Wageningen University and Research

Master Thesis

A statistical model of the biospheric carbon cycle for CarbonTracker

Author:
Rolf Van Der Vleugel

Supervisors:
Ir. G. B. Koren
Prof. Dr. W. Peters

February 12, 2019

Contents

1 Introduction

The rate of anthropogenic carbon dioxide (CO_2) emissions have reached an all-time high since pre-industrial times, due to population and economic growth (Pachauri et al., 2014). The natural carbon sinks are not able to keep up with these CO_2 emissions causing a global atmospheric CO_2 increase (Ballantyne et al., 2012). The higher atmospheric CO_2 concentration leads to more absorption of longwave radiation, causing a global temperature increase. CO_2 is the most important driver for climate change, as its radiative forcing compromises 80% of the total increase in radiative forcing by greenhouse gases from 1750 until 2011 (Myhre et al., 2013). According to the Intergovernmental Panel on Climate Change (IPCC), the period from 1983 to 2012 is seen as the warmest 30 year period of the past 1400 years (Pachauri et al., 2014). Moreover, 2015 was the first year anthropogenic activity led to an 1 °C global temperature increase compared to pre-industrial levels (Hawkins et al., 2017).

Next to anthropogenic CO_2 emissions, the atmospheric carbon balance is linked to the biosphere and hydrosphere (Quéré et al., 2015). Within the biosphere, the Amazon Basin takes up and releases twice the amount of carbon per year compared to the worldwide anthropogenic fossil fuel emissions through photosynthesis and respiration (Phillips et al., 2009). The Amazon basin stores a vast amount of carbon and acts as long-term net biomass sink (Gatti et al., 2014; Brienen et al., 2015). Compared to the 1990s, the biomass mortality of the Amazon Basin has increased throughout as tree mortality increased, while above-ground biomass growth rate is recently levelling off (Brienen et al., 2015). A driver for this Amazon Basin carbon decrease is very likely linked to larger interannual climate variabilities (Brienen et al., 2015).

In 2005, one of the most intense droughts of the past hundred years hit the Amazon Basin. Droughts in Amazonia are mostly caused by El Niño-Southern Oscillation (ENSO), however this drought was caused by anomalously high tropical North Atlantic sea surface temperatures (Phillips et al., 2009). This extreme weather event is linked to climate change, which is likely to increase the severity and frequency of extreme weather events (Frank et al., 2015). The 2005 event led to increased moisture stress and eventual forest biomass loss. The drought reversed Amazonia from long-term carbon sink to a carbon source. And showed an exceptional global atmospheric CO_2 growth rate, which may be partially linked to the Amazon drought (Phillips et al., 2009).

During the anomalously long-term Amazon drought in 2010, Gatti et al. (2014) measured the vertical profile of carbon monoxide (CO) and CO_2 in the atmosphere above the Amazon Basin. From these measurements a more accurate seasonal and annual carbon balance was set up. The results of the measurements showed a carbon loss of around 0.5 petagrams of carbon in 2010, while the next year an anomalously wet year occurred, which showed a neutral carbon balance. The feedback mechanisms between climate and tropical land carbon pools are one of the key uncertainties to predict global climate (Cox et al., 2013; Gatti et al., 2014).

Therefore, additional constrains on carbon exchange were added by Laan-Luijkx et al. (2015), which include satellite observations of fire hot spots, burned area and CO. With the use of the data assimilation system CarbonTracker, results confirmed a reduction of carbon uptake by vegetation in 2010 (Laan-Luijkx et al., 2015). CarbonTracker makes use of inverse modelling in which four surface models in combination with atmospheric CO_2 observations is able to optimize global surface CO_2 fluxes. This includes for example the optimization of biospheric fluxes and are optimized as follows (Peters et al., 2007, 2010a; Van der Laan-Luijkx et al., 2017):

$$F(x,y,t) = \lambda_R \cdot F_{bio}(x,y,t) + \lambda_R \cdot F_{oce}(x,y,t) + F_{fossil}(x,y,t) + F_{fire}(x,y,t) \quad (1)$$

Here $F(x,y,t)$ represents the total carbon fluxes at each longitude, latitude and for each time step. $F_{bio}(x,y,t)$ and $F_{oce}(x,y,t)$ are biospheric and oceanic a priori carbon fluxes respectively, obtained from pre-calculated space-time biosphere and ocean models. $F_{fossil}(x,y,t)$ and $F_{fire}(x,y,t)$ represent the carbon fluxes of the fossil fuel and fire emissions respectively, and λ_R represents the scaling factor of each region which is the final assimilation product to determine the CarbonTracker carbon fluxes. This approach assumes the coupling of fluxes between gridboxes in one region and therefore suggests the vegetation response to climate conditions to be captured correctly. Such data assimilation systems are computational expensive because of its size and complexity (see Section 2) and for studies on longer times scales, simplification of its components could help to make faster progress and reveal more detail of carbon fluxes within ecosystems such as the net ecosystem exchange (NEE). This is the net carbon flux of photosynthesis and respiration (see Section 2). Moreover, with the use of a lightweight biosphere model, NEE can be resolved at a higher spatial resolution, showing different responses within a single ecosystem.

This research therefore aims to decrease the computing costs of CarbonTracker, which uses the computational expensive Simple-Biosphere-Model-Carnegie-Ames Stanford Approach (SiB-CASA) as input for the biospheric carbon fluxes. In my project, I will build a lightweight biosphere model which parameterizes the SiBCASA NEE output, allowing us to simulate biospheres' CO_2 exchange at a much smaller computational burden.

The lightweight model will focus on seasonal and interannual variability (IAV) as well as long-term trends, and at its most basic is a simple statistical model. Here the seasonal cycle of NEE will be captured by using the least square regression model of Thoning et al. (1989), which he used to smoothen atmospheric CO_2 observations. The NEE IAV will be based on the work of Rödenbeck et al. (2018), which uses the sensitivity of carbon fluxes versus temperature anomalies to capture the IAV based on (Cox et al., 2013). The parameters used in it will statistically link the behaviour of the biosphere to essential climate variables such as temperature and precipitation, which helps to compare similar statistical relations typically derived by climate models.

This leads us to our research questions:

1. To what extent can we reproduce biospheric CO_2 NEE fluxes from SiBCASA by a (temperature or precipitation dependent) statistical model inspired by Thoning et al. (1989) and Rödenbeck et al. (2018) for different spatial and temporal scales?

2. To what extent can the (temperature or precipitation dependent) statistical model save computational costs at different spatial and temporal scales?

3. What is the sensitivity of atmospheric CO_2 in forward atmospheric simulations using TM5 to replace SiBCASA fluxes with our (temperature or precipitation dependent) statistical model for different station locations?

To answer our research questions, Section 2 will first give background information about the seasonal, interannual and long-term carbon trends. Moreover, the biospheric carbon cycle, climate-carbon models and inverse modelling will be further elaborated in this section. In Section 3 we answer our first research question and start with the construction of the temperature or precipitation dependent statistical model after which we analyze and interpret our results. In Section 4, we answer our second research question with the use of experiments, which can spatially and temporally save computational costs. In Section 5, we answer our third research question, in which we run an atmospheric transport model with different inputs with SiBCASA as reference. Section 6 will discuss the results and elaborate on possible recommendations and follow up research. Section 7 will finally include the conclusions of this thesis.

2 Background information

2.1 Seasonal and interannual variability of the carbon cycle

To understand the trends and variability of terrestrial carbon exchange, observational records of CO_2 are held since the late 1950s at Mauna Loa (Hawaii) (NOAA a, 2016). The CO_2 concentration at observation stations is described by the quantity mole fractions, which is defined as the number of CO_2 moles in a given number of molecules of dry air in parts per million (ppm) (NOAA b, 2018). Normally, we express gas concentrations as the amount of molecules per cubic meter. However, to quantify the addition and removal of carbon to the atmosphere, we can not rely on the concentration gas per cubic meter. It namely depends mainly on temperature and pressure, and secondly depends on how much the relative abundance of each gas has been diluted by water vapour and is highly variable. When we apply the dry air mole fraction, the measurement is thus not affected by different measuring heights or seasonal temperature fluctuations. This quantitatively informs us about the removals and emissions of CO_2.

From the CO_2 time series, we can derive a seasonal pattern, regulated by photosynthesis and respiration (Figure 1). The larger landmass on the Northern hemisphere accommodates more biomass, and has higher photosynthetic uptake of CO_2 compared to the southern hemisphere. This gives an CO_2 maximum in the winter of the northern hemisphere and a minimum in the summer of the northern hemisphere, which explains the seasonal cycle (Thoning et al., 1989). We can also derive a long-term CO_2 concentration trend from the time series, which is caused by anthropogenic activity, however the long-term trend of CO_2 does not follow a perfect linear function as IAV plays a role as well.
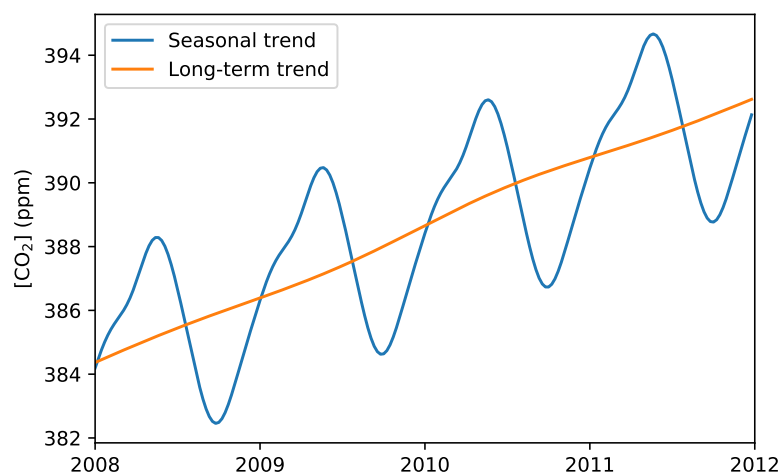

Figure 1: Seasonal and long-term trend of CO_2 at Mauna Loa from 2008 until 2012 derived from NOAA c (2018) with the methods of Section 3.2.2. The CO_2 concentration is expressed in mole fraction.

For IAV in terrestrial carbon fluxes, three main drivers can be designated. These are (1) metabolic changes in natural processes, such as ecosystem respiration, due to climate variations (ENSO, droughts), (2) natural disturbances (fires, storms) and (3) land-use change (Houghton, 2000). This research only focuses on climate variations as variability in land-use change is expected to be small for two reasons. First of all, land-use change only significantly varies on local scale and generally less on regional or global scale. Secondly, only half the CO_2 emissions due to land use change are released into the atmosphere in the first year of the disturbance, while the remaining fraction will be released more slowly in subsequent years and thus contribute less to IAV. Land-use changes likely account for only 5-10% of the global IAV in observed CO_2. Natural disturbances are also likely to only account for 5-10% of the interannual variations. Therefore climate variations due to natural processes, are the main factors for interannual terrestrial carbon flux variability.

The general oscillation pattern of the growth rate of CO_2 can be explained by an El Niño (warming phase of ENSO), followed by a La Niña (cooling phase of ENSO), (Figure 2), (Thoning et al., 1989). In 1997-1998 and 2015-2016, two of the three strongest El Niño events occurred and led to a strong growth rate of CO_2 (L'Heureux et al., 2017). In 2005 not an ENSO event, but likely the long-term drought in the Amazon Basin caused a strong increase in the CO_2 growth rate, due to a lower respiration rate and a higher photosynthetic rate of the biosphere (Phillips et al., 2009).
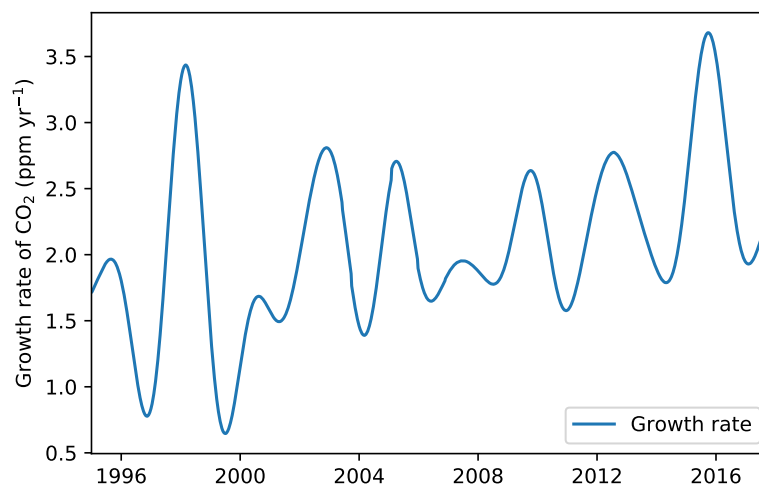
Figure 2: Growth rate of CO_2 at Mauna Loa from 2008 until 2012 derived from NOAA c (2018) with the methods of Section 3.2.2. The CO_2 concentration is expressed in mole fraction.

The decomposition of the CO_2 time series in the seasonal, long-term and interannual trends give more insights in corresponding processes and their magnitude. This method will also be applied for the NEE output fluxes of SiBCASA. The methods of the decomposition of the time series is elaborated in Section 3.2.2.

2.2 Carbon-climate relations in the biosphere

The photosynthesis process of vegetation is a powerful mechanism for carbon exchange fluxes in the atmosphere-biosphere carbon pools (Bonan, 2015). The process includes CO_2, water and absorbs light to produce carbohydrates, which serves as energy source for the plants. When the stomata of the plants absorb the CO_2, water is simultaneously lost due to transpiration. Leaf respiration is due to the oxidation of organic compounds for growth and maintain plant function.

In general, higher temperatures increase the enzymatic activity of leaves, which increases photosynthesis, respiration and water transpiration (Figure 3), (Bonan, 2015). At temperatures between 20 and 35 °Celsius, an enzymatic maximum for the photosynthesis is reached, which applies for the uptake of photon and CO_2 uptake as well (Figure 3). However, when temperature increases too much, the photosynthesis rate drops, as result of an excessive amount of water lost through transpiration during photosynthesis. This results in a net unfavourable trade of water versus energy. The leaf respiration keeps increasing as temperature increases due to temperature induced enzyme activity until the leaf dies off. Microbial and root respiration follow the same pattern as the leaf respiration.
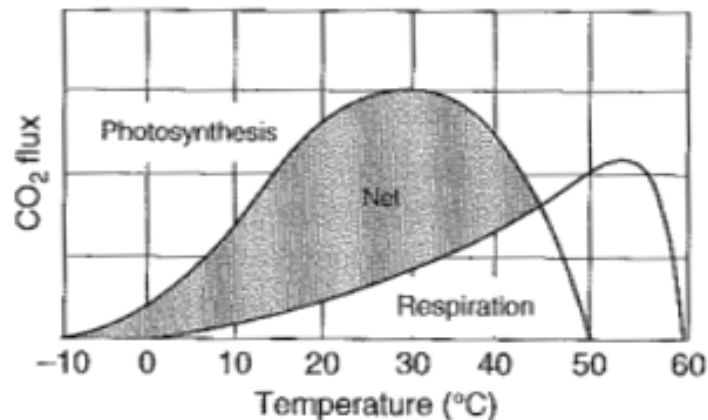

Figure 3: effect of temperature on leaf photosynthesis and respiration(Bonan, 2015). Note that this is a general leaf response.

