biggest factors in the 40% increase in DOC export to the Gulf of Mexico (Ren et al., 2016). Another recent study (Tian et al., 2015), found an increase in DIC export from eastern North America to the Atlantic Ocean from 1901-2008 but no significant trend in DOC or POC. They demonstrated that climate change and increasing atmospheric CO_2 had a significant positive effect on long-term C export while land-use change had a substantial negative impact.

It is important to note that we can have greater confidence in the historic trend (until 2005), as the future changes are reliant on the skill of Earth System model predictions and of course on the accuracy of the RCP 6.0 scenario. There are for example, large uncertainties associated with the future CO_2 fertilization effect (Schimel et al., 2015) and the majority of land surface models, ORCHILEAK included in its current form, do not account for the effect of nutrient limitation on plant growth meaning that estimates of land C uptake may be too large (Goll et al., 2017). There are also considerable uncertainties associated with future climate projections in the Congo basin (Haensler et al., 2013). However, in most cases the future trends that we find are more or less continuations of the historic trends, which already represent substantial changes to the magnitude of many fluxes.

Moreover, we do not account methane fluxes from Congo wetlands, estimated at 1.6 to 3.2 Tg (CH_4) per year (Tathy et al., 1992), and instead assume that all C is evaded in the form of CO_2. Another limitation is the lack of accounting for bespoke peatland dynamics in the ORCHILEAK model. ORCHILEAK is able to represent the general reduction in C decomposition in water-logged soils and indeed Hastie et al. (accepted) demonstrated that

increasing the maximum floodplain extent in the Amazon Basin led to an increase in NEP despite fueling aquatic CO_2 evasion because of the effect of reducing soil heterotrophic respiration. Moreover, ORCHILEAK uses a "poor soils" forcing file (Fig. S3 a) based on the Harmonized World Soil Database (FAO/IIASA/ISRIC/ISS-CAS/JRC, 2009), which prescribes reduced decomposition rates in low pH and nutrient soils such as Podzols and Arenosols. The effect of the "poor soils" forcing can clearly be seen in the spatial distribution of the soil C stock in Fig. S3, b where the highest C storage coincides with the highest proportion of poor soils. Interestingly, this does not include the Cuvette Centrale wetlands (Fig. 1), an area which was recently identified as containing the world's largest intact tropical peatland and a stock of around 30 Pg C (Dargie at al., 2017). A relatively simple potential improvement that could be made to ORCHILEAK would be the development of a new tailored "poor soils" forcing file for the Congo Basin, perhaps informed by the Soil Grids database (Hengl et al., 2014), to better represent the Cuvette Centrale. This could in turn, be validated and/or calibrated against the observations of Dargie et al. (2017). A more long-term aim could be the integration/ coupling of the ORCHIDEE-PEAT module with ORCHILEAK. ORCHIDEE- PEAT (Qiu et al., in review) represents peat as an independent sub-grid hydrological soil unit in which peatland soils are characterized by peat-specific hydrological properties and multi-layered transport of C and water. Thus far, it has only been applied to northern peatlands, and calibrating it to tropical peatlands, along with integrating it within ORCHILEAK would require considerable further model development, but would certainly be a valuable longer-term aspiration. This could also be applied across the tropical region and would allow us to comprehensively explore the implications of climate change and land-use change for tropical peatlands.

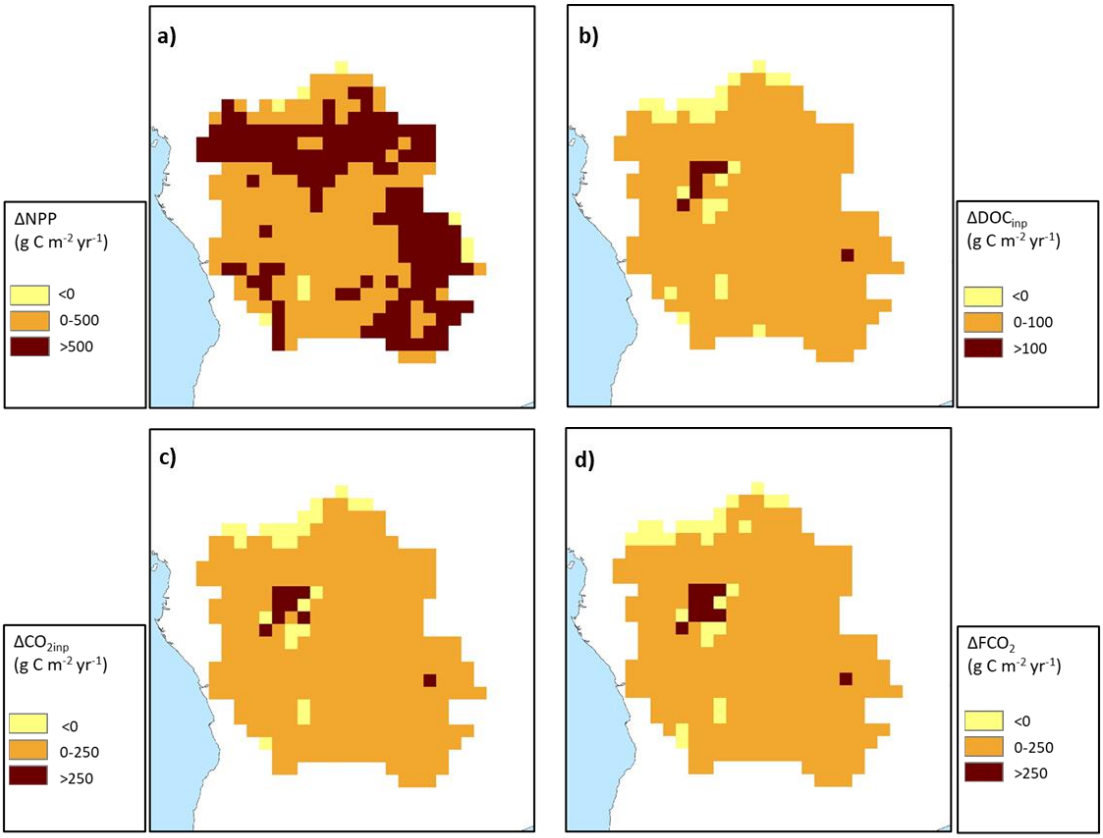
4.5. Supporting Information

Chapter 4 Table S 1: Performance statistics for modelled versus observed seasonality of discharge on the Congo at Brazzaville

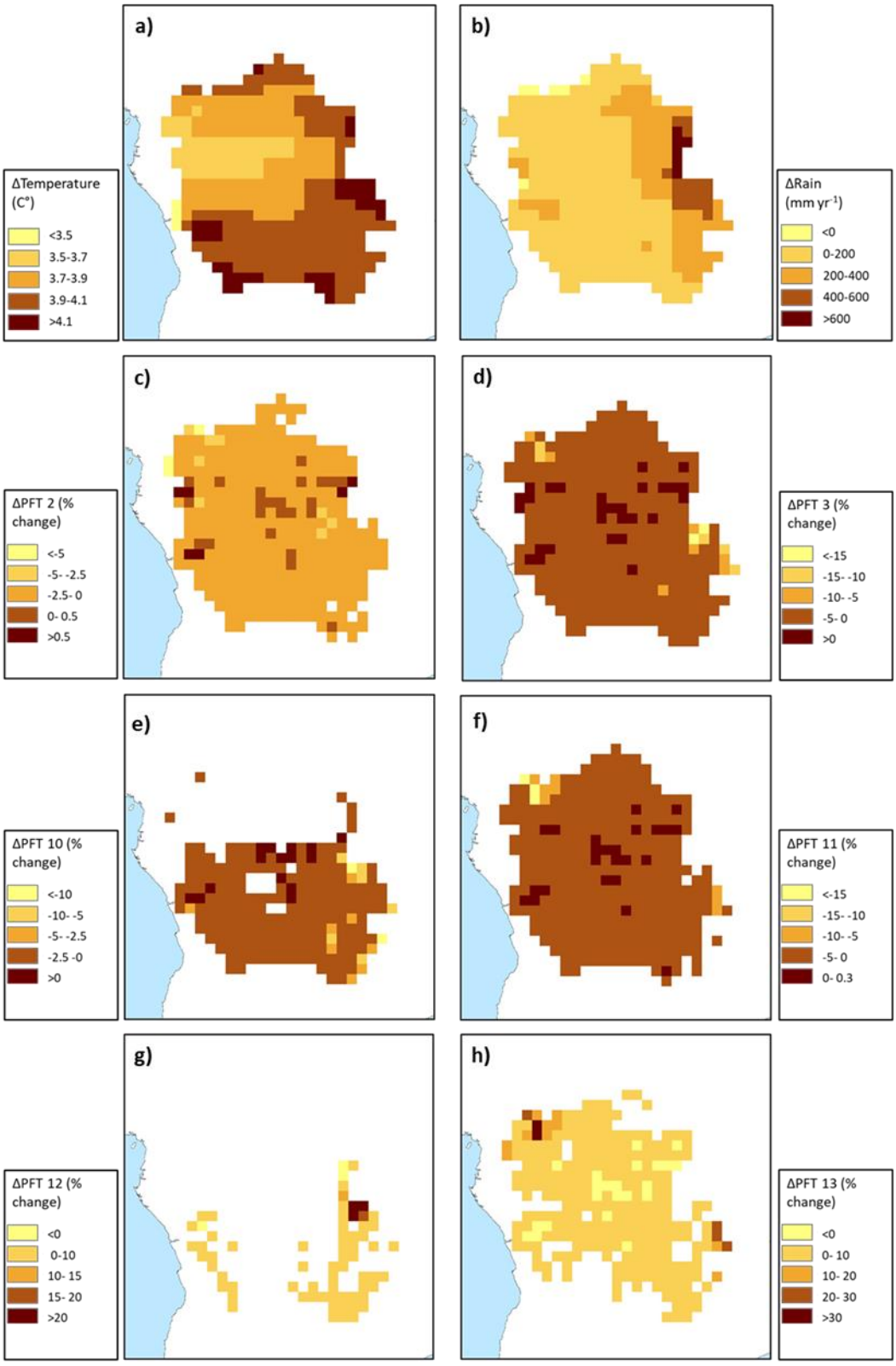
Climate forcing	RSME	NSE	R^2	Mean monthly discharge ($m^3 s^{-1}$)
ISIMIP	29%	0.20	0.23	38,944
Princeton GPCC	40%	-0.25	0.20	49,784
GSWP3	46%	-4.13	0.04	24,880
CRUNCEP	65%	-15.94	0.01	16,394
Observed (HYBAM)				40,080

Chapter 4 Table S 2: Pearson correlation coefficient (r) between detrended carbon fluxes and detrended climate variables (NCC climate data)

	SHR	Aquatic CO_2 evasion	Lateral C	NEP	Rain	Temp.	MEI
NPP	-0.48	0.68	0.72	0.90	0.64	-0.57	-0.09
SHR		-0.41	-0.48	-0.71	-0.32	0.76	0.04
Aquatic CO_2 evasion			0.92	0.41	0.87	-0.30	-0.21
Lateral C				0.52	0.81	-0.38	-0.15
NEP					0.40	-0.74	-0.01
Rain						-0.31	-0.26
Temp.							0.03

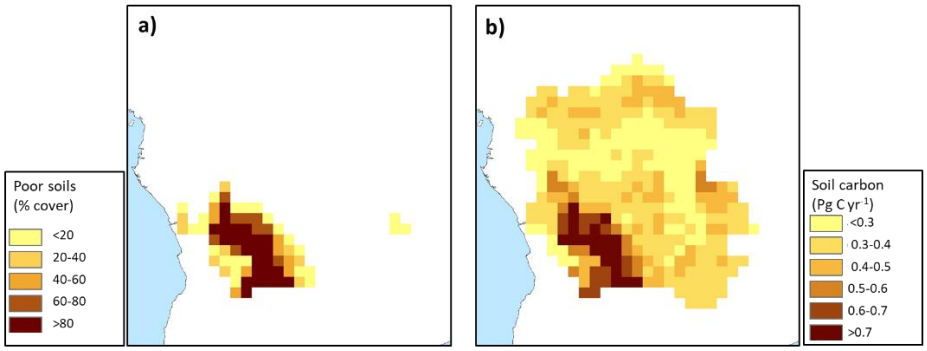


Chapter 4 Figure S 1:Change (Δ, 2099 minus 1861) in the spatial distribution of a) terrestrial NPP, b) DOC leaching into the aquatic system, c) CO_2 leaching into the aquatic system and d) aquatic CO_2 evasion. All at a resolution of 1°



Chapter 4 Figure S 2: Change (Δ, 2099 minus 1861) in the spatial distribution of the principal climate and land-use drivers across the Congo Basin; a) mean annual temperature in °C, b) mean annual rainfall in mm yr^{-1}, c)-h) mean annual maximum vegetated fraction for PFTs 2,3, 10,11,12 and 13. All at a resolution of 1°.

Chapter 4 Table S 3: Past (1861-1890), present-day (1981-2010) and future (2070-2099) mean values for important climate and land-use drivers across the Congo Basin								
Period	Temp.	Rain.	PFT2	PFT3	PFT10	PFT11	PFT12	PFT13
1861-1890	24.0	1451	0.263	0.375	0.154	0.254	0.015	0.014
1981-2010	25.2	1526	0.255	0.359	0.154	0.255	0.038	0.030
2070-2099	28.2	1654	0.258	0.362	0.147	0.245	0.039	0.037



Chapter 4 Figure S 3: a) The "poor soils" forcing file, which prescribes the spatial distribution of low pH and low nutrient level soils such as Podzols, Arenosols, or soils located in black-water swamps (FAO/IIASA/ISRIC/ISS-CAS/JRC, 2009), where decomposition rates are reduced. b) spatial distribution of simulated total carbon stored in soils for the present day (1981-2020).

5. Conclusions and outlook

5.1. Major conclusions, outputs and implications

Previous research has shown a close relationship between the terrestrial and aquatic carbon (C) cycle, namely that part of the C fixed via terrestrial NPP that is exported to the aquatic system, driven largely by the hydrological cycle (Cole et al., 2007; Battin et al., 2009; Tranvik et al., 2009; Regnier et al., 2013). In turn, it has been demonstrated that the land-ocean aquatic continuum (LOAC) is an active component of the global C cycle, in which C can not only be transported laterally as dissolved organic carbon (DOC), particulate organic carbon (POC) and dissolved CO_2 but also mineralized and evaded back to the atmosphere as CO_2, or buried in lakes and floodplains (Cole et al., 2007; Battin et al., 2009; Tranvik et al., 2009; Regnier et al., 2013). A number of hotspot areas of aquatic CO_2 evasion and TOC export to the coast have previously been identified (Meybeck et al., 2006; Raymond et al., 2013; Lauerwald et al., 2015) but there are considerable gaps in our knowledge, particularly associated with understanding and accounting for the temporal and spatial variation of aquatic C fluxes at regional to global scales, which we know from local scale studies, to be substantial. The limits of our current understanding are reflected in the large uncertainty bands associated with global estimates of these fluxes (Raymond et al., 2013; Lauerwald et al., 2015; Mendonça et al., 2017) and in the fact that the terrestrial C cycle remains the sink/source with the largest uncertainty in the most recent Global Carbon Budget (Le Quéré at al., 2018[b]).

In this thesis, three particularly important regional hotspots of LOAC activity were identified, where significant gaps in our understanding remain; the boreal region, the Amazon Basin and the Congo Basin. After formulating research questions for each, both empirical and process-based modelling approaches were applied to explore the spatial and temporal dynamics of aquatic C cycling and connectivity with the terrestrial C cycle. In doing so, this thesis provides

new insights into the interaction between the terrestrial and aquatic C cycles and their perturbation across interannual to centennial timescales.

For the boreal region (published in Hastie et al., 2018), an empirical model at the 0.5° grid scale was developed which explains 56% of the variation in the partial pressure of CO_2 (pCO_2) of boreal lakes using just three drivers; terrestrial net primary productivity (NPP, positive relationship), lake area (negative relationship) and precipitation (negative relationship), confirming the close connection between the terrestrial and aquatic C cycle in the boreal region. Not only is terrestrial NPP an important driver of aquatic pCO_2 but the smaller the lake area, the greater the pCO_2 because of the higher circumference to volume ratio and the thus increased impact of allochthonous C inputs and net-heterotrophy. Using this approach, the first high resolution maps of boreal lake pCO_2 and CO_2 evasion were created, providing a new estimate for total evasion from boreal lakes of 189 (74–347) Tg C yr^{-1}, which is more than double the previous best estimate (Raymond et al., 2013). Perhaps most importantly, our estimate is better constrained than the previous global estimate of Raymond et al. (2013) and gas exchanged velocity *k* is identified as the greatest source of uncertainty in the new estimate of boreal lake CO_2 evasion. Our new estimate could therefore be integrated into the global C budget of the LOAC and in turn the overall land C budget.

The boreal region is predicted to undergo substantial climate change over the 21st century and to be highly sensitive to this change (Gauthier et al., 2015; Intergovernmental Panel on Climate Change (IPCC)., 2013; Koven, 2013; Price et al., 2011). As such, the regression model was also used along with future projections of terrestrial NPP and precipitation, to predict future lake CO_2 evasion under Representative Concentration Pathway (RCP) 2.6 and RCP 8.5, and it was found that even under the most conservative pathway (RCP 2.6) CO_2 evasion from boreal lakes may increase 38% by 2100. This increase is more or less proportionate to the growth in

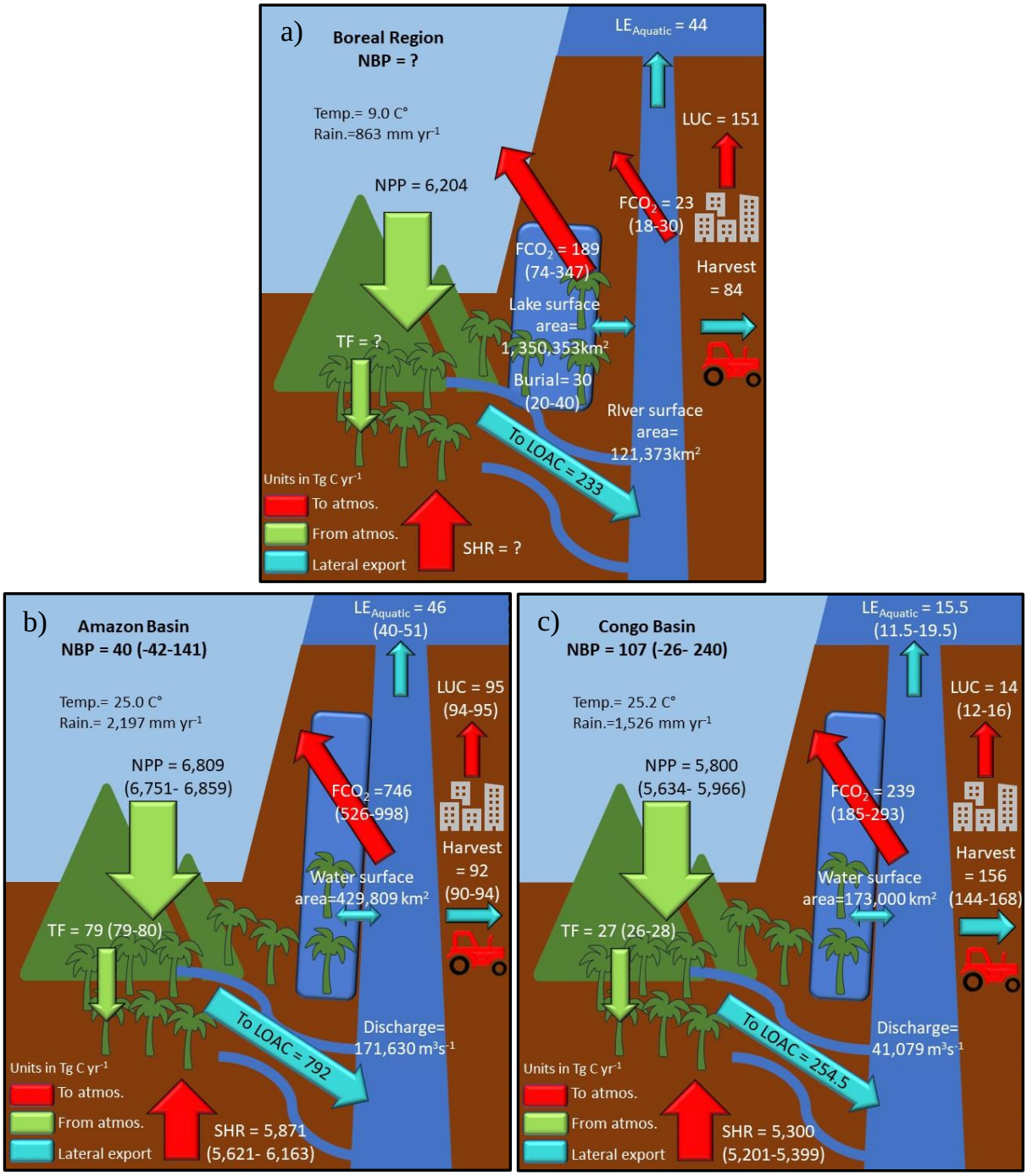
terrestrial NPP and thus the proportion of NPP lost to the aquatic system remains relatively constant.

For the Amazon Basin (accepted for publication in *Global Change Biology*), the ORCHILEAK land surface model (Lauerwald et al., 2017) driven by a newly developed wetland forcing file, was applied to explore the interannual variation of the aquatic C fluxes and their impact on the overall net ecosystem productivity of the Amazon. Based on this new wetland forcing and two climate forcing datasets, it was demonstrated that an average of 11.6% of NPP is lost to the river-floodplain system each year. It was also shown that this flux is highly variable at the interannual timescale; greatest during wet years and lowest during droughts. However, at the same time overall net ecosystem productivity (NEP) and C sequestration is highest during wet years, partly due to reduced decomposition rates in water-logged floodplain soils. It is years with the lowest discharge and floodplain inundation, often associated with El Nino events, that have the lowest NEP and the highest total (terrestrial plus aquatic) CO_2 emissions back to atmosphere. Furthermore, it was found that aquatic C fluxes display greater variation than terrestrial C fluxes, and that this variation significantly dampens the interannual variability in NEP of the Amazon basin. Previous research had already shown substantial decreases in terrestrial net primary productivity (NPP), and in turn C uptake from the atmosphere at interannual timescales as a result of climatic extremes and the ENSO cycle, the 2005 and 2010 droughts for example (Zhao & Running, 2010; Potter et al, 2011; Doughty et al., 2015 and Feldpausch et al., 2016). However, these studies did account for LOAC fluxes and the results of this thesis suggest that the interannual variation in aquatic C fluxes may act to moderate, at least partially, for the concurrent variation in the terrestrial C cycle. If this dynamic between the terrestrial and aquatic C cycle holds for the entire tropical region, it could have even greater implications for the land C cycle and its interannual variation.

Finally, ORCHILEAK was applied to the Congo Basin to investigate the evolution of the integrated aquatic and terrestrial C fluxes from 1861 to the present day, and in turn to 2099 under RCP 6.0. It was shown that terrestrial and aquatic fluxes increase substantially over time, both over the historical period and into the future, and that these increases are largely driven by atmospheric CO_2. Interestingly the proportion of terrestrial NPP lost to the LOAC also rises from 3% in 1861 to 5% in 2099 and this trend is driven not only by atmospheric CO_2 but also by climate change, namely an increase in precipitation, which in turn drives an increase in discharge. This is in contrast to the boreal region where we predict that the proportion of NPP exported to the LOAC will remain relatively constant, and to the Amazon, where Lauerwald et al. (in submission), actually show a decrease in the proportion of terrestrial NPP exported to the aquatic system over time due to differences in projected climate change. While there are decadal fluctuations in precipitation and discharge in the Amazon, total values across the basin remain basically unchanged in 2099 compared to 1861 (Lauerwald et al., in submission).

In figure 1, the present-day C budgets for the three different regions are shown and there is a number of interesting comparisons to be made. Firstly, the percentage of NPP transferred to the LOAC is variable. It was estimated that 3.8%, 11.6% and 4.4% of terrestrial NPP is transported to the aquatic system each year for the boreal region, the Amazon basin and the Congo basin, respectively. The values for the Congo and boreal regions are comparable to the global mean of 4.7% estimated by Regnier et al. (2013), though a recent review of Drake et al. (2018) postulated that the global mean could be as high as 9%, much closer to our estimate for the Amazon Basin. Figure 2 displays the spatial variation of CO_2 evasion rate (g C yr^{-1} m^{-2} grid surface area) for the three different regions at a resolution of 0.5°. The boreal region only shows CO_2 evasion from lakes, though as outlined in chapter 2 (and indeed figure 1 of this section) CO_2 evasion from lakes dominates the boreal LOAC budget. The Amazon and Congo regions show CO_2 evasion from rivers and wetlands. The Amazon Basin exhibits by far the highest

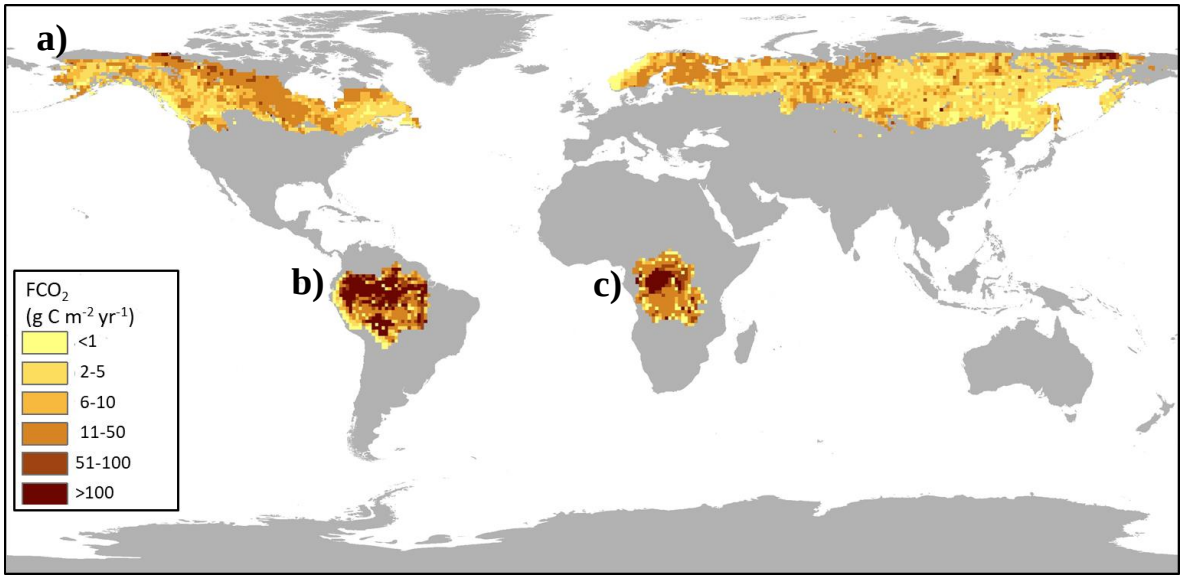
areal CO_2 fluxes, followed by the Congo Basin, though the latter is heavily dominated by one region; the Cuvette Centrale wetlands.