Photosynthesis and respiration are linked to the carbon balance as follows (figure 4), (Schaefer et al., 2008);

$$NEE = R_H + R_A - GPP \qquad (2)$$

$$NPP = R_A - GPP \qquad (3)$$

NEE is the net ecosystem exchange of CO_2 at which a positive NEE indicates net CO_2 flux from the biosphere into the atmosphere. All fluxes which thus tend to increase the CO_2 concentration, will have a positive value, which is called the "atmospheric perspective". Heterotrophic respiration (R_H) is the carbon dioxide release by microorganisms during the decay of organic material and autotrophic respiration (R_A) is the release by plants during growth and maintenance. GPP is the gross primary production, which resembles the total carbon uptake from the atmosphere to the biosphere. Net primary production (NPP) is the total flux of the vegetation and does not include the respiration of the soil microbes.
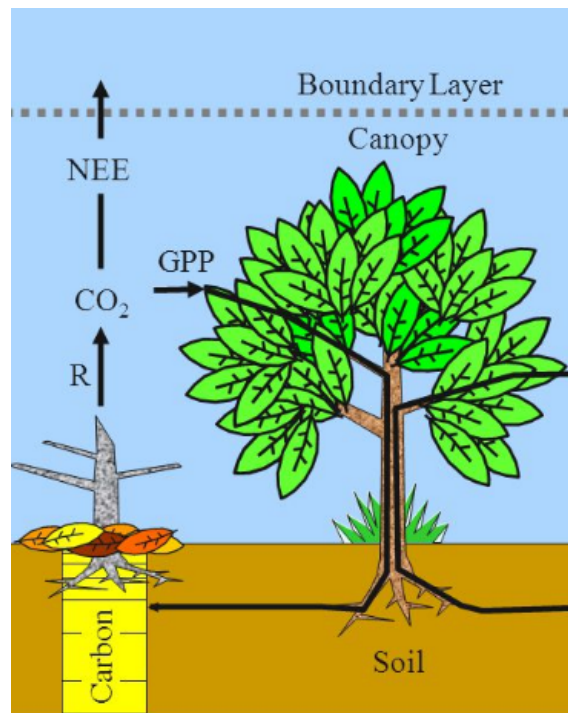

Figure 4: Visualization of the terrestrial carbon cycle (Schaefer et al., 2008). Note that R is the total of R_H and R_A.

The increase of CO_2 and global surface temperature as result of anthropogenic activity (see Section 2.1) will stimulate plants' photosynthesis rate, which is called the CO_2 fertilization effect. This causes the dynamic carbon equilibrium between biosphere and atmosphere to shift towards a larger biosphere carbon storage (Bonan, 2015). The future climate could also enhance plant decomposition, which improves the nutrient availability of plants. Extra nitrogen mineralization stimulates the mobilization of nitrogen from the decomposition process of soil organic matter, which enhances carbon sequestration, however is not included in SiBCASA and thus not represented in our statistical model. The statistical model will use the NEE output fluxes of SiBCASA and try to represent these seasonal and IAV. We will not further analyze the CO_2 fertilization effect as we focus on a ten year timescale, which is too short to investigate its effect on the terrestrial carbon fluxes.

2.3 Carbon-climate models

The trends and variability of terrestrial carbon exchange is observed by local measurements of eddy covariance towers (Baldocchi et al., 2001). The observation sites are part of a global measurements network called FLUXNET. Herewith we can study the spatial and temporal patterns of ecosystem-scale of CO_2 exchange among others. However, observations in especially the tropics are scarce and to quantify regional and global carbon fluxes, we need coupled climate-carbon-cycle models, which can estimate the future atmospheric CO_2 pathways and its impact on global climate. In a coupled climate-carbon-cycle model, the ocean and land carbon cycles feed back on the atmosphere, for example through CO_2 and heat fluxes, to contribute to further climate change (Cox et al., 2013). For an uncoupled model, changes in the ocean and land carbon reservoirs are not fed back into the atmosphere to contribute to further climate change .

The magnitude of the carbon flux from the biosphere remains uncertain according to the results of the Coupled Climate Carbon Cycle Model Intercomparison Project (C[4]MIP) (Cox et al., 2013). These large uncertainties can mainly be explained by vegetation response to warming and variations in rainfall predictions in the Amazon Basin. To explain the effects of atmospheric CO_2 and temperature increase (T_T) across the wide range of models and responses simulated in the C[4]MIP experiments, the effects of temperature and CO_2 as forcing of changes in land carbon uptake are often separated (Friedlingstein et al., 2006). This separation linearizes the response of the coupled climate models such that the total simulated change is the result of changes in atmospheric temperature (T_T) and CO_2 mixing ratio. For example, the tropical land carbon storage (ΔC_{LT}) can be expressed as (Cox et al., 2013);

$$\Delta C_{LT} = \beta_{LT} \Delta C_a + \gamma_{LT} \Delta T_T \tag{4}$$

Note that "LT" specifically applies to land in the tropics (Wenzel et al., 2014). Here β_{LT} and γ_{LT} are the land carbon storage sensitivities to direct CO_2 effects and climate change induced effects, and typical units are GtC/ppm and GtC/K respectively. These sensitivity factors can easily be compared across many different models. To determine the γ_{LT}, first an uncoupled climate-carbon model (CLTU) determines the β_{LT}, and hereafter with the use of a coupled climate-carbon model C_{LT}^c the γ_{LT} is calculated (Wenzel et al., 2014);

$$\gamma_{LT} = \frac{\Delta C_{LT}^c + \Delta C_{LT}^u}{\Delta T_T^c} \tag{5}$$

γ_{LT} is a key uncertainty in climate-carbon models and Cox et al. (2013) presented an emergent constraint on γ_{LT} across models from the C[4]MIP study. He suggested that short term processes which affect CO_2 fluxes especially in the tropics, dominate the long-term CO_2 response as well, and therefore the currently observed response of CO_2 mixing ratios to variations in observed temperatures is a measure of γ_{LT} (Figure 2). With the use of CO_2 and temperature measurements, a tight constraint on the γ_{LT} is provided by this relationship (Figure 5).
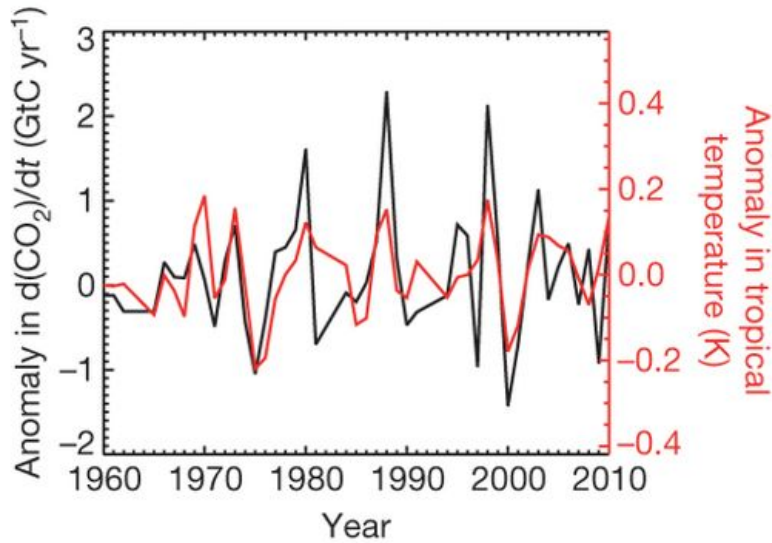
Figure 5: Annual anomalies in CO_2 growth rate (black) and tropical temperature (red) versus year from 1960 until 2010 (Cox et al., 2013).

These emergent constrains will help to develop a coupled carbon-climate model that is more reliable for projections, as it is closer to observational data than the full ensemble of models (Borodina et al., 2017). Temperature and precipitation anomalies have a strong climate co-variance, and as tropical terrestrial carbon exchange is the largest uncertainty due to mostly droughts, precipitation and temperature anomalies will be applied to capture the climate sensitivity (Berg et al., 2015). Observational data can thus be used to develop model constrains and deepen the understanding of the carbon cycle. This forms the base of the approach we will take in this research, targeting the climate sensitivity parameter to capture the IAV in the NEE time series with the use of T or P anomalies. These will from here be referenced as γ_{NEE-T} and γ_{NEE-P}.

2.4 Modelling the carbon cycle

In forward carbon cycle modelling, surface forcing by land and ocean carbon fluxes, initial conditions for CO_2, and weather patterns that drive atmospheric transport are combined to simulate CO_2 mixing ratios. The carbon (in terms of CO_2) fluxes are represented as output of the biospheric SiBCASA model and the CO_2 mixing ratios are the output variables of the global chemical transport model TM5 (Figure 6).

SiBCASA consists of the model SiB, which contains biophysical and photosynthesis calculations and the CASA model, which focuses on the biogeochemistry aspects. The SIBCASA hybrid model has the capabilities to estimate terrestrial carbon and biomass fluxes from diurnal to decadal timescales and is further elaborated in Section 3.2.1. A detailed description of chemical and biological processes that determine vegetation response is available in Schaefer et al. (2008); Sellers et al. (1996a,b).

The TM5 model can amongst others be applied for inversion studies and allows for high spatial resolution within a coarser global grid and ensures regional estimates to be consistent with global constraints. TM5 simulates vertical diffusion, convection and advection in both the free troposphere and boundary layer. The European Centre for Medium-Range Weather Forecasts (ECMWF) meteorological data fields drive the model, but are first pre-processed into coarser grids to retrieve a flow which conserves the tracer mass (CarbonTracker Team, 2017). More elaborate information is given in Section 5.2.1.

With inverse modelling, atmospheric observations can be integrated into further analysis of the carbon cycle. With this method, initial conditions, or the surface forcing (as controlled by input parameters in the land- and ocean flux models) can be optimized, in this case, to decrease the uncertainty of the SiBCASA CO_2 fluxes. However, computational restrictions typically cause inverse modelling studies, such as Gurney et al. (2002), to be limited to a small number of large regions for which CO_2 fluxes can be estimated across the globe. With the use of a more advanced data-assimilation system, Peters et al. (2007, 2010b) was able to overcome these computational limitations and called the inverse modelling system "CarbonTracker". With the use of CarbonTracker regional CO_2 fluxes are now regularly estimated (Laan-Luijkx et al., 2015).
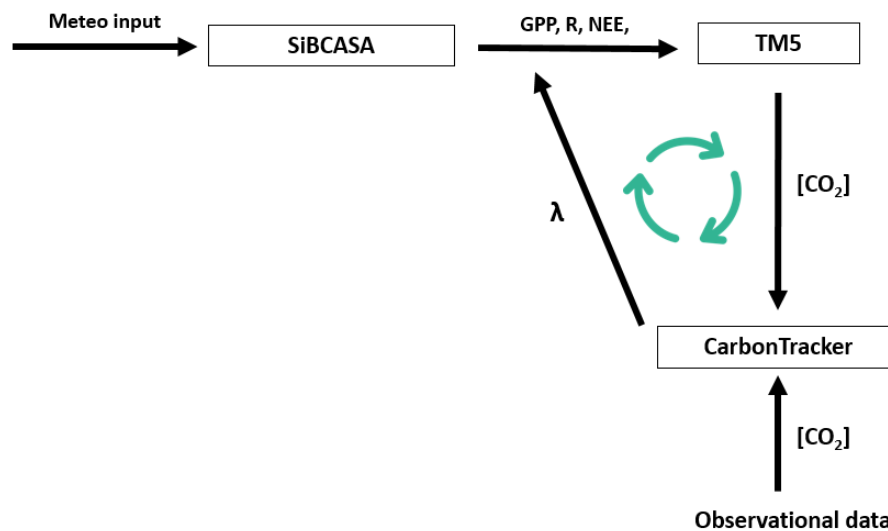

Figure 6: Simplified CarbonTracker framework, in which the fire, fossil and ocean emissions are not taken into account. The final CarbonTracker assimilation product λ is computational expensive to calculate on longer timescales.

For the optimization in inverse modelling, an a-priori estimate of the unknowns (e.g. SiB-CASA's output CO_2 fluxes) is needed to optimize the λ's, the final assimilation product (Figure 6. A prior is the initial distribution of these unknown parameters and is used in the Bayesian analysis. Starting from the distribution of the parameter of interest, the posterior distribution is estimated with observational data to provide the model with extra information about the parameter. The parameters to be estimated can be CO_2 surface fluxes themselves (as for CarbonTracker on the SiBCASA output) or scaling factors on a-priori calculated surface fluxes to make them larger or smaller (Peters et al., 2007), or even biosphere model parameters that control the surface fluxes themselves (Friedlingstein et al., 2006). However, we focus on longer timescales to capture the seasonal and IAV in the NEE time series, which is computational costly as the scaling factor λ needs to be calculated weekly (Koren and Peters, 2018). A promising method to save computational costs is the optimization of climate-relevant parameters which was recently pioneered by Rödenbeck et al. (2018). (Rödenbeck et al., 2018) applied linear regression of temperature (taken as climate proxy) and NEE anomalies fitted to atmospheric CO_2 data to capture NEE IAV with the use of γ_{NEE-T} and showed interpretable and distinct patters along seasons and latitudes. We will not use an atmospheric inversion to constrain our statistical model, but still use the method of Rödenbeck et al. (2018) as guideline further explained in Section 3.2.2.

The statistical model can be implemented within the CarbonTracker iteration process during the inversion run, in which γ_{NEE-T} or γ_{NEE-P} will be optimized instead of the output fluxes as for SiBCASA (Section 3.1. This ensures that a redistribution of the output fluxes is avoidable and therefore the output CO_2 fluxes can independently be optimized for each area. Additionally, we will try to decrease the computational costs (Section 4), lead to a possibility to optimize the statistical model CO_2 fluxes on a higher spatial resolution compared to the SiBCASA inversion run when more measurement sites are set-up.

3   Construction and performance analysis of the lightweight statistical model

3.1   Introduction

Biosphere models are uncertain about effects of climate change and its corresponding extreme weather events on the carbon cycle (Cox et al., 2013). Moreover, IAV such as the 2010 drought in the Amazon for example, caused a different allocation of carbon within the biosphere, resulting in a larger uncertainty within the biosphere models (Doughty et al., 2015). The uncertainties can be reduced with CarbonTracker, which makes use of ensemble data assimilation (Peters et al., 2007). The high computational costs of inverse modelling and SiBCASA limit the spatial resolution and lead to the indirect optimization of the carbon output fluxes of SiBCASA and not the internal model's parameters. Therefore this chapter will use a statistical model to resemble the NEE output signal of SiBCASA. This model is based on the work of Thoning et al. (1989) whom constructed a general statistical curve fit function (fit function) originally set-up to smooth out the observed CO_2 concentration and will be used to reproduce the seasonal NEE cycle. This leads us to our first research question, which determines our baseline performance:

1. To what extent can we reproduce biospheric seasonal NEE fluxes from the generic fit function inspired by Thoning et al. (1989) for different timescales and regions?

To improve the fit function, we introduce an extra climate variable which will be applied to capture the IAV within the NEE signal. For the tropics, temperature anomalies could possibly better explain the NEE IAV as the link with ENSO is well established (Chiang and Sobel, 2002). Strong El Niño years can namely cause a warming of 0.5 - 1°C when averaging the monthly mean temperature over a global tropical strip of 20°S-20°N. The application of using temperature anomalies as climate proxy originates from Rödenbeck et al. (2018) as they state that it also represents other co-varying environmental anomalies. The fit function with temperature anomalies as climate proxy will from here be referred as the T-fit function and includes the climate sensitivity parameter γ_{NEE-T}. The construction of the T-fit function will be explained in the method section .The main goal of this chapter is to quantify how much extra NEE variability can be captured with the T-fit function compared with the original fit function. With the focus on capturing IAV in the NEE signal within the tropical and semi-arid to arid climate zones. This leads us to our second research question of this chapter:

2. To what extent can we capture IAV in the NEE signal reproduced by the T-fit function inspired by Thoning et al. (1989) and Rödenbeck et al. (2018) for different timescales and regions?

We first need to understand the spatial and temporal variability of the NEE output of SiBCASA and relate this to the respiration and photosynthesis term. Hereafter we will analyze the performance of the fit function and compare it with the T-fit function. To analyze the temporal variability we use the grid boxes in which the Amazon Tall Tower Observatory (ATTO) and the Zotino Tall Tower Observatory (ZOTTO) are located. These are two observatory stations which measure CO_2 among others and can be used to validate the output fluxes of SiBCASA in follow-up research(Andreae et al., 2015). The stations are respectively located in a pristine area in the middle of the Amazon Basin and in the boreal climate zone of Siberia, Russia.

These locations are also selected because of their large climatological differences, which should show differences in the significance of the the seasonal and interannual cycle (Andreae et al., 2015; Heimann et al., 2014). In the Northern part of South America, the largest worldwide near-surface temperature effect of ENSO for the months January, February and March are found in the Northern part of South America (Davey et al., 2014). While local weather stations over Northern Eurasia show near-surface temperature IAV during the summer months (Heimann et al., 2014). As temperature and precipitation are linked with the respiration and photosynthesis rate, potential interannual NEE variability can be expected for both cases (Berg et al., 2015).

The third focus of this chapter is to check the hypothesis of Rödenbeck et al. (2018) regarding their use of temperature anomalies as climate proxy. As mentioned in Section 1, the 2005 and 2010 droughts in the Amazon caused the region to turn into a carbon source instead of sink Phillips et al. (2009); Gatti et al. (2014). It therefore seems that precipitation anomalies are possibly a better climate proxy for the tropics. In this climate zone, temperature is not the GPP limiting factor, while for the less wet parts, precipitation can be. Therefore this chapter will also explore the role of precipitation anomalies as climate proxy instead of temperature anomalies for the Amazon region, which leads to our main research question:

3. To what extent can we capture IAV in the NEE signal reproduced by the P-fit function inspired by Thoning et al. (1989) and Rödenbeck et al. (2018) for different timescales and regions?

Also for semi-arid to arid climate zones photosynthesis and respiration are driven by precipitation (anomalies). In the Australian desert, rainfall is highly erratic year-to-year and a large shares falls within a few months (Australian Bureau of Statistics, 2006; Australian Government - Bureau of Meteorology, 2016). The timing of these precipitation peaks varies every year and shows large spatial variability within the region as precipitation mostly occurs in heavy thunderstorms (Australian Bureau of Statistics, 2006). Therefore also for the Australian desert the role of precipitation anomalies as climate proxy will be explored. Precipitation (P) will be constructed using the same method as the T-fit and we will therefore answer research question three for the P-fit function as well.

The section will finally deepen the understanding of γ_{NEE-T} and γ_{NEE-P}, with a spatial and temporal analysis to capture its variability and try to link it to natural processes. Herewith we will compare our findings of γ_{NEE-T} with (Rödenbeck et al., 2018)'s findings, which gives the possibility to validate SiBCASA. This leads us to our final research question:

4. How do γ_{NEE-T} and γ_{NEE-P} vary over time and space and which natural processes can we link to these climate sensitivity parameters?

3.2   Methods

In this subsection, we will first give detailed information of SiBCASA. Hereafter, we will explain and construct the (T/P-)fit function after which we formulate the growth rate and long-term trend. Next, we explain the selection of the regions and gridboxes, which form the base of the results interpretation. Finally we explain the statistical measures we apply throughout the analysis.

3.2.1   SiBCASA

The biosphere module of CarbonTracker called SiBCASA (Schaefer et al., 2008) consists of the model SiB (Sellers et al., 1996a), which contains biophysical and photosynthesis calculations and the CASA model (Potter et al., 1995), which focuses on the biogeochemistry aspects. The SiBCASA hybrid model consists of a detailed description of chemical and biological processes that determine vegetation responses. SiBCASA contains all carbon stored in biomass and has the capabilities to estimate terrestrial carbon and biomass fluxes from diurnal to decadal timescales (Schaefer et al., 2008).

SiBCASA is driven by 23 climatological drivers such as radiation, temperature and precipitation. The meteorological data fields are used from the European Centre for Medium Range Weather Forecasts (ECWF) ERA-interim reanalysis (CarbonTracker Team, 2017). The carbon fluxes output is represented as GPP, R_A, R_H, NEE and NPP and is generated as monthly output over a 1x1 degree worldwide spatial grid (Schaefer et al., 2008). SiBCASA also provides output every three hours, but is not currently necessary to use as we are interested in seasonal, interannual and long-term trends. The higher temporal resolution will be useful when the parameters of the fit function will be optimized during the inversion run.

The aim of SiBCASA is to give a good physical description of the in and outgoing biosphere fluxes as atmospheric measurements only show the effect of the net carbon flux (CarbonTracker Europe, 2017). The net yearly carbon flux is namely relatively small, in the order of a few Petagrams ($Pg = 10^{15}$), compared to the annual GPP, R_A and R_H, which are in the order of 120 Pg. The model gives information on these fluxes on for example a daily and seasonal temporal scale and will be used as input in CarbonTracker as first guess.

The SiBCASA data we use is the 'sibcasa_zeus_run_1851-2011.nc' and only uses the data ranging from 2001 until 2011. The data set includes monthly data without the month December in the year 2010. For this run, standard settings are applied.

3.2.2 The fit function

The fit function will be based on the so-called "CCGCRV" program, which is a digital filtering curve fitting program developed late 1980s by atmospheric scientist Kirk Thoning (Thoning et al., 1989; Pickers and Manning, 2015). He originally developed the program to analyze the seasonal, interannual and long-term atmospheric CO_2 trend over time Thoning et al. (1989), which can be found at the National Oceanic and Atmospheric Administration (NOAA f, 2016). Within the CCGCRV a time-dependent function was set-up to fit data-sets with a linear least square regression. This fit function can also be used as starting point to represent the NEE output signal.

The fit function does not include other climate-dependent variables and therefore interannual variability, driven by ENSO, will not be taken into account. The performance of the fit function compared to the SiBCASA NEE signal can quantitatively show which climate zones are affected by IAV. This is because SiBCASA is driven by meteorological data and therefore includes the effect of ENSO for example.

The fit function makes use of the Fourier Series, which is a general signal processing method. It includes a summation of four harmonic terms, which consist of a sine function with different frequencies, amplitudes and phases. These will resemble the seasonal cycle of the carbon fluxes and move around an average given by three polynomial values. These indicate the initial, mid-term and long-term average of a signal over time. The fit function also uses four harmonic terms to represent the seasonal cycle of the NEE signal and is therefore represented as follows:

$$C(t) = a_1 + a_2 t + a_3 t^2 + \sum_{n=1}^{4} C_n \big(\sin(2n\pi t + \phi_n) \big) \tag{6}$$

Here C(t) resembles the different carbon fluxes, a_1, a_2 and a_3 are the polynomial values, in which a_1 is the initial value, a_2 is the linear trend and a_3 is the exponential trend of the fitted NEE signal. C_n represents the amplitude and ϕ_n represents the phase of the wave. For each harmonic two parameters are given by the CCGCRV program, which can be used to calculate the phase and amplitude. In total 11 parameters are presented to represent the NEE signal in this case. The parameters thus represent constants when curve fitting is applied. However, these parameters are only constant for a single NEE signal per grid box. To analyze the phase and amplitude variability for each grid box and climate zone, first the amplitude and phase need to be derived from the 11 parameters given by the CCGCRV program.

3.2.3 Reconstruction of the fit function

The first step is to reconstruct the fit function as we need to understand how these parameters relate to equation 6. As already mentioned, the first three parameters consist of the three polynomial terms, while the following two parameters, C_1 and ϕ_1, correspond to the cosine and sine of the first harmonic. To determine the amplitude and phase of the first harmonic, the following relations from the CCGCRV function are derived:

$$\phi = \arccos\left(\frac{a_n}{C_n}\right)\frac{180}{\pi}$$ (7)

$$\phi = \arcsin\left(\frac{a_{n+1}}{C_n}\right)\frac{180}{\pi}$$ (8)

Here a_n and a_{n+1} resembles the parameters with n as 4,6,8,10, in which n = 4 resembles the first harmonic term. From this relation, the phase of the first harmonic can independently from the amplitude be determined:

$$\phi = \arctan\left(\frac{a_4}{a_3}\right)\frac{180}{\pi}$$ (9)

By substituting equation 9 into equation 7 and 8, the amplitude of the first harmonic is represented as:

$$C_n = \frac{a_3}{\cos(\pi\phi/180)}$$ (10)

$$C_n = \frac{a_4}{\sin(\pi\phi/180)}$$ (11)

The amplitude and phase of the other harmonic terms are derived in similar fashion.

3.2.4 Construction of the T and P-fit function

To add the NEE IAV to the curve fit function, Rödenbeck et al. (2018) introduces the γ_{NEE-T} (gC m^{-2}yr^{-1}K^{-1}) variable. This parameters represents climate sensitivity, whereby changes in temperature anomalies impact the NEE as explained in Section 2.3. To estimate the NEE response to climate anomalies, Rödenbeck et al. (2018) uses a linear relation between temperature and NEE which is resolved with regression coefficients. The linear regression is hereafter combined with an atmospheric inversion to decompose the atmospheric CO_2 signal in seasonal, interannual and decadal timescales. Their inversion constrains the climate-carbon cycle and can quantify the contribution of different climate zones to NEE response. In our case we will not use inversion constrains due to time limitations, but we can still use this method to obtain γ_{NEE-T} and compare it with the results of Rödenbeck et al. (2018). Implementing γ_{NEE-T} in the fit function gives:

$$T - fit_{NEE} = fit_{NEE} + \gamma_{NEE-T}(T - T_{LT+Seas+Deca+Trend}) \tag{12}$$

Here T represents the SiBCASA monthly averaged input temperature data from the ECMWF ERA-interim reanalysis and is spatially dependent. The $T_{LT+Seas+Deca+Trend}$ represents the sum of the longterm (LT), seasonal(Seas), decadal (Deca) and trend (Trend) (Rödenbeck et al., 2018). Subtracting this term from T, gives the IAV of the temperature. In this case the fit function of equation 6 is applied to the temperature signal, which does not include any IAV. And with the use of the original spatial temperature input of SiBCASA, the following equation for temperature IAV can be derived:

$$T_{IAV} = T - T_{fit} \tag{13}$$

The T_{IAV} represents the residual of the line fitting function of the fit function. Implementing this equation into equation 12, gives:

$$T - fit_{NEE} = fit_{NEE} + \gamma_{NEE-T} \cdot T_{IAV} \tag{14}$$

The T-fit_{NEE} is in μ gC mol m^{-2} s-1 and T_{IAV} in K. For the P-fit function, the same method will be applied and therefore we end up with:

$$P - fit_{NEE} = fit_{NEE} + \gamma_{NEE-P} \cdot P_{IAV} \tag{15}$$

Here γ_{NEE-P} is in (μgC mol m^{-2} s^{-1} mm^{-1}h) and P_{IAV} in mm h^{-1}.

3.2.5   Construction of $\gamma_{\mathbf{NEE-T}}$ and $\gamma_{\mathbf{NEE-P}}$

Our climate sensitivity γ consists of 12 parameters, in which each parameter corresponds to the conditional mean of all residuals of a certain month over a time series. Thus a parameter value for the month January is obtained by using linear regression to all January residuals over the time period. These values are obtained by linear regression of the residuals of NEE and T calculated with equation 13 and for NEE applies:

$$NEE_{IAV} = NEE - NEE_{fit} \tag{16}$$

Both residuals are thus the remaining IAV, which the fit function is not able to capture as it only focuses on capturing the seasonal cycle with its harmonics and long-term trend with the use of the polynomial terms.
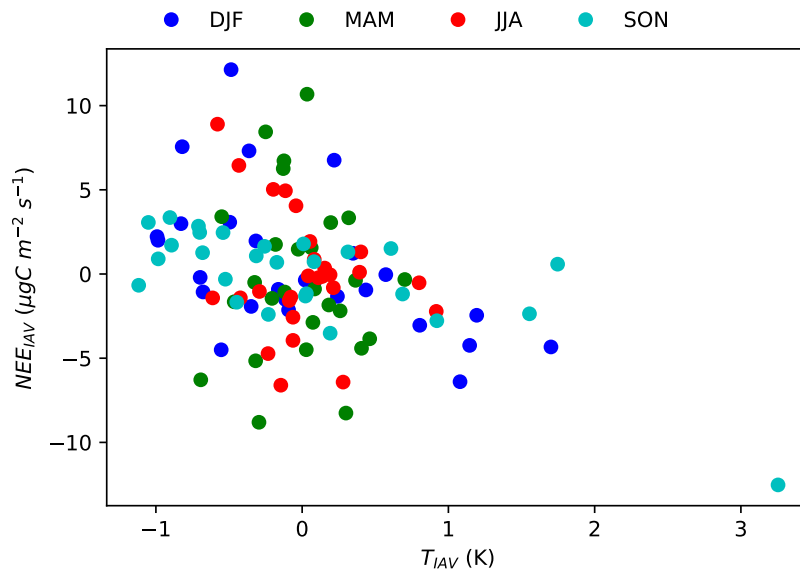

Figure 7: Monthly NEE_{IAV} versus NEE_{IAV} from 2001 until 2010 for the ATTO gridbox. The period DJF therefore contains 27 data points resembling a climate sensitivity relations between NEE_{IAV} and T_{IAV}.

On a yearly timescale the γ_{NEE-T} values do not seem to show any clear relation, however from a monthly and seasonal perspective a pattern can be derived (Figure 7). From a visualization perspective we show a three month cluster of residual data points, but do not that for our results we will apply linear regression to monthly residuals of NEE and T. To obtain the γ_{NEE-P} monthly values, we will perform the same method only this time with the use of precipitation residuals.

Precipitation consists of roughly two mechanisms: convective and stratiform precipitation, which mainly differ in growth time of the rain droplets and magnitude of vertical air flows in a cloud system (Houze, 1993). For stratiform precipitation, updraft is weak and droplets grow with the use of condensation and/or deposition. This precipitation has a low intensity and is fairly homogeneous spatially distributed.

Convective precipitation is caused by from a strong updraft due to buoyancy and grow by accretion, which is the collection by raindrops at the bottom of the cloud. They growth of the droplets coincides with cloud formation and precipitation is formed in short time span. These rain showers have a very local character and are therefore heterogeneous spatially distributed.

SiBCASA makes distinction between the two precipitation mechanisms, in this research we will define precipitation as a sum of convective and stratiform precipitation, which is therefore the total precipitation.

3.2.6   Growth rate and long-term trend

The growth rate and long-term trend within the CCGCRV function are derived with a low-pass filter on the residuals to remove the short-term variations (Thoning et al., 1989). Note that it is not clearly documented how this residual is established as it does not represent the raw time series minus the fit function for each lat/lon grid point. The filter makes use of the following decreasing exponential filter:

$$H(f) = \exp\left[-c_1 \left(\frac{f}{f_c}\right)^p\right] \tag{17}$$

Here c_1 is equal to $\ln(2)$ and resembles a normalizing coefficient, for which H(f)=0.5 at f=f_c. The variable f resembles the signal's frequency (cycles yr^{-1}), f_c is the cut-off frequency (cycles yr^{-1}) and p is an integer with value 4. For smoothing the data, a short-term cut-off value is used, in which $f_c = f_s$ and to remove the remaining seasonal oscillation $f_c = f_1$. The cut-off values used in Thoning et al. (1989) are 7.3 cycles yr^{-1} (50 days) and 0.55 cycles yr^{-1} (667 days) for f_1 and f_s respectively, as atmospheric mixing takes 2-3 months in the northern hemisphere. The long-term trend (Trend$_{LT}$) and growth rate (gr) are determined by man2html (2002):

$$Trend_{LT} = C(t)_{polynomial} + H(f_1) \tag{18}$$

$$gr = \frac{d}{dt}\left(C(t)_{polynomial} + H(f_1)\right) \tag{19}$$

Here C(t)$_{polynomial}$ represents only the polynomial part.

3.2.7 Grid box and region selection

To determine the NEE climate zone spatial and temporal variability, we have selected one gridbox in three different climate zones. The first climate zone is equatorial fully humid (Af) according to the Køppen-Geiger climate classification Kottek et al. (2006) and will from here be referenced as the tropical climate region. Within this region ATTO is located at 2° south, 59° east, which will be used as reference gridbox. The second selected climate zone has a snow climate, fully humid with cool summers and cold winters (Dfc) and will be referenced as the boreal region. In this region we selected the ZOTTO gridbox at 61° north, 109° west as reference gridbox. The third climate zone we will further look into has a arid climate and warm summer(BWh), in which the gridbox is located in the Australian desert at 22 ° south and 135° west. This gridbox is selected as it is located in the middle of the Australian desert.

To compare different climate zones, we will select the tropics and extratropics. The tropics are defined as every gridbox between 23° south and 23° north, while the extratropics are defined as 30° north until 31° north.