Chapter 5 Figure 1: Integrated mean annual terrestrial and aquatic carbon budgets for the present day for a) the boreal region, b) the Amazon Basin and Congo Basin, where NBP is net biome production, NPP is terrestrial net primary productivity, TF is throughfall, SHR is soil heterotrophic respiration, FCO_2 is aquatic CO_2 evasion, LOAC is carbon leakage to the land-ocean aquatic continuum (FCO_2 + $LE_{Aquatic}$), and $LE_{Aquatic}$ is the export carbon flux to the coast. See chapter 2 for full details of the additional references used to derive the boreal region budget.

Perhaps the most surprising difference between the regions is the dramatically higher rate of NPP export in the Amazon compared to the Congo, but there are a number of differences in the drivers which can explain this. Mean annual rainfall is 44% greater in the Amazon, while mean annual discharge is 4 times higher. Moreover, 7.7% of the surface of the Amazon Basin is water compared to only 4.7% in the Congo. As a result, the Amazon also has higher simulated mean pCO_2 values than the Congo, in line with observations from field studies (Borges et al., 2015[b]). As previously discussed, our results suggest that the difference in the connectivity of the terrestrial and aquatic environment between the two tropical regions has reduced slightly and will continue to do so over time, as the percentage of NPP lost to the aquatic system is increasing in the Congo basin. In the boreal region, 6.9% of the surface is covered by water, though this is heavily dominated by lakes with substantially lower pCO_2 values, as well as far lower gas exchange velocities due to lower turbulence than rivers. Ultimately, lakes have considerably longer residence times than rivers, which accounts for much of these differences.

Another interesting difference is the ratio of CO_2 evasion to the C export flux to the ocean (combined flux of DOC and DIC). For the boreal region CO_2 evasion is only 5 times the flux to the coast, while the ratio for the Amazon and Congo is similar, 16 times, and 15 times respectively. This concurs with previous studies showing generally higher DOC concentrations in boreal (Campeau and Giorgio, 2014; Hutchins et al., 2019) rivers compared to the Amazon (Rasera et al., 2013) and Congo (Borges et al., 2015[a]). For the global scale, Regnier et al. (2013) put the ratio of CO_2 evasion to C export to the coast at approximately 2:1, while Drake at al. (2018), estimate that it could be as high as 5:1. However, it is important to note that neither of these studies account for C evaded directly from inundated floodplains or wetlands as CO_2 (Regnier et al,. 2013 did include CH_4 evasion from wetlands), which accounts for the majority of total CO_2 evasion in the Amazon and Congo basins in this thesis.

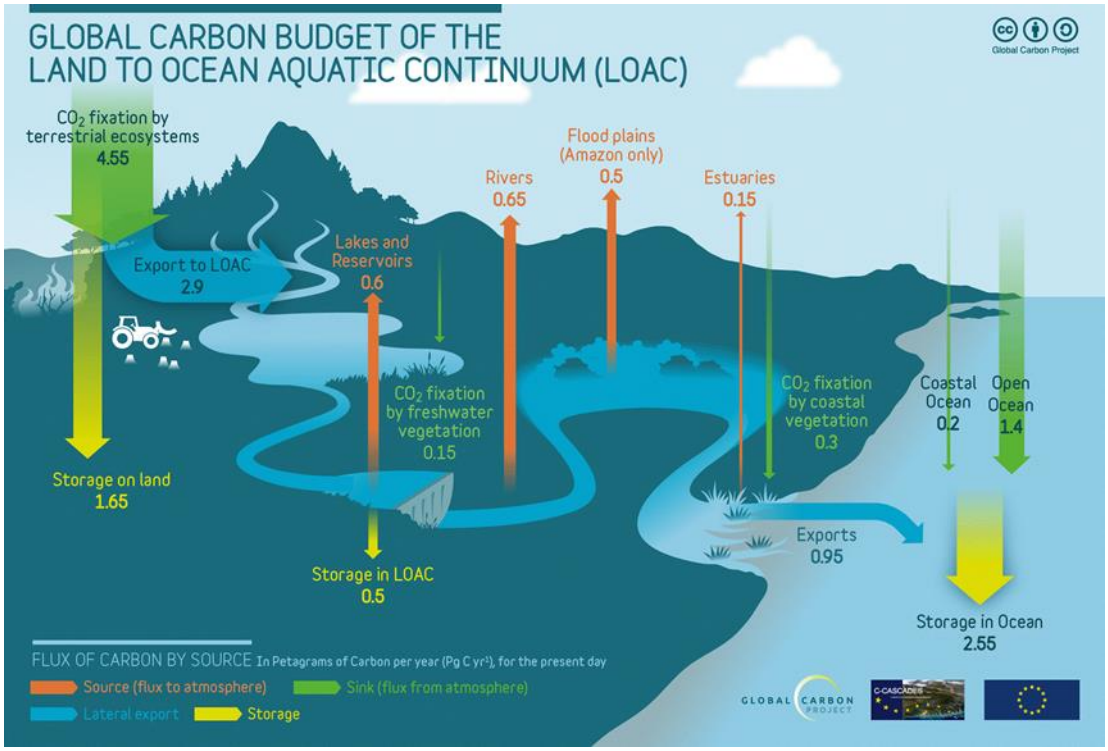


Chapter 5 Figure 2: Estimated present-day spatial variation (0.5°) in CO_2 evasion from a)-boreal lakes, b) the rivers and wetlands of the Amazon Basin and c) the rivers and wetlands of the Congo Basin.

5.1.1. Integrating the LOAC into the Global Carbon Atlas

Arguably the most novel aspect of our results lies in quantifying the spatial variation in aquatic C fluxes both between biomes and within biomes (Fig. 2), as well as showing that these fluxes are highly temporally variable, changing across interannual and centennial timescales with climate and atmospheric CO_2, and perhaps most importantly that the proportion of NPP lost to the LOAC is in turn highly variable. Though the results of this thesis have also been directly applied to produce an LOAC C budget for the Global Carbon Atlas (Fig. 3). The Global Carbon Atlas is an outreach initiative of the Global Carbon Project, namely an online *"platform to explore and visualize the most up-to-date data on carbon fluxes resulting from human activities and natural processes"*. Figure 3 shows the global C budget for the LOAC after incorporating the latest research, including the results from the first two chapters of this thesis. We estimate that a total of 2.9 Pg C yr^{-1} is exported from the terrestrial environment to the LOAC, equivalent to the entire land sink estimated in the latest Global Carbon Budget (Le Quéré et al., 2018[b]) and greater than the ocean sink. After subtracting export to the LOAC, we calculate that only

1.7 Pg C yr^{-1} is stored on land, with a further 0.5 Pg C yr^{-1} being stored in the LOAC producing a total storage of 2.2 Pg C yr^{-1} (Fig. 3). Note that this budget does not incorporate our results for the Congo. These would have a negligible impact on the global CO_2 evasion estimate for rivers but would increase evasion from floodplains by approximately 0.2 Pg C yr^{-1}. However, it can be debated whether wetlands should be considered as terrestrial or aquatic. Moreover, some would even argue that at the global scale the land C cycle can effectively be considered as nearly closed, in other words it is unimportant whether CO_2 is emitted from the terrestrial environment or whether it is first exported to the LOAC before being emitted as CO_2 evasion. This remains an open question that will only be tackled by concurrently running process-based models such as ORCHILEAK with and without the incorporation of LOAC fluxes, which a new publication should help to answer (Lauerwald., in submission). Nevertheless, it is undeniably important from a perspective of understanding the land C cycle, in turn how it has been historically perturbed, and crucially how climate change, land-use change and hydrological management will affect these C fluxes in the future. Ultimately it is our understanding of these processes which will help to inform governmental decision making. Moreover, we estimate that almost 1 Pg C yr^{-1} is transported from land to the coast each year (Fig. 3).



Chapter 5 Figure 3: The Global Carbon Budget for the Land to Ocean Aquatic Continuum for the present day (C-CASCADES, 2019). Values are based on the following literature sources: Regnier et al., 2013; Bauer et al., 2013; Laruelle et al., 2013; Laruelle et al 2014 ; Lauerwald et al., 2015; Laruelle et al 2017 ; Maavara et al., 2017; Hastie et al., 2018; Hastie et al., accepted; Roobaert et al., in preparation.

5.2. The advantages and limitations of empirical based approaches versus ORCHILEAK

In this thesis two very different methodologies were used; an empirical approach was applied to boreal lakes in chapter 2 while the process based ORCHILEAK model was used in the two tropical basins, and both methodologies have their strengths and weaknesses. The empirical approach has the advantage of being driven directly by all of the available "observations" of pCO_2 and allowed a relatively objective assessment of the uncertainty using Monte Carlo simulations (comprising 10,000 runs) to quantify the uncertainty associated with the calculation of pCO_2 based on the b-estimates of the three independent variables from the multilinear regression. This was followed by further Monte Carlo simulations (again 10,000 runs) to estimate the uncertainty associated with the calculation of CO_2 evasion, which allowed us to conclude that gas exchange velocity k is the largest source of uncertainty in the estimate

of total CO_2 evasion from boreal lakes (Hastie et al., 2018). Indeed, our estimate of the mean annual CO_2 evasion for the present day of 189 Tg C yr^{-1} comes with a considerable uncertainty range of between 74 Tg C yr^{-1} and 347 Tg C yr^{-1} (5^{th} and 95^{th} percentiles). Provided that the statistical model used for the upscaling is sufficiently powerful (which was the case here), data-driven approaches are arguably the best methods for estimating the regional to global scale spatial variation of aquatic CO_2 evasion for the present day, as well as a range of uncertainty. However, there are considerable limitations associated with empirical models.

As previously discussed, the vast majority of pCO_2 "observations" are indirectly calculated from alkalinity, pH, and water temperature, a method which is highly error prone in low-pH and low-alkalinity waters (Abril et al., 2015) and can lead to an overestimation of pCO_2 (Wallin et al., 2014), though this problem was at least partly addressed by discarding any calculations of pCO_2 undertaken at a pH $\leq$ 5.4. It was also found that the calculation of CO_2 evasion is highly dependent on the estimate of gas exchange velocity *k*, for which numerous parametrizations exist and indeed this was the largest source of uncertainty. Moreover, while we applied our empirical model to predict future CO_2 evasion under two RCP scenarios, these estimates were ultimately driven and therefore reliant on outputs of process-based earth system models (future projections of terrestrial NPP and precipitation). It is also important to reiterate the fact that this approach relied on the considerable assumption that the relationship between the environmental drivers observed for the present day holds into the future.

A land surface model such as ORCHILEAK allows for a wider range of research questions to be investigated, particularly those associated with temporal variation and crucially it allows integration of the aquatic C cycle within the terrestrial. In chapter 3 one of the major findings was that the interannual variation in aquatic C fluxes moderates the corresponding variation in terrestrial C fluxes at the regional scale of the Amazon Basin, something that would be much more difficult, if not impossible, to investigate using only observations or empirical

approaches. With models such as ORCHILEAK, one can also apply different scenarios of atmospheric CO_2, climate and land-use change amongst other boundary conditions, to investigate how the C cycle will react to different perturbations. Indeed, in this thesis ORCHILEAK was applied to the Congo Basin in chapter 4 for this very purpose. Interestingly, in the ORCHILEAK model, CO_2 evasion is relatively insensitive to changes in the gas exchange velocity *k* (Lauerwald et al., 2017), in contrast to the empirical approach where evasion changes linearly with *k*. Rather, CO_2 evasion in ORCHILEAK is sensitive to CO_2 inputs and in-stream decomposition. However, process-based models are also open to criticism for excluding important processes, an issue discussed in greater detail in the proceeding chapters. Moreover, models like ORCHILEAK are dependent on observed data of sufficient quality and spatio-temporal resolution to calibrate and validate their outputs. In summary, despite their respective limitations empirical, process-based models and field studies all have to be employed in our attempts to better understand the global C cycle.

5.3. Current methodological limitations and future research avenues

Looking at the uncertainty ranges in our regional C budgets of Figure 1, it is clear that LOAC research is a field in which considerable data and methodological limitations remain. The following section discusses the most pressing limitations and uncertainties associated with ORCHILEAK and the current database of observations, as well as some proposals to address some of these limitations.

5.3.1. Application of ORCHILEAK to other regions

In terms of modelling, in the immediate future ORCHILEAK could be applied across the entire tropical region in its current form. The Gumbricht dataset (Gumbricht et al., 2017) utilized in both Amazon and Congo studies spans the entire tropical region and therefore it would be a relatively simple task to create pan-tropical wetland forcing files. Moreover, a recent study (Tootchi et al., 2019) developed a high resolution (500m) composite wetland database at the

global scale, which accounts for both groundwater/ rainfed wetlands, as well as floodplains. This could potentially be used to develop wetland forcing files for ORCHILEAK both at the global scale, and for non-tropical regional studies.

Perhaps more difficult would be finding a model calibration which provides sufficiently good performance (against observed data) across the tropics. In order to improve the hydrological performance of ORCHILEAK in the Congo, the model constants which dictate residence time of the water reservoirs were substantially reduced, which suggests that finding a universal calibration might not be so straightforward. However, more fundamental to the poorer performance of the Congo hydrological performance against observations, was the climate forcing datasets, all of which tested in this study overestimate the seasonality of the Congo hydrological cycle, irrespective of calibration. As such, it may be worthwhile trying to improve one of the present-day climate forcing datasets such as Princeton GPCC (which already incorporates some observations), by correcting them with the latest observed data, something which has previously been done for the Amazon Basin (Guimberteau et al., 2012) for NCC (Ngo-Duc et al., 2005). Alternatively, there may already be climate forcing datasets with sufficiently good performance across the tropical region that were not tested in this analysis. Applying ORCHILEAK to the entire tropical region would allow us to test whether the moderating effect of the aquatic C fluxes on the interannual variation of net ecosystem production (NEP), that was found in the Amazon (chapter 3), applies more widely. The results of a potential pan-tropics study could have substantial implications for understanding and quantifying the overall land C budget for the tropical region. Another relatively simple improvement, discussed in the previous chapter, would be the development and validation of an improved “poor soils” forcing file, which prescribes reduced decomposition rates in low pH and nutrient soils such as Podzols and Arenosols based on the Harmonized World Soil Database (FAO/IIASA/ISRIC/ISS-CAS/JRC, 2009). The new version could be based on the

Soil Grids database (Hengl et al., 2014), which appears to be able to better represent tropical peatland soils. A modified version of ORCHILEAK has also been recently applied to Arctic rivers, while it is currently in the process of being calibrated to large European rivers.

5.3.2. Model Improvements

In terms of land surface models, ORCHILEAK is at the cutting edge of integrating the LOAC into the terrestrial C cycle but still it suffers from considerable limitations. As discussed in the previous chapter, ORHCILEAK fails to account for tropical peatlands. A peatland sink of 30 Pg C (Dargie at al., 2017) was recently discovered in the Congo Basin while a smaller but still considerable peatland sink of 3 Pg C (Draper et al., 2014) was found in the Amazon. This is an emerging research field and thus it is likely that further tropical peatlands will be discovered over the coming years. As such it is becoming more and more important to incorporate the specific dynamics of tropical peatlands into land surface models, especially as recent research suggests that tropical peatlands are less resilient to drainage than northern peatlands (Evans et al., 2014). Recent research on South-East Asian catchments has also shown that a large proportion the C exported from tropical peatlands to estuaries is respired before reaching the coast (Müller et al., 2016).

ORCHIDEE- PEAT (Qiu et al., accepted, 2019, Appendix 8.2) is a new model branch of ORCHIDEE and represents peat as an independent sub-grid hydrological soil unit in which peatland soils are characterized by peat-specific hydrological properties and multi-layered transport of C and water. It has thus far only been applied to northern latitudes but again could be merged with ORCHILEAK and potentially recalibrated to represent tropical peatlands. A recent study applied the CLIMBER-LPJ model, a version of the LPJ-DGVM to simulate the long-term persistence of peatlands globally (including tropical) and found that tropical peatlands are more stable than northern peatlands over millennial timescales (Treat et al., 2019). However, this study only incorporated very few tropical peatland cores in its calibration

and moreover uses a grassland PFT as the sole land-cover type for peatlands, an approach which does not reflect the reality of most tropical peatlands where many are overlain by forest PFTs such as palm and pole forest (Draper et al., 2014).

Other potential medium-long term aspirations could be the implementation of lakes and reservoirs into ORCHILEAK, the representation of methanogenesis and methane (CH_4) evasion, the effects of waterlogged soils on terrestrial NPP, the explicit representation of aquatic plants, and the representation of weathering derived DIC fluxes, as well as erosion derived POC fluxes. Wetlands are the world's largest natural source of methane comprising one-third of total emissions and this flux is projected to increase with climate change (Zhang et al., 2017). Moreover, the world is undergoing a second boom in dam construction and this time it will be concentrated in the tropical region; in the Amazon basin alone 140 dams under construction or already in operation, and a further 288 planned (Latrubesse et al., 2017). These dams will have direct impacts on the C retention efficiency within the LOAC (Maavara et al., 2017) as well as implications for hydrology and ecology, and this is therefore another important area for research efforts to be focused on. In its current form, ORCHILEAK does not simulate the lateral transport of POC, though it does account for the decomposition of submerged litter as a substantial source of DOC and dissolved CO_2 to the water column; in other words, POC from submerged litter is assumed to decompose locally in ORCHILEAK. Increasing soil erosion due to land-use change represents another substantial perturbation of the C fluxes to the LOAC (Naipal et al. 2018). Once again, another version of ORCHIDEE developed to represent POC erosion fluxes (Naipal et al., 2018) could potentially be merged with ORCHILEAK to address this limitation. Moreover, another aspect of the C cycle which is currently missing in earth system models such as ORCHIDEE is primary production in freshwaters. Indeed, there is no PFT in ORCHIDEE to represent aquatic plants such as

macrophytes. Globally, freshwater photosynthetic fixation could be as high as 0.3 Pg C yr^{-1} (Regnier et al., 2013) so this is a significant current limitation which needs to be addressed.

Finally, ORCHILEAK does not account for the effect of nutrient limitation on the C cycle. A 2012 study using the JSBACH model found that globally, the implementation of the Nitrogen (N) and Phosphorus (P) and coupling with the C cycle, reduced estimates of the land C uptake from 1810-2100 by 25% (Goll et al., 2012). In the tropics, accounting for the P cycle had the greater effect, as tropical biomes tend to be dominated by old, highly weathered soils with low P availability. The study showed that for the present day, accounting for P limitation reduced the NPP of tropical plant functional types (PFTs) by 10-20%, though the study also predicted that even P limitation would become less important in the tropics by 2100 (Goll et al., 2012). This is due to increasing temperatures eventually lowering NPP and in turn nutrient demands. A new version of ORCHIDEE, which represents both the P and N cycle was recently developed (Goll et al., 2017) and therefore integrating the transfer of N and P along the LOAC into ORCHILEAK could be possible in the relatively near future, though merging the two model branches and implementing the aquatic fluxes of N and P would require considerable time and resources. Nevertheless, the N and P cycles also require that we integrate the terrestrial, aquatic and export fluxes in the same boundless way as has been represented for C in ORCHILEAK, i.e. not restricted to their impact on the C cycle. This approach would also allow us to investigate the long-term perturbations of the N and P cycles right along the LOAC.

5.3.3. Data for calibration and validation

In terms of observations, there are large geographic regions in which little or no data is available such as large parts of Siberia, as discussed in chapter 2. Moreover, even in areas with extensive datasets, observations are rarely made at sufficiently frequent intervals to capture the full spectrum of the temporal variation in pCO_2 which has been shown to fluctuate significantly even over the course of a day. For example, a recent study (Xu et al., 2019) conducted fieldwork

in a subtropical eutrophic lake and found that mean pCO_2 was more than 4 times higher at 07:00 (1113 µatm) than at 17:00 (205 µatm) meaning that the same lake changed from a substantial C source to sink over the course of a day. pCO_2 has also been found to vary at extremely high spatial resolutions. A 2018 study on an oligotrophic boreal lake, found that CO_2 evasion was highest 5m from the lake shore and 12 times higher than values at 15m from the shore (Spafford & Risk, 2018). Moreover, they found that each distance from the shore (5, 10 and 15m) had its own distinct temporal variability in CO_2 exchange with the atmosphere. Studies such as these further highlight the incredible challenge that we face in accounting for spatial and temporal variation at larger scales and resolutions. Additionally, the vast majority of pCO_2 data is indirectly calculated from pH, alkalinity and water temperature and CO_2 fluxes are similarly rarely measured directly, meaning that these are considerable sources of error (Abril et al., 2015). As previously noted, the estimation of gas exchange velocity *k* is a large source of uncertainty and hence more direct observations of GHG fluxes would be invaluable for both training empirical models with, as well as for the calibration and validation of process-based models.

5.4. Closing remarks

With so many current methodological limitations and gaps in our understanding, one of the greatest challenges lies in the prioritization of future research. One could improve models indefinitely, incorporating more and more processes at higher and higher temporal and spatial resolutions until there is theoretically little difference between the modelled and the real world. However, how can we know that our results are realistic unless we have sufficient observed data with which to compare them to? Moreover, how can we be sure that we are not missing some crucial unknown biogeochemical process if we do not continue to undertake fundamental research in the field? Model developments have to be complemented by data collection in the field, as well as remote sensing; indeed, this thesis used field data to both drive an empirical

model and to calibrate/ validate a process-based model, while remote sensing data was used to create novel forcing files. Furthermore, as previously discussed, the C cycle varies at an extremely high temporal and spatial resolution and therefore it is likely that modelling and remote sensing will always be reliant upon field data to confine and test the limits of their skill and power. Current land surface models such as ORCHILEAK are simply not capable of simulating some of the extremely high spatio-temporal variation observed in field studies and in these cases, we are even more reliant upon field investigations.

Finding the balance between putting resources into further model development versus remote sensing versus fieldwork requires detailed discussion and collaboration. One such project that aims to achieve this kind of collaboration, and in which I am involved, is the Readiness of ICOS for Necessities of Integrated Global Observations (RINGO) project, led by the Integrated Carbon Observation System (ICOS) organization. RINGO aims to expand the current network of 130 greenhouse gas measurement stations in Europe whilst at the same time improving the quality of data and access to potential users. A major part of this project is to identify geographic gaps in the current network of stations. This is being achieved through a series of workshops involving many countries and crucially, biogeochemical researchers operating at various different scales, from local observational studies to global models. This allows for a unique transfer of knowledge that would be impossible if it were only modellers or similarly only observational scientists. If we are to address the significant remaining challenges in our understanding of the C cycle, these collaborations are nothing short of essential.