The selection of the semi-arid to arid region in Australia will be clarified in section 3.3. The boreal and tropical regions are selected because of the large climatological differences in which ATTO and ZOTTO are located respectively. Both stations are set up because, in the last decade, observational data was limited by Eddy Covariance flux tower observations, which only take the micro-meteorology processes into account and can only measure the local CO_2 footprint (Andreae et al., 2015). A large carbon flux variability was found with the use of these local 'snapshots of reality'. Therefore long-term and continuous studies are necessary, in which ATTO and ZOTTO play a crucial role to obtain more observational data as input for CarbonTracker amongst others. Both towers namely stand over 300 meters tall and are especially useful as they can measure regional influences and processes with a concentration footprint in the order of 1000 km (Heimann et al., 2014; Andreae et al., 2015). Now regional measured CO_2 can be used as input for the the inversion model and can therefore better optimize the regional biospheric carbon fluxes.

3.2.8 Statistical measures

To analyze the results we will make use of the average ($\langle$ NEE $\rangle$), standard deviation (σ_{NEE}), Pearson correlation coefficient (p_{xy}) and the root mean square error (RMSE) as depicted below:

$$\langle NEE \rangle = \frac{1}{N} \sum_{i=1}^{N} NEE_i \tag{20}$$

$$\sigma_y = \sqrt{\frac{1}{N-1} \sum_{i=1}^{N} \left(y_i - \langle x_i \rangle \right)^2} \tag{21}$$

$$p_{xy} = \frac{n(\sum\limits_{i=1}^{N} x_i y_i) - (\sum\limits_{i=1}^{N} x_i)(\sum\limits_{i=1}^{N} y_i)}{\sigma_x \sigma_y} \tag{22}$$

$$RMSE = \sqrt{\frac{\sum\limits_{n=1}^{N} (y_i - x_i)^2}{N}} \tag{23}$$

Here NEE represents any NEE time signal, which can be the signal of SiBCASA or the (T/P-)fit function and N is the total amount of time steps. y_i represents the NEE signal of the (T/P-)fit function and x_i represents the NEE signal of SiBCASA.

3.3 Results and discussion

To answer our three research questions stated in the introduction, we will divide the results section in four subsections which will each try to explain one research question. Before we answer our research questions we will first introduce you to spatial and temporal variability of NEE and show the distinction between seasonal, interannual and long-term variability in the NEE time series with the use of the original SiBCASA output. Hereafter we will determine the performance of the fit-function, which should capture the seasonality of the NEE time series. Next, we will analyze the performance of the T/P- fit function in a similar fashion and compare them to each other. The results section will finally elaborate on the climate sensitivity parameter in which we biophysically interpret the parameter and compare it with literature.

3.3.1 Analysis of the SiBCASA NEE output

Figure 8 shows large spatial variability for the global NEE distribution with a global average NEE of -0.58 $\mu gC\ m^{-2} s^{-1}$. The largest global negative net carbon flux can be found in the tropical latitudes as climate variables such as temperature, light and precipitation are most beneficial for photosynthesis and dominates the respiration term (Welp et al., 2011). The average NEE in the tropics is -2.02 $\mu gC\ m^{-2} s^{-1}$ and almost four times as high compared to the extratropics, as NEE is -0.51 $\mu gC\ m^{-2} s^{-1}$ on average.

The extratropics also show a negative NEE, but is relatively lower compared to the tropics as light and temperature are the limiting factor. While in the desert no significant NEE fluxes are found as precipitation is the limiting factor. The positive NEE in Australia indicates a larger respiration rate compared to the photosynthesis rate (Haverd et al., 2017; Cleverly et al., 2013).

These worldwide NEE variabilities show that the spatial climate conditions are important to take into account the interannual variability of NEE. Interestingly is the variability of NEE on small scale within a climate zone, in for example the Amazon basin and Australian desert, in which a NEE variability of a factor 5-10 can be observed. In Australia this could be explained by local heavy rainfall events (Cleverly et al., 2013; Australian Government - Bureau of Meteorology, 2019). To gain a better understanding of the NEE fluxes for constructing the fit function, NEE on different temporal scales will be analyzed next for the ATTO and ZOTTO.
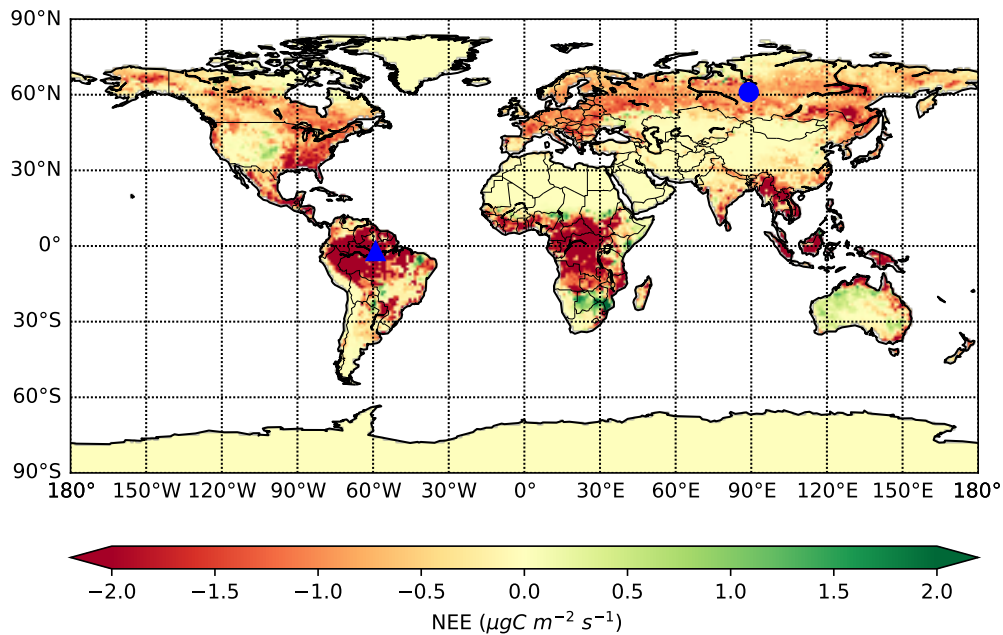

Figure 8: Yearly averaged SiBCASA NEE output from 2001-2011. The more negative the NEE values, the larger the net carbon taken up by the biosphere from the atmosphere and vice verse. The blue dot indicates the location of ZOTTO and the blue triangle indicates the location of ATTO. The tropics show large negative NEE fluxes as climatology is favorable.

Figure 2 shows large temporal NEE variability for ATTO and ZOTTO. The ATTO site shows relative large IAV compared to its seasonal trend. The seasonal trend is determined by the movement of the intertropical convergence zone (ITCZ), which roughly follows the solar equator (Fu and Li, 2004; Li and Fu, 2004). The ITCZ is accompanied with heavy thunderstorms and causes the wet season in the tropics. During the wet season, the tropics are PAR limited due to the thick cloud cover, which causes a decrease in photosynthesis resulting in a negative NEE. For ATTO, located just south of the equator, the wet season is from October until March. From April until September, the ITCZ is located at the northern hemisphere and therefore precipitation decreases. This period is called the dry season, in which cloud cover is less thick and fresh leaves increase the photosynthetic capacity leading to negative NEE (Myneni et al., 2007). The deep rooting depth of the Amazon vegetation ensures that in regular years no soil moisture stress occurs (Fu and Li, 2004).

However, during the 2005 (figure 9) ,2010 and 2015/2016 Amazon droughts, the dry period was extended, which did cause soil moisture stress (Phillips et al., 2009; Gatti et al., 2014; van Schaik et al., 2018; Jiménez-Muñoz et al., 2016; Lewis et al., 2011). Here the 2005 and 2010 droughts were caused by anonymously high sea surface temperatures, which are identified as a 1-in-100-year event and the 2015/2016 drought was caused by an El Niño. During these periods, GPP decreased and let to a carbon flux from the biosphere into the atmosphere. These are examples of IAV and are represented in the NEE time series as well.

In SiBCASA, the 2005 drought seems to be recognizable, however the 2010 drought is not. The Amazon is not homogeneously impacted during the droughts as especially the Eastern part of the Amazon receives less precipitation in general and the Northwestern part is very wet. The precipitation input data at ATTO (not shown) do not suggest precipitation anomalies in this area. With the use of the climate sensitivity parameter γ_{NEE-P}, possible responses of these droughts at the ATTO gridbox could possible explained and will be further looked into in upcoming results subsections.

The tropics have a long-term negative NEE trend of around 5 $\mu gC\ m^{-2}s^{-1}$ and does not show a trend over the 2001-2011 period as the time span is too short to observe the CO_2 fertilization effect. The calculated growth shows large interannual fluctuations with a maximum of almost 5 $\mu gC\ m^{-2}s^{-1}$.

Note that the long-term trend does not exactly represent what was expected in figure 9. The long-term trend namely still shows yearly fluctuations, because the current standard cut-off value (f_c) is set as 667 days. The trend line therefore takes into account the irregularities between two years, which causes the small seasonal cycle in the long-term trend. For a better representation of the long-term trend a larger time interval could be used, but will also effect the resemblance of the interannual variability as it is the derivative of the interannual trend (see section 3.3.2). A filter of 1643 days (4,5 years) can for example be applied for the long-term filter, which is the average time interval for an El Niño to occur (Chang and Zebiak, 2015). In this case the variability caused by ENSO in the long-term trend will be filtered out, but will also largely filter out ENSO in the interannual trend (depicted as growth rate). Therefore a trade-off between the two trends in terms of f_c will need to be taken into account. A cut-off value of 667 days was eventually chosen to show the interannual trend in detail and still have a decent representation of the long-term trend.
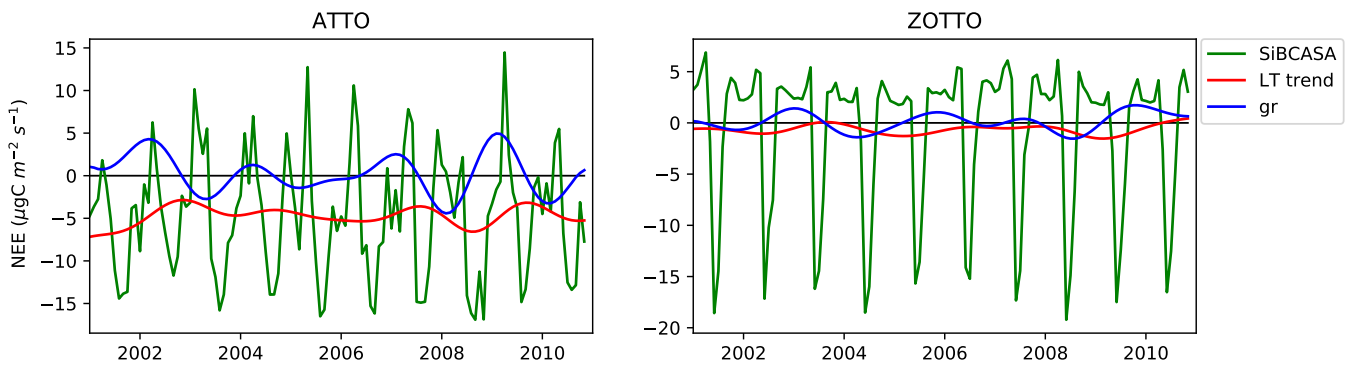
Figure 9: Monthly averaged raw time series of the SiBCASA NEE output as represented as the green line for ATTO and ZOTTO grid boxes. The red and blue line corresponds to the long-term trend and interannual variability of the ccgcrv program respectively.

The NEE time series at ZOTTO is dominated by the seasonal cycle as the boreal region is light and temperature limited (9). In the summer months of the Northern hemisphere NEE becomes negative as temperature and daylight increases and becomes positive in winter. IAV is low as the distance to the origin of El niño's is large and therefore hardly impacts the ZOTTO and its corresponding boreal region. In 2010, a severe drought over European Russia had taken place (Arpe et al., 2012), which could have impacted ZOTTO. But it seems that ZOTTO is located too far East to show any impacts of the drought on the NEE signal by SiBCASA. The IAV and long-term trend at ZOTTO is therefore low.

3.3.2 Fit function performance analysis

To asses the performance of the fit function we will first show how the fit function looks like for ATTO and ZOTTO and hereafter check its performance with the use of the Pearson correlation coefficient and RMSE.
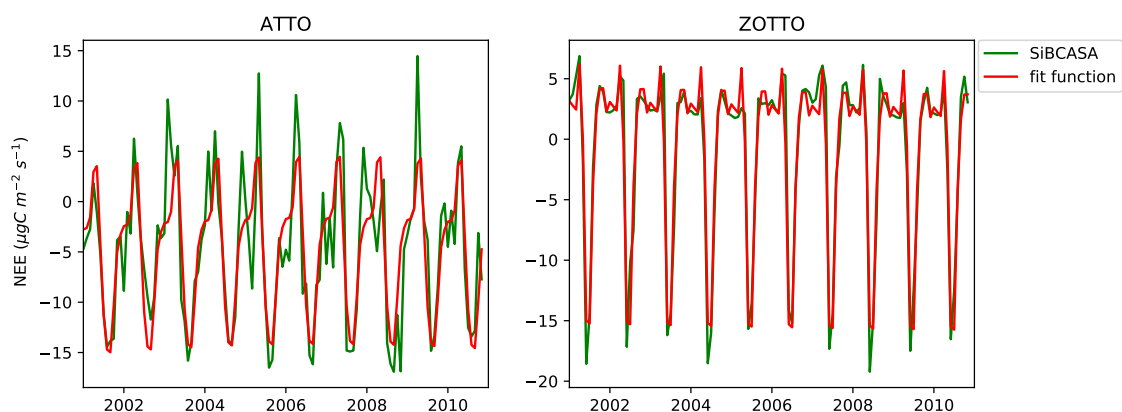

Figure 10: Visualization of the fit function for ATTO and ZOTTO from 2001 until 2011 as defined in section 3.2.2. The fit function is able to capture the seasonal cycle of both gridboxes.

The fit function is able to capture the largest portion of the NEE signal for ZOTTO and has difficulty capturing the NEE signal of the ATTO gridbox as expected. The large IAV at ATTO results in a RMSE of 3.84 $\mu gC\ m^{-2}s^{-1}$ and correlation coefficient of 0.84 For ZOTTO we obtain a RMSE of 1.80 $\mu gC\ m^{-2}s^{-1}$ and a correlation coefficient of 0.97. The IAV at ATTO seems to result in a phase and amplitude change a the fit function is not able to capture both. The signal shows a large seasonal cycle, but seemingly also a smaller second seasonal cycle, which we hypothesize, could be due to the location of ATTO, which is near the equator. This causes the ITCZ to pass by twice and therefore seems to cause a small and shorter second dry and wet season, which is affected by IAV as well. The second time the ITCZ crosses the equator, the NEE could again be lowered as PAR is limited due to thick cloud cover, which could possible explain the second NEE peak.

For ZOTTO the fit function is able to capture the seasonal cycle well and IAV only causes small changes in amplitude in the winter and summer months. This therefore show that for ZOTTO a small set of parameters is already able to explain the signal and show that for these regions performing research at longer time scales is possible and that we could save computational costs as the final product of CarbonTracker, λ does not have to be calculated any more. However, for the ATTO this is not the case. Therefore we will now zoom into the Australian gridbox to gain a better understanding of the fit function and the correlation coefficient performance lack in this region.
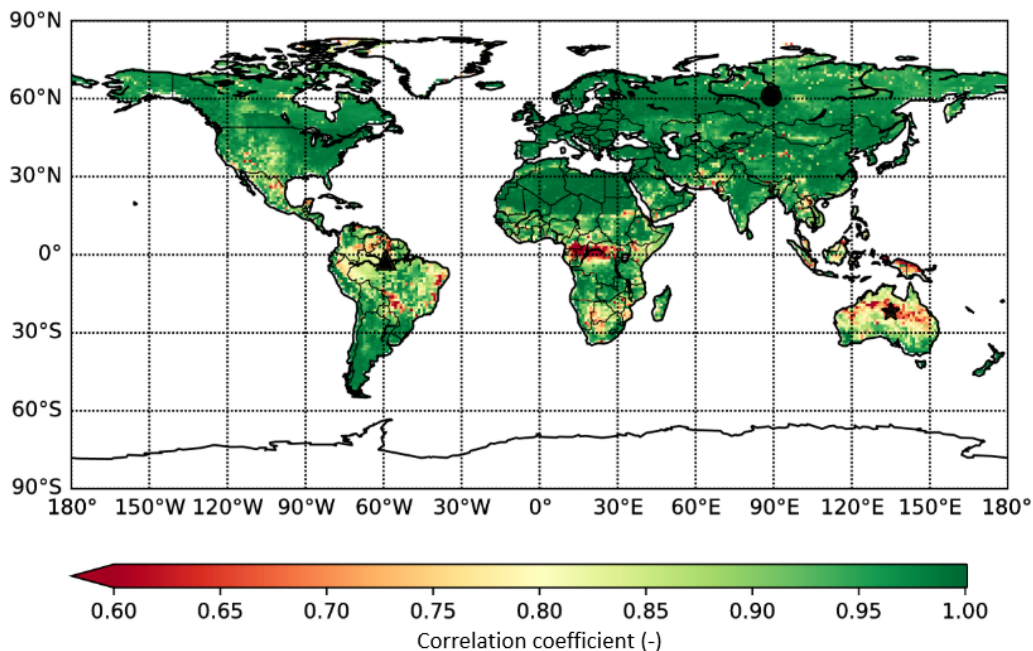

Figure 11: Performance of the fit function compared to the raw SiBCASA output signal represented in Pearson correlation coefficient. The value 0 represents no correlation and 1 represents a perfect correlation. The correlation coefficient is determined from the SiBCASA output and fit function from 2001-2011. High correlation coefficient in the extratropics and reasonable in the tropics and Australian desert.

The average global correlation of the fit function is 0.93 and therefore the seasonality trend is captured well (Figure 11). In the tropics the correlation coefficient is 0.89 and in the extratropics the correlation is higher than average with 0.97 as expected. The tropics and semi-arid to arid regions such as the Amazon basin and Australian desert show the lowest captured seasonality trend due to high IAV in the NEE signal. The correlation coefficient, especially in the extratropics, suggest that applying a simple fit function with only 11 parameters is a valid method to capture the trend of NEE signal when it is dominated by the seasonal cycle. However, while the correlation coefficient is still quite high in the tropics, we have seen at the ATTO gridbox (Figure 10) that the RMSE is still significantly large.
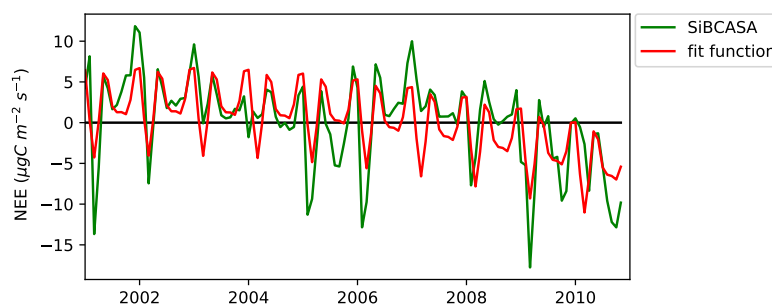

Figure 12: NEE time series of Australian desert gridbox with high IAV in the NEE signal of SiBCASA.

The phase and amplitude of the NEE signal at the Australian desert is dominated by IAV (Huxman et al., 2004). The large NEE IAV can be explained by the erratic year to year rainfall Australian Bureau of Statistics (2006). Microbial activity is triggered at small precipitation events and saturates at moderate events. While photosynthetic activity only takes place when relative large precipitation events or multiple small events take place Huxman et al. (2004). The different responses cause semi-arid areas to act as carbon sources as also seen in figure 8.

The erratic year to year rainfall is caused by four main processes (Australian Bureau of Statistics, 2006). Firstly, the Southward extension of the monsoon through, which position varies interannually and when located south the Northern part of the Australian desert receives precipitation. Secondly, tropical cyclones, which occur multiple times in a regular year and cross the Australian coast from the north to penetrate well inland. This causes highly erratic and intensive rain showers.Parts of the Australian desert receive rain up to 40 % due to this atmospheric process. Thirdly, the Northwest Cloudbands, which transport moisture from the warm Indian Ocean over Australia in a Northwest to Southeasterly fashion. These three processes mainly cause the interannual NEE variability in the Australian desert. In 2010/211, the effect of the fourth process took place in which moisture is taken inland due to a strong La Niña (Poulter et al., 2014). During this period record-braking rain record turned the Australian desert green and caused a large GPP fluxes, which led to strong negative NEE.

This region thus shows how γ_{NEE-P} could explain a large portion of IAV in the NEE signal. When only using the fit function we obtain a correlation coefficient of 0.72 and a RMSE of 3.76 $\mu gC\ m^{-2}s^{-1}$. Now we again would like to zoom out and look at the RMSE to further increase our understanding of the performance of the fit function.
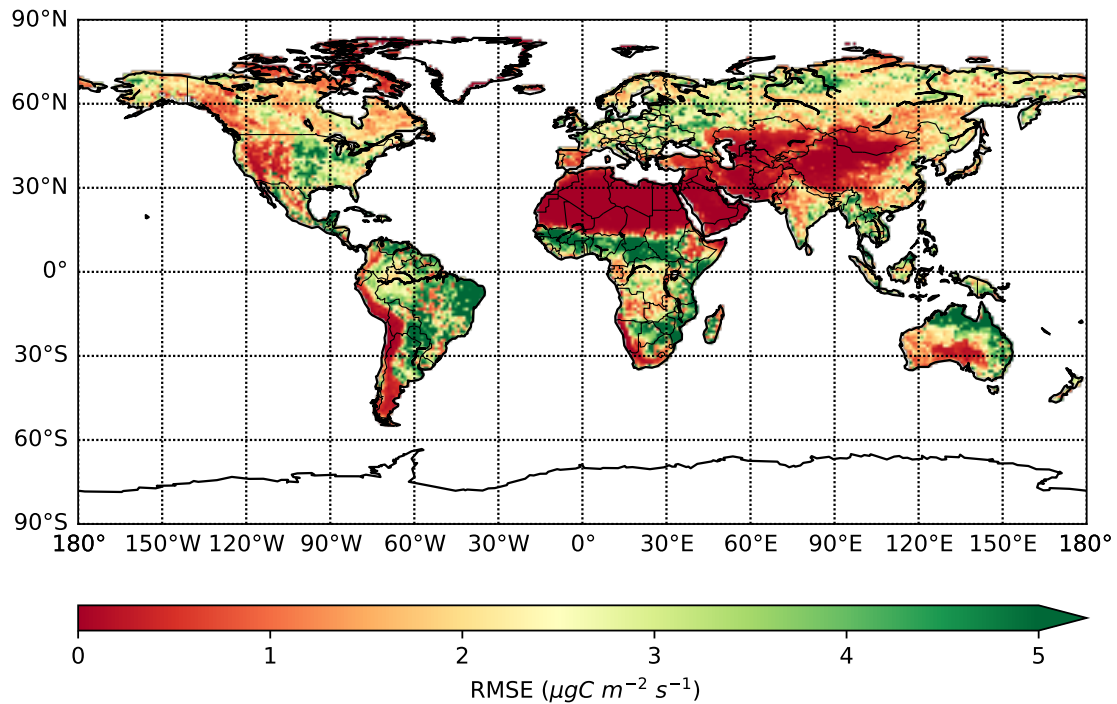

Figure 13: Performance of the fit function compared to the raw SiBCASA output signal represented in RMSE. The RMSE is determined from the SiBCASA output and fit function from 2001-2011. Tropics show large RMSE as IAV in the NEE time series is large as well, for the extratropics RMSE is low.

The RMSE is especially high in the tropics as NEE fluxes are large and IAV within the NEE signal is dominant (figure 13). This leads to RMSE of 3.13 $\mu gC\ m^{-2}s^{-1}$ in the tropics, while globally the RMSE is 2.19 $\mu gC\ m^{-2}s^{-1}$ and the extratropics have a RMSE of 1.77 $\mu gC\ m^{-2}s^{-1}$. The arid regions show the lowest RMSE as there is hardly any NEE signal and therefore residuals of the fit function are very low. In the semi-arid area of Australia we do see an RMSE of around 2 $\mu gC\ m^{-2}s^{-1}$, which is high as NEE fluxes are relatively small as shown in figure 12. To achieve a better performance we will in the next section add the climate sensitivity term, γ.

3.3.3 T/P-fit function performance analysis

To evaluate the performance of T/P-fit function, we will first show the its implementation temporally and use the spatial correlation coefficient and RMSE metric to obtain its performance.
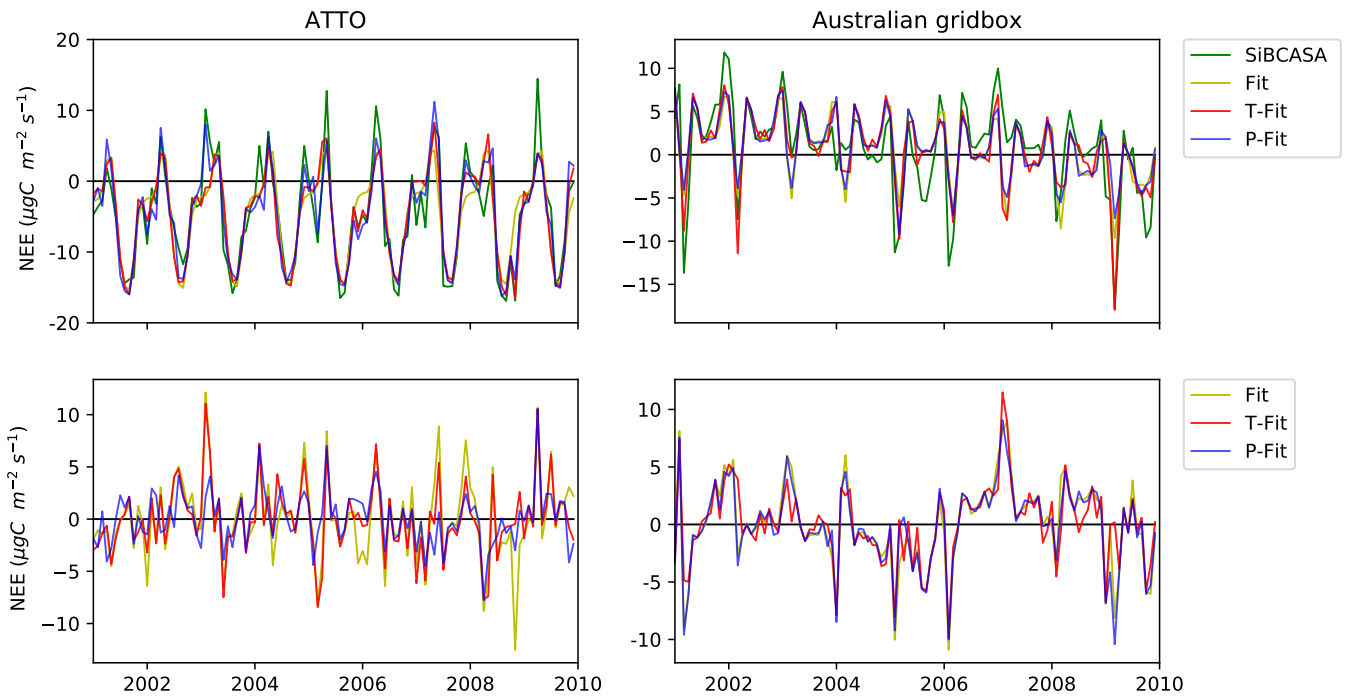
Figure 14: Top: Time series of the (T/P-)fit functions of the ATTO and Australian gridbox from 2001 until 2010. Bottom: residuals of (T/P-)fit functions.

Figure 14 shows the original NEE signal and the fit and T-fit functions with their corresponding residuals. Both methods show a similar performance of the original signal, whereas the T-fit function captures the original signal slightly better.

Both the T and P fit function show a performance increase in terms of correlation coefficient compared to the fit function. The T-fit function shows a correlation coefficient of 0.88 and 0.78 for ATTO and the Australian gridbox respectively, which is an increase of 0.04 and 0.08 respectively. The P-fit shows an increase of 0.10 and 0.02 for ATTO and the Australian gridbox respectively, which lead to a 0.93 and 0.73 correlation coefficient respectively.

The T and P fit function also show RMSE decrease at both gridboxes and therefore increase the performance compared to the fit function. The T-fit function now has a RMSE of 3.38 $\mu gC\ m^{-2}s^{-1}$ and 3.24 $\mu gC\ m^{-2}s^{-1}$ for ATTO and the Australian gridbox respectively, in which an RMSE of 66.3 % and 96.9 % remains respectively compared to the fit function. The RMSE of the P-fit shows a remain 86.5 % and 88.5 % compared to the fit function, which lead to a RMSE of 2.59 $\mu gC\ m^{-2}s^{-1}$ and 3.55 $\mu gC\ m^{-2}s^{-1}$ for ATTO and ZOTTO respectively.

The T and P-fit function thus show a performance increase for both gridboxes, however residuals remain significantly large. The interannual phase and amplitude change at both gridboxes is not resolved. Especially for the Australian gridbox we expected a high performance increase of the γ_{NEE-P} as NEE in the Australian desert is strongly dependent erratic precipitation events (Huxman et al., 2004) as shown in Figure 12. However, the P-fit function only shows a small RMSE decrease, which is in contrast to the ATTO gridbox in which the RMSE significantly decreases. Notewhorty is the large difference of the RMSE between the T and P-fit functions at the ATTO gridbox. We expected a similar performance as precipitation and temperature have a strong climate covariance, which drives the carbon cycle (Jung et al., 2017; Berg et al., 2015). Jung et al. (2017) states that on a local scale the IAV of GPP and R is dominated by water availability and to a lesser extent also for NEE. Regionally and globally, NEE IAV is driven by temperature anomalies. This seemingly contradiction could be due to two outbalancing water effects: (1) NEE variability is dampened due to the compensation of temporal and local water-driven R and GPP, and (2) spatially, the anomalies in water availability also compensate. This leads to interannual carbon fluctuations dominated by the temperature signal.

Thus on a gridbox scale, which is roughly 110 km by 110 km, it could be possible that the IAV of precipitation is dampened and therefore we do not observe the performance increase as expected for the P-fit function in Australia. To increase our understanding of the performance of the T and P-fit functions we our next step is to check the spatial performance and afterwards elaborate on the climate sensitivity parameter.
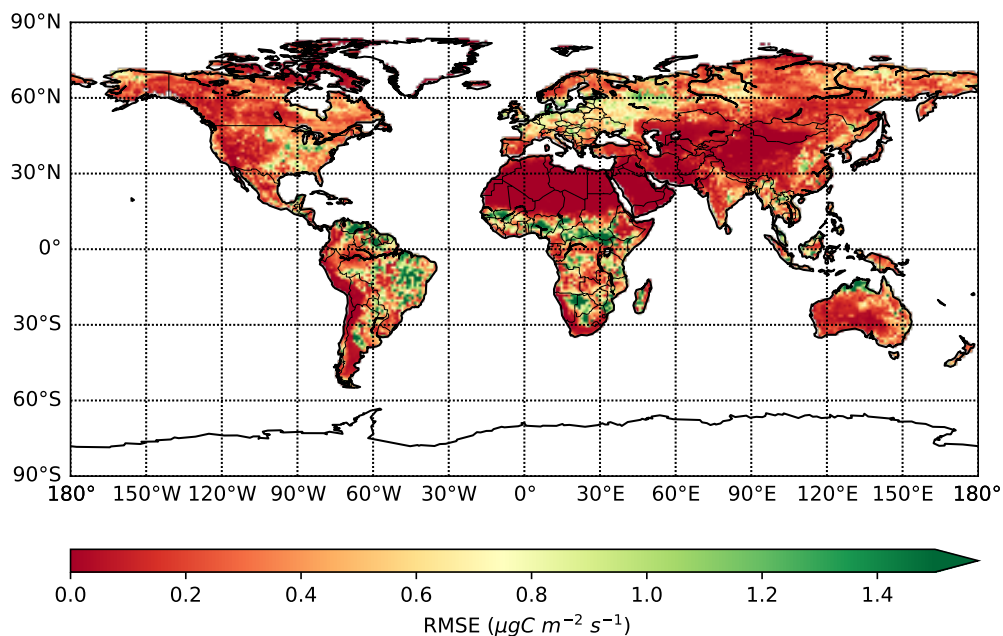

Figure 15: Spatial difference of the RMSE of the fit function - T-fit function from 2001 until 2010. Higher RMSE values indicate a performance increase of the T-fit function, which is mainly in the tropics.