6. References

Abril, G., Bouillon, S., Darchambeau, F., Teodoru, C. R., Marwick, T. R., Tamooh, F., … Borges, A. V. (2015). Technical Note: Large overestimation of pCO_2 calculated from pH and alkalinity in acidic, organic-rich freshwaters. *Biogeosciences*, *12*(1), 67–78. http://doi.org/10.5194/bg-12-67-2015

Abril, G., Martinez, J.-M., Artigas, L. F., Moreira-Turcq, P., Benedetti, M. F., Vidal, L., … Roland, F. (2013). Amazon River carbon dioxide outgassing fuelled by wetlands. *Nature*, *505*, 395. Retrieved from http://dx.doi.org/10.1038/nature12797

Adam, H., Ronny, L., Gesa, W., Sebastian, S., Charles, V., & Pierre, R. (2017). CO2 evasion from boreal lakes: Revised estimate, drivers of spatial variability, and future projections. *Global Change Biology*, *24*(2), 711–728. http://doi.org/10.1111/gcb.13902

Akaike, H. (1998). A New Look at the Statistical Model Identification. In E. Parzen, K. Tanabe, & G. Kitagawa (Eds.), *Selected Papers of Hirotugu Akaike* (pp. 215–222). New York, NY: Springer New York. http://doi.org/10.1007/978-1-4612-1694-0_16

Algesten, G., Sobek, S., Bergström, A.-K., Ågren, A., Tranvik, L. J., & Jansson, M. (2004). Role of lakes for organic carbon cycling in the boreal zone. *Global Change Biology*, *10*(1), 141–147. http://doi.org/10.1111/j.1365-2486.2003.00721.x

Almeida, R. M., Pacheco, F. S., Barros, N., Rosi, E., & Roland, F. (2017). Extreme floods increase CO2 outgassing from a large Amazonian river. *Limnology and Oceanography*, *62*(3), 989–999. http://doi.org/10.1002/lno.10480

Amiotte Suchet, P., Probst, J.-L., & Ludwig, W. (2003). Worldwide distribution of continental rock lithology: Implications for the atmospheric/soil CO2 uptake by continental weathering and alkalinity river transport to the oceans. *Global Biogeochemical Cycles*, *17*(2). http://doi.org/10.1029/2002GB001891

Andreoli, R. V., de Souza, R. A., Kayano, M. T., & Candido, L. A. (2012). Seasonal anomalous rainfall in the central and eastern Amazon and associated anomalous oceanic and atmospheric patterns. *International Journal of Climatology*, *32*(8), 1193–1205. http://doi.org/10.1002/joc.2345

Asner, G. P., Townsend, A. R., & Braswell, B. H. (2000). Satellite observation of El Niño effects on Amazon Forest phenology and productivity. *Geophysical Research Letters*, *27*(7), 981–984. http://doi.org/10.1029/1999GL011113

Aufdenkampe, A. K., Mayorga, E., Raymond, P. A., Melack, J. M., Doney, S. C., Alin, S. R., … Yoo, K. (2011). Riverine coupling of biogeochemical cycles between land, oceans, and atmosphere. *Frontiers in Ecology and the Environment*, *9*(1), 53–60. http://doi.org/10.1890/100014

Battin, T. J., Luyssaert, S., Kaplan, L. A., Aufdenkampe, A. K., Richter, A., & Tranvik, L. J. (2009). The boundless carbon cycle. *Nature Geoscience*, *2*, 598. Retrieved from https://doi.org/10.1038/ngeo618

Bauer, J. E., Cai, W.-J., Raymond, P. A., Bianchi, T. S., Hopkinson, C. S., & Regnier, P. A. G. (2013). The changing carbon cycle of the coastal ocean. *Nature*, *504*, 61. Retrieved from https://doi.org/10.1038/nature12857

Becker, M., Papa, F., Frappart, F., Alsdorf, D., Calmant, S., da Silva, J. S., … Seyler, F. (2018). Satellite-based estimates of surface water dynamics in the Congo River Basin. *International Journal of Applied Earth Observation and Geoinformation*, *66*, 196–209. http://doi.org/https://doi.org/10.1016/j.jag.2017.11.015

BOND-LAMBERTY, B. P., & THOMSON, A. M. (2014). A Global Database of Soil Respiration Data, Version 3.0. ORNL Distributed Active Archive Center. https://doi.org/10.3334/ornldaac/1235

Borges, A. V, Abril, G., Darchambeau, F., Teodoru, C. R., Deborde, J., Vidal, L. O., … Bouillon, S. (2015)[b]. Divergent biophysical controls of aquatic CO2 and CH4 in the World's two largest rivers. *Scientific Reports*, *5*, 15614. http://doi.org/10.1038/srep15614

Borges, A. V, Darchambeau, F., Teodoru, C. R., Marwick, T. R., Tamooh, F., Geeraert, N., … Bouillon, S. (2015)[a]. Globally significant greenhouse-gas emissions from African inland waters. *Nature Geoscience*, *8*, 637. Retrieved from https://doi.org/10.1038/ngeo2486

Botta, A., Ramankutty, N., & Foley, J. A. (2002). Long-term variations of climate and carbon fluxes over the Amazon basin. *Geophysical Research Letters*, *29*(9), 33–34. http://doi.org/10.1029/2001GL013607

Bouillon, S., Yambélé, A., Gillikin, D. P., Teodoru, C., Darchambeau, F., Lambert, T., & Borges, A. V. (2014). Contrasting biogeochemical characteristics of the Oubangui River and tributaries (Congo River basin). *Scientific Reports*, *4*, 5402. Retrieved from https://doi.org/10.1038/srep05402

Butman, D., Stackpoole, S., Stets, E., McDonald, C. P., Clow, D. W., & Striegl, R. G. (2016). Aquatic carbon cycling in the conterminous United States and implications for terrestrial carbon accounting. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(1), 58–63. http://doi.org/10.1073/pnas.1512651112

Cai, W.-J. (2011). Estuarine and Coastal Ocean Carbon Paradox: CO2 Sinks or Sites of Terrestrial Carbon Incineration? Annual Review of Marine Science, 3(1), 123–145. https://doi.org/10.1146/annurev-marine-120709-142723

Camino-Serrano, M., Guenet, B., Luyssaert, S., Ciais, P., Bastrikov, V., De Vos, B., … Janssens, I. A. (2018). ORCHIDEE-SOM: modeling soil organic carbon (SOC) and dissolved organic carbon (DOC) dynamics along vertical soil profiles in Europe. *Geosci. Model Dev.*, *11*(3), 937–957. http://doi.org/10.5194/gmd-11-937-2018

Campeau, A., & del Giorgio, P. A. (2014). Patterns in CH4 and CO2 concentrations across boreal rivers: Major drivers and implications for fluvial greenhouse emissions under climate change scenarios. *Global Change Biology*, *20*(4), 1075–1088. http://doi.org/10.1111/gcb.12479

Catalan, N., Marce, R., Kothawala, D. N., & Tranvik, L. J. (2016). Organic carbon decomposition rates controlled by water retention time across inland waters. *Nature Geosci*, *9*(7), 501–504. Retrieved from http://dx.doi.org/10.1038/ngeo2720

Chmiel, H. E., Kokic, J., Denfeld, B. A., Einarsdóttir, K., Wallin, M. B., Koehler, B., … Sobek, S. (2016). The role of sediments in the carbon budget of a small boreal lake. *Limnology and Oceanography*, *61*(5), 1814–1825. http://doi.org/10.1002/lno.10336

Chou, C., Chiang, J. C. H., Lan, C.-W., Chung, C.-H., Liao, Y.-C., & Lee, C.-J. (2013). Increase in the range between wet and dry season precipitation. *Nature Geoscience*, *6*, 263. Retrieved from http://dx.doi.org/10.1038/ngeo1744

Ciais, P., Piao, S.-L., Cadule, P., Friedlingstein, P., & Chédin, A. (2009). Variability and recent trends in the African terrestrial carbon balance. *Biogeosciences*, *6*(9), 1935–1948. http://doi.org/10.5194/bg-6-1935-2009

Ciais, P., Sabine, C., Bala, G., Bopp, L., Brovkin, V., Canadell, J., . . . (2013). Carbon and Other Biogeochemical Cycles. In T. F.. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S. K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex, and P. M. Midgley (Eds.), Climate change 2013: The physical science basis. Contribution of working group I to the fifth assessment report of the intergovernmental panel on climate change (pp. 465–570). Cambridge University Press, Cambridge.

Ciais, P., Gasser, T., Lauerwald, R., Peng, S., Raymond, P. A., Wang, Y., Zhu, D. (2017). Observed regional carbon budgets imply reduced soil heterotrophic respiration. Nature, in review.

CIESIN, and CIAT. (2005). Gridded population of the world version 3 (GPWv3): POPULATION Grids CIESIN, Columbia University. New York, Palisades, N. Y. https://doi.org/10.7927/h4639mpp

Cochonneau, G., Sondag, F., Guyot, J.-L., Geraldo, B., Filizola, N., Fraizy, P., Laraque, A., Magat, P., Martinez, J.-M., Nor iega, L., Oliveira, E., Ordonez, J., Pombosa, R., Seyler, F., Sidgwick, J., and Vauchel, P.: The environmental observation and research project, ORE HYBAM, and the rivers of the Amazon basin, in: Climate Variability and Change – Hydrological Impacts, IAHS Publ. 308, edited by: Demuth, S., Gustard, A., Planos, E., Scatena, F., and Servat, E., IAHS Press, UK, 44–50, 2006

Cole, J. J., Prairie, Y. T., Caraco, N. F., McDowell, W. H., Tranvik, L. J., Striegl, R. G., … Melack, J. (2007). Plumbing the Global Carbon Cycle: Integrating Inland Waters into the Terrestrial Carbon Budget. *Ecosystems*, *10*(1), 172–185. http://doi.org/10.1007/s10021-006-9013-8

Cole, J. J., & Caraco, N. F. (1998). Atmospheric exchange of carbon dioxide in a low-wind oligotrophic lake measured by the addition of SF6. *Limnology and Oceanography*, *43*(4), 647–656. http://doi.org/10.4319/lo.1998.43.4.0647

Cole, J. J., Caraco, N. F., Kling, G. W., & Kratz, T. K. (1994). Carbon Dioxide Supersaturation in the Surface Waters of Lakes. *Science*, *265*(5178), 1568–1570. http://doi.org/10.1126/science.265.5178.1568

Cox, P. M., Betts, R. A., Jones, C. D., Spall, S. A., & Totterdell, I. J. (2000). Acceleration of global warming due to carbon-cycle feedbacks in a coupled climate model. *Nature*, *408*, 184. Retrieved from https://doi.org/10.1038/35041539

Coynel, A., Seyler, P., Etcheber, H., Meybeck, M., & Orange, D. (2005). Spatial and seasonal dynamics of total suspended sediment and organic carbon species in the Congo River. *Global Biogeochemical Cycles*, *19*(4). http://doi.org/10.1029/2004GB002335

Curtarelli, M. P., Ogashawara, I., de Araújo, C. A. S., Lorenzzetti, J. A., Leão, J. A. D., Alcântara, E., & Stech, J. L. (2016). Carbon dioxide emissions from Tucuruí reservoir (Amazon biome): New findings based on three-dimensional ecological model simulations. *Science of The Total Environment*, *551*(Supplement C), 676–694. http://doi.org/https://doi.org/10.1016/j.scitotenv.2016.02.001

Dargie, G. C., Lewis, S. L., Lawson, I. T., Mitchard, E. T. A., Page, S. E., Bocko, Y. E., & Ifo, S. A. (2017). Age, extent and carbon storage of the central Congo Basin peatland complex. *Nature*, *542*, 86. Retrieved from https://doi.org/10.1038/nature21048

de Fátima F. L. Rasera, M., Krusche, A. V, Richey, J. E., Ballester, M. V. R., & Victória, R. L. (2013). Spatial and temporal variability of pCO2 and CO2 efflux in seven Amazonian Rivers. *Biogeochemistry*, *116*(1), 241–259. http://doi.org/10.1007/s10533-013-9854-0

de Rosnay, P., Polcher, J., Bruen, M., & Laval, K. (2002). Impact of a physically based soil water flow and soil-plant interaction representation for modeling large-scale land surface processes. *Journal of Geophysical Research: Atmospheres*, *107*(D11), ACL 3-1-ACL 3-19. http://doi.org/10.1029/2001JD000634

D'Elia, A. H., Liles, G. C., Viers, J. H., & Smart, D. R. (2017). Deep carbon storage potential of buried floodplain soils. *Scientific Reports*, *7*(1), 8181. http://doi.org/10.1038/s41598-017-06494-4

Denfeld, B. A., Baulch, H. M., del Giorgio, P. A., Hampton, S. E., & Karlsson, J. (2018). A synthesis of carbon dioxide and methane dynamics during the ice-covered period of northern lakes. *Limnology and Oceanography Letters*, *3*(3), 117–131. http://doi.org/10.1002/lol2.10079

Denfeld, B. A., Kortelainen, P., Rantakari, M., Sobek, S., & Weyhenmeyer, G. A. (2016). Regional Variability and Drivers of Below Ice CO2 in Boreal and Subarctic Lakes. *Ecosystems*, *19*(3), 461–476. http://doi.org/10.1007/s10021-015-9944-z

Dominic, V., & T., P. Y. (2013). The ecosystem size and shape dependence of gas transfer velocity versus wind speed relationships in lakes. *Canadian Journal of Fisheries and Aquatic Sciences*, *70*(12), 1757–1764. http://doi.org/10.1139/cjfas-2013-0241

d'Orgeval, T., Polcher, J., & de Rosnay, P. (2008). Sensitivity of the West African hydrological cycle in ORCHIDEE to infiltration processes. *Hydrol. Earth Syst. Sci.*, *12*(6), 1387–1401. http://doi.org/10.5194/hess-12-1387-2008

dos Santos, M. A., Damázio, J. M., Rogério, J. P., Amorim, M. A., Medeiros, A. M., Abreu, J. L. S., … Rosa, L. P. (2017). Estimates of GHG emissions by hydroelectric reservoirs: The Brazilian case. *Energy*, *133*(Supplement C), 99–107. http://doi.org/https://doi.org/10.1016/j.energy.2017.05.082

Doughty, C. E., Metcalfe, D. B., Girardin, C. A. J., Amézquita, F. F., Cabrera, D. G., Huasco, W. H., … Malhi, Y. (2015). Drought impact on forest carbon dynamics and fluxes in Amazonia. *Nature*, *519*, 78. Retrieved from https://doi.org/10.1038/nature14213

Downing, J. A., Cole, J. J., Duarte, C. M., Middelburg, J. J., Melack, J. M., Prairie, Y. T., … Tranvik, L. J. (2012). Global abundance and size distribution of streams and rivers. *Inland Waters*, *2*(4), 229–236. http://doi.org/10.5268/IW-2.4.502

Drake, T. W., Raymond, P. A., & Spencer, R. G. M. (2018). Terrestrial carbon inputs to inland waters: A current synthesis of estimates and uncertainty. *Limnology and Oceanography Letters*, *3*(3), 132–142. http://doi.org/10.1002/lol2.10055

Drake, T. W., Wickland, K. P., Spencer, R. G. M., McKnight, D. M., & Striegl, R. G. (2015). Ancient low–molecular-weight organic acids in permafrost fuel rapid carbon dioxide production upon thaw. *Proceedings of the National Academy of Sciences* , *112*(45), 13946–13951. http://doi.org/10.1073/pnas.1511705112

Draper, F. C., Roucoux, K. H., Lawson, I. T., Mitchard, E. T. A., Coronado, E. N. H., Lähteenoja, O., … Baker, T. R. (2014). The distribution and amount of carbon in the largest peatland complex in Amazonia. *Environmental Research Letters*, *9*(12), 124017. http://doi.org/10.1088/1748-9326/9/12/124017

Evans, C. D., Page, S. E., Jones, T., Moore, S., Gauci, V., Laiho, R., … Garnett, M. H. (2014). Contrasting vulnerability of drained tropical and high-latitude peatlands to fluvial loss of stored carbon. *Global Biogeochemical Cycles*, *28*(11), 1215–1234. http://doi.org/10.1002/2013GB004782

Evans, W., Hales, B., & Strutton, P. G. (2013). pCO2 distributions and air–water CO2 fluxes in the Columbia River estuary. *Estuarine, Coastal and Shelf Science*, *117*, 260–272. http://doi.org/https://doi.org/10.1016/j.ecss.2012.12.003

FAO/IIASA/ISRIC/ISS-CAS/JRC: Harmonized World Soil Database (version 1.1), FAO, Rome, 2009.

Fasullo, J. T., Otto-Bliesner, B. L., & Stevenson, S. (2018). ENSO's Changing Influence on Temperature, Precipitation, and Wildfire In a Warming Climate. *Geophysical Research Letters*, *0*(ja). http://doi.org/10.1029/2018GL079022

Fekete, B. M., Vörösmarty, C. J., & Grabs, W. (2002). High-resolution fields of global runoff combining observed river discharge and simulated water balances. *Global Biogeochemical Cycles*, *16*(3), 10–15. http://doi.org/10.1029/1999GB001254

Feldpausch, T. R., Phillips, O. L., Brienen, R. J. W., Gloor, E., Lloyd, J., Lopez-Gonzalez, G., … Vos, V. A. (2016). Amazon forest response to repeated droughts. *Global Biogeochemical Cycles*, *30*(7), 964–982. http://doi.org/10.1002/2015GB005133

Fisher, J. B., Sikka, M., Sitch, S., Ciais, P., Poulter, B., Galbraith, D., … Malhi, Y. (2013). African tropical rainforest net carbon dioxide fluxes in the twentieth century. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *368*(1625), 20120376. http://doi.org/10.1098/rstb.2012.0376

Foley, J. A., Botta, A., Coe, M. T., & Costa, M. H. (2002). El Niño–Southern oscillation and the climate, ecosystems and rivers of Amazonia. *Global Biogeochemical Cycles*, *16*(4), 20–79. http://doi.org/10.1029/2002GB001872

Friedlingstein, P., Cox, P., Betts, R., Bopp, L., von Bloh, W., Brovkin, V., … Zeng, N. (2006). Climate–Carbon Cycle Feedback Analysis: Results from the C4MIP Model Intercomparison. *Journal of Climate*, *19*(14), 3337–3353. http://doi.org/10.1175/JCLI3800.1

Frieler, K., Lange, S., Piontek, F., Reyer, C. P. O., Schewe, J., Warszawski, L., … Yamagata, Y. (2017). Assessing the impacts of 1.5 °C global warming – simulation protocol of the Inter-Sectoral Impact Model Intercomparison Project (ISIMIP2b). *Geosci. Model Dev.*, *10*(12), 4321–4345. http://doi.org/10.5194/gmd-10-4321-2017

Fu, R., Yin, L., Li, W., Arias, P. A., Dickinson, R. E., Huang, L., … Myneni, R. B. (2013). Increased dry-season length over southern Amazonia in recent decades and its implication for future climate projection. *Proceedings of the National Academy of Sciences*, *110*(45), 18110–18115. http://doi.org/10.1073/pnas.1302584110

Gaillardet, J., Dupré, B., Louvat, P., & Allègre, C. J. (1999). Global silicate weathering and CO2 consumption rates deduced from the chemistry of large rivers. *Chemical Geology*, *159*(1), 3–30. http://doi.org/https://doi.org/10.1016/S0009-2541(99)00031-5

Gatti, L. V, Gloor, M., Miller, J. B., Doughty, C. E., Malhi, Y., Domingues, L. G., … Lloyd, J. (2014). Drought sensitivity of Amazonian carbon balance revealed by atmospheric measurements. *Nature*, *506*, 76. Retrieved from http://dx.doi.org/10.1038/nature12957

Gauthier, S., Bernier, P., Kuuluvainen, T., Shvidenko, A. Z., & Schepaschenko, D. G. (2015). Boreal forest health and global change. *Science*, *349*(6250), 819–822. http://doi.org/10.1126/science.aaa9092

Global Land Cover 2000 database. (2003). European Commission, Joint Research Centre, 2003. http://bioval.jrc.ec.europa.eu/glc2000/glc 2000.php.

GLOBE Task Team and others (Hastings, David A., Paula K. Dunbar, Gerald M. Elphingstone, Mark Bootz, Hiroshi Murakami, Hiroshi Maruyama, Hiroshi Masaharu, Peter Holland, John Payne, Nevin A. Bryant, Thomas L. Logan, J.-P. Muller, Gunter Schreier, and John S. MacDonald), eds. (1999). The Global Land One-kilometer Base Elevation (GLOBE) Digital Elevation Model, Version 1.0. National Oceanic and Atmospheric Administration, National Geophysical Data Center, 325 Broadway, Boulder, Colorado 80305-3328, U.S.A. Digital data base on the World Wide Web (http://www.ngdc.noaa.gov/mgg/to po/globe.html) and CD-ROMs.

Gloor, M., Barichivich, J., Ziv, G., Brienen, R., Schöngart, J., Peylin, P., … Baker, J. (2015). Recent Amazon climate as background for possible ongoing and future changes of Amazon humid forests. *Global Biogeochemical Cycles*, *29*(9), 1384–1399. http://doi.org/10.1002/2014GB005080

Gloor, M., Brienen, R. J. W., Galbraith, D., Feldpausch, T. R., Schöngart, J., Guyot, J.-L., … Phillips, O. L. (2013). Intensification of the Amazon hydrological cycle over the last two decades. *Geophysical Research Letters*, *40*(9), 1729–1733. http://doi.org/10.1002/grl.50377

Gloor, M., Gatti, L., Brienen, R., Feldpausch, T. R., Phillips, O. L., Miller, J., … Lloyd, J. (2012). The carbon balance of South America: a review of the status, decadal trends and main determinants. *Biogeosciences*, *9*(12), 5407–5430. http://doi.org/10.5194/bg-9-5407-2012

Goll, D. S., Brovkin, V., Parida, B. R., Reick, C. H., Kattge, J., Reich, P. B., … Niinemets, Ü. (2012). Nutrient limitation reduces land carbon uptake in simulations with a model of combined carbon, nitrogen and phosphorus cycling. *Biogeosciences*, *9*(9), 3547–3569. http://doi.org/10.5194/bg-9-3547-2012

Goll, D. S., Vuichard, N., Maignan, F., Jornet-Puig, A., Sardans, J., Violette, A., … Ciais, P. (2017). A representation of the phosphorus cycle for ORCHIDEE (revision 4520). *Geosci. Model Dev.*, *10*(10), 3745–3770. http://doi.org/10.5194/gmd-10-3745-2017

Grace J. (2016) The Amazon Carbon Balance: An Evaluation of Methods and Results. In: Nagy L., Forsberg B., Artaxo P. (eds) Interactions Between Biosphere, Atmosphere and Human Land Use in the Amazon Basin. Ecological Studies (Analysis and Synthesis), vol 227. Springer, Berlin, Heidelberg

Gudasz, C., Bastviken, D., Steger, K., Premke, K., Sobek, S., & Tranvik, L. J. (2010). Temperature-controlled organic carbon mineralization in lake sediments. *Nature*, *466*(7305), 478–481. Retrieved from http://dx.doi.org/10.1038/nature09186

Guimberteau, M., Drapeau, G., Ronchail, J., Sultan, B., Polcher, J., Martinez, J.-M., … Vauchel, P. (2012). Discharge simulation in the sub-basins of the Amazon using ORCHIDEE forced by new datasets. *Hydrology and Earth System Sciences*, *16*(3), 911–935. http://doi.org/10.5194/hess-16-911-2012

Gumbricht, T., Roman-Cuesta, R. M., Verchot, L., Herold, M., Wittmann, F., Householder, E., … Murdiyarso, D. (2017). An expert system model for mapping tropical wetlands and peatlands reveals South America as the largest contributor. *Global Change Biology*, *23*(9), 3581–3599. http://doi.org/10.1111/gcb.13689

Haensler, A., Saeed, F., & Jacob, D. (2013). Assessing the robustness of projected precipitation changes over central Africa on the basis of a multitude of global and regional climate projections. *Climatic Change*, *121*(2), 349–363. http://doi.org/10.1007/s10584-013-0863-8

Hanson, P. C., Bade, D. L., Carpenter, S. R., & Kratz, T. K. (2003). Lake metabolism: Relationships with dissolved organic carbon and phosphorus. *Limnology and Oceanography*, *48*(3), 1112–1119. http://doi.org/10.4319/lo.2003.48.3.1112

Hartmann, J., Jansen, N., Dürr, H. H., Kempe, S., & Köhler, P. (2009). Global CO2-consumption by chemical weathering: What is the contribution of highly active weathering regions? *Global and Planetary Change*, *69*(4), 185–194. http://doi.org/https://doi.org/10.1016/j.gloplacha.2009.07.007

Hastie, A., Lauerwald, R., Weyhenmeyer, G., Sobek, S., Verpoorter, C., & Regnier, P. (2018). CO2 evasion from boreal lakes: Revised estimate, drivers of spatial variability, and future projections. *Global Change Biology*, *24*(2), 711–728. http://doi.org/10.1111/gcb.13902

Hastie, A., Lauerwald, R., Ciais, P., Regnier, P (accepted). Aquatic carbon fluxes dampen the overall variation of net ecosystem productivity in the Amazon basin: An analysis of the interannual variability in the boundless carbon cycle**.** *Global Change Biology, accepted*

Heathcote, A. J., Anderson, N. J., Prairie, Y. T., Engstrom, D. R., & del Giorgio, P. A. (2015). Large increases in carbon burial in northern lakes during the Anthropocene. *Nature Communications*, *6*, 10016. Retrieved from http://dx.doi.org/10.1038/ncomms10016

Hengl, T., de Jesus, J. M., MacMillan, R. A., Batjes, N. H., Heuvelink, G. B. M., Ribeiro, E., … Gonzalez, M. R. (2014). SoilGrids1km — Global Soil Information Based on Automated Mapping. *PLOS ONE*, *9*(8), 1–17. http://doi.org/10.1371/journal.pone.0105992

HESS, L. L., MELACK, J. M., NOVO, E. M. L. M., BARBOSA, C. C. F., & GASTIL, M. (2015). LBA-ECO LC-07 JERS-1 SAR Flooded Wetlands and Vegetation, Amazon Basin: 1995-1996. ORNL Distributed Active Archive Center. http://doi.org/10.3334/ornldaac/1284

Hess, L. L., Melack, J. M., Novo, E. M. L. M., Barbosa, C. C. F., & Gastil, M. (2003). Dual-season mapping of wetland inundation and vegetation for the central Amazon basin. *Remote Sensing of Environment*, *87*(4), 404–428. http://doi.org/https://doi.org/10.1016/j.rse.2003.04.001

Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., & Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, *25*(15), 1965–1978. http://doi.org/10.1002/joc.1276

Holgerson, M. A., & Raymond, P. A. (2016). Large contribution to inland water CO2 and CH4 emissions from very small ponds. *Nature Geosci*, *9*(3), 222–226. Retrieved from http://dx.doi.org/10.1038/ngeo2654

HUMBORG, C., MÖRTH, C.-M., SUNDBOM, M., BORG, H., BLENCKNER, T., GIESLER, R., & ITTEKKOT, V. (2010). CO2 supersaturation along the aquatic conduit in Swedish watersheds as constrained by terrestrial respiration, aquatic respiration and weathering. *Global Change Biology*, *16*(7), 1966–1978. http://doi.org/10.1111/j.1365-2486.2009.02092.x

Hurtt, G. C., Chini, L. P., Frolking, S., Betts, R. A., Feddema, J., Fischer, G., … Wang, Y. P. (2011). Harmonization of land-use scenarios for the period 1500--2100: 600 years of global gridded annual land-use transitions, wood harvest, and resulting secondary lands. *Climatic Change*, *109*(1), 117. https://doi.org/10.1007/s10584-011-0153-2

Hutchins, R. H. S., Prairie, Y. T., & del Giorgio, P. A. (2019). Large-Scale Landscape Drivers of CO2, CH4, DOC, and DIC in Boreal River Networks. *Global Biogeochemical Cycles*, *33*(2), 125–142. http://doi.org/10.1029/2018GB006106

Huth, E. R. and M. C. and A. R. and R. F. and F. T. and A. (2018). The importance of forest structure for carbon fluxes of the Amazon rainforest. *Environmental Research Letters*, *13*(5), 54013. Retrieved from http://stacks.iop.org/1748-9326/13/i=5/a=054013

Intergovernmental Panel on Climate Change (IPCC), Climate Change. (2013). The physical science basis. In T. F. Stocker, et al. (Eds.), Contribution of working group I to the fifth assessment report of the intergovernmental panel on climate change. Cambridge: Cambridge University Press

Jones Jr., J. B., Stanley, E. H., & Mulholland, P. J. (2003). Long-term decline in carbon dioxide supersaturation in rivers across the contiguous United States. *Geophysical Research Letters*, *30*(10). http://doi.org/10.1029/2003GL017056

Joyce, L. A., Price, D. T., McKenney, D. W., Martin Siltanen, R. ., Papadopol, P., Lawrence, K., & Coulson, D. P. (2011). High resolution interpolation of climate scenarios for the conterminous USA and Alaska derived from general circulation model simulations. *Gen. Tech. Rep. RMRS-GTR-263*, 87. Retrieved from https://www.scopus.com/inward/record.uri?eid=2-

s2.0-84856470197%7B&%7DpartnerID=40%7B&%7Dmd5=3d4ceaaaf79cff6ad542c53f729a2067

Kicklighter, D. W., Hayes, D. J., McClelland, J. W., Peterson, B. J., McGuire, A. D., & Melillo, J. M. (2013). Insights and issues with simulating terrestrial DOC loading of Arctic river networks. *Ecological Applications*, *23*(8), 1817–1836. http://doi.org/10.1890/11-1050.1

Kim, H. (2017). *Global Soil Wetness Project Phase 3 Atmospheric Boundary Conditions (Experiment 1)* [Data set]. Data Integration and Analysis System (DIAS). https://doi.org/10.20783/DIAS.501

KORTELAINEN, P., RANTAKARI, M., HUTTUNEN, J. T., MATTSSON, T., ALM, J., JUUTINEN, S., … MARTIKAINEN, P. J. (2006). Sediment respiration and lake trophic state are important predictors of large CO2 evasion from small boreal lakes. *Global Change Biology*, *12*(8), 1554–1567. http://doi.org/10.1111/j.1365-2486.2006.01167.x

Kortelainen, P., Rantakari, M., Pajunen, H., Huttunen, J. T., Mattsson, T., Juutinen, S., … Martikainen, P. J. (2013). Carbon evasion/accumulation ratio in boreal lakes is linked to nitrogen. *Global Biogeochemical Cycles*, *27*(2), 363–374. http://doi.org/10.1002/gbc.20036

Kosten, S., Roland, F., Da Motta Marques, D. M. L., Van Nes, E. H., Mazzeo, N., Sternberg, L. da S. L., … Cole, J. J. (2010). Climate-dependent CO2 emissions from lakes. *Global Biogeochemical Cycles*, *24*(2), n/a-n/a. http://doi.org/10.1029/2009GB003618

Koven, C. D. (2013). Boreal carbon loss due to poleward shift in low-carbon ecosystems. *Nature Geosci*, *6*(6), 452–456. Retrieved from http://dx.doi.org/10.1038/ngeo1801

Landschützer, P., Gruber, N., Bakker, D. C. E., Schuster, U., Nakaoka, S., Payne, M. R., … Zeng, J. (2013). A neural network-based estimate of the seasonal to inter-annual variability of the Atlantic Ocean carbon sink. *Biogeosciences*, *10*(11), 7793–7815. http://doi.org/10.5194/bg-10-7793-2013

Lapierre, J.-F., & del Giorgio, P. A. (2012). Geographical and environmental drivers of regional differences in the lake pCO2 versus DOC relationship across northern landscapes. *Journal of Geophysical Research: Biogeosciences*, *117*(G3), n/a-n/a. http://doi.org/10.1029/2012JG001945

Lapierre, J.-F., Seekell, D. A., & del Giorgio, P. A. (2015). Climate and landscape influence on indicators of lake carbon cycling through spatial patterns in dissolved organic carbon. *Global Change Biology*, *21*(12), 4425–4435. http://doi.org/10.1111/gcb.13031

LARSEN, S., ANDERSEN, T. O. M., & HESSEN, D. A. G. O. (2011). Climate change predicted to cause severe increase of organic carbon in lakes. *Global Change Biology*, *17*(2), 1186–1192. http://doi.org/10.1111/j.1365-2486.2010.02257.x

Laruelle, G. G., Dürr, H. H., Lauerwald, R., Hartmann, J., Slomp, C. P., Goossens, N., & Regnier, P. A. G. (2013). Global multi-scale segmentation of continental and coastal waters from the watersheds to the continental margins. *Hydrol. Earth Syst. Sci.*, *17*(5), 2029–2051. http://doi.org/10.5194/hess-17-2029-2013

Laruelle, G. G., Lauerwald, R., Pfeil, B., & Regnier, P. (2014). Regionalized global budget of the CO2 exchange at the air-water interface in continental shelf seas. *Global Biogeochemical Cycles*, *28*(11), 1199–1214. https://doi.org/10.1002/2014GB004832

Laruelle, G. G., Landschützer, P., Gruber, N., Tison, J.-L., Delille, B., & Regnier, P. (2017). Global high-resolution monthly pCO2 climatology for the coastal ocean derived from neural network interpolation. *Biogeosciences*, *14*(19), 4545–4561. http://doi.org/10.5194/bg-14-4545-2017

Laruelle, G. G., Cai, W.-J., Hu, X., Gruber, N., Mackenzie, F. T., & Regnier, P. (2018). Continental shelves as a variable but increasing global sink for atmospheric carbon dioxide. *Nature Communications*, *9*(1), 454. http://doi.org/10.1038/s41467-017-02738-z

Latrubesse, E. M., Arima, E. Y., Dunne, T., Park, E., Baker, V. R., d'Horta, F. M., … Stevaux, J. C. (2017). Damming the rivers of the Amazon basin. *Nature*, *546*, 363. Retrieved from http://dx.doi.org/10.1038/nature22333

Laudon, H., & Buffam, I. (2008). Impact of changing DOC concentrations on the potential distribution of acid sensitive biota in a boreal stream network. *Hydrol. Earth Syst. Sci.*, *12*(2), 425–435. http://doi.org/10.5194/hess-12-425-2008

Lauerwald, R., Regnier, P., Camino-Serrano, M., Guenet, B., Guimberteau, M., Ducharne, A., … Ciais, P. (2017). ORCHILEAK (revision 3875): a new model branch to simulate carbon transfers along the terrestrial--aquatic continuum of the Amazon basin. *Geoscientific Model Development*, *10*(10), 3821–3859. http://doi.org/10.5194/gmd-10-3821-2017

Lauerwald, R., Hartmann, J., Moosdorf, N., Kempe, S., & Raymond, P. A. (2013). What controls the spatial patterns of the riverine carbonate system? — A case study for North America. *Chemical Geology*, *337–338*, 114–127. http://doi.org/https://doi.org/10.1016/j.chemgeo.2012.11.011

Lauerwald, R., Laruelle, G. G., Hartmann, J., Ciais, P., & Regnier, P. A. G. (2015). Spatial patterns in CO2 evasion from the global river network. *Global Biogeochemical Cycles*, *29*(5), 534–554. http://doi.org/10.1002/2014GB004941

Lauerwald, R., Regnier, P, Guenet., B, Friedlingstein,.P, Ciais., P: Coupled evolution of riverine carbon fluxes and the terrestrial C sink in the Amazon basin 1861-2099, in submission

Le Quéré, C., Andrew, R. M., Canadell, J. G., Sitch, S., Korsbakken, J. I., Peters, G. P., … Zaehle, S. (2016). Global Carbon Budget 2016. *Earth Syst. Sci. Data*, *8*(2), 605–649. http://doi.org/10.5194/essd-8-605-2016

Le Quéré, C., Andrew, R. M., Friedlingstein, P., Sitch, S., Pongratz, J., Manning, A. C., … Zhu, D. (2018). Global Carbon Budget 2017. *Earth Syst. Sci. Data*, *10*(1), 405–448. http://doi.org/10.5194/essd-10-405-2018

Le Quéré, C., Moriarty, R., Andrew, R. M., Canadell, J. G., Sitch, S., Korsbakken, J. I., … Zeng, N. (2015). Global Carbon Budget 2015. *Earth Syst. Sci. Data*, *7*(2), 349–396. http://doi.org/10.5194/essd-7-349-2015

Le Quéré, C., Andrew, R. M., Friedlingstein, P., Sitch, S., Hauck, J., Pongratz, J., Pickers, P. A., Korsbakken, J. I., Peters, G. P., Canadell, J. G., Arneth, A., Arora, V. K., Barbero, L.,

Bastos, A., Bopp, L., Chevallier, F., Chini, L. P., Ciais, P., Doney, S. C., Gkritzalis, T., Goll, D. S., Harris, I., Haverd, V., Hoffman, F. M., Hoppema, M., Houghton, R. A., Hurtt, G., Ilyina, T., Jain, A. K., Johannessen, T., Jones, C. D., Kato, E., Keeling, R. F., Goldewijk, K. K., Landschützer, P., Lefèvre, N., Lienert, S., Liu, Z., Lombardozzi, D., Metzl, N., Munro, D. R., Nabel, J. E. M. S., Nakaoka, S.-I., Neill, C., Olsen, A., Ono, T., Patra, P., Peregon, A., Peters, W., Peylin, P., Pfeil, B., Pierrot, D., Poulter, B., Rehder, G., Resplandy, L., Robertson, E., Rocher, M., Rödenbeck, C., Schuster, U., Schwinger, J., Séférian, R., Skjelvan, I., Steinhoff, T., Sutton, A., Tans, P. P., Tian, H., Tilbrook, B., Tubiello, F. N., van der Laan-Luijkx, I. T., van der Werf, G. R., Viovy, N., Walker, A. P., Wiltshire, A. J., Wright, R., Zaehle, S., and Zheng, B.: Global Carbon Budget 2018, Earth Syst. Sci. Data, 10, 2141-2194, https://doi.org/10.5194/essd-10-2141-2018, 2018[b].

Le, T. P. Q., Le, N. Da, Dao, V. N., Rochelle-Newall, E., Nguyen, T. M. H., Marchand, C., … Phung, T. X. B. (2018). Change in carbon flux (1960--2015) of the Red River (Vietnam). *Environmental Earth Sciences*, *77*(18), 658. http://doi.org/10.1007/s12665-018-7851-2

Lehner, B., & Döll, P. (2004). Development and validation of a global database of lakes, reservoirs and wetlands. *Journal of Hydrology*, *296*(1–4), 1–22. http://doi.org/https://doi.org/10.1016/j.jhydrol.2004.03.028

Lewis, S. L., Brando, P. M., Phillips, O. L., van der Heijden, G. M. F., & Nepstad, D. (2011). The 2010 Amazon Drought. *Science*, *331*(6017), 554. http://doi.org/10.1126/science.1200807

Lewis, S. L., Lopez-Gonzalez, G., Sonké, B., Affum-Baffoe, K., Baker, T. R., Ojo, L. O., … Wöll, H. (2009). Increasing carbon storage in intact African tropical forests. *Nature*, *457*, 1003. Retrieved from https://doi.org/10.1038/nature07771

Li, M., Peng, C., Zhou, X., Yang, Y., Guo, Y., Shi, G., & Zhu, Q. (n.d.). Modeling Global Riverine DOC Flux Dynamics From 1951 to 2015. *Journal of Advances in Modeling Earth Systems*, *0*(0). http://doi.org/10.1029/2018MS001363

Li, W., Zhang, P., Ye, J., Li, L., & Baker, P. A. (2011). Impact of two different types of El Niño events on the Amazon climate and ecosystem productivity. *Journal of Plant Ecology*, *4*(1–2), 91–99. Retrieved from http://dx.doi.org/10.1093/jpe/rtq039

Liu, H., Zhang, Q., Katul, G. G., Cole, J. J., Chapin, F. S., & MacIntyre, S. (2016). Large {CO}2effluxes at night and during synoptic weather events significantly contribute to {CO}2emissions from a reservoir. *Environmental Research Letters*, *11*(6), 64001. http://doi.org/10.1088/1748-9326/11/6/064001

Ludwig, W., Probst, J.-L., & Kempe, S. (1996). Predicting the oceanic input of organic carbon by continental erosion. *Global Biogeochemical Cycles*, *10*(1), 23–41. http://doi.org/10.1029/95GB02925

Maavara, T., Lauerwald, R., Regnier, P., & Van Cappellen, P. (2017). Global perturbation of organic carbon cycling by river damming, *8*, 15347. Retrieved from http://dx.doi.org/10.1038/ncomms15347

Maberly, S. C., Barker, P. A., Stott, A. W., & De Ville, M. M. (2012). Catchment productivity controls CO2 emissions from lakes. *Nature Climate Change*, *3*, 391. Retrieved from https://doi.org/10.1038/nclimate1748

MacIntyre, S., Jonsson, A., Jansson, M., Aberg, J., Turney, D. E., & Miller, S. D. (2010). Buoyancy flux, turbulence, and the gas transfer coefficient in a stratified lake. *Geophysical Research Letters*, *37*(24), n/a-n/a. http://doi.org/10.1029/2010GL044164

Mackenzie, F. T., Lerman, A. & Ver, L. M. in Geological Perspectives of Global Climate Change. AAPG Studies in Geology Vol. 47 (eds L.C. Gerhard, W.E. Harrison, & B.M. Hanson) 51-82 (American Association of Petroleum Geologists, 2001).

Marengo, J. A., Tomasella, J., Alves, L. M., Soares, W. R., & Rodriguez, D. A. (2011). The drought of 2010 in the context of historical droughts in the Amazon region. *Geophysical Research Letters*, *38*(12). http://doi.org/10.1029/2011GL047436

Marescaux, A., Thieu, V., Borges, A. V., & Garnier, J. (2018). Seasonal and spatial variability of the partial pressure of carbon dioxide in the human-impacted Seine River in France. *Scientific Reports*, *8*(1), 13961. http://doi.org/10.1038/s41598-018-32332-2

Marotta, H., Duarte, C. M., Pinho, L., & Enrich-Prast, A. (2010). Rainfall leads to increased pCO_2 in Brazilian coastal lakes. *Biogeosciences*, *7*(5), 1607–1614. http://doi.org/10.5194/bg-7-1607-2010

Marotta, H., Duarte, C. M., Sobek, S., & Enrich-Prast, A. (2009). Large CO2 disequilibria in tropical lakes. *Global Biogeochemical Cycles*, *23*(4), n/a-n/a. http://doi.org/10.1029/2008GB003434

Masui, T., Matsumoto, K., Hijioka, Y., Kinoshita, T., Nozawa, T., Ishiwatari, S., … Kainuma, M. (2011). An emission pathway for stabilization at 6 Wm−2 radiative forcing. *Climatic Change*, *109*(1), 59. http://doi.org/10.1007/s10584-011-0150-5

Mayorga, E., Seitzinger, S. P., Harrison, J. A., Dumont, E., Beusen, A. H. W., Bouwman, A. F., … Drecht, G. Van. (2010). Global Nutrient Export from WaterSheds 2 (NEWS 2): Model development and implementation. *Environmental Modelling & Software*, *25*(7), 837–853. http://doi.org/https://doi.org/10.1016/j.envsoft.2010.01.007

McDonald, C. P., Rover, J. A., Stets, E. G., & Striegl, R. G. (2012). The regional abundance and size distribution of lakes and reservoirs in the United States and implications for estimates of global lake extent. *Limnology and Oceanography*, *57*(2), 597–606. http://doi.org/10.4319/lo.2012.57.2.0597

McDonald, C. P., Stets, E. G., Striegl, R. G., & Butman, D. (2013). Inorganic carbon loading as a primary driver of dissolved carbon dioxide concentrations in the lakes and reservoirs of the contiguous United States. *Global Biogeochemical Cycles*, *27*(2), 285–295. http://doi.org/10.1002/gbc.20032

Medlyn, B. E. (2011). Comment on {\textquotedblleft}Drought-Induced Reduction in Global Terrestrial Net Primary Production from 2000 Through 2009{\textquotedblright}. *Science*, *333*(6046), 1093. http://doi.org/10.1126/science.1199544

Meinshausen, M., Smith, S. J., Calvin, K., Daniel, J. S., Kainuma, M. L. T., Lamarque, J.-F., … van Vuuren, D. P. P. (2011). The RCP greenhouse gas concentrations and their extensions from 1765 to 2300. *Climatic Change*, *109*(1), 213. http://doi.org/10.1007/s10584-011-0156-z

MELACK, J. M., HESS, L. L., GASTIL, M., FORSBERG, B. R., HAMILTON, S. K., LIMA, I. B. T., & NOVO, E. M. L. M. (2011). LBA-ECO LC-07 Monthly Mean Flooded Wetlands Habitat, Central Amazon Basin: 1979-1996. ORNL Distributed Active Archive Center. http://doi.org/10.3334/ornldaac/1049

Mendonça, R., Müller, R. A., Clow, D., Verpoorter, C., Raymond, P., Tranvik, L. J., & Sobek, S. (2017). Organic carbon burial in global lakes and reservoirs. *Nature Communications*, *8*(1), 1694. http://doi.org/10.1038/s41467-017-01789-6

Meybeck, M., Dürr, H. H., & Vörösmarty, C. J. (2006). Global coastal segmentation and its river catchment contributors: A new look at land-ocean linkage. *Global Biogeochemical Cycles*, *20*(1), n/a-n/a. http://doi.org/10.1029/2005GB002540

Miralles, D. G., Holmes, T. R. H., De Jeu, R. A. M., Gash, J. H., Meesters, A. G. C. A., & Dolman, A. J. (2011). Global land-surface evaporation estimated from satellite-based observations. *Hydrol. Earth Syst. Sci.*, *15*(2), 453–469. http://doi.org/10.5194/hess-15-453-2011

Morales-Pineda, M., Cózar, A., Laiz, I., Úbeda, B., & Gálvez, J. Á. (2014). Daily, biweekly, and seasonal temporal scales of pCO2 variability in two stratified Mediterranean reservoirs. *Journal of Geophysical Research: Biogeosciences*, *119*(4), 509–520. http://doi.org/10.1002/2013JG002317

Moreira-Turcq, P., Seyler, P., Guyot, J. L., & Etcheber, H. (2003). Exportation of organic carbon from the Amazon River and its main tributaries. *Hydrological Processes*, *17*(7), 1329–1344. http://doi.org/10.1002/hyp.1287

Morgan, M. G., Adams, P. J., & Keith, D. W. (2006). Elicitation of Expert Judgments of Aerosol Forcing. *Climatic Change*, *75*(1), 195–214. http://doi.org/10.1007/s10584-005-9025-y

Müller, D., Warneke, T., Rixen, T., Müller, M., Mujahid, A., Bange, H. W., & Notholt, J. (2016). Fate of terrestrial organic carbon and associated CO2 and CO emissions from two Southeast Asian estuaries. *Biogeosciences*, *13*(3), 691–705. http://doi.org/10.5194/bg-13-691-2016

Naipal, V., Ciais, P., Wang, Y., Lauerwald, R., Guenet, B., & Van Oost, K. (2018). Global soil organic carbon removal by water erosion under climate change and land use change during AD 1850–2005. *Biogeosciences*, *15*(14), 4459–4480. http://doi.org/10.5194/bg-15-4459-2018

Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I — A discussion of principles. *Journal of Hydrology*, *10*(3), 282–290. http://doi.org/https://doi.org/10.1016/0022-1694(70)90255-6

Nemani, R. R., Keeling, C. D., Hashimoto, H., Jolly, W. M., Piper, S. C., Tucker, C. J., … Running, S. W. (2003). Climate-Driven Increases in Global Terrestrial Net Primary Production from 1982 to 1999. *Science*, *300*(5625), 1560–1563. http://doi.org/10.1126/science.1082750

Ngo-Duc, T., Polcher, J., & Laval, K. (2005). A 53-year forcing data set for land surface models. *Journal of Geophysical Research: Atmospheres*, *110*(D6). http://doi.org/10.1029/2004JD005434

Nishina, K., Ito, A., Beerling, D. J., Cadule, P., Ciais, P., Clark, D. B., … Yokohata, T. (2014). Quantifying uncertainties in soil carbon responses to changes in global mean temperature and precipitation. *Earth Syst. Dynam.*, *5*(1), 197–209. http://doi.org/10.5194/esd-5-197-2014

Nobre, C. A., Sampaio, G., Borma, L. S., Castilla-Rubio, J. C., Silva, J. S., & Cardoso, M. (2016). Land-use and climate change risks in the Amazon and the need of a novel sustainable development paradigm. *Proceedings of the National Academy of Sciences*, *113*(39), 10759–10768. http://doi.org/10.1073/pnas.1605516113

O'Loughlin, F., Trigg, M. A., Schumann, G. J.-P., & Bates, P. D. (2013). Hydraulic characterization of the middle reach of the Congo River. *Water Resources Research*, *49*(8), 5059–5070. http://doi.org/10.1002/wrcr.20398

Pan, S., Dangal, S. R. S., Tao, B., Yang, J., & Tian, H. (2015). Recent patterns of terrestrial net primary production in Africa influenced by multiple environmental changes. *Ecosystem Health and Sustainability*, *1*(5), 1–15. http://doi.org/10.1890/EHS14-0027.1

Pan, Y., Birdsey, R. A., Fang, J., Houghton, R., Kauppi, P. E., Kurz, W. A., … Hayes, D. (2011). A Large and Persistent Carbon Sink in the World{\textquoteright}s Forests. *Science*, *333*(6045), 988–993. http://doi.org/10.1126/science.1201609

Pangala, S. R., Enrich-Prast, A., Basso, L. S., Peixoto, R. B., Bastviken, D., Hornibrook, E. R. C., … Gauci, V. (2017). Large emissions from floodplain trees close the Amazon methane budget. *Nature*, *552*, 230. Retrieved from https://doi.org/10.1038/nature24639

Pekel, J.-F., Cottam, A., Gorelick, N., & Belward, A. S. (2016). High-resolution mapping of global surface water and its long-term changes. *Nature*, *540*, 418. Retrieved from https://doi.org/10.1038/nature20584

Perga, M.-E., Maberly, S. C., Jenny, J.-P., Alric, B., Pignol, C., & Naffrechoux, E. (2016). A century of human-driven changes in the carbon dioxide concentration of lakes. *Global Biogeochemical Cycles*, *30*(2), 93–104. http://doi.org/10.1002/2015GB005286

Peter, H., Singer, G. A., Preiler, C., Chifflard, P., Steniczka, G., & Battin, T. J. (2014). Scales and drivers of temporal pCO2 dynamics in an Alpine stream. *Journal of Geophysical Research: Biogeosciences*, *119*(6), 1078–1091. http://doi.org/10.1002/2013JG002552

Phillips, O. L., Aragão, L. E. O. C., Lewis, S. L., Fisher, J. B., Lloyd, J., López-González, G., … Torres-Lezama, A. (2009). Drought Sensitivity of the Amazon Rainforest. *Science*, *323*(5919), 1344–1347. http://doi.org/10.1126/science.1164033

Potapov, P., Hansen, M. C., Stehman, S. V, Loveland, T. R., & Pittman, K. (2008). Combining {MODIS} and Landsat imagery to estimate and map boreal forest cover loss. *Remote Sensing of Environment*, *112*(9), 3708–3719. http://doi.org/https://doi.org/10.1016/j.rse.2008.05.006

Potter, C., Klooster, S., & Genovese, V. (2012). Net primary production of terrestrial ecosystems from 2000 to 2009. *Climatic Change*, *115*(2), 365–378. http://doi.org/10.1007/s10584-012-0460-2

Potter, C., Klooster, S., Hiatt, C., Genovese, V., & Castilla-Rubio, J. C. (2011). Changes in the carbon cycle of Amazon ecosystems during the 2010 drought. *Environmental Research Letters*, *6*(3), 34024. http://doi.org/10.1088/1748-9326/6/3/034024

Prentice, I. C., & Lloyd, J. (1998). C-quest in the Amazon Basin. *Nature*, *396*, 619. Retrieved from https://doi.org/10.1038/25224

Price, D. T., McKenney, D. W., Joyce, L. A., Siltanen, R. M., Papadopol, P., & Lawrence, K. (2011). High-resolution interpolation of climate scenarios for canada derived from general circulation model simulations. Nat. Resour. Can., Can. For. Serv., North. For. Cent., Edmonton, AB.

Prigent, C., Papa, F., Aires, F., Rossow, W. B., & Matthews, E. (2007). Global inundation dynamics inferred from multiple satellite observations, 1993–2000. *Journal of Geophysical Research: Atmospheres*, *112*(D12). http://doi.org/10.1029/2006JD007847

Qiu, C., Zhu, D., Ciais, P., Guenet, B., Peng, S., Krinner, G., … Hastie, A. (2018). Modelling northern peatlands area and carbon dynamics since the Holocene with the ORCHIDEE-PEAT land surface model (SVN r5488). *Geosci. Model Dev. Discuss.*, *2018*, 1–45. https://doi.org/10.5194/gmd-2018-256

R Core Team. (2013). R: A language and environment for statistical computing. [Available at http://www.r-project.org.]

Rasera, M. F. F. L., Krusche, A. V., Richey, J. E., Ballester, M. V. R., and Victória, R. L. (2013). Spatial and temporal variability of pCO2 and CO2 efflux in seven Amazonian Rivers. *Biogeochemistry*, *116*(1), 241–259. https://doi.org/10.1007/s10533-013-9854-0

Rantakari, M., & Kortelainen, P. (2005). Interannual variation and climatic regulation of the CO2 emission from large boreal lakes. *Global Change Biology*, *11*(8), 1368–1380. http://doi.org/10.1111/j.1365-2486.2005.00982.x

Rasilo, T., Prairie, Y. T., & del Giorgio, P. A. (2015). Large-scale patterns in summer diffusive CH4 fluxes across boreal lakes, and contribution to diffusive C emissions. *Global Change Biology*, *21*(3), 1124–1139. http://doi.org/10.1111/gcb.12741

Raymond, P. A., Hartmann, J., Lauerwald, R., Sobek, S., McDonald, C., Hoover, M., … Guth, P. (2013). Global carbon dioxide emissions from inland waters. *Nature*, *503*(7476), 355–359. Retrieved from http://dx.doi.org/10.1038/nature12760

Read, J. S., Hamilton, D. P., Desai, A. R., Rose, K. C., MacIntyre, S., Lenters, J. D., … Wu, C. H. (2012). Lake-size dependency of wind shear and convection as controls on gas exchange. *Geophysical Research Letters*, *39*(9), n/a-n/a. http://doi.org/10.1029/2012GL051886

Regnier, P., Friedlingstein, P., Ciais, P., Mackenzie, F. T., Gruber, N., Janssens, I. A., … Thullner, M. (2013). Anthropogenic perturbation of the carbon fluxes from land to ocean. *Nature Geosci*, *6*(8), 597–607. Retrieved from http://dx.doi.org/10.1038/ngeo1830

Reiman, J. H., & Xu, Y. J. (2018). Diel Variability of pCO2 and CO2 Outgassing from the Lower Mississippi River: Implications for Riverine CO2 Outgassing Estimation. *Water*, *11*(1). http://doi.org/10.3390/w11010043

Ren, W., Tian, H., Cai, W.-J., Lohrenz, S. E., Hopkinson, C. S., Huang, W.-J., … He, R. (2016). Century-long increasing trend and variability of dissolved organic carbon export from the Mississippi River basin driven by natural and anthropogenic forcing. *Global Biogeochemical Cycles*, *30*(9), 1288–1299. http://doi.org/10.1002/2016GB005395

Resplandy, L., Keeling, R. F., Rödenbeck, C., Stephens, B. B., Khatiwala, S., Rodgers, K. B., … Tans, P. P. (2018). Revision of global carbon fluxes based on a reassessment of oceanic and riverine carbon transport. *Nature Geoscience*, *11*(7), 504–509. http://doi.org/10.1038/s41561-018-0151-3

Reynolds, C. A., Jackson, T. J., & Rawls, W. J. (2000). Estimating soil water-holding capacities by linking the Food and Agriculture Organization Soil map of the world with global pedon databases and continuous pedotransfer functions. *Water Resources Research*, *36*(12), 3653–3662. http://doi.org/10.1029/2000WR900130

RICHEY, J. E., VICTORIA, R. L., HEDGES, J. I., DUNNE, T., MARTINELLI, L. A., MERTES, L., & ADAMS, J. (2008). Pre-LBA Carbon in the Amazon River Experiment (CAMREX) Data. ORNL Distributed Active Archive Center. http://doi.org/10.3334/ornldaac/904

Richey, J. E., Hedges, J. I., Devol, A. H., Quay, P. D., Victoria, R., Martinelli, L., & Forsberg, B. R. (1990). Biogeochemistry of carbon in the Amazon River. *Limnology and Oceanography*, *35*(2), 352–371. http://doi.org/10.4319/lo.1990.35.2.0352

Richey, J. E., Melack, J. M., Aufdenkampe, A. K., Ballester, V. M., & Hess, L. L. (2002). Outgassing from Amazonian rivers and wetlands as a large tropical source of atmospheric CO2. *Nature*, *416*, 617. Retrieved from http://dx.doi.org/10.1038/416617a

Rödig E, Cuntz M, Rammig A, Fischer R, Taubert F and Huth A. (2018). The importance of forest structure for carbon fluxes of the Amazon rainforest. *Environmental Research Letters*, *13*(5), 54013. Retrieved from http://stacks.iop.org/1748-9326/13/i=5/a=054013

Roehm, C. L., Prairie, Y. T., & del Giorgio, P. A. (2009). The pCO2 dynamics in lakes in the boreal region of northern Québec, Canada. *Global Biogeochemical Cycles*, *23*(3), n/a-n/a. http://doi.org/10.1029/2008GB003297

Rueda-Delgado, G., Wantzen, K. M., and Tolosa, M. B.: Leaflitter decomposition in an Amazonian floodplain stream: Effects of seasonal hydrological changes, J. North Am. Benthol. Soc., 25, 233–249, https://doi.org/10.1899/0887 3593(2006)25[233:LDIAAF]2.0.CO;2, 2006.