The T-fit function improves the representation of the NEE time series compared to the fit function across the whole globe. In the desert areas, no improvement is made as the NEE signal is already close to zero. In the tropics the biggest improvements are made as the T-fit function decreases the RMSE by 0.51 $\mu gC\ m^{-2}s^{-1}$ on average and globally the RMSE decreases by 0.35 $\mu gC\ m^{-2}s^{-1}$ on average. This leads to an average RMSE of 2.61 $\mu gC\ m^{-2}s^{-1}$ and 1.81 $\mu gC\ m^{-2}s^{-1}$ for the tropical and global average respectively.

The P-fit function (not shown), has a RMSE decrease of 0.33 $\mu gC\ m^{-2}s^{-1}$ on a global average and 0.54 $\mu gC\ m^{-2}s^{-1}$ for the tropics on average. This leads to an average RMSE of 2.59 $\mu gC\ m^{-2}s^{-1}$ and 1.83 $\mu gC\ m^{-2}s^{-1}$ for the tropical and global average respectively. The P-fit function thus slightly outperforms the T-fit function in the tropics, but globally the P-fit function performs slightly worse. On this level the co-variance of precipitation and temperature IAV seems to be linked performance wise as stated by Jung et al. (2017). To understand how small or large these RMSE's are we will in chapter 5 transform them to CO_2 mole fractions and compare them with the SiBCASA output.

3.3.4 $\gamma_{\mathbf{NEE-T/P}}$ interpretation and comparison
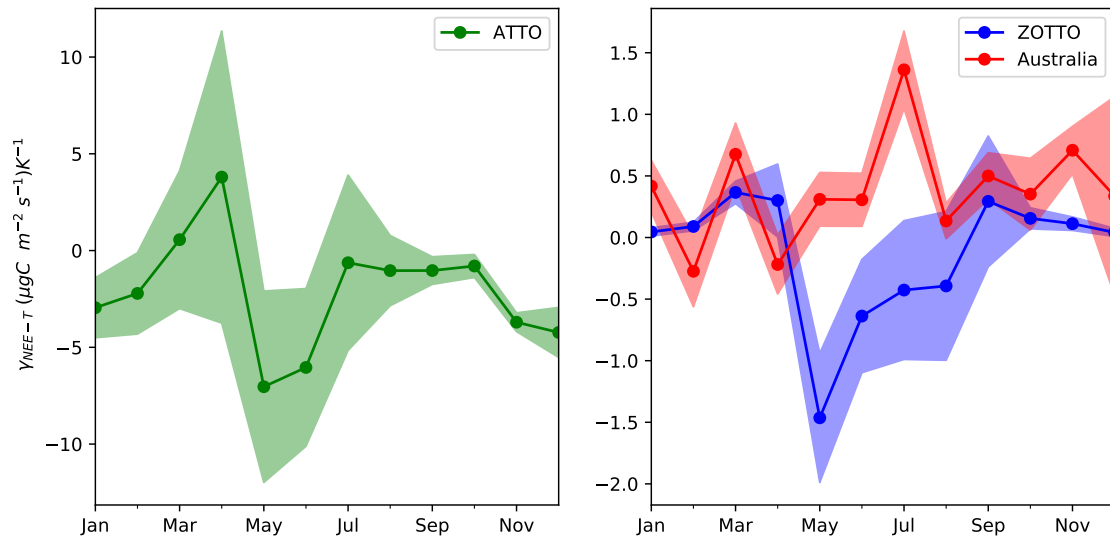

Figure 16: Time series of γ_{NEE-T} for the ATTO gridbox (left panel) and the ZOTTO and Australian gridboxes (right panel) from 2001 until 2010. Here corresponding shading indicates the uncertainty of the $gamma_{NEE-T}$ value within one σ and shows a clear seasonal pattern for all gridboxes.

All gridboxes show a distinct seasonal pattern for γ_{NEE-T}, in which a positive γ resembles positive NEE fluxes when temperatures are higher than average and vice verse (Figure 16). For ATTO the begin of the dry season, in May and June, we observe negative NEE fluxes, which could be due to the fact that during the dry season, PAR is not the limiting factor as cloud cover is low. Higher than normal temperatures could herewith increase the GPP and

thus result in a negative NEE (Fu and Li, 2004).

During the months March and April the ITCZ is on top of the ATTO gridbox as the wet season transitions into the dry season on the southern hemisphere. When temperatures are higher than average, it could lead to extra evapotranspiration, extra moisture and thus a thicker cloud cover (Fu and Li, 2004). This eventually leads to a decrease in PAR availability and therefore decreases the GPP, which causes a positive NEE peak during this period. However, the large variability in during these months also show that there is a possible negative γ_{NEE-T} during this period which would reject our possible explanation. During September and October the ITCZ surpasses ATTO again and here we do not see a clear γ_{NEE-T} response.

During the wet season in the months September until March, we see a negative trend at ATTO. When higher temperatures than average occur, we expect a positive γ_{NEE-T} as explained in previous paragraph. Therefore the processes behind the seasonality of γ_{NEE-T} is difficult to explain and the large variability shows that the value of γ_{NEE-T} at ATTO has a large uncertainty.From this single gridbox we hypothesize that γ_{NEE-T} is not clearly linked to natural processes and its corresponding IAV. However, to support this claim, we will have to check the global spatial perspective first.

In the Australian desert, higher than average temperatures cause a positive γ_{NEE-T} annually. Temperatures near 40 ° Celsius decrease a plants activity and therefore shuts down photosynthesis and root respiration (Bonan, 2015). The net NEE increase is due to microorganisms in the soil which have an enzymatic optimum at such high temperatures. The noisy behaviour over the months could possible be explained by the highly variable precipitation which also effects surface air temperature (Huxman et al., 2004).

At the ZOTTO gridbox, we observe a strong NEE peak in May as higher than average temperatures lead to an earlier start of the spring season. Therefore the plants activity increases with a net larger photosynthetic carbon uptake compared to respiration release (Bonan, 2015). γ_{NEE-T} is negative for the whole growing season at ZOTTO, which lasts until August. Hereafter temperature anomalies have a small impact on the NEE fluxes as the whole region is covered in snow.

The ATTO gridbox shows the largest γ_{NEE-T}, due to the seemingly much larger influence of ENSO and NEE fluxes. The monthly variability relative to its yearly cycle is comparable for the ATTO and ZOTTO box during the growth season in the boreal region. The large monthly variability is one of the reasons, the T-fit function only slightly decreases the RMSE in figure 14.
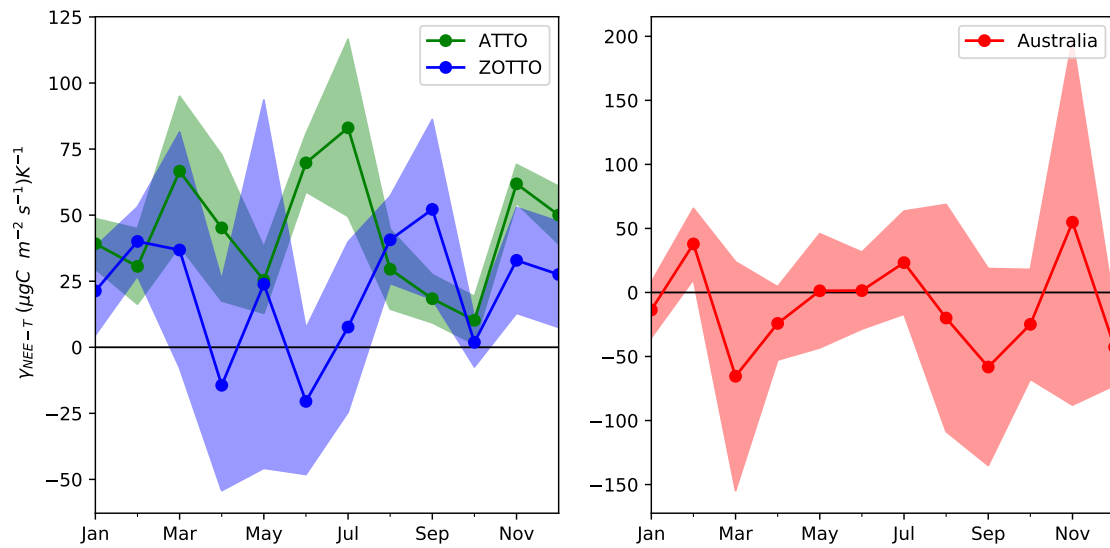
Figure 17: Monthly values for γ_{NEE-P} for the ATTO, ZOTTO and Australian gridbox calculated with linear regression over a period of 2001-2010. Here corresponding shading indicates the uncertainty of the γ_{NEE-P} value within one σ, which is large and a clear seasonal pattern is missing.

For all three gridboxes the link between γ_{NEE-P} and their natural processes is difficult to explain (Figure 17). ATTO depicts a seasonal cycle, however all monthly γ_{NEE-P} are positive, which means that higher than normal precipitation leads to an increase of carbon flux from the biosphere into the atmosphere. This would indicate that extra precipitation, thus extra cloud cover limits the PAR throughout the year. However, we know that an El Niño may extend the dry season leading to soil moisture stress. In this case higher than normal precipitation in especially the months July, August (end of dry season) should show a negative correlation with the γ_{NEE-P}.

At ZOTTO, the NEE is light and temperature limited and an interpretation of γ_{NEE-P} is difficult. We do see that the uncertainty of γ_{NEE-P} during spring and the start of summer as the no correlation between precipitation and NEE anomalies is found. At the Australian gridbox we would expect a strong negative γ_{NEE-P} as the semi-arid region is water limited. higher than normal precipitation should therefore result in negative NEE.

These results indicate that mathematically IAV is captured in the signal of the T and P-fit function, however we can not statistically link them to natural processes. To decrease the uncertainty of the γ's a larger time series of NEE could solve the problem. Currently, climate anomalies in some months due to an El Niño may be underrepresented. If we for example have thirty years of NEE time series, we have an extensive amount of ENSO captured in the NEE signal. This also increases the likelihood of an El Niño and a La Niña to take place during each month. This approach could lead to a better representation of the climate sensitivity parameter γ.

A second method to improve the γ's is to optimize them with the use of inverse modelling as performed in Rödenbeck et al. (2018). Visualizing the γ's as in the Hovmöller diagram of (Rödenbeck et al., 2018). In this case we plot the the latitude versus the monthly γ's values, which is an average of all longitudinal γ's at each latitude as depicted below. Here we depict (Rödenbeck et al., 2018) Hovmöller diagram and our Hovmöller diagram of γ_{NEE-T}.
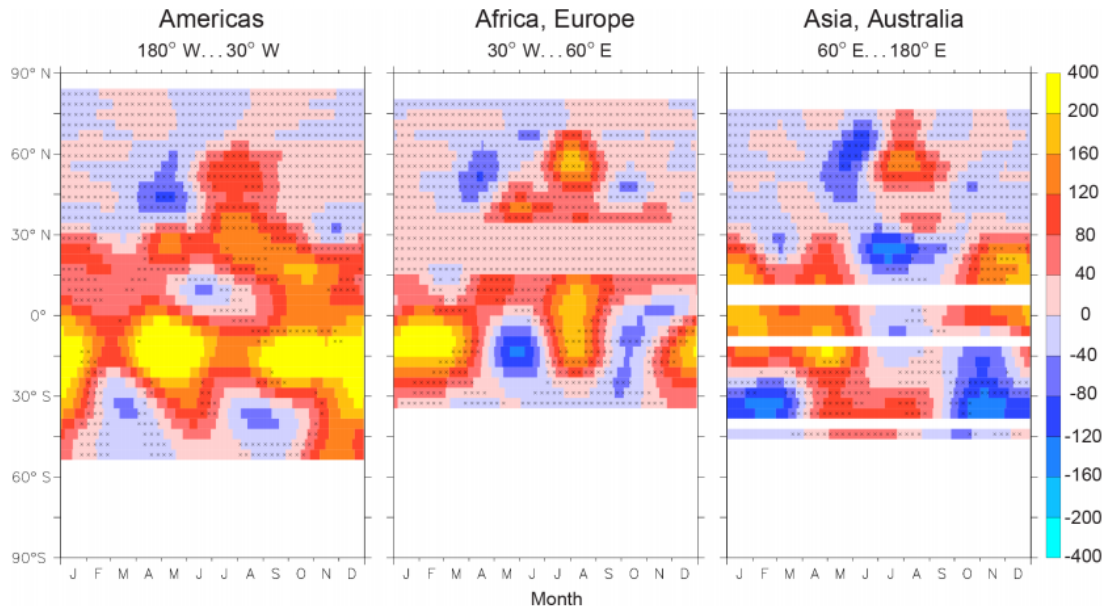

Figure 18: Interannual climate sensitivity γ_{NEE-T} (gC m^{-2} yr^{-1}K^{-1}) of the Americas (180 °W...30 °W) shown as Hovmöller diagrams: longitudinal averages of γ_{NEE-T} are plotted as colour over latitude (vertical) and month of the year (horizontal). The stippling indicates robustness see Rödenbeck et al. (2018) for more detail.

Focusing on the America's, the climate sensitivity parameter γ_{NEE-T} of Rödenbeck et al. (2018) are able to illustrate the photosynthesis rate being temperature limited in spring for the northern extratropical land shown as negative γ_{NEE-T} (Figure 18). These higher than average temperatures lead to a higher photosynthesis rate and therefore a more negative NEE. The NEE flux is defined as negative when atmospheric CO_2 concentrations decrease and vice versa, which is called the 'atmospheric perspective' and will be used throughout this thesis. In May and June we see the opposite effect, in which warmer summers enhance the respiration rate and tend to be drier leading to a lower photosynthesis rate and therefore a net positive γ_{NEE-T}.

In the tropical region, positive γ_{NEE-T} is dominant, however it also shows less systematic variations. The missing stippling indicates results of limited robustness and therefore have a larger uncertainty, see Rödenbeck et al. (2018) for more details.

The Southern extratropics show a similar γ_{NEE-T} pattern as the northern extratropics, but is less robust due to the relative small land mass. For Europe, Africa and Asia, Australia, we also see a γ_{NEE-T} , which can be linked to natural processes. Thus, without computing the complex ecosystem and plant processes, a large fraction of interannual NEE variations is captured, while considerably decreasing the computational costs.

Figure 19: Interannual climate sensitivity γ_{NEE-T} (gC m^{-2} yr^{-1}K^{-1}) shown as Hovmöller diagrams: longitudinal averages of γ_{NEE-T} are plotted as colour over latitude (vertical) and month of the year (horizontal).