Sanders, L. M., Taffs, K. H., Stokes, D. J., Sanders, C. J., Smoak, J. M., Enrich-Prast, A., … Marotta, H. (2017). Carbon accumulation in Amazonian floodplain lakes: A significant component of Amazon budgets? *Limnology and Oceanography Letters*, *2*(1), 29–35. http://doi.org/10.1002/lol2.10034

Santini, Z. Z. and W. B. and B. M. and C. V. R. and P. W. and W. L.-C. and J.-L. G. and W. (2016). Projected increases in the annual flood pulse of the Western Amazon. *Environmental Research Letters*, *11*(1), 14013. Retrieved from http://stacks.iop.org/1748-9326/11/i=1/a=014013

Sawakuchi, H. O., Neu, V., Ward, N. D., Barros, M. de L. C., Valerio, A. M., Gagne-Maynard, W., … Richey, J. E. (2017). Carbon Dioxide Emissions along the Lower Amazon River. *Frontiers in Marine Science*, *4*, 76. http://doi.org/10.3389/fmars.2017.00076

Schimel, D., Stephens, B. B., & Fisher, J. B. (2015). Effect of increasing CO2 on the terrestrial carbon cycle. *Proceedings of the National Academy of Sciences*, *112*(2), 436–441. http://doi.org/10.1073/pnas.1407302112

Sheffield, J., Goteti, G., & Wood, E. F. (2006). Development of a 50-Year High-Resolution Global Dataset of Meteorological Forcings for Land Surface Modeling. *Journal of Climate*, *19*(13), 3088–3111. http://doi.org/10.1175/JCLI3790.1

Shirokova, L. S., Pokrovsky, O. S., Kirpotin, S. N., Desmukh, C., Pokrovsky, B. G., Audry, S., & Viers, J. (2013). Biogeochemistry of organic carbon, CO2, CH4, and trace elements in thermokarst water bodies in discontinuous permafrost zones of Western Siberia. *Biogeochemistry*, *113*(1), 573–593. http://doi.org/10.1007/s10533-012-9790-4

SOBEK, S., ALGESTEN, G., BERGSTRÖM, A.-K., JANSSON, M., & TRANVIK, L. J. (2003). The catchment and climate regulation of pCO2 in boreal lakes. *Global Change Biology*, *9*(4), 630–641. http://doi.org/10.1046/j.1365-2486.2003.00619.x

Sobek, S., Tranvik, L. J., & Cole, J. J. (2005). Temperature independence of carbon dioxide supersaturation in global lakes. *Global Biogeochemical Cycles*, *19*(2), n/a-n/a. http://doi.org/10.1029/2004GB002264

Sobek, S., Tranvik, L. J., Prairie, Y. T., Kortelainen, P., & Cole, J. J. (2007). Patterns and regulation of dissolved organic carbon: An analysis of 7,500 widely distributed lakes. *Limnology and Oceanography*, *52*(3), 1208–1219. http://doi.org/10.4319/lo.2007.52.3.1208

SoilGrids. (2014). http://soilgrids.org/index.html

Spafford, L., & Risk, D. (2018). Spatiotemporal Variability in Lake-Atmosphere Net CO2 Exchange in the Littoral Zone of an Oligotrophic Lake. *Journal of Geophysical Research: Biogeosciences*, *123*(4), 1260–1276. http://doi.org/10.1002/2017JG004115

Spencer, R. G. M., Hernes, P. J., Dinga, B., Wabakanghanzi, J. N., Drake, T. W., & Six, J. (2016). Origins, seasonality, and fluxes of organic matter in the Congo River. *Global Biogeochemical Cycles*, *30*(7), 1105–1121. http://doi.org/10.1002/2016GB005427

Striegl, R. G., Kortelainen, P., Chanton, J. P., Wickland, K. P., Bugna, G. C., & Rantakari, M. (2001). Carbon dioxide partial pressure and 13C content of north temperate and boreal lakes at spring ice melt. *Limnology and Oceanography*, *46*(4), 941–945. http://doi.org/10.4319/lo.2001.46.4.0941

Suchet, P. A., & Probst, J. L. (1995). A global model for present-day atmospheric/soil CO2 consumption by chemical erosion of continental rocks (GEM-CO2). *Tellus B*, *47*(1-2), 273–280. http://doi.org/10.1034/j.1600-0889.47.issue1.23.x

Sullivan, M. J. P., Talbot, J., Lewis, S. L., Phillips, O. L., Qie, L., Begne, S. K., … Zemagho, L. (2017). Diversity and carbon storage across the tropical forest biome. *Scientific Reports*, *7*, 39102. Retrieved from https://doi.org/10.1038/srep39102

T., N., J., P., & K., L. (2005). A 53-year forcing data set for land surface models. *Journal of Geophysical Research: Atmospheres*, *110*(D6). http://doi.org/10.1029/2004JD005434

Tathy, J. P., Cros, B., Delmas, R. A., Marenco, A., Servant, J., & Labat, M. (1992). Methane emission from flooded forest in central Africa. *Journal of Geophysical Research: Atmospheres*, *97*(D6), 6159–6168. http://doi.org/10.1029/90JD02555

Telmer, K., & Veizer, J. (1999). Carbon fluxes, pCO2 and substrate weathering in a large northern river basin, Canada: carbon isotope perspectives. *Chemical Geology*, *159*(1–4), 61–86. http://doi.org/https://doi.org/10.1016/S0009-2541(99)00034-0

Tian, H., Melillo, J. M., Kicklighter, D. W., McGuire, A. D., Helfrich III, J. V. K., Moore III, B., & Vörösmarty, C. J. (1998). Effect of interannual climate variability on carbon storage in Amazonian ecosystems. *Nature*, *396*, 664. Retrieved from https://doi.org/10.1038/25328

Tian, H., Yang, Q., Najjar, R. G., Ren, W., Friedrichs, M. A. M., Hopkinson, C. S., & Pan, S. (2015). Anthropogenic and climatic influences on carbon fluxes from eastern North America to the Atlantic Ocean: A process-based modeling study. *Journal of Geophysical Research: Biogeosciences*, *120*(4), 757–772. http://doi.org/10.1002/2014JG002760

Tootchi, A., Jost, A., & Ducharne, A. (2019). Multi-source global wetland maps combining surface water imagery and groundwater constraints. *Earth Syst. Sci. Data*, *11*(1), 189–220. http://doi.org/10.5194/essd-11-189-2019

Tranvik, L. J., Downing, J. A., Cotner, J. B., Loiselle, S. A., Striegl, R. G., Ballatore, T. J., … Weyhenmeyer, G. A. (2009). Lakes and reservoirs as regulators of carbon cycling and climate. *Limnology and Oceanography*, *54*(6part2), 2298–2314. http://doi.org/10.4319/lo.2009.54.6_part_2.2298

Treat, C. C., Kleinen, T., Broothaerts, N., Dalton, A. S., Dommain, R., Douglas, T. A., … Brovkin, V. (2019). Widespread global peatland establishment and persistence over the last 130,000 y. *Proceedings of the National Academy of Sciences*, 201813305. http://doi.org/10.1073/PNAS.1813305116

University, C. for I. E. S. I. N.-C.-C., FAO, U. N. F. and A. P.-, & CIAT, C. I. de A. T.-. (2005). Gridded Population of the World, Version 3 (GPWv3): Population Count Grid. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). Retrieved from http://dx.doi.org/10.7927/H4639MPP

Vachon, D., & Prairie, Y. T. (2013). The ecosystem size and shape dependence of gas transfer velocity versus wind speed relationships in lakes. Canadian Journal of Fisheries and Aquatic Sciences, 70(12), 1757–1764. https://doi.org/10.1139/cjfas-2013-0241

Valentini, R., Arneth, A., Bombelli, A., Castaldi, S., Cazzolla Gatti, R., Chevallier, F., … Scholes, R. J. (2014). A full greenhouse gases budget of Africa: synthesis, uncertainties, and vulnerabilities. *Biogeosciences*, *11*(2), 381–407. http://doi.org/10.5194/bg-11-381-2014

van der Laan-Luijkx, I. T., van der Velde, I. R., Krol, M. C., Gatti, L. V, Domingues, L. G., Correia, C. S. C., … Peters, W. (2015). Response of the Amazon carbon balance to the 2010 drought derived with CarbonTracker South America. *Global Biogeochemical Cycles*, *29*(7), 1092–1108. http://doi.org/10.1002/2014GB005082

van der Werf, G. R., Randerson, J. T., Giglio, L., van Leeuwen, T. T., Chen, Y., Rogers, B. M., … Kasibhatla, P. S. (2017). Global fire emissions estimates during 1997–2015. *Earth Syst. Sci. Data Discuss.*, *2017*, 1–43. http://doi.org/10.5194/essd-2016-62

van Vuuren, D. P., Edmonds, J., Kainuma, M., Riahi, K., Thomson, A., Hibbard, K., … Rose, S. K. (2011). The representative concentration pathways: an overview. *Climatic Change*, *109*(1), 5. https://doi.org/10.1007/s10584-011-0148-z

Ver, L. M. B., Mackenzie, F. T. & Lerman, A. Biogeochemical responses of the carbon cycle to natural and human perturbations: Past, present, and future. Am. J. Sci. 299, 762-801, (1999).

Verhegghen, A., Mayaux, P., de Wasseige, C., & Defourny, P. (2012). Mapping Congo Basin vegetation types from 300 m and 1 km multi-sensor time series for carbon stocks and forest areas estimation. *Biogeosciences*, *9*(12), 5061–5079. http://doi.org/10.5194/bg-9-5061-2012

Verpoorter, C., Kutser, T., Seekell, D. A., & Tranvik, L. J. (2014). A global inventory of lakes based on high-resolution satellite imagery. *Geophysical Research Letters*, *41*(18), 6396–6402. http://doi.org/10.1002/2014GL060641

Verpoorter, C., Kutser, T., & Tranvik, L. (2012). Automated mapping of water bodies using Landsat multispectral data. *Limnology and Oceanography: Methods*, *10*(12), 1037–1050. http://doi.org/10.4319/lom.2012.10.1037

Viovy, N. (2018). CRUNCEP Version 7 - Atmospheric Forcing Data for the Community Land Model. Boulder, CO: Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory. Retrieved from http://rda.ucar.edu/datasets/ds314.3/

Vörösmarty, C. J., Fekete, B. M., Meybeck, M., & Lammers, R. B. (2000). Geomorphometric attributes of the global system of rivers at 30-minute spatial resolution. *Journal of Hydrology*, *237*(1), 17–39. http://doi.org/https://doi.org/10.1016/S0022-1694(00)00282-1

Wallin, M. B., Löfgren, S., Erlandsson, M., & Bishop, K. (2014). Representative regional sampling of carbon dioxide and methane concentrations in hemiboreal headwater streams reveal underestimates in less systematic approaches. *Global Biogeochemical Cycles*, *28*(4), 465–479. http://doi.org/10.1002/2013GB004715

Wang, W., Ciais, P., Nemani, R. R., Canadell, J. G., Piao, S., Sitch, S., … Myneni, R. B. (2013). Variations in atmospheric CO2 growth rates coupled with tropical temperature. *Proceedings of the National Academy of Sciences*, *110*(32), 13061–13066. http://doi.org/10.1073/pnas.1219683110

Wang, X., Ciais, P., Wang, Y., & Zhu, D. (2018). Divergent response of seasonally dry tropical vegetation to climatic variations in dry and wet seasons. *Global Change Biology*, *24*(10), 4709–4717. http://doi.org/10.1111/gcb.14335

Weyhenmeyer, G. A., Kortelainen, P., Sobek, S., Müller, R., & Rantakari, M. (2012). Carbon Dioxide in Boreal Surface Waters: A Comparison of Lakes and Streams. *Ecosystems*, *15*(8), 1295–1307. Retrieved from http://www.jstor.org/stable/23325685

Weyhenmeyer, G. A., Kosten, S., Wallin, M. B., Tranvik, L. J., Jeppesen, E., & Roland, F. (2015). Significant fraction of CO2 emissions from boreal lakes derived from hydrologic inorganic carbon inputs. *Nature Geosci*, *8*(12), 933–936. Retrieved from http://dx.doi.org/10.1038/ngeo2582

Wilkinson, G. M., Buelo, C. D., Cole, J. J., & Pace, M. L. (2016). Exogenously produced CO2 doubles the CO2 efflux from three north temperate lakes. *Geophysical Research Letters*, *43*(5), 1996–2003. http://doi.org/10.1002/2016GL067732

Williams, C. A., Hanan, N. P., Neff, J. C., Scholes, R. J., Berry, J. A., Denning, A. S., & Baker, D. F. (2007). Africa and the global carbon cycle. *Carbon Balance and Management*, *2*, 3. http://doi.org/10.1186/1750-0680-2-3

Wolter, K., & Timlin, M. S. (2011). El Niño/Southern Oscillation behaviour since 1871 as diagnosed in an extended multivariate ENSO index (MEI.ext). *International Journal of Climatology*, *31*(7), 1074–1087. http://doi.org/10.1002/joc.2336

Xu, Y. J., Xu, Z., & Yang, R. (2019). Rapid daily change in surface water pCO2 and CO2 evasion: A case study in a subtropical eutrophic lake in Southern USA. *Journal of Hydrology*, *570*, 486–494. https://doi.org/https://doi.org/10.1016/j.jhydrol.2019.01.016

Yin, S., Li, X., & Wu, W. (2017). Comparative analysis of NPP changes in global tropical forests from 2001 to 2013. *IOP Conference Series: Earth and Environmental Science*, *57*(1), 12009. Retrieved from http://stacks.iop.org/1755-1315/57/i=1/a=012009

Zarfl, C., Lumsdon, A. E., Berlekamp, J., Tydecks, L., & Tockner, K. (2015). A global boom in hydropower dam construction. *Aquatic Sciences*, *77*(1), 161–170. http://doi.org/10.1007/s00027-014-0377-0

Zhang, Z., Zimmermann, N. E., Stenke, A., Li, X., Hodson, E. L., Zhu, G., … Poulter, B. (2017). Emerging role of wetland methane emissions in driving 21st century climate change. *Proceedings of the National Academy of Sciences*, *114*(36), 9647–9652. http://doi.org/10.1073/pnas.1618765114

Zhao, M., Heinsch, F. A., Nemani, R. R., & Running, S. W. (2005). Improvements of the MODIS terrestrial gross and net primary production global data set. *Remote Sensing of Environment*, *95*(2), 164–176. http://doi.org/https://doi.org/10.1016/j.rse.2004.12.011

Zhao, M., & Running, S. W. (2010). Drought-Induced Reduction in Global Terrestrial Net Primary Production from 2000 Through 2009. *Science*, *329*(5994), 940–943. http://doi.org/10.1126/science.1192666

Zscheischler, J., Mahecha, M. D., Avitabile, V., Calle, L., Carvalhais, N., Ciais, P., … Reichstein, M. (2017). Reviews and syntheses: An empirical spatiotemporal description of the global surface–atmosphere carbon fluxes: opportunities and data limitations. *Biogeosciences*, *14*(15), 3685–3703. http://doi.org/10.5194/bg-14-3685-2017

Zulkafli, Z., Buytaert, W., Manz, B., Rosas, C. V., Willems, P., Lavado-Casimiro, W., Guyot, J.-L., and Santini, W.: Projected increases in the annual flood pulse of the Western Amazon, Environ. Res. Lett., 11, 14013, doi:10.1088/1748-9326/11/1/014013, 2016.

7. Acknowledgements

I would first like to thank Pierre Regnier and Ronny Lauerwald for their guidance, advice and patience throughout my PhD. I have learned so much from both of you over the last few years, not least how much hard work and persistence is really required in academia! Hopefully we can continue to collaborate in the coming years, if not in academia then certainly in whisky tasting! I would also like to thank Gesa Weyhenmeyer, Sebastian Sobek, Philippe Ciais and Charles Verpoorter for their advice and extremely helpful comments. Thank you to all of my fellow ESRS, number 1 right through to number 15, what a ride it has been! (TOTEM…TOTEM…TOTEM….). Thank you to all of my colleagues in Brussels for your good humour and interesting conversations in the canteen; Christel van Eck, Steeve Bonneville, Philip Pika, Goulven Laruelle, Alizée Roobaert, Jean-Philippe Blouet, Nathalie Roevros, Sandra Arndt and Céline Gommet. Special thank you also to Emily Mainetti, the heart and soul of the C-CASCADES project and without whom we would all have been lost! Biggest thank you to my amazing wife Marketa who kept me sane throughout and has the patience and heart of a saint. Thank you also to my incredible family for all their support. Finally, I would like to thank the glorious golden city of Brussels, perhaps not the prettiest city in the world but certainly one of the most interesting, diverse and stimulating!

8. Appendices

8.1. Methods section of Lauerwald et al. (2017)

8.1.1. Model developments

ORCHILEAK is based on the recent model branch ORCHIDEE-SOM (Camino Serrano, 2015) which relies on a novel module representing the vertical distribution of soil organic carbon (SOC) and associated transport and reaction processes. These processes include the production, consumption, adsorption/desorption and transport of DOC within the soil column as well as DOC exports from the soil column by drainage and surface runoff. In this study, the module is upgraded to represent DOC cycling in tropical rain forests, in particular by adding fluxes of DOC from the atmosphere and canopy with throughfall and by distinguishing soil carbon processes on non-flooded and flooded soils, including the direct input of DOC and CO_2 from the decomposition of submerged litter and soil carbon to the water column. The trunk-version of ORCHIDEE, as well as the branch ORCHIDEE-SOM, includes a river routing module (Guimberteau et al., 2012; Polcher, 2003) that simulates the lateral transfer of water from one grid to another, representing the river channel as well as connected wetlands. Here, this routing module has been upgraded with a tracer transport equation to simulate the fluxes of DOC and CO_2 along the fluvial network, distinguishing two pools of DOC, labile and refractory DOC. In addition, the representation of the floodplain dynamics is improved in this study to better reproduce the seasonal flooding in the Amazon basin, which is a major controlling factor of the water (Guimberteau et al., 2012) and carbon flow dynamics along the river network (Richey et al., 1990). ORCHIDEE can be run at different spatial and temporal resolutions. Here, in line with Guimberteau et al. (2012), the model runs for calibration and model testing were performed at 1° spatial resolution over the period 1980-2000, using the regional climate and wetland forcing for the Amazon from Guimberteau et al. (2012), forcing of land cover and land use change after Belward et al. (1999), Olson et al. (1983) and Hurtt et al. (2006), river flow directions from Vörösmarty et al. (2000), as well as soil parameters after Reynolds et al. (1999) and the Harmonized World Soil Data base (FAO/IIASA/ISRIC/ISS-CAS/JRC, 2009) compiled by Guenet et al., in prep. The necessary forcing data are listed in table 1. As temporal resolution, we use the default 30-minute time-step for all vertical exchanges of water, carbon and energy between atmosphere, vegetation, and soils, and the default 1-day time step for the lateral routing of water. In the following, the model description will be based on these spatial and temporal resolutions. To obtain initial soil carbon pools which are in steady-state with the

model set-up for the 1980-2000 period, the model was first run for 5000 years, looping over the full set of climate forgings and using the land use and an atmospheric pCO_2 as representative for the year 1980. The terrestrial C pools simulated during this initialization phase were subsequently used for the simulation over the period 1980-2000 with changing land cover and increasing atmospheric pCO_2.This section starts with the representation of the soil hydrology and the river routing scheme in ORCHIDEE and ORCHILEAK (section 2.1). Here, we give an overview of the features that are shared between the original version of ORCHIDEE (the configuration used by Guimberteau et al., 2012) and ORCHILEAK and we then highlight the improvements that have been implemented in ORCHILEAK. In the second part, the mathematical formulation of DOC production and leaching from the soil as well as transport and transformation of DOC and CO_2 along the fluvial network is described (section 2.2).

8.1.1.1. Hydrology

Like most land surface schemes of ESMs, ORCHIDEE distinguishes two kinds of surface hydrology processes: (i) the water budget processes, which are mostly vertical and control the partitioning of precipitation into evapotranspiration, infiltration, production of surface runoff and drainage (section 2.1.1); (ii) the horizontal transfer, or routing, of grid-based simulated surface runoff and drainage along the river network (section 2.1.2, with improvements described in 2.1.3).

Water budget and soil hydrology

In the vegetation canopy, rainfall is partitioned between interception loss and throughfall according to the leaf area index (LAI). The throughfall (possibly increased by snowmelt in cold climates and by return flow from the floodplains, cf. section 2.1.2) is then further subdivided into infiltration into the soil and surface runoff produced by infiltration excess. In ORCHIDEE, the infiltration rate depends on precipitation rates, local slope, and vegetation and is limited by the hydraulic conductivity of the soil, which defines a Hortonian surface runoff (D'Orgeval et al. 2008). The corresponding parameterization is tightly linked to the soil moisture redistribution scheme, which is ruled by the Richards equation, solved here over a 2 m soil profile, using an 11-layer discretization, with layers of geometrically increasing depth (de Rosnay et al. 2002; Campoy et al. 2013). The redistribution of soil moisture is controlled by the soil hydraulic properties, transpiration and evaporation within the soil column, and a gravitational drainage at the soil bottom. All these processes are simulated at a 30 min time

step and a 1° resolution. In addition, a bottom return flow feeding the soil is also accounted for in presence of swamps, simulated at the daily time-step of the routing scheme (section 2.1.2).

Table 1. List of forcing data needed to run ORCHILEAK. See text for explanations and Fig. 6 for an overview.

Variable	Spatial resolution	Temporal resolution	Data source
		Forcing data	
Rainfall	1°	6 hours	Guimberteau et al. (2012), replaced original NCC data
Snowfall	1°	6 hours	NCC (Ngo-Duc et al., 2005)
Air Temperature (close to surface)	1°	6 hours	NCC (Ngo-Duc et al., 2005)
Incoming shortwave radiation	1°	6 hours	NCC (Ngo-Duc et al., 2005)
Incoming longwave radiation	1°	6 hours	NCC (Ngo-Duc et al., 2005)
Air pressure (close to surface)	1°	6 hours	NCC (Ngo-Duc et al., 2005)
Wind speed (10 m above surface)	1°	6 hours	NCC (Ngo-Duc et al., 2005)
Relative humidity (close to surface)	1°	6 hours	NCC (Ngo-Duc et al., 2005)
Soil texture class	0.5°	-	Reynolds et al. (1999)
Soil pH	0.5°	-	after HWSD v 1.1 (FAO et al., 2009)
Soil bulk density	0.5°	-	after HWSD v 1.1 (FAO et al., 2009)
Poor soils	0.5°	-	This study after HWSD v 1.1 (FAO et al., 2009)
Land cover (and change)	0.5°	annual	after Belward et al. (1999), Olson et al. (1983) and Hurtt et al. (2006)
Stream flow directions	0.5°	-	STN-30p (Vorosmarty et al., 2000)
Topographic index ($Topo_{grid\ x}$)	0.5°	-	STN-30p (Vorosmarty et al., 2000)
Floodplains ($\%flood_{max}$)	0.5°	-	After Guimberteau et al. (2012)
Swamps ($\%swamp$)	0.5°	-	After Guimberteau et al. (2012)
River surface areas (A_{river})	0.5°	-	Lauerwald et al. (2015)
10^{th}, 50^{th}, 90^{th} percentile of the stream reservoir	1°	-	derived from pre-runs with ORCHIDEE (see text)
95^{th} percentile of water table height over flood plain	1°	-	derived from pre-runs with ORCHIDEE (see text)

Routing of water along the river network, floodplains and swamps

The river routing module simulates the water exports from the soil column as river discharge along a distributed routing scheme, and it is possible to simulate lateral flows at a higher spatial resolution than the rest of the model to better describe the borders of watersheds within each grid-box and the directions of incoming / outgoing water from distinct basins (Fig.2). For that, each ORCHIDEE grid cell x is divided into multiple subunits named “basins”. As in our case, we run simulations at 1° resolution and use a routing scheme at 0.5° resolution (Vörösmarty et al., 2000), each grid cell is simply subdivided into four basins (Fig. 2). Note that all information derived from the forcing files or computed in the other modules has the resolution of the grid cell and is then downscaled to the basins within the routing module. In the following, variables

at grid scale are denoted by the index 'grid x', while information at basin scale are denoted by the index '*i*'. For a full overview of the variables and the system of indices used here, consult Table A.1 in the appendix.

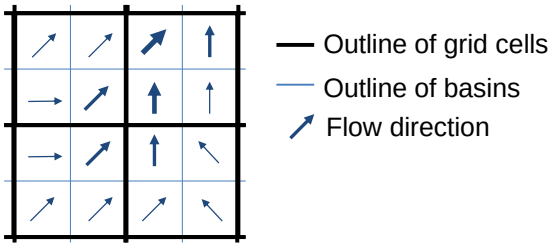


Figure 2. Schematic representation of 4 ORCHIDEE grids *x* at 1 degree spatial resolution for a simulation using a river routing scheme running at 0.5-degree resolution.