If we compare the results of (Rödenbeck et al., 2018) with our results we see that we have a sign change in the γ_{NEE-T} (Figure 18 and 19). This confirms our hypothesis in which our γ_{NEE-T} is mathematically correct, but can unfortunately not be directly linked with natural processes, which (Rödenbeck et al., 2018) is able to show. We do see that in the tropics the γ_{NEE-T} becomes more pronounced as in the Amazon γ_{NEE-T} is more negative.
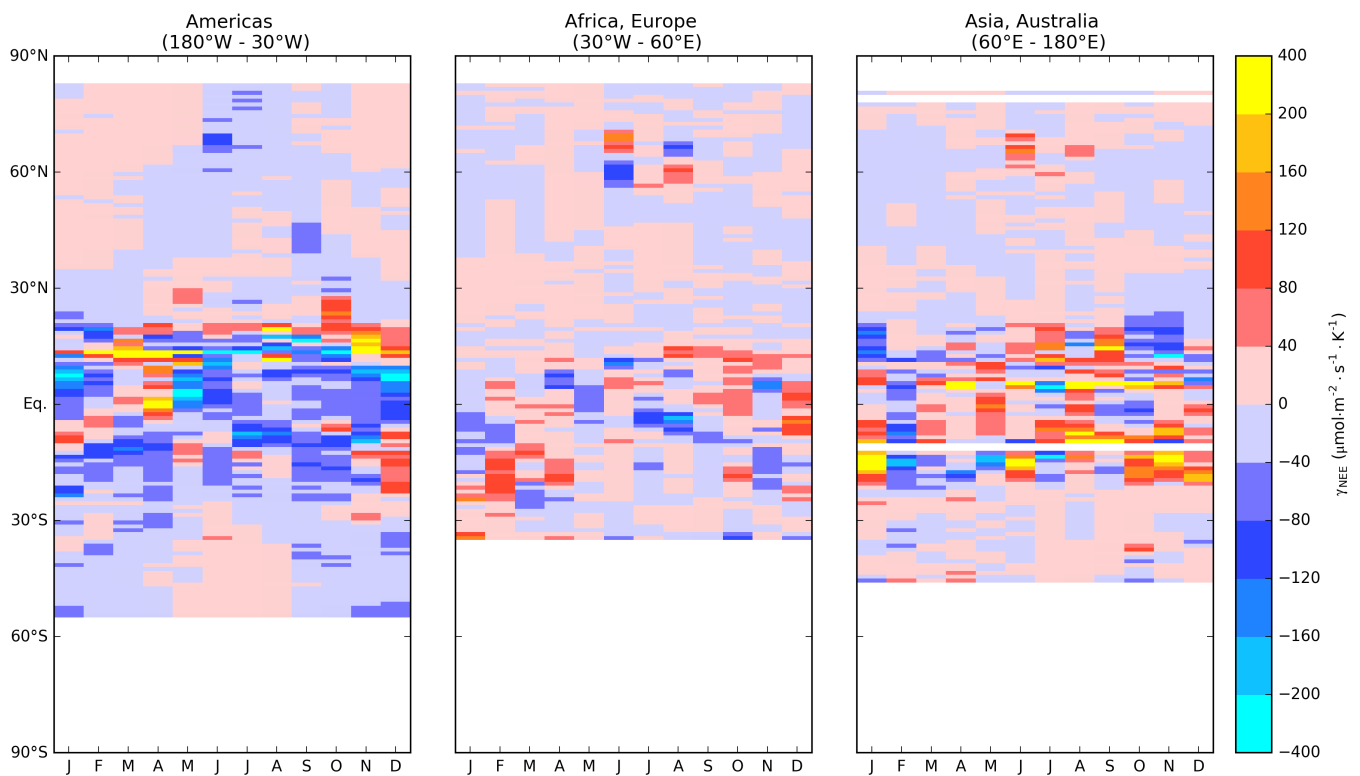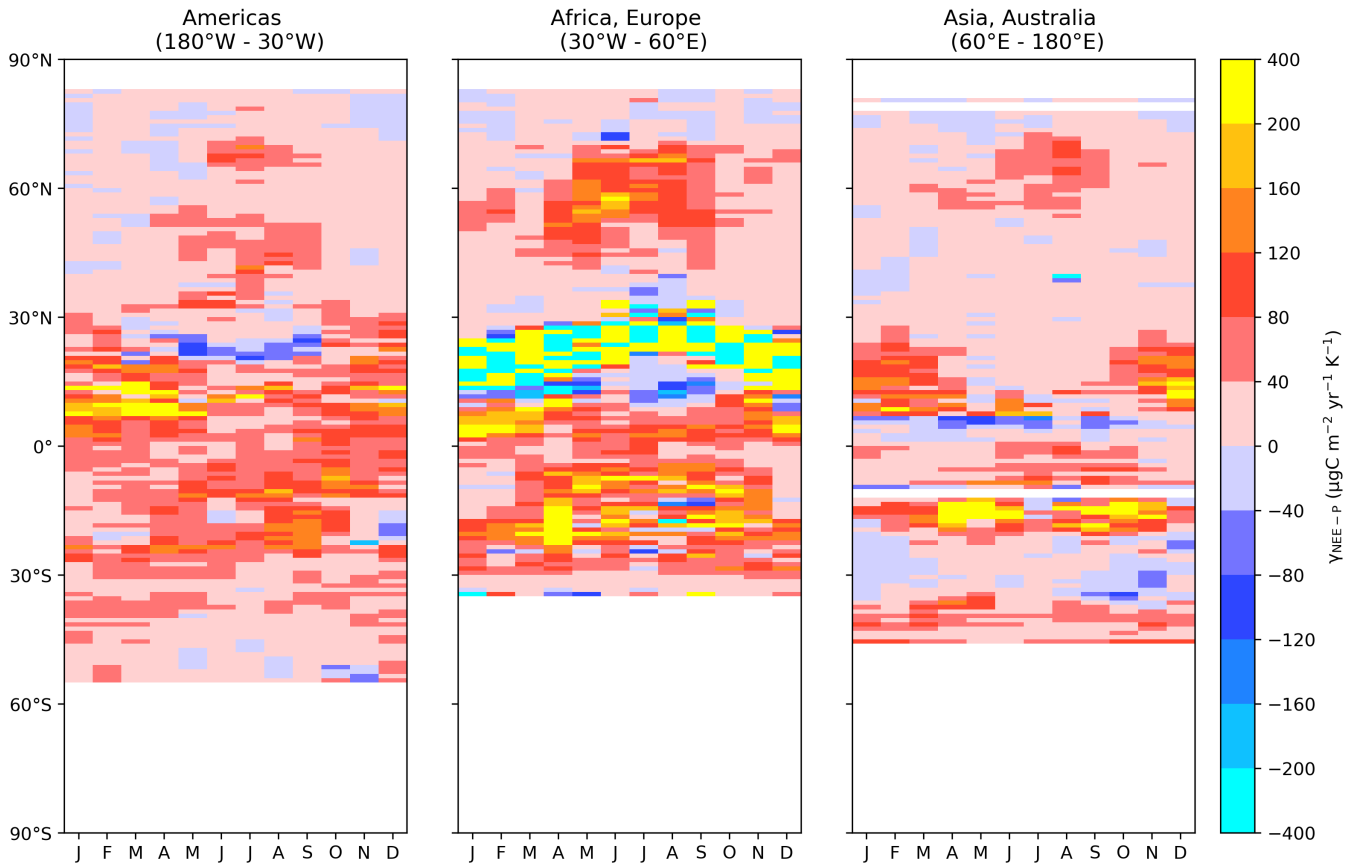
Figure 20: Interannual climate sensitivity γ_{NEE-P} (μgC m$^{-2}yr^{-1}$ mm^{-1}h) shown as Hovmöller diagrams: longitudinal averages of γ_{NEE-P} are plotted as colour over latitude (vertical) and month of the year (horizontal).

For γ_{NEE-P} we also do not observe a clear seasonal pattern (Figure 20). We do see mostly positive γ_{NEE-P}, which indicates that higher than normal monthly precipitation results in positive NEE, which does not seem correct. This could possible be explained by the too low response of the soil moisture stress in SiBCASA (van Schaik et al., 2018; Koren et al., 2018). Laan-Luijkx et al. (2015) showed that SiBCASA did not show any effect on the carbon cycle during the 2010 drought in the Amazon, which we have also seen in figure 9. (van Schaik et al., 2018) therefore implemented a hydrological model, which directly couples the soil moisture balance with the surface hydrology. With the use of this method (van Schaik et al., 2018) was able to capture the drought response in the Amazon. This could thus possible be a method increase our mathematical performance of γ_{NEE-P} and perhaps be able to link it to natural processes. This could also increase the performance and interpretability of γ_{NEE-T} as they are closely correlated (Jung et al., 2017).

3.4   Conclusion

Reproducing the biospheric seasonal cycle is possible with the generic fit function inspired by Thoning et al. (1989) for the extratropics, and performs reasonably for the tropics and Australian desert. Capturing the IAV with the use of the method of Rödenbeck et al. (2018) is possible as the performance of the T/P-fit function is for both cases better compared to the fit function. The performance especially increases in the tropics, however the interpretability of the climate sensitivity parameters γ_{NEE-T} and γ_{NEE-P} has shown difficult. γ_{NEE-T} is not comparable with literature and both γ's can hardly be linked to natural processes.

4 Computational cost savings: flux analysis

4.1 Introduction

In the last section we constructed the T and P-fit function, which reasonably capture the IAV of the SiBCASA NEE output. Optimizing the climate sensitivity parameter $\gamma_{NEE-T/P}$ can increase the performance of the T and P-fit function, however does require a CarbonTracker run, which is computational costly as we focus on longer timescales. Therefore we set up experiments, which will indicate if saving computational costs is possible and to what extent. Herewith we will only apply our experiments to the T-fit function as it slightly performance better compared to the P-fit function in terms of global RMSE, while the global terrestrial CO_2 sink is comparable.

In our first experiment we will decrease the amount of monthly NEE data from SiBCASA needed to still resemble the T-fit function. Instead of applying the T-fit function to ten years of monthly NEE data to resemble a ten year time series, we will use only the first three or five years of the dataset. Herewith, we apply the least square regression method of the T-fit function and obtain the polynomial, amplitude, phase and climate sensitivity parameters to reconstruct the original ten year NEE time series. In this case we thus only use $3/10$ and $5/10$ of the monthly NEE dataset. This method will be called the extended 3 or 5 yr T-fit function, the ext3 and ext5-T-fit function respectively. This leads us to our first research questions:

1. To what extent can the ext3 and ext5-T-fit function resemble the original T-fit function in terms of seasonal and IAV of the NEE signal?

Here the performance loss will be indicated by the RMSE and the computational savings will be compared with the original costs. So, if we for example only use half the data for the representation of the original T-fit, the computational savings are a factor two. Note that, computational savings temporally are more beneficial compared to spatial savings as less SiBCASA input data has to be downloaded, which is thus more cost-efficient than decreasing the spatial resolution twofold. With this remark in mind, the second experiment will focus on decreasing the computational costs by lowering the spatial resolution.

This will be performed for a 4° by 6° (latitude x longitude) region, which we define as a small region. Herewith we apply the T-fit function to one of the gridboxes located in the small region. The other gridboxes located in the same small region will be assigned the same parameter values to reconstruct the NEE time series. Therefore all gridboxes within a small region will be spatial and temporal homogeneous. So we only need SiBCASA NEE output for 1 out the 24 gridboxes in the small region, and therefore computationally save a factor 24. We will also check if this method can be applied for a Transcom region, which is an area of land mass that roughly covers 1/11th of the total land mass and contains 1000 land gridboxes. This could thus potentially lead to a computational savings of a factor 1000, taking into consideration that the performance is reasonable. This leads us to our second research question:

2. To what extent can a single gridbox represent a Transcom or small region in terms of the representation of seasonal and IAV in the NEE signal?

In our third and final experiment, we will first select the best representable gridbox within each Transcom region and apply the fit function to capture its parameter values. Note applying a T-fit function was preferable, however time resources were limited. Hereafter, we assign the parameter values to the other corresponding gridboxes within the Transcom region resulting in a spatial and temporal homogeneous NEE Transcom region. For the selection method, we use all gridboxes and therefore do not save computational costs spatially. The selection method will be based on selecting a gridbox, which represents the total NEE fluxes of each Transcom region as further elaborated in Section 4.2.4. With this selection method we represent the total NEE fluxes of all 11 Transcom regions with the use of only 11 gridboxes, the method will be defined as the extended Transcom fit function.

The computational costs will namely be similarly saved as in experiment 1 as in this method we apply an extended fit function to the 11 gridboxes and for example use the first 4 out of 9 years to represent the NEE time series. This method will show a lower performance compared to the extended fit function as we only use 11 instead of all gridboxes as for experiment 1. The aim of this method is to check the performance when using only such a limited amount of gridboxes and will show the performance benchmark of the fit function. In this experiment we also include the cumulative error experiment and will be further explained in Section 4.2 and 4.3.

In this section we will first further introduce our methods of our computational saving experiments after which the results will separately show the representation of the original (T-)fit function and its performance to check whether the experiment is an efficient method to save computational costs. After we also discuss the results in the same section, we wrap up with the conclusion.

4.2 Methods

4.2.1 Transcom regions

The Transcom regions are set up by the Atmospheric Tracer Transport Model Intercomparison Project (Transcom) to diagnose and quantify the uncertainty of inverse calculations of the global carbon budget, which are a result of simulated atmospheric transport errors (DeFries and Townshend, 1994). The Transcom regions were set up in 11 land and 12 ocean regions.
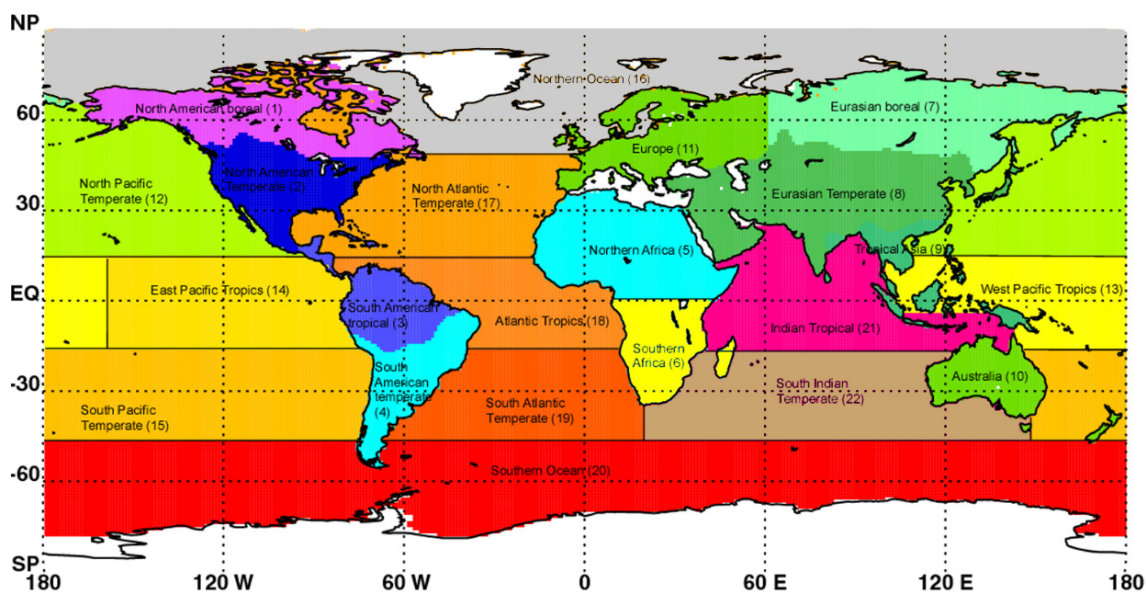

Figure 21: Transcom regions with their corresponding name , which include 11 land Transcom regions and 12 ocean Transcom regions(DeFries and Townshend, 1994).

The Transcom region mask will be used to answer our first research question. Do note that the distinction between land and ocean gridboxes is not equal to those of SiBCASA. SiBCASA has more land gridboxes, but we will only use those indicated by the Transcom region mask, which therefore does not take into account all SiBCASA gridboxes. This should be taken into account when implementing the NEE fluxes in a forward transport model such as TM5.

The third and seventh Transcom regions include ATTO and ZOTTO respectively and will be named Amazon and Boreal Russia region throughout.

4.2.2 Experiment 1: Temporal extension method

To save computational costs we will set up an experiment in which we will decrease the temporal costs of SiBCASA. We can for example take the first five years of a ten year NEE time series and extract its eleven parameters with the use of the CCGCRV program of (Thoning et al., 1989). This enables us to reconstruct the fit function in a similar fashion, in which we use our obtained parameters to construct a new fit function and only use half the data set for parameterization.

4.2.3 Experiment 2: Gridbox selection of small region

To select a gridbox in a small region (6° by 4 °, lat/lon), which will eventually represent the small region, we will select the center gridbox. To determine the center box, we average the latitude and longitude coordinates of every land box within the small region. If all gridboxes are land gridboxes, the southwestern center box will be selected. If the center box is an ocean box, we will select the the first box we come across working our way from southwest to northeast.
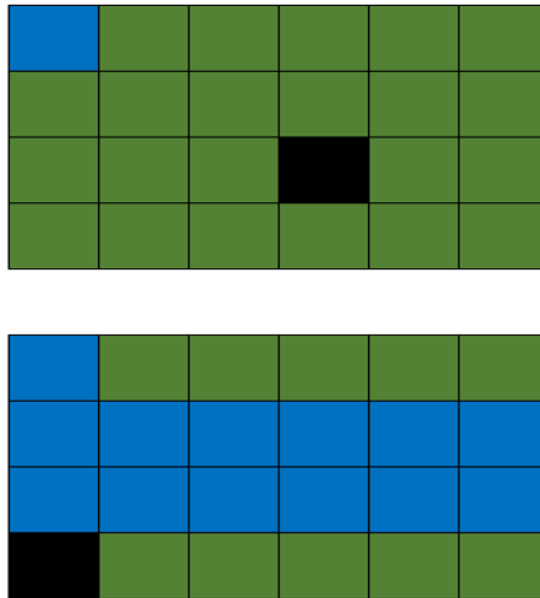
Figure 22: Visualization of the gridbox selection method within a small region. Here represented as a 4° by 6 ° (lat/lon) grid. Here the green boxes represents a land gridbox, the blue gridbox represents an ocean gridbox and the black gridbox resembles the selected gridbox.

4.2.4 Experiment 3: optimization selection method

To select the best gridbox to represent a Transcom region, we will make use of an optimization method. In this method the gridbox with the smallest difference in NEE signal compared to the average NEE of the Transcom region is selected. These are the steps taken to select the optimized gridbox for each Transcom region:

$$NEE_{rel\_region} = \Big(\sum_{n=1}^{N_{gridbox}} \sum_{t=1}^{N_{timestep}} NEE(n,t) * (A_{gridbox}/A_{gridbox_{max}}) \Big)/N_{gridbox} \qquad (24)$$

$$NEE_{box}(n) = \Big(\sum_{t=1}^{N_{timestep}} NEE(n,t)(A_{gridbox}/A_{gridbox_{max}}) \Big) \qquad (25)$$

$$n_{opt} = \underset{n}{\mathrm{ArgMin}}(NEE_{rel_{region}} - NEE_{box}(n)) \qquad (26)$$

$$NEE_{box\_opt} = NEE_{box}(n_{opt}) \qquad (27)$$

$$Performance = NEE_{box\_opt}/NEE_{rel\_region} \qquad (28)$$

Here NEE_{rel_region} is the relative NEE for the Transcom region, which represents the average gridbox value of the Transcom region. $N_{gridbox}$ is the amount of gridboxes in a Transcom region, $N_{timestep}$ is the total amount of time steps, $NEE(n,t)$ is the NEE signal of a gridbox, $A_{gridbox}$ is the total area of the gridbox and $A_{gridbox_{max}}$ is the total maximum area of a single box in the Transcom region. The fraction $A_{gridbox}$ / $A_{gridbox_{max}}$ is the relative size of each gridbox and is needed to take into account the different sizes of the region. This allows use to get a well presented average NEE flux per m^2 of the Transcom region. Note that we take the sum of all gridboxes and divide it by the total amount of gridboxes in the Transcom region instead of applying an average. This is due to the fact that the average NEE signal of a Transcom region is dampened due to changes in phase in the NEE signal as further elaborated in the Section 4.3.

Next we obtain the total sum of all separate gridboxes ($NEE_{box}(n)$), again applied with the fraction of the area. Afterwards we use a minimization function between NEE_{rel_region} and $NEE_{box}(n)$ to obtain the box number with the smallest NEE difference (n_{opt}). This allows us to select the optimized box (NEE_{box_opt}).

The optimization selection will be run for the whole time series (9 years) and for 1,2 and 4 years. This will show if it is more important to select a high performance in the first year(s) or a good resemblance after 9 years. This will be checked as for an atmospheric transport models, small errors in the early stages propagate over time a result in large errors, which is know has the cumulative propagation method. It is therefore key to know how well the optimized selection works after different time steps. Moreover, if the selected box after for example 2 years still works properly after 9 years it indicates that temporal cost savings are possible as well.

4.3 Results and discussion

4.3.1 Experiment 1: extended fit function experiment

The extended fit experiment aims to decrease the computational costs about 2-3 threefold and will first focus on the performance of the seasonal cycle with the use of the ext3 and ext5-fit function. Instead of applying the fit function for 9 years to the SiBCASA output NEE signal, we will fit it for the first three and five years. With the use of the methods from Section 2 we can obtain the set of parameters of the first three and five years. Hereafter both time series are extended back to the original 9 years with the earlier obtained set of parameters. After the removal of the second and third polynomial, the trend, we end up with our final computational savings method; the ext3-det-T-fit function and show its global performance loss versus the T-fit function in terms of RMSE.
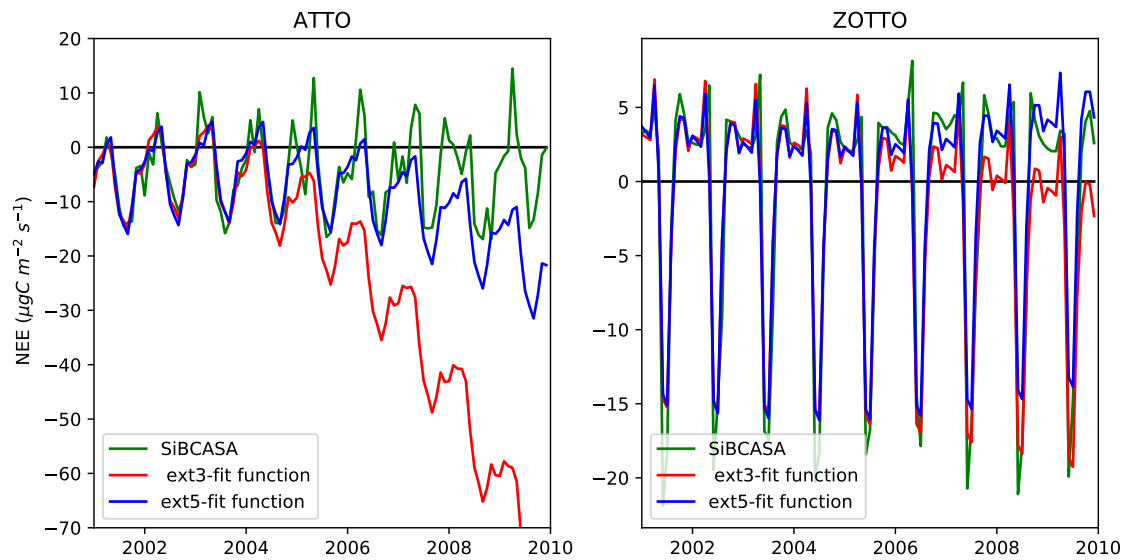

Figure 23: Time series of SiBCASA (green lines), ext3 (red lines) and ext5-fit-function (blue lines). The extended fit functions are applied over monthly SiBCASA NEE output for the ATTO and ZOTTO gridbox, in which the polynomial terms are badly parameterized. The fit function is shown for comparison.

The linear and exponential trend is not correctly parameterized as a seasonal cycle around 0 μgC m^{-2} s^{-1} is not upheld over the years for the extended fit functions (Figure 23). The parameterized trend of the first 3 or 5 years is crucial to determine the overall trend, due to the cumulative error propagation.Therefore the worse parameterized trend of the ext3-fit-function compared to the ext5-fit-function causes a large increasing error over time for both the ATTO and ZOTTO locations.

The trend offset for both extended fit functions is mainly caused by the IAV in the parameterized years. Therefore ATTO shows a larger trend offset as IAV is more pronounced in the NEE signal compared to ZOTTO. The extended fit functions are able to fit the four harmonics, amplitude and trend well compared to the fit function as expected.

The ext5-fit-function is most suitable to decrease the computational costs as the RMSE decrease of ext5-fit-function outweighs the computational costs decrease of the ext3-fit (Figure 23). The ext3-fit shows a RMSE increase of 30.74 μgC m^{-2} s^{-1} for ATTO and 2.44 μgC m^{-2} s^{-1} for ZOTTO with a factor three computational costs decrease both compared to SiBCASA. For the ext5-fit we obtain a RMSE increase of 7.87 μgC m^{-2} s^{-1} for ATTO and a 2.03 μgC m^{-2} s^{-1} for ZOTTO with a factor 1.8 computational costs decrease both compared to SiB-CASA. To increase the performance of the extended fit function we will remove the linear and exponential trend. This gives us an extended fit function without a trend. This computational saving method is called the ext3-det-T-fit function:

$$Ext3 - det - T - fitfunction = a_1 + \sum_{n=1}^{4} C_n \big(\sin(2n\pi t + \phi_n) + T_{IAV} \cdot \gamma_{NEE-T} \big) \quad (29)$$

Do note that the γ_{NEE-T} will be calculated from linear regression with only 3 data points as γ_{NEE-T} contains a parameter value for each month. For the mean initial value (a_1), the amplitude (C_n) and the phase (ϕ_n) only the first three years of data is used to parameterize the ext3-det-T-fit function.
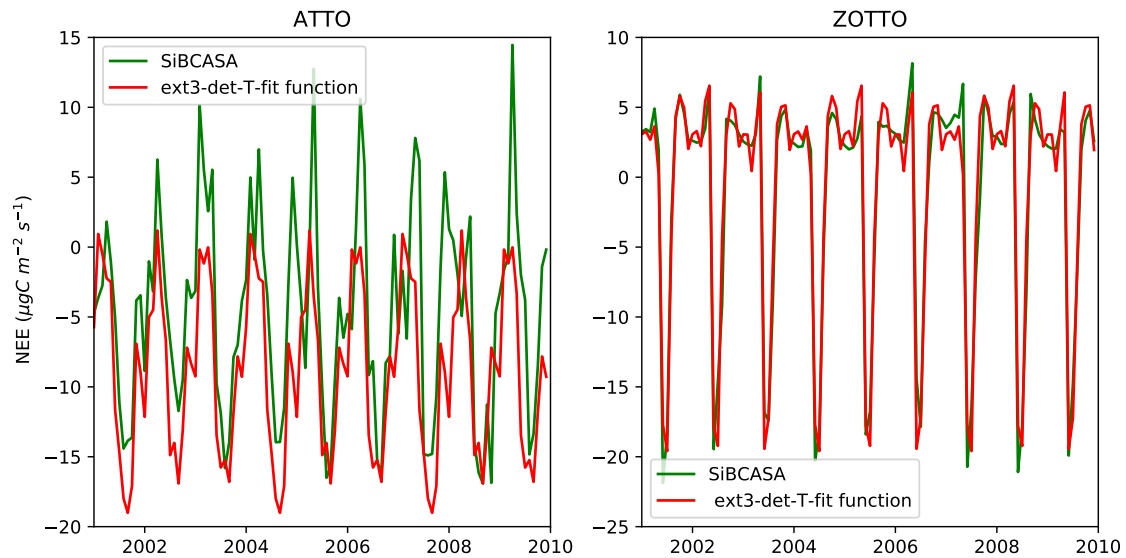

Figure 24: Time series of ext3-det-T-fit function (red line) compared to the SiBCASA output (green line) for ATTO and ZOTTO gridbox with a structural underestimation of the ext3-det-T-fit function for ATTO.

The ext3-det-T-fit function improved the extended fit functions, however does show an structural underestimation for ATTO (Figure 24). The small seasonal cycle of the dry and wet season in the first year in terms of NEE causes the ext3-det-T-fit function to structurally underestimate the signal. Especially for ZOTTO, the ext3-det-T-fit function now cycles around 0 μgC m^{-2} s^{-1}, while the seasonal cycle, compared to the extended fit functions, is still captured well. The RMSE of the ext3-det-T-fit at ATTO is 6.26 μgC m^{-2} s^{-1} and for ZOTTO is 1.86 μgC m^{-2} s^{-1} both compared to SiBCASA. The downside of this method is that we can not determine the CO2 fertilization effect, which is determined by the least square regression of the polynomials of the original T-fit. The CO_2 fertilization effect takes place at a timescale of decades (Cox et al., 2013) and is therefore not the main focus as we currently try to decrease the computational costs at a 9 year timescale.
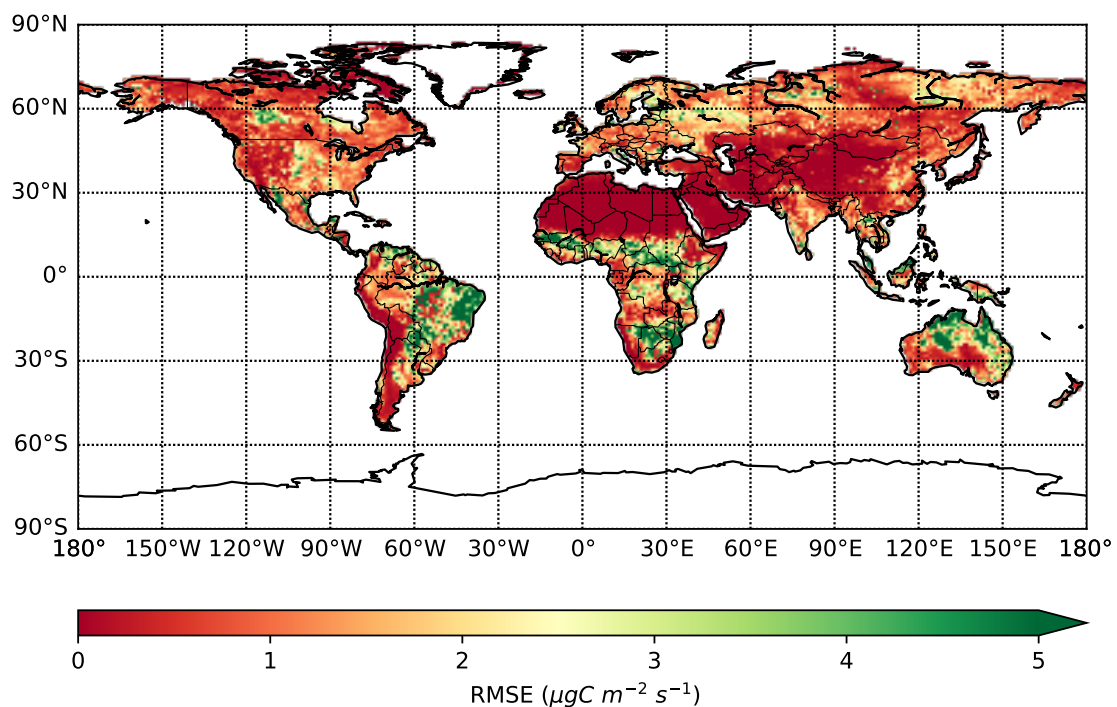
Figure 25: Global RMSE of the ext3-det-T-fit function - T-fit function, which is the performance loss of the ext3-det-T-fit function compared to the T-fit function. The highest RMSE difference, the largest performance loss, is mainly depicted in the tropics.

The highest RMSE difference between the RMSE of the ext3-det-T-fit function and the T-fit function, compared to SiBCASA, is in the tropics (Figure 25), due to the large, not well represented IAV in the NEE time series (see Figure 24). We also observe high RMSE values in the extratropics such as Eastern Europe for example, in which the seasonal cycle is generally dominant. Two droughts between 2001 and 2004 in this area possible decreased the GPP and caused IAV in the NEE signal (Spinoni et al., 2015). As we selected these years to parameterize the signal, the seasonal cycle deviates and causes a low performance.

The Northern Australian desert shows a large RMSE as well taking into account the relative small NEE fluxes. Interannual precipitation anomalies caused by soutward extension of the monsoon through, tropical cyclones, Northwest Cloudbands and ENSO cause this IAV in the signal as previous explained