The river routing aggregates the 30' surface runoff and drainage computed by the soil hydrology module to the daily time step *t* of this module. As shown in Fig. 3, surface runoff and drainage initially feed a 'fast' ($S_{fast,H2O}$) and a 'slow' ($S_{slow,H2O}$) water reservoir, respectively (Eqs. 1,2). The proportions of runoff ($F_{RO,H2O,grid\ x,t}$) and drainage ($F_{DR,H2O,grid\ x,t}$) assigned to each basin *i* within the grid *x* are scaled to the area of the basin ($A_{total,i}$) relative to that of the grid cell ($A_{total,grid\ x.}$). S_{fast} and S_{slow} have distinct linear response time scales in each basin of the simulation domain, which are defined by a topographic index $Topo_{grid\ x}$ extracted from a forcing file (values range between 1 and 4 in our study area) and a factor τ which translates $Topo_{grid\ x}$ into a water residence time of each reservoir (Eqs. 3,4). Following the calibration of Guimberteau et al. (2012), both τ_{fast} and τ_{slow} are set to a value of 3.0 days. The river reservoir (S_{river}) in each basin *i* is mainly fed by the outflows of S_{fast}, S_{slow}, and S_{river} of the basins *i-1* lying immediately upstream (Eqs. 5,6,7), but can, in addition, interact with two kinds of hydraulic sub-systems, the floodplains and the swamps, the maximum extent of which are defined by forcing files. Swamps are intended to mimic ground water fed wetlands. Where swamps are present, a constant fraction of the upstream inflow F_{up} (Eq. 7), which is scaled to the areal proportion of swamps (*%swamp*) in a given basin *i*, is diverted from the S_{river} and added to the bottom of the soil column of the *grid x* containing the basin *i* ($F_{up2swamp}$, Eq. 8). Contrarily to the floodplains, the swamps are not represented by an explicit water body (S_{flood}). In the original version of ORCHIDEE, if floodplains are present, all the water coming from upstream not diverted to swamps is first directed to the floodplains ($F_{up2flood}$, Eqs. 9,10, see section 2.1.3 for an improved representation). S_{flood} then sustains a delayed return flow ($F_{flood\ out,H2O}$) to the river reservoir of the same basin *i* (Eqs. 11,12). The water balance of the S_{flood} is in addition controlled by input from throughfall ($F_{WD,H2O}$), evaporation ($F_{flood2atm,H2O}$) , or infiltration into

the soil ($F_{flood2soil,H2O}$) in the floodplain (Eq. 11), depending on the temporarily changing areal extent of the inundation *%flood*. The values of τ_{river} and τ_{flood} used by *Guimberteau et al.* (2012) are 0.24 days and 2.5 days, respectively. Note that both $F_{flood\ out}$ (Eq. 12) and $F_{river\ out}$ (Eq. 6) are dependent on %flood as well. For further details see the publications of *d'Orgeval et al.* (2008) and *Guimberteau et al.* (2012).

$$S_{fast,H_2O,i,t+1} = S_{fast,H_2O,i,t} + F_{RO,H_2O,grid\ x,t} \cdot \frac{A_{total,i}}{A_{total,grid\ x}} - F_{fast\ out,H_2O,i,t} \quad (1)$$

$$S_{slow,H_2O,i,t+1} = S_{slow,H_2O,i,t} + F_{DR,H_2O,grid\ x,t} \cdot \frac{A_{total,i}}{A_{total,grid\ x}} - F_{slow\ out,H_2O,i,t} \quad (2)$$

$$F_{fast\ out,H_2O,i,t} = \frac{S_{fast,H_2O,i,t}}{\tau_{fast} \cdot Topo_{grid\ x}} \quad (3)$$

$$F_{slow\ out,H_2O,i,t} = \frac{S_{slow,H_2O,i,t}}{\tau_{slow} \cdot Topo_{grid\ x}} \quad (4)$$

$$S_{river,H_2O,i,t+1} = S_{river,H_2O,i,t} + F_{up2river,H_2O,i,t} + F_{flood\ out,H_2O,i,t} - F_{river\ out,H_2O,i,t} \quad (5)$$

$$F_{river\ out,H_2O,i,t} = \frac{S_{river,H_2O,i,t}}{\tau_{river} \cdot Topo_{grid\ x}} \cdot \left(1 - \sqrt{\%flood_{i,t}}\right) \quad (6)$$

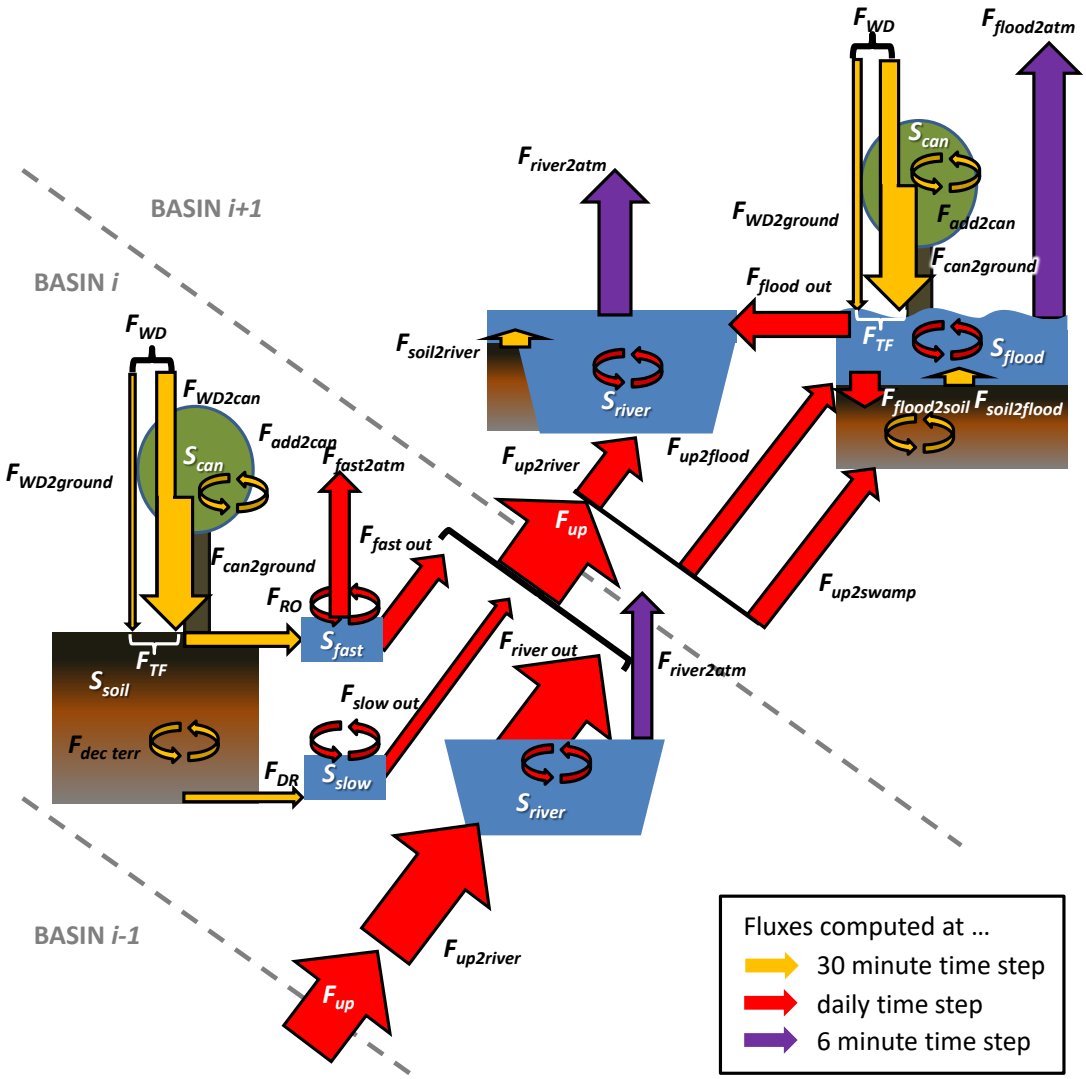


Figure 3. Simulated flows of water and C along the vegetation-soil-aquatic continuum. For reasons of simplicity, the fluxes (*F*) and storages (*S*) are characterized by subscripts indicating path or environmental compartment only (see Table A.1). Basin *i-1* is the basin upstream of basin *i*, basin *i+1* is the basin downstream of basin *i*. In this hypothetical example, swamps and floodplains are only present in basin *i+1*. The depiction of water and soil-river C fluxes in basins *i+1* and *i-1* were omitted for reasons of readability. Straight arrows represent water and C fluxes between the canopy (S_{can}), soil (S_{soil}), fast (S_{fast}), slow (S_{slow}), river (S_{river}) and flood (S_{flood}) reservoirs. Circular arrows represent carbon transformations within the reservoirs. See text for further details.

$$F_{up,H_2O,i,t} = \sum_{i-1} \left(F_{fast\ out,H_2O,i-1,t} + F_{slow\ out,H_2O,i-1,t} + F_{river\ out,H_2O,i-1,t} \right) \quad (7)$$

$$F_{up2swamp,H_2O,i,t} = f_{swamp} \cdot F_{up,H_2O,i,t} \cdot \%swamp_i \quad (8)$$

$$F_{up2flood,H2O,i,t} = F_{up,H2O,i,t} - F_{up2swamp,H_2O,i,t} \quad (9)$$

$$F_{up2river,H2O,i,t} = F_{up,H2O,i,t} - F_{up2swamp,H_2O,i,t} - F_{up2flood,H_2O,i,t} \quad (10)$$

$$S_{flood,H_2O,i,t+1} = S_{flood,H_2O,i,t} + F_{up2flood,H_2O,i,t} - F_{flood2soil,H_2O,i,t} + \left(F_{TF,H_2O,i,t} \cdot \%flood_{i,t} - F_{flood2atm,H_2O,i,t}\right) - F_{i,flood\ out,H_2O,t} \quad (11)$$

$$F_{flood\ out,H_2O,i,t} = \frac{S_{flood,H_2O,i,t}}{\tau_{flood} \cdot Topo_{grid\ x}} \cdot \%flood_{i,t} \quad (12)$$

Improved floodplain dynamics

Seasonal flooding in the Amazon is a major control of the hydraulic and C dynamics of the river system (Abril et al., 2014; Melack et al., 2009; Rasera et al., 2013; Richey et al., 1990, 2002). This is particularly true in the central basin where the extent of flooded areas can increase from 4 to 16% of the total area (Hamilton et al., 2002; Hess et al., 2003; Richey et al., 2002). In the following, we first present how flooding is simulated in the trunk-version of ORCHIDEE, summarizing mainly the work of D'Orgeval et al. (2008) and Guimberteau et al. (2012); next we describe improvements in simulated floodplain dynamics undertaken for ORCHILEAK in this study. Flooding is generally simulated in the temporal resolution of the routing module, in the default setting used in this study at the daily time-step.

Original trunk version

When floodplains are present in a given basin, all water inputs from upstream basins (F_{up}) which are not infiltrating in swamps ($F_{up2swamp}$) are routed to S_{flood} instead of S_{river} (Eq. 9). After floodplain and river reservoirs have been updated with in- and outflows for each basin (Eqs. 5,11), the inundated fraction *%flood* is calculated firstly for each grid-cell, and secondly for each basin within the grid cell. This sequential procedure is necessary, because the maximum floodable proportion ($\%flood_{max}$), which is prescribed by the forcing file, is given at the resolution of the grid cells. *%flood* per grid *x* is calculated from the total water storage in the floodplain reservoirs ($S_{flood,H2O,grid\ x,t}$, Eq. 13) of all basins *i* contained in that grid cell, assuming a slightly convex slope of the floodable area (Eqs. 14,15), as this shape is typical of large lowland rivers like the Amazon (Hamilton et al., 2002; Huggett, 2016). In the original version of ORCHIDEE (Fig. 5), the computation is performed as follows: first, a potential fraction of flooded area ($\%flood_{pot}$) is calculated based on the total area of the grid cell ($A_{total,grid\ x}$) and a potential water level height on the floodplain (*floodcri*, set to 2m by default) for which it is assumed that the whole grid cell is inundated (Eq. 14, Fig 5). The maximum flooded proportion

(*%flood*$_{max}$) of the grid cell is defined by values reported in the PRIMA forcing file (see below), that is, *%flood* cannot exceed *%flood*$_{max}$ (Eq. 15). Second, the actual water level over the floodplain area (*floodh*) is calculated from *%flood* and the water storage in the floodplain reservoir $S_{flood,H2O}$ (Eq. 16). Finally, the *%flood* of each basin *i* within the grid *x* is calculated based on the $S_{flood,H2O}$ of the basin compared to that of the grid box and A_{total} of the basin *i* compared to A_{total} of grid *x* (Eq. 17).

$$S_{flood,H2O,grid\ x,t} = \sum_i S_{flood,H_2O,i,t} \tag{13}$$

$$\%flood_{pot,grid\ x,t} = \left(\frac{S_{grid\ x,flood,H_2O,t} \cdot 3}{A_{total,grid\ x} \cdot floodcri}\right)^{\frac{2}{3}} \tag{14}$$

$$\%flood_{grid\ x,t} = min\left(\%flood_{pot,grid\ x,t}, \%flood_{max,grid\ x}\right) \tag{15}$$

$$floodh_{grid\ x,t} = \frac{2}{3} \cdot floodcri \cdot \sqrt{\%flood_{grid\ x,t}} + \frac{S_{flood,H_2O,grid\ x,t}}{\left(\%floodh_{grid\ x,t} \cdot A_{total,grid\ x}\right)} \tag{16}$$

$$\%flood_{i,t} = \%flood_{grid\ x,t} \cdot \frac{\left(\frac{S_{flood,H_2O,i,t}}{S_{flood,grid\ x,H2O,t}}\right)}{\left(\frac{A_{total,i}}{A_{total,grid\ x}}\right)} \tag{17}$$

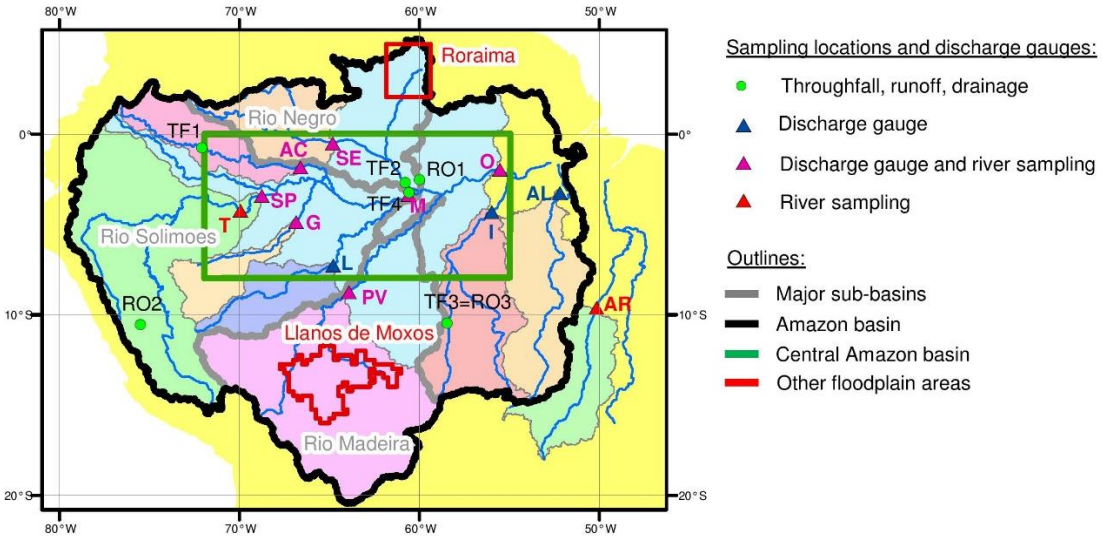


Figure 4. Overview of the Amazon Basin, with highlighted boundaries (thick grey) between the three major sub-basins (R. Solimoes, Madeira and Negro). The central Amazon basin (green box) and the sampling locations discussed in this study are also shown. River sampling locations and discharge gauges include: Rio Japura at Acanaui (AC), Rio Xingu at Altamira (AL),

Rio Araguaia (AR), Rio Jurua at Gaviao (G), Rio Tapajos at Itaituba (I), Rio Purus at Labrea (L), Rio Solimoes at Manacapuru (M), Amazon River at Obidos (O), Rio Madeira at Porto Velho (PV), Rio Negro at Serrinha (SE), Rio Solimoes at Sao Paulo de Olivenca (SP) and Tabatinga (T). The contributing areas are shown by the different colour codes on the map, except for location T as it is very similar to location SP. The remaining ungauged terrestrial area is represented in yellow. Sampling locations for throughfall DOC are indicated by "TF" and report data from Tobon et al. (2004) (TF1), Filoso et al. (1999) (TF2), Johnson et al. (2006) (TF3) and Williams et al. (1997) (TF4). Sampling location for DOC concentration in surface runoff and/or head waters are indicated by "RO" and report data from Waterloo et al. (2006) (RO1), Saunders et al. (2006) (RO2) and Johnson et al. (2006) (RO3). The red box and red line represent large floodplain areas outside the central Amazon basin for which observations are available.

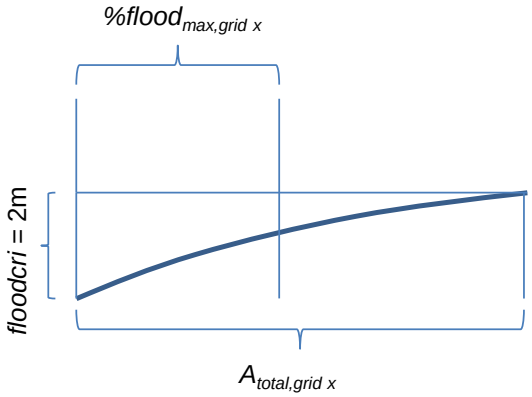


Figure 5. Schematic representation of the floodplain dynamics in the trunk version of ORCHIDEE. The bold line corresponds to the assumed shape of the floodplain. In ORCHILEAK, *floodcri* is replaced by $floodh_{95th}$, which represent the 95th percentile of the water level above the floodplain (*floodh*) over the simulation period 1980-2000.

Table 2. Data sets used for model evaluation.

Variable	Spatial resolution	Temporal resolution	Data source
Discharge	Multiple locations	bi-weekly	ORE-HYBAM (Cochc al., 2006)
Discharge	Multiple locations	average monthly values	GRDC (Global Runc Center)
Inundation in the Central Amazon basin	-	seasonality	Richey et al. (2002) af et al., 2003)
Inundation in Roraima and Llanos de Moxos wetland areas	-	multi-year time-series of monthly values	Hamilton et al. (2
Soil Organic Carbon stocks	1:5,000,000	-	HWSD v 1.1 (FAO et
Water temperature	Multiple locations	bi-weekly	ORE-HYBAM (Cochc al., 2006)
Riverine DOC concentrations and fluxes	Multiple locations	Irregular time-series	CAMREX (Richey et a ORE-HYBAM (Cochc al., 2006), Moreira-Tu (2003)
Seasonality of CO_2 evasion from Central Amazon Basin	-	Seasonality with average monthly values	Richey et al. (20
CO_2 evasion rates from the river surface at different sampling locations	-	Multiple values during high and low flow periods	Rasera et al. (20

The PRIMA forcing file was introduced by Guimberteau et al. (2012) to represent the maximum spatial extent of swamps and floodplains at the scale of the entire Amazon basin.

The available global wetland (swamps and floodplains) forcings (Lehner and Döll, 2004) are underrepresenting swamp and floodplain areas in this region, and were thus not sufficient to simulate water retention needed to reproduce the hydrograph of the Amazon River. The PRIMA dataset was obtained using the maximum floodable areas derived from satellite imagery (Prigent et al., 2007), after subtraction of the vegetated proportion reported by Martinez and le Toan (2007). The vegetated part of the maximum floodable area was assigned to 'swamp' areas, which, as stated above, does not include a specific water body in ORCHIDEE.

Changes in ORCHILEAK

Although water retention in floodplains was validated by reproducing the water height over the floodplains (Guimberteau et al., 2012), the seasonality in flooded areas extent is still not well captured in the trunk version. Furthermore, according to the PRIMA forcing, the maximum floodable area in the central Amazon basin is < 5%, while according to Richey et al. (2002) the areal proportion of inundated area is comprised between 4 and 16%, leaving a temporarily flooded proportion of 12%. For the simulations with ORCHILEAK, we merged back the swamp and floodplain areas, thus relying directly on the maximum inundated area of Prigent et al. (2007), while, at the same time, keeping swamp areas as zone of return flow from the river to the bottom layer of the soil column (Fig. 6). With this modified forcing, $\%flood_{max}$ increases to 10% within the Central Amazon basin, in better agreement with observations.

To improve the representation of seasonal flooding using updated values of $\%flood_{max}$, the original equations to calculate the inflow of water to the floodplains and the extent of flooded area in each grid cell were altered as follows. Firstly, floodplains are now only inundated when a threshold in river discharge is exceeded ($F_{up\ lim}$, Eq. 18), and it is only the excess part of the river discharge that contributes to the flooding while the remainder is directly entering the next river reservoir (Eq. 19). The threshold is defined for each grid by the median river reservoir water storage of each grid cell over the simulation period (1980-2000), which is derived in a first simulation with flooding deactivated, and then used as a forcing file for the model (Fig. 6). The choice of the median as threshold provides the advantage of a robust statistical measure and is similar to threshold of 90% of long-term mean discharge used by Vörösmarty et al. (1989) for the Amazon. This modification assumes that a fraction of river water continues to be transported by the river instead of being entirely diverted to the floodplains.

$$F_{up2flood,H_2O,grid\ x,t} = max(F_{up,H_2O,grid\ x,t} - F_{up2swamp,H_2O,grid\ x,t} - F_{up\ lim,H_2O,grid\ x}, 0) \quad (18)$$

$$F_{up2river,H_2O,grid\ x,t} = min(F_{up,H_2O,grid\ x,t} - F_{up2swamp,H_2O,grid\ x,t}, F_{up\ lim,H_2O,grid\ x}) \quad (19)$$

$$\%flood_{pot,grid\ x,t} = \left(\frac{S_{flood,H_2O,grid\ x,t} \cdot 3}{Area_{total,grid\ x} \cdot floodh_{grid\ x,95th}}\right)^{\frac{2}{3}} \quad (20)$$

$$floodh_{grid\ x,t} = \frac{2}{3} \cdot floodh_{grid\ x,95th} \cdot \sqrt{\%flood_{grid\ x,t}} + \frac{S_{flood,H_2O,grid\ x,t}}{(\%flood_{max,grid\ x} \cdot A_{total,grid\ x})} \quad (21)$$

While the default value for *floodcri*, as to be used in global modelling, was set to 2 m in the trunk version, this value is not applicable to the Amazon, where water levels of up to 12 m have been reported in the Central Amazon floodplain (Trigg et al., 2009). Thus, instead of using a single value for *floodcri* as previously done, we now first compute for each grid cell the 95th percentile of all simulated water level heights over the floodplain area for the simulation period 1980-2000 (*floodh*$_{95th}$, Eq. 21, cf. Fig. 5). We used the regional data set of monthly inundated areas from Hamilton et al. (2011) for validation in the Roraima and Llanos de Moxos wetland areas, which covers part of our simulation period. For inundation in the central Amazon basin, we used the data from Hess et al. (2003) as summarized in Richey et al. (2002) for validation.

$$F_{river\ out,H_2O,i,t} = \frac{S_{river,H_2O,i,t}}{\tau_{river} \cdot Topo_{grid\ x}} \quad (22)$$

Following the changes in the flooding scheme, we recalibrated two parameters in order to reproduce the monthly discharges from the Amazon and its major tributaries: 1) We decrease the water residence time on the floodplains by changing τ_{flood} from 2.5 days as used by Guimberteau et al. (2012) to 1.4 days (Eq. 12); and 2) we halved the proportion of water diverted to swamps by setting f_{swamp} from 0.2 to 0.1 (Eq. 8), while using the same forcing for %swamp as Guimberteau et al. (2012). In addition, because %flood can now take values close to 100% in some areas, we modified the equation to calculate the outflow from the river reservoir, which is not decreased anymore depending of %flood (Eq. 22). The simulated river discharges were validated against gauging data from ORE-HYBAM (Cochonneau et al., 2006) and mean monthly discharges provided by the Global Runoff Data Centre (GRDC, n.d.).

In ORCHILEAK, for the purpose of calculating CO_2 evasion from the river network, the river reservoir is now assigned a surface area as well (A_{river}). The base surface area A_{river} ($A_{river\ basic}$) per grid cell is extracted from a forcing file derived from the global river surface maps of Lauerwald et al. (2015). Following the findings by Rasera et al. (2013), we assume that the surface area of small rivers ($A_{river\ small}$, width < 100m) can increase by about 20% from low to high water stages, whereas the area of larger rivers ($A_{river\ large}$, width ≥ 100m) increases by about 10%. Assuming the 10th and 90th percentile of $S_{river,H2O}$ over the simulation period 1980-2000 ($S_{river,H2O,grid\ x,10th}$, $S_{river,H2O,grid\ x,90th}$, Fig. 6) as representative for the low and high water stages, an actual A_{river} ($A_{river\ act}$) is calculated at each time-step depending on $S_{river,H20}$ (Eqs. 23-26). As the A_{river} forcings likely underestimate the total A_{river} (Lauerwald et al., 2015), it is assumed that $A_{river\ basic}$ represent A_{river} at low water stage. $A_{river\ act}$ per basin *i* is calculated from A_{river} per grid *x* containing that basin, scaling to the square root of $S_{river,H2O}$, because $S_{river,H2O}$ is linearly related to discharge (Eq. 27) and it was empirically shown that stream width scales roughly with the square root of discharge (Raymond et al., 2012, 2013). Assuming that stream length does not change significantly, the relative change in stream width equals the relative change in $A_{river\ act}$.

$$A_{river\ basic,grid\ x} = A_{river\ small,grid\ x} + A_{river\ large,grid\ x} \tag{23}$$

If $S_{river,H_2O,grid\ x,t} \leq S_{river,H_2O,grid\ x,10th}$:

$$A_{river\ act,grid\ x,t} = A_{river\ basic,grid\ x} \tag{24}$$

If $S_{river,H_2O,grid\ x,10th} < S_{river,H_2O,grid\ x,t} < S_{river,H_2O,grid\ x,90th}$:

$$A_{river\ act,grid\ x,t} = \left(1 + \frac{S_{river,H_2O,grid\ x,t} - S_{river,H_2O,grid\ x,10th}}{S_{river,H_2O,grid\ x,90th}} \cdot 0.2\right) \cdot A_{river\ small,grid\ x} + \left(1 + \frac{S_{river,H_2O,grid\ x,t} - S_{river,H_2O,grid\ x,10th}}{S_{river,H_2O,grid\ x,90th}} \cdot 0.1\right) \cdot A_{river\ large,grid\ x} \tag{25}$$

If $S_{river,H_2O,grid\ x,t} \geq S_{river,H_2O,grid\ x,90th}$:

$$A_{river\ act,grid\ x,t} = 1.2 \cdot A_{river\ small,grid\ x} + 1.1 \cdot A_{river\ large,grid\ x} \tag{26}$$

$$A_{river\ act,i,t} = A_{river\ act,grid\ x,t} \cdot \frac{\sqrt{S_{river,H_2O,i,t}}}{\sqrt{S_{river,H_2O,grid\ x,t}}} \tag{27}$$

The difference between $A_{river\ act}$ and $A_{river\ basic}$ gives a seasonally flooded area directly adjacent to the river ($\%flood_{river}$, Eqs. 28, 29). This flooded area induced by changes in water levels in the river was then added to the total flooded proportion of soils ($\%flood_{total}$, Eqs. 30,31). Note, however, that for the calculation of C inputs from flooded soils to the water column (section 2.3), S_{flood} and S_{river} need again to be distinguished.

$$\%flood_{river,grid\ x,t} = \frac{A_{river\ act,grid\ x,t} - A_{river\ basic,grid\ x}}{A_{total,grid\ x}} \quad (28)$$

$$\%flood_{river,i,t} = \frac{\left(A_{river\ act,grid\ x,t} - A_{river\ basic,grid\ x}\right) \cdot \sqrt{S_{river,H_2O,i,t}}}{A_{total,i} \cdot \sqrt{S_{river,H_2O,grid\ x,t}}} \quad (29)$$

$$\%flood_{total,grid\ x,t} = \%flood_{grid\ x,t} + \%flood_{river,grid\ x,t} \quad (30)$$

$$\%flood_{total,i,t} = \%flood_{i,t} + \%flood_{river,i,t} \quad (31)$$

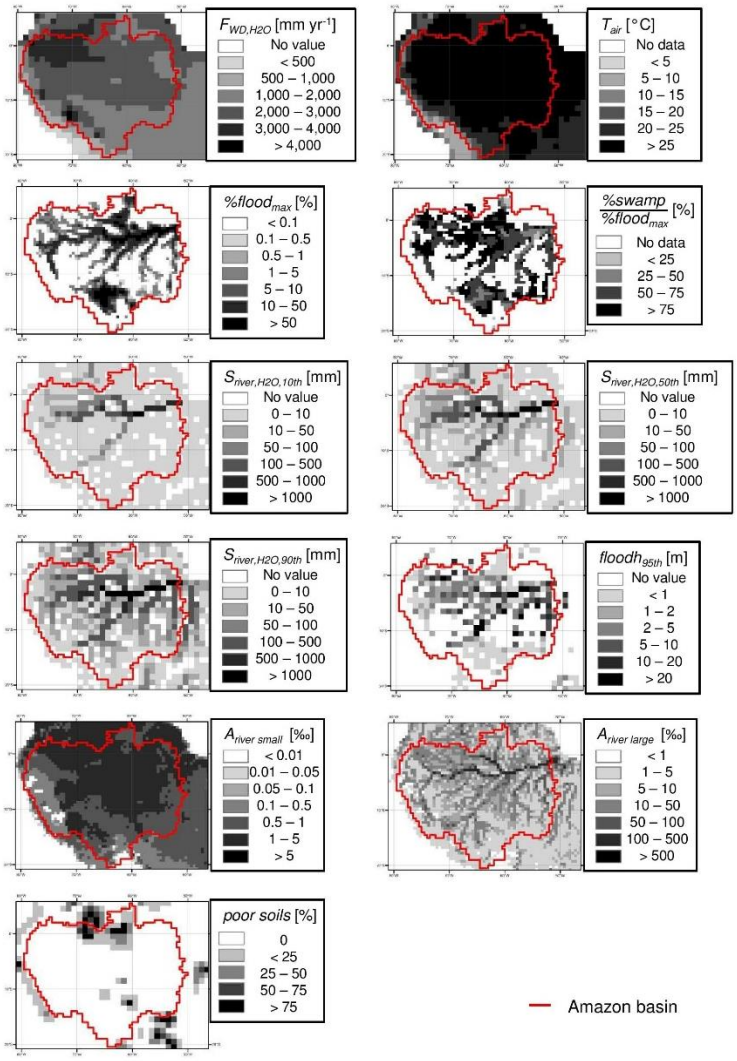


Figure 6. Overview of forcing files (cf. Table 2). Climatic forcings comprise, among others, variables like precipitation ($F_{WD,H2O}$) and air temperature (T_{air}). The climatic forcings used here are based on the NCC ((Ngo-Duc et al., 2005) data set,

only $F_{WD,H2O}$ was replaced by a regional data set created by Guimberteau et al. (2012). The forcing of maximum floodable areas $\%flood_{max}$ was adopted from Guimberteau et al. (2012) after merging swamp areas ($\%swamp$) into $\%flood_{max}$. Simulations of inundation in ORCHILEAK are based on 10th, 50th and 90th percentile of water storage in the river reservoir S_{river} ($S_{river,H2O,10th}$, $S_{river,H2O,50th}$, $S_{river,H2O,90th}$), here given in mm which equals kg H_2O m^{-2} assuming a density of water of 10^{-3} kg m^{-3}, and the 95th percentile of water table level over the floodplains *floodh* ($floodh_{95th}$), all derived from simulation results over the period 1980 to 2000. Surface areas of small (width < 100 m) and large (width ≥ 100 m) rivers ($A_{river\ small}$, $A_{river\ large}$) are taken from Lauerwald et al. (2015). Of importance for representation of DOC cycling in watersheds of black water rivers is the identification of '*poor soils*' (Podzols, Arenosols and soils in black water swamps), which we derived from the Harmonized World Soil Database (HWSD, FAO/IIASA/ISRIC/ISS-CAS/JRC, 2009) and *%swamp*.

8.1.1.2. Carbon dynamics along the vegetation-soil-aquatic continuum

Overview of the DOC transport scheme

DOC and CO_2 are exported through the terrestrial-water interface by runoff (F_{RO}) and drainage (F_{DR}), respectively (Fig. 3). Part of the terrestrial DOC stems from throughfall ($F_{TF} = F_{WD2ground} + F_{can2ground}$, see below), the other part stems from the decomposition of litter and soil organic carbon ($F_{dec\ terr}$). DOC exports from flooded areas to the river network are another important source, because F_{TF} and the decomposition of submerged litter and soil carbon in the floodplains ($F_{soil2flood}$) add directly to the DOC storage in the overlying water column and, from there, a delayed flux ($F_{flood\ out}$) feeds S_{river}. In addition, streams and rivers extend laterally during high flow periods (see section 2.1.3) and there is thus a direct input of DOC from litter and SOC decomposition on/in seasonally inundated soils immediately adjacent to the stream bed into S_{river} ($F_{soil2river}$). DOC and CO_2 are transported as passive tracers with the fluxes of water through the different reservoirs of the routing scheme (see section 2.1) and can feed back into the soil system via two mechanisms: 1) re-infiltration from the floodplain reservoir into the first layer of the soil column ($F_{flood2soil}$); 2) infiltration of DOC into the bottom layer of the soil column entrained with water entering swamps ($F_{up2swamp}$) (Fig. 3). In addition, DOC is mineralized to CO_2 in transit and CO_2 is evading to the atmosphere from the water surface. Depending on the relative magnitude between inputs, outputs and in-situ transformations, the storage of DOC in canopy, soil, fast, slow, river and floodplain reservoirs (S_{can}, S_{soil}, S_{fast}, S_{slow}, S_{river}, and S_{flood}) can thus increase or decrease over different time periods. For the routing of DOC, we distinguish two pools, a labile and a refractory pool. Like the cycling of water and C in vegetation and soils, the allochthonous inputs of DOC from S_{can} and S_{soil} into the inland water network (F_{RO}, F_{DR}, $F_{soil2flood}$, $F_{soil2river}$, see Fig 3) are computed at a temporal resolution of 30 minutes and at the spatial resolution of the grid cell. The lateral transfer between the S_{fast}, S_{slow}, S_{river} and S_{flood} and the transformation of C within those storage reservoirs are only

simulated at a daily time step and at the spatial resolution of the basin. Therefore, to simulate the lateral transfers, the allochthonous DOC and CO_2 inputs are first aggregated over 48 30-minute time steps until one full day is over. The fluxes from the water column back into the soil column ($F_{flood2soil}$,$F_{up2swamp}$ in Fig. 3) are simulated at the daily time-step of the routing module, but are used as inputs in the soil carbon module, which runs at a 30 minute temporal resolution. This is achieved by downscaling the daily fluxes uniformly over the 48 30 minute time-steps of the following day of simulation. The evasion of CO_2 from river and floodplain water surface ($F_{river2atm}$, $F_{flood2atm}$) is also simulated at the daily time-step of the routing module, but to approximate the continuous interplay of CO_2 inputs and CO_2 evasion controlling the water-air gradient in CO_2 partial pressures (pCO_2) a much shorter time-step of 6 minutes is used, and the CO_2 inputs to the water column are thus uniformly distributed over the 240 6-minutes time-step contained in each day. The following subsections describe in more detail the simulation of DOC in precipitation and throughfall (section 2.2.2), production of DOC and its export through the terrestrial-aquatic interface (section 2.2.3), CO_2 inputs through the terrestrial-aquatic interface (2.2.4), and in-transit DOC mineralisation and CO_2 evasion along the inland water network (section 2.2.5).

DOC in precipitation and throughfall

Reported average rain DOC concentrations in the Amazon basin are significant with 1.3 to 3.9 mg C L^{-1} (Table 5, in most temperate regions average concentrations < 1 mg C L^{-1} are common, see Michalzik et al. 2001), of the same magnitude as observed concentrations in white and clear water rivers of the region (Moreira-Turcq et al., 2003). The spatial variation in rain DOC concentration is unknown and we thus assumed a constant value of 2.4 mg C L^{-1} throughout the Amazon basin, from the average of reported literature values (Table 5). Observed average DOC concentrations in throughfall are higher than in precipitation because of the DOC enrichment of leaf-intercepted water due to evaporation losses and dissolution of organic carbon from leaf-leachates and dry deposition. Reported annual throughfall DOC flux (F_{TF}) in the Amazonian rain forest varies little, from 14.8 to 19.0 g C m^{-2} yr^{-1} (see Table 5). The temporal variability in throughfall DOC concentrations is mainly controlled by the amount of throughfall, which acts as a dilution factor, and by the duration of preceding dry periods, which favours the accumulation of soluble organic C on the canopy (Johnson et al., 2006). Here, we used the time-series data on throughfall DOC fluxes in South Amazonia from Johnson et al. (2006) to set up and calibrate a simple model of throughfall DOC fluxes.

In ORCHILEAK, the wet deposition of DOC, F_{WD} is calculated from precipitation and the prescribed constant concentration of 2.4 mg C L^{-1}, which also equals the minimum throughfall concentration in the time-series by Johnson et al. (2006). For each of the 13 ORCHIDEE plant functional types (PFTs) which are potentially present in a grid cell, the wet deposition of DOC onto the canopy (F_{WD2can}) and the direct precipitation of DOC onto the ground ($F_{WD2ground}$) directly scales to the corresponding water fluxes simulated in the hydrology module. According to our simulation, F_{WD} contributes to only about one third of the F_{TF} at our calibration site (14.9 g C $m^{-2}yr^{-1}$ (Johnson et al., 2006)). Thus we assumed that the unaccounted flux of 10 g C m^{-2} yr^{-1} must originate from dry deposition onto the canopy or leaf leachates. We further assumed that this dry addition of soluble organic carbon ($F_{add2can}$) does not vary over time and scales to the leaf biomass (which, in the model, is directly related to leaf area). Based on the simulated leaf biomass of 457±1 g C m^{-2} for tropical rain forests at the field-site location, we calibrated $F_{add2can}$ at $6*10^{-5}$ g C per day and per g C in the leaf biomass (Eq. 32). For agricultural and grass lands, we set $F_{add2can}$ to zero.

Whenever intercepted water from the canopy falls to the ground ($F_{can2ground}$), the related flux of DOC ($F_{can2ground}$) will empty the storage of DOC in the canopy (S_{can}) at once unless a maximum concentration DOC_{max} of 100 mg DOC kg_{H2O}^{-1} (Eq. 33) in $F_{can2ground}$ is exceeded. This value corresponds to the maximum concentration observed by Johnson et al. (2006). Beyond this threshold, $F_{can2ground}$ is set as the product of the water flux and the maximum concentration, and the DOC in excess is assumed to remain in the canopy reservoir S_{can}. This threshold prevents unreasonably high DOC concentrations in the first throughfall events after dry periods and allows simulation of progressive depletion of the S_{can} reservoir after a time of significant DOC accumulation. At each 30 min time step, F_{WD2can}, $F_{add2can}$ and $F_{can2ground}$ are calculated and subsequently used to update the DOC storage in the canopy at each *grid x* and PFT *v* (Eq. 34).

$$F_{add2can,DOC,grid\ x,v,t} = leaf\ biomass_{grid\ x,v,t} \cdot 10^{-5} \frac{dt}{day} \tag{32}$$

$$F_{can2ground,DOC,grid\ x,v,t} = \max(F_{can2ground,H_2O,grid\ x,v,t} \cdot 0.1\ g\ kg^{-1}H_2O, S_{can,DOC,grid\ x,v,t}) \tag{33}$$

$$\begin{aligned} S_{can,DOC,grid\ x,v,t+1} &= S_{can,DOC,grid\ x,v,t} + F_{WD2can,DOC,grid\ x,v,t} + F_{add2can,DOC,grid\ x,v,t} \\ &\quad + F_{can2ground,DOC,grid\ x,v,t} \end{aligned} \tag{34}$$

$$F_{TF,DOC,grid\ x,v,t} = F_{WD2ground,DOC,grid\ x,v,t} + F_{can2ground,DOC,grid\ x,v,t} \quad (35)$$

$$F_{TF,DOC_{lab},grid\ x,v,t} = F_{TF,DOC_{ref},grid\ x,v,t} = 0.5 \cdot F_{TF,DOC,grid\ x,v,t} \quad (36)$$

F_{TF} is calculated as the sum of the non-intercepted wet deposition $F_{WD2ground}$ and $F_{can2ground}$ (Eq 35). Based on the range of values reported in the literature (Aitkenhead-Peterson et al., 2003), we assume that half of the DOC reaching the ground is labile (DOC_{lab}) while the other half is refractory (DOC_{ref}) (Eq. 36). F_{TF} then infiltrates into the topsoil or adds to S_{flood} in areas where it falls on inundated land (see section 2.2.4).

Production and export of soil DOC through the terrestrial-aquatic interface

ORCHILEAK is largely based on ORCHIDEE-SOM, the new soil carbon module simulating microbial production and consumption of DOC, its adsorption and desorption onto/from mineral surfaces, the vertical advective and diffusive fluxes of DOC within the soil profile and the exports of DOC from the soil via surface runoff and drainage (Camino Serrano, 2015). Consistent with the soil hydrology module (Campoy et al., 2013; de Rosnay et al., 2002), the carbon dynamics are resolved using a discretization of a 2m-soil profile into 11 layers geometrically increasing in depth and running at a 30 minutes time-step (Camino Serrano, 2015).

DOC is produced from the decomposition of litter and soil organic carbon (SOC) (Eqs. 37-40), and consumed by further decomposition (Eqs. 41,42). Here, the soil carbon module has been modified to better represent the soil DOC dynamics in the Amazon. First, decomposition on non-flooded ($F_{dec\ terr}$) and flooded ($F_{dec\ flood}$) soils is distinguished, with decomposition rates of the litter, SOC and DOC pools 3 times slower when soils are flooded (Rueda-Delgado et al., 2006). Second, in 'poor soils' characterized by low pH and low nutrient levels such as Podzols, Arenosols or soils located in black water swamps (referred to as Igapo in the Amazon basin), decomposition rates are significantly reduced. Here, we assume a reduction by a factor of 2, following findings from the literature (Bardy et al., 2011; Vitousek and Hobbie, 2000; Vitousek and Sanford, 1986). This feature was implemented in the model by adding a layer defining the areal proportion of 'poor soils' in the soil-forcing file. The spatial distribution of Podzols and Arenosols was derived from the Harmonized World Soil Data base (FAO/IIASA/ISRIC/ISS-CAS/JRC, 2009). To determine the spatial distribution of Igapo forest soils, we used the PRIMA forcing for swamps in combination with the boundaries of the Rio Negro catchment as derived from the 0.5° river network (Fig. 6).

$$F_{dec\ terr,SOC\ pool,grid\ x,v,l,t} = S_{soil,SOC\ pool,grid\ x,v,l,t} \cdot \frac{k_{SOC\ pool}}{1 + \%poorsoils_{grid\ x}} \cdot \left(1 - \%flood_{total,grid\ x,t}\right) \quad (37)$$

$$F_{dec\ flood,SOC\ pool,grid\ x,v,l,t} = S_{soil,SOC\ pool,grid\ x,v,l,t} \cdot \frac{k_{SOC\ pool}}{1 + \%poorsoils_{grid\ x}} \cdot \frac{\%flood_{total,grid\ x,t}}{3} \quad (38)$$

$$F_{dec\ terr,litter\ pool,grid\ x,v,l,t} = S_{soil,litter\ pool,grid\ x,v,l,t} \cdot \frac{k_{litter\ pool}}{1 + \%poorsoils_{grid\ x}} \cdot \left(1 - \%flood_{total,grid\ x,t}\right) \quad (39)$$

$$F_{dec\ flood,litter\ pool,grid\ x,v,l,t} = S_{soil,litter\ pool,grid\ x,v,l,t} \cdot \frac{k_{litter\ pool}}{1 + \%poorsoils_{grid\ x}} \cdot \frac{\%flood_{total,grid\ x,t}}{3} \quad (40)$$

$$F_{dec\ terr,DOC\ pool,grid\ x,v,l,t} = S_{soil,DOC\ pool,grid\ x,v,l,t} \cdot \frac{k_{DOC\ pool}}{1 + \%poorsoils_{grid\ x}} \cdot \left(1 - \%flood_{total,grid\ x,t}\right) \quad (41)$$

$$F_{dec\ flood,DOC\ pool,grid\ x,v,l,t} = S_{soil,DOC\ pool,grid\ x,v,l,t} \cdot \frac{k_{DOC\ pool}}{1 + \%poorsoils_{grid\ x}} \cdot \frac{\%flood_{total,grid\ x,t}}{3} \quad (42)$$

The soil carbon module distinguishes 3 different pools of DOC depending on the source material: active, slow and passive (Camino Serrano, 2015). The DOC derived from the active SOC pool and metabolic litter is assigned to the active DOC pool, while the DOC derived from the slow and passive SOC pools are assigned to the slow and passive DOC pools, respectively (Eqs. 43-45). A part of DOC derived from structural plant litter, which is related to the lignin structure of the litter pool (Krinner et al., 2005), is allocated to the slow DOC pool, while the remainder feeds the active DOC pool. The proportion of the decomposed litter and SOC that is transformed into DOC instead of CO_2 depends on the carbon use efficiency (*CUE*), set here to a value of 0.5 (Manzoni et al., 2012). Taken that the same residence time for the slow and passive DOC pools is used in ORCHIDEE-SOM (Camino Serrano, 2015), we merge these two pools when computing throughfall and lateral transport of DOC. Thus, the labile pool is identical to the active pool of the soil carbon module, while the refractory pool combines the slow and passive pools. The labile ($F_{TF,DOClab}$) and refractory ($F_{TF,DOCref}$) proportions of throughfall DOC are added to the active and slow DOC pools of the first soil layer, respectively.

$$\Delta S_{soil,DOC\ active,grid\ x,v,t} = \sum_{l=1}^{11} \Big(\big(F_{dec\ terr,litter\ str,grid\ x,v,l,t} + F_{dec\ flood,litter\ str,grid\ x,v,l,t} \big) \cdot \big(1 - \%lignin_{grid\ x,v,l,t} \big) + F_{dec\ terr,litter\ met,grid\ x,v,l,t} + F_{dec\ flood,litter\ met,grid\ x,v,l,t} + F_{dec\ terr,SOC\ active,grid\ x,v,l,t} + F_{dec\ flood,SOC\ active,grid\ x,v,l,t} \Big) \cdot CUE - \sum_{l=1}^{11} \big(F_{dec\ terr,DOC\ labile,grid\ x,v,l,t} + F_{dec\ flood,DOC\ labile,grid\ x,v,l,t} \big) + F_{TF,DOC_{lab},grid\ x,v,t} \cdot \big(1 - \%flood_{grid\ x,t} \big) - F_{RO,DOC\ active,grid\ x,v,t} - F_{DR,DOC\ active,grid\ x,v,t} - F_{Flood\ inp,DOC\ active,grid\ x,v,t} \quad (43)$$

$$\Delta S_{soil,DOC\ slow,grid\ x,v,t} = \sum_{l=1}^{11} \Big(\big(F_{dec\ terr,litter\ str,grid\ x,v,l,t} + F_{dec\ flood,litter\ str,grid\ x,v,l,t} \big) \cdot \%lignin_{grid\ x,v,l,t} + F_{dec\ terr,SOC\ slow,grid\ x,v,l,t} + F_{dec\ flood,SOC\ slow,grid\ x,v,l,t} + F_{dec\ terr,SOC\ passive,grid\ x,v,l,t} + F_{dec\ flood,SOC\ passive,grid\ x,v,l,t} \Big) \cdot CUE - \sum_{l=1}^{11} \big(F_{dec\ terr,DOC\ slow,grid\ x,v,l,t} + F_{dec\ flood,DOC\ slow,grid\ x,v,l,t} \big) + F_{TF,DOC_{ref},grid\ x,v,t} \cdot \big(1 - \%flood_{grid\ x,t} \big) - F_{RO,DOC\ slow,grid\ x,v,t} - F_{DR,DOC\ slow,grid\ x,v,t} - F_{soil2flood,DOC\ slow,grid\ x,v,t} \quad (44)$$

$$\Delta S_{soil,DOC\ passive,grid\ x,v,t} = \sum_{l=1}^{11} \big(F_{dec\ terr,SOC\ passive,grid\ x,v,l,t} + F_{dec\ flood,SOC\ passive,grid\ x,v,l,t} \big) \cdot CUE - \sum_{l=1}^{11} \big(F_{dec\ terr,DOC\ passive,grid\ x,v,l,t} + F_{dec\ flood,DOC\ passive,grid\ x,v,l,t} \big) - F_{RO,DOC\ passive,grid\ x,v,t} - F_{DR,DOC\ passive,grid\ x,v,t} - F_{soil2flood,DOC\ passive,grid\ x,v,t} \quad (45)$$

Alongside with decomposition, DOC is lost from the soil column through lateral exports with surface runoff and / or drainage, which occur at the top and bottom of the soil column, respectively. The DOC export by drainage at the bottom of the soil is proportional to the DOC concentration in the deepest (11^{th}) soil layer (Eq. 46). Surface runoff occurs when the maximum infiltration rate is exceeded, beyond which the excess water does not enter the soil column anymore. Because the first soil layers are extremely thin, it is assumed here that surface runoff can entrain DOC from the first five layers of the soil column, which together have a thickness of 4.5 cm (Eq. 47). In each basin, the DOC release is proportional to the mean DOC

concentration in this zone of the soil column as well as to the areal extent of the saturated zone around headwaters, as detailed below. To simulate the DOC production in flooded areas, we assume that the DOC produced from the decomposition of litter and SOC within these same 5 topsoil layers adds directly to the DOC storage in the overlying surface water body S_{flood} (see Fig. 3, Eqs. 48-50). Accordingly, the inputs of DOC to the non-flooded soils via $F_{dec\ terr}$ are estimated using the non-flooded proportion of the grid cell (1-$\%flood_{i,t}$) (Eqs. 37, 39, 41).

$$F_{DR,DOC\ pool,grid\ x,v,t} = min(F_{DR,H2O,i,v,t} \cdot \frac{S_{soil,DOC\ pool,grid\ x,v,l=11,t}}{S_{soil,H_2O,grid\ x,v,l=11,t}}, S_{soil,DOC\ pool,grid\ x,v,l=11,t}) \quad (46)$$

$$F_{RO,DOC\ pool,grid\ x,v,t} = min(F_{RO,H2O,grid\ x,v,t} \cdot \frac{\sum_{l=1}^{5} S_{soil,DOC\ pool,grid\ x,v,l,t}}{\sum_{l=1}^{5} S_{soil,H_2O,grid\ x,v,l,t}} \cdot red_{RO,grid\ x,t}, \sum_{l=1}^{5} S_{soil,DOC\ pool,grid\ x,v,l,t}) \quad (47)$$

$$F_{soil2flood,DOC\ active,grid\ x,v,t} = \sum_{l=1}^{5} \left(F_{dec\ flood,litter\ str,grid\ x,v,l,t} \cdot \left(1 - \%lignin_{grid\ x,v,l,t}\right) + F_{dec\ flood,litter\ met,grid\ x,v,l,t} + F_{dec\ flood,SOC\ active,grid\ x,v,l,t}\right) \cdot CUE \quad (48)$$

$$F_{soil2flood,DOC\ slow,grid\ x,v,t} = \sum_{l=1}^{5} \left(F_{dec\ flood,litter\ str,grid\ x,v,l,t} \cdot \%lignin_{grid\ x,v,l,t} + F_{dec\ flood,SOC\ slow,grid\ x,v,l,t}\right) \cdot CUE \quad (49)$$

$$F_{soil2flood,DOC\ passive,grid\ x,v,t} = \sum_{l=1}^{5} F_{dec\ flood,SOC\ passive,grid\ x,v,l,t} \cdot CUE \quad (50)$$

The usually higher DOC concentration in the topsoil compared to the subsoils is mainly due to the higher inputs of plant litter into and onto the topsoil. However, DOC is efficiently transported between the soil layers along with the vertical flow of water through the soil matrix ($F_{soil\ adv}$, Eqs. 51-52). Therefore, a part of the DOC exported with the drainage is not produced in-situ but rather originates from percolation across the entire soil column. The vertical DOC transport within the soils, as well as for the export of DOC with surface runoff are not directly computed as the product of water flux and DOC concentration. Instead, a reduction factor (red_{DOC}) is applied to account for the effect of preferential vertical flow paths, e.g. along macrospores produced by the root system (Karup et al., 2016), and zones of reduced flow rates which increase the DOC residence time in the remaining parts of the soil. Only in “poor soils”

the flow of DOC is not reduced relative to the flow of water (no reduction, Eq. 54). This allows to account for their poor filtering capacity which is the cause of the very high DOC concentrations in groundwater below Podzols and black water swamps (Brinkmann, 1984; McClain et al., 1997). While the effect of preferential flow path should be envisioned as a general concept in ORCHILEAK, the introduction of 'poor soils' is specific to tropical black water systems. It remains to be shown in future work how their effects will have to be parametrized in other climate zones, for instance in the Boreal zone where Podzols are abundant.

$$F_{soil\ adv,DOC\ pool,grid\ x,v,l\rightarrow l+1,t} = max\left(F_{soil\ adv,H_2O,grid\ x,v,l\rightarrow l+1,t} \cdot \frac{S_{soil,DOC\ pool,grid\ x,v,l,t}}{S_{soil,H_2O,grid\ x,v,l,t}} \cdot red_{DOC,grid\ x}, 0\right) \quad (51)$$

$$F_{soil\ adv,DOC\ pool,grid\ x,v,l\rightarrow l-1,t} = max\left(F_{soil\ adv,H_2O,grid\ x,v,l\rightarrow l-1,t} \cdot \frac{S_{soil,DOC\ pool,grid\ x,v,l,t}}{S_{soil,H_2O,grid\ x,v,l,t}} \cdot red_{DOC,grid\ x}, 0\right) \quad (53)$$

$$red_{RO,grid\ x,v,t} = red_{DOC,grid\ x} \cdot red_{connect,grid\ x,t} \quad (53)$$

$$red_{DOC,grid\ x} = (1 - \%poorsoils) \cdot red_{DOC,base} + \%poorsoils \quad (54)$$

$$red_{connect,grid\ x,t} = min\left(\frac{\left(S_{fast,H2O,grid\ x,t} + S_{slow,H2O,grid\ x,t}\right)^{0.5}}{S_{fast+slow,H_2O,ref}{}^{0.5}}, 1\right) \quad (55)$$

DOC exports with surface runoff is even further reduced, because the riverine DOC mostly derives from saturated soils in direct vicinity to surface waters (Idir et al., 1999). As we do not have direct information on the density of headwater streams at small scale and the extent of the saturated, riparian zone, the reduction in DOC exports with surface runoff ($red_{connect}$) was scaled to the storage of water in S_{fast} and S_{slow} (Eq. 55). We assumed these reservoirs to represent the water stored in groundwater and headwater streams (S_{river} being attributed to wider water bodies due to the coarse resolution (0.5°) of the river network). Next, based on model calibration, we set a threshold value for the sum of $S_{fast,H2O}$ and $S_{slow,H2O}$ ($S_{fast+slow,H_2O,ref}$) at which a 100% connection between top soils and headwaters is achieved. When $S_{fast+slow,H_2O,ref}$ does not reach the threshold, a lower proportion of topsoil is in connection with the headwaters. Consistent with our approach in section 2.1.3, we assumed here that the

extent of saturated soils around headwaters (i.e. the connected topsoils) increases linearly with the square root of the sum of $S_{fast,H2O}$ and $S_{slow,H2O}$. Finally, the maximum amount of DOC that can be exported through surface runoff and drainage is limited by the storage of DOC in the top and bottom soil layers (Eqs. 46-47).

Export of dissolved CO_2 through the soil-aquatic network interface

Although mineralization of litter, SOC, DOC in the soil are simulated in ORCHIDEE, the CO_2 partial pressure in the soil air and soil solution of the different layers is not represented. Thus, we implemented simple estimates of these soil-derived CO_2 inputs in order to reproduce the observed CO_2 evasion fluxes from the water surface of the fluvial network. For simulating the export of CO_2 with surface runoff and drainage, we use fixed concentrations of 20 mg C L^{-1} (pCO_2 of 50,000 µatm at 25°C) and 2 mg C L^{-1} (pCO_2 of 5,000 µatm at 25°C), respectively, derived from reported literature values (Davidson et al., 2010; Johnson et al., 2008; Saunders et al., 2006). The lateral exports of CO_2 dissolved in soil water are then calculated by multiplying these CO_2 concentrations with the water fluxes from surface runoff and drainage simulated at half-hourly time-step in the soil hydrology module (Eqs. 56,57). Next, the computed lateral fluxes of CO_2 exported out of soils are subtracted from the total soil respiration and the remainder, by far the dominant fraction (Davidson et al., 2010), is assumed to evade directly to the atmosphere through the topsoil (Eq. 58). Carbonate chemistry and export of alkalinity are neglected.

$$F_{RO,CO_2,grid\ x,t} = F_{RO,H_2O,grid\ x,t} \cdot w_{RO,CO_2} \tag{56}$$

$$F_{DR,CO_2,grid\ x,t} = F_{DR,H_2O,grid\ x,t} \cdot w_{DR,CO_2} \tag{57}$$

$$F_{soil2atm,CO_2,grid\ x,t} = F_{soil\ resp,CO_2,grid\ x,t} - F_{RO,CO_2,grid\ x,t} - F_{DR,CO_2,grid\ x,t} \tag{58}$$

$$\begin{aligned} F_{soil2flood,CO_2,grid\ x,t} &= \sum_{v=1}^{13}\left(\sum_{l=1}^{11}\left(\left(\sum_{litter\ pool} F_{dec\ flood,litter\ pool,grid\ x,v,l,t}\right.\right.\right. \\ &\left.\left.\left. + \sum_{SOC\ pool} F_{dec\ flood,SOC\ pool,grid\ x,v,l,t} + \sum_{DOC\ pool} F_{dec\ flood,DOC\ pool,grid\ x,v,l,t}\right)\right.\right. \\ &\left.\left. \cdot (1 - CUE)\right) + F_{root\ respiration,grid\ x,v,t} \cdot \%flood_{grid\ x,t}\right) \end{aligned} \tag{59}$$

In floodplains, mineralisation of submerged litter and soil carbon are considered to be sources of CO_2 to S_{flood} (Eq. 59). In addition, we allocated the root respiration in inundated areas to the "CO_2 inputs to S_{flood}" term. The lateral transfer of CO_2 by advection and the re-infiltration of dissolved CO_2 into swamps and on floodplains are simulated following the approach implemented for DOC (Fig. 3, and preceding subsections).

Carbon transport and transformation along the inland water network

Transport and transformation of terrestrially derived C in the river system are implemented into the river routing module. The lateral transport of DOC and CO_2 between reservoirs are assumed to be proportional to the water fluxes, that is, the exports from each reservoir to the next have the same concentration of DOC and CO_2 as in the reservoir from which they originate (Eq. 60). The same holds true for infiltration on the floodplains ($F_{flood2soil}$, Eq. 61). The inputs from upstream F_{up} are the sum of $F_{fast\ out}$, $F_{slow\ out}$, $F_{river\ out}$ of all basins *i-1* lying directly upstream (Eq. 62), and inputs into swamps ($F_{up2swamp}$, Eq. 63), S_{flood} ($F_{up2flood}$, Eq. 64) and S_{river} ($F_{up2river}$, Eq. 65) have all the same concentrations as F_{up}.

$$F_{*\ out,C\ spec,i,t} = F_{*\ out,H_2O,i,t} \cdot \frac{S_{*,C\ spec,i,t}}{S_{*,H2O,i,t}} \tag{60}$$

*: 'fast', 'slow', 'stream', or 'flood' reservoir; C spec: DOC_{lab}, DOC_{ref}, CO_2

$$F_{flood2soil,C\ spec,i,t} = F_{flood2soil,H_2O,t} \cdot \frac{S_{flood,C\ spec,i,t}}{S_{flood,H_2O,i,t}} \tag{61}$$

$$F_{up,C\ spec,i,t} = \sum_{i-1} (F_{fast\ out,C\ spec,i-1,t} + F_{slow\ out,C\ spec,i-1,t} + F_{river\ out,C\ spec,i-1,t} \tag{62}$$

$$F_{up2swamp,C\ spec,t} = F_{up2swamp,H_2O,t} \cdot \frac{F_{up,C\ spec,i,t}}{F_{up,H_2O,i,t}} \tag{63}$$

$$F_{up2flood,C\ spec,i,t} = F_{up2flood,H_2O,t} \cdot \frac{F_{up,C\ spec,i,t}}{F_{up,H_2O,i,t}} \tag{64}$$

$$F_{up2river,C\ spec,i,t} = F_{up,C\ spec,i,t} - F_{up2swamp,C\ spec,i,t} - F_{up2flood,C\ spec,i,t} \tag{65}$$

As discussed above, in the routing scheme, we distinguish two pools of DOC: the labile (DOC_{lab}), which corresponds to the active DOC pool of the soil carbon module, and the

refractory pool (DOC_{ref}), which combines the slow and passive pool of the soil carbon module. For each pool, the DOC stocks in S_{fast} and S_{low} are then updated from the balance between the C inputs simulated in the soil carbon module at 30 minute time-step and aggregated to the one day time step of the routing module, and the outflows of C which are proportional to the water fluxes (Eqs. 66, 67). S_{river} in basin i is augmented by the sum of outflows from the fast, slow and river reservoirs of the basins located directly upstream (i-1), minus the flows diverted to the subsoil of swamps and into floodplains (Eq. 68). The floodplains (S_{flood}) receive inputs from upstream ($F_{up2flood}$) and transfer C to the river reservoir ($F_{flood\ out}$) and via infiltration into the soil ($F_{flood2soil}$) (Eq. 69). The inputs of DOC from the decomposition of inundated SOC and litter are added to S_{river} and S_{flood} according to their contribution to the total fraction of inundated soil ($\%flood_{total}$).

$$S_{fast,C\ spec,i,t+1} = S_{fast,C\ spec,i,t} + F_{RO,C\ spec,i,t} - F_{fast\ out,C\ spec,i,t} \quad (66)$$

$$S_{slow,C\ spec,i,t+1} = S_{slow,C\ spec,i,t} + F_{DR,C\ spec,i,t} - F_{slow\ out,C\ spec,i,t} \quad (67)$$

$$S_{river,C\ spec,i,t+1} = S_{river,C\ spec,i,t} + F_{up2river,C\ spec,i,t} + F_{flood\ out,C\ spec,i,t} - F_{river\ out,C\ spec,i,t} + \sum_{v=1}^{13} \left(F_{soil2flood,C\ spec,grid\ x,v,t}\right) \cdot \frac{dt}{day} \cdot \frac{\%flood_{river,i,t} \cdot A_{total,i}}{\%flood_{total,grid\ x,t} \cdot A_{total,grid\ x}} \quad (68)$$

$$S_{flood,C\ spec,i,t+1} = S_{flood,C\ spec,i,t} + F_{up2flood,C\ spec,i,t} - F_{flood2soil,C\ spec,i,t} + F_{TF,C\ spec,i,t} \cdot \%flood_{i,t} - F_{flood\ out,C\ spec,i,t} + \sum_{v=1}^{13} \left(F_{soil2flood,C\ spec,grid\ x,v,t}\right) \cdot \frac{dt}{day} \cdot \frac{\%flood_{i,t} \cdot A_{total,i}}{\%flood_{total,grid\ x,t} \cdot A_{total,grid\ x}} \quad (69)$$

For Eqs. 68, 69: $F_{soil2flood}$ only for DOC; for CO_2, see Eqs. 83, 84

At each daily time-step, after the lateral transfers along the flow path have been calculated, DOC decomposition and CO_2 evasion within the river and floodplain reservoirs are simulated. The continuous CO_2 production and CO_2 evasion from the aquatic network are computed using a much finer integration time step of 1/240 day (6 min) than the one of the river routing scheme to ensure precision of our numerical scheme. In addition, CO_2 inputs from the decomposition from flooded SOC and litter are also added at the same time-step to represent the continuous additions of CO_2 during the water-atmosphere gas exchange.

For each 6-min time step, the pCO_2 in the water column is calculated from the concentration of dissolved CO_2 and the temperature dependent solubility of CO_2 (K_{CO2}) (Eq. 70). The water

temperature (T_{water}) needed to calculate K_{CO2} (Telmer and Veizer, 1999) (Eq. 71) is derived from the average air temperature close at the ground (T_{ground}) over the whole one-day-time-step of the routing scheme (Eq. 72, R^2=0.56, σ=0.91°C). This equation was empirically derived using values from the ORE-HYBAM dataset (Cochonneau et al., 2006) observed at a 10 day interval over the years 1999 and 2000 at 3 sampling locations (Fig. 7, see Fig. 4 for location). As the linear fits for each sampling location are quite similar (Fig. 6 a), we consider the prediction equation derived for the total of observed data as representative. Note that the slope is quite similar to that (0.82) found by Lauerwald et al. (2015) for average monthly T_{water} using a global data set. Furthermore, we investigated whether the correlations could be improved by introducing a time-lag between T_{water} and T_{ground}, as suggested in the literature (Ducharne, 2008; Van Vliet et al., 2011). However, no significant improvement could be achieved (Fig. 7 b), and we thus maintained Eq. (72) as predictor of water temperature.

$$pCO_{2H_2O,*,i,t} = \frac{S_{*,CO_2,i,t}}{S_{*,H_2O,i,t} \cdot 12.011 \cdot K_{CO_2}} \quad (70)$$

* stands for slow, fast, river, flood

$$K_{CO_2,i,t} = 10^{\left(2.22 \cdot 10^{-6} \cdot T_{water,grid\,x,t}{}^3 + 1.91 \cdot 10^{-5} \cdot T_{water,grid\,x,t}{}^2 + 1.63 \cdot 10^{-2} \cdot T_{water,grid\,x,t} - 1.11\right)} \quad (71)$$

$$T_{water,grid\,x,t} = 6.13°C + 0.80 \cdot T_{ground,grid\,x,t} \quad (72)$$

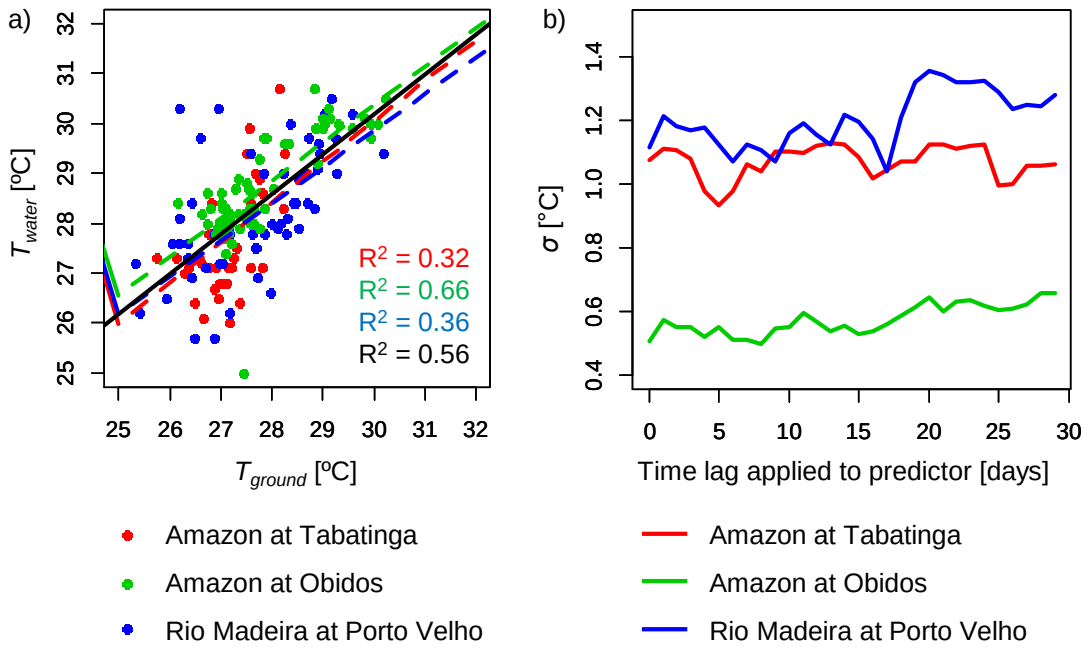


Figure 7. Predictability of water temperature (T_{water}) from simulated ground temperature (T_{ground}). a) Linear regressions between T_{water} and T_{ground} recorded on the same day. The black line represents the linear fit through all data combined, while the coloured dashed lines represent the linear fits per sampling location. b) Changes in RMSE (σ) of the prediction equation per sampling location after applying different time lags to the predictor, T_{ground}.

The same water temperature is used for the calculation of the Schmidt number (*SC*) (Wanninkhof, 1992) (Eq. 73), which is needed to calculate the actual gas exchange velocity from the standard conditions k_{600} (Eqs. 74, 75). We used distinct values of k_{600} for rivers ($k_{river,600}$), and for swamps ($k_{swamp,600}$) to account for the reduced effect of the wind in flooded forests. The value $k_{swamp,600}$ = 0.65 m d^{-1} is taken from Richey et al. (2002) while the value for $k_{river,600}$ = 3.5 m d^{-1} corresponds to the average of the values reported in Alin et al. (2011). For the calculation of $k_{flood,600}$ on the floodplains, we assumed that open floodplains have the same gas exchange velocity than the rivers, while within flooded forests (represented by *%swamp*), the gas exchange velocity is set to $k_{swamp,600}$. As the gas exchange is calculated for the whole floodplain, and is thus a combination of open-water floodplain and swamps, the average k_{flood} is calculated according to the vegetated and open proportions (Eq. 75). In rivers and floodplains, the CO_2 evasion is calculated based on the pCO_2, the gas exchange velocity, and the surface water area available for gas exchange, which changes at the daily time-step (Eqs. 76, 77). The maximum possible CO_2 evasion per time-step is constrained by the amount of dissolved CO_2 in excess to the hypothetical equilibrium with the atmospheric pCO_2. For S_{fast}, for which a surface area is not known, full equilibration with the atmosphere is assumed (Eq. 78). For S_{slow}, which we consider as groundwater storage even though a ground water table itself is not simulated, no gas exchange is assumed.

$$SC_{i,t} = 1911 - 118.11 \cdot T_{water,grid\ x,t} + 3.453 \cdot {T_{water,grid\ x,t}}^2 - 0.0413 \cdot {T_{water,grid\ x,t}}^3 \quad (73)$$

$$k_{river,i,t} = k_{river,600} \cdot \sqrt{\frac{600}{SC_{i,t}}} \quad (74)$$

$$k_{flood,i,t} = \left(\left(1 - \frac{\%_{swamp}}{\%flood_{max}}\right) \cdot k_{river,600} + \left(\frac{\%_{swamp}}{\%flood_{max}}\right) \cdot k_{swamp,600}\right) \cdot \sqrt{\frac{600}{SC_{i,t}}} \quad (75)$$

$$F_{river2atm,CO_2,i,t} = \min(K_{CO_2,i,t} \cdot \left(pCO_{2\ river,i,t} - pCO_{2\ atm,t}\right) \cdot 12.011 \cdot A_{river\ act,i,t} \cdot k_{river,CO_2,i,t} \cdot \frac{dt}{\text{day}} \cdot 10^3, K_{CO_2,i,t} \cdot \left(pCO_{2\ river,i,t} - pCO_{2\ atm,t}\right) \cdot 12.011 \cdot S_{river,H_2O,i,t} \cdot 10^3) \quad (76)$$

$$F_{flood2atm,CO_2,i,t} = \min(K_{CO_2,i,t} \cdot \left(pCO_{2\ flood,i,t} - pCO_{2\ atm,t}\right) \cdot 12.011 \cdot A_{flood,i,t} \cdot k_{flood,CO_2,i,t} \cdot \frac{dt}{\text{day}} \cdot 10^3, K_{CO_2,i,t} \cdot \left(pCO_{2\ flood,i,t} - pCO_{2\ atm,t}\right) \cdot 12.011 \cdot S_{flood,H_2O,i,t} \cdot 10^3) \quad (77)$$

$$F_{fast2atm,CO_2,i,t} = K_{CO_2,i,t} \cdot \left(pCO_{2\ fast,i,t} - pCO_{2\ atm,t}\right) \cdot 12.011 \cdot S_{fast,H_2O,i,t} \cdot 10^3 \quad (78)$$

The instream decomposition of terrestrial DOC is calculated using base rate constants for labile and refractory DOC, k_{DOClab} = 0.3 day^{-1} and k_{DOCref} = 0.01 day^{-1}, respectively (Eqs. 79, 80). These values correspond to half-live times of 2 days and 80 days respectively. The value for k_{DOClab} is thus in agreement with Devol and Hedges (2001) who conclude that DOC_{lab} in the Amazon river must have a very short half-life of hours to a few days. k_{DOCref} also corresponds to the lower range of respiration rates found for Rio Solimoes of 0.2 µM h^{-1} (Amon and Benner, 1996) if an average concentration of about 5 mg C L^{-1} is assumed (cf. Moreira-Turcq et al., 2003). We assumed that the values for the rate constants are valid for an average T_{water} of 28°C (consistent with experiments of Amon and Benner, 1996 and the average temperature simulated here) and apply a temperature sensitivity factor on decomposition rates after Hanson et al. (2011) (Eqs. 79,80).

$$S_{*,\mathrm{DOC}_{lab},i,t+1} = S_{*,\mathrm{DOC}_{lab},i,t} - S_{*,\mathrm{DOC}_{lab},i,t} \cdot \frac{k_{DOC_{lab}} \cdot dt}{day} \cdot 1.073^{(T_{water,i,t}-28)} \quad (79)$$

$$S_{*,\mathrm{DOC}_{ref},i,t+1} = S_{*,\mathrm{DOC}_{ref},i,t} - S_{*,\mathrm{DOC}_{ref},i,t} \cdot \frac{k_{DOC_{ref}} \cdot dt}{day} \cdot 1.073^{(T_{water,i,t}-28)} \quad (80)$$

*: fast, slow, river, flood

$$S_{slow,\mathrm{CO2,i,t+1}} = S_{slow,\mathrm{CO2,i,t+1}} + S_{slow,\mathrm{DOC}_{lab},i,t} \cdot \frac{k_{DOC_{lab}} \cdot dt}{day} + S_{slow,\mathrm{DOC}_{ref},i,t} \cdot \frac{k_{DOC_{ref}} \cdot dt}{day} \quad (81)$$

$$S_{fast,\mathrm{CO2,i,t+1}} = S_{fast,\mathrm{CO2,i,t+1}} + S_{fast,\mathrm{DOC}_{lab},i,t} \cdot \frac{k_{DOC_{lab}} \cdot dt}{day} + S_{fast,\mathrm{DOC}_{ref},i,t} \cdot \frac{k_{DOC_{ref}} \cdot dt}{day} - F_{fast2atm,CO_2,i,t} \quad (82)$$

$$S_{river,\mathrm{CO2,i,t+1}} = S_{river,\mathrm{CO2,i,t+1}} + S_{river,\mathrm{DOC}_{lab},i,t} \cdot \frac{k_{DOC_{lab}} \cdot dt}{day} + S_{river,\mathrm{DOC}_{ref},i,t} \cdot \frac{k_{DOC_{ref}} \cdot dt}{day} - F_{river2atm,CO_2,i,t} + \sum_{v=1}^{13} \left(F_{soil2flood,CO_2,grid\ x,v,t}\right) \cdot \frac{dt}{day} \cdot \frac{\%flood_{river,i,t} \cdot A_{total,i}}{\%flood_{total,grid\ x,t} \cdot A_{total,grid\ x}} \quad (83)$$

$$S_{flood,\mathrm{CO2,i,t+1}} = S_{floos,\mathrm{CO2,i,t+1}} + S_{flood,\mathrm{DOC}_{lab},i,t} \cdot \frac{k_{DOC_{lab}} \cdot dt}{day} + S_{flood,\mathrm{DOC}_{ref},i,t} \cdot \frac{k_{DOC_{ref}} \cdot dt}{day} - F_{flood2atm,CO_2,i,t} + \sum_{v=1}^{13} \left(F_{soil2flood,CO_2,grid\ x,v,t}\right) \cdot \frac{dt}{day} \cdot \frac{\%flood_{i,t} \cdot A_{total,i}}{\%flood_{total,grid\ x,t} \cdot A_{total,grid\ x}} \quad (84)$$

At each 6-min time-step, the CO_2 produced from the decomposition of DOC is added to the relevant reservoirs (Eqs. 81-84). For S_{fast}, S_{river}, and S_{flood}, the amount of evading CO_2 is subtracted from the CO_2 stocks (Eqs. 82-84). For S_{river} and S_{flood}, the inputs of CO_2 from the decomposition of inundated SOC and litter are added to these reservoirs, based on the relative contribution of swollen rivers ($\%flood_{river}$) and floodplains ($\%flood$) on the total fraction of inundated soils ($\%flood_{total}$) (Eqs. 83-84).

8.2. Modelling northern peatlands area and carbon dynamics since the Holocene with the ORCHIDEE-PEAT land surface model (SVN r5488)

Chunjing Qiu[1], Dan Zhu[1], Philippe Ciais[1], Bertrand Guenet[1], Shushi Peng[2], Gerhard Krinner[3], Ardalan Tootchi[4], Agnès Ducharne[4], Adam Hastie[5],

1. Laboratoire des Sciences du Climat et de l'Environnement, UMR8212, CEA-CNRS-UVSQ F-91191 Gif sur Yvette, France
2. Sino-French Institute for Earth System Science, College of Urban and Environmental Sciences, Peking University, 100871 Beijing, China
3. CNRS, Université Grenoble Alpes, Institut de Géosciences de l'Environnement (IGE), F-38000 Grenoble, France
4. Sorbonne Université, CNRS, EPHE, Milieux environnementaux, transferts et interaction dans les hydrosystèmes et les sols, Metis, F-75005 Paris, France
5. Department of Geoscience, Environment and Society, Université Libre de Bruxelles, 1050 Bruxelles, Belgium

Correspondence: Chunjing Qiu (chunjing.qiu@lsce.ipsl.fr)

Abstract

The importance of northern peatlands in the global carbon cycle has been recognized, especially for long-term changes. Yet, the complex interactions between climate and peatland hydrology, carbon storage and area dynamics make it challenging to represent these systems in land surface models. This study describes how peatland are included as an independent sub-grid hydrological soil unit (HSU) into the ORCHIDEE-MICT land surface model. The peatland soil column in this tile is characterized by multi-layered vertical water and carbon transport, and peat-specific hydrological properties. The cost-efficient version of TOPMODEL and the scheme of peatland initiation and development from the DYPTOP model, are implemented and adjusted, to simulate spatial and temporal dynamics of peatland. The model is tested across a range of northern peatland sites and for gridded simulations over the Northern Hemisphere (>30 °N). Simulated northern peatland area (3.9 million km^2), peat carbon stock (463 PgC) and peat depth are generally consistent with observed estimates of peatland area (3.4 – 4.0 million km^2), peat carbon (270 – 540 PgC) and data compilations of peat core depths. Our results show that both net primary production (NPP) and heterotrophic respiration (HR) of northern peatlands increased over the past century in response to CO_2 and climate change. NPP increased more rapidly than HR, and thus net ecosystem production (NEP) exhibited a positive trend, contributing a cumulative carbon storage of 11.13 Pg C since 1901, most of it being realized after the 1950s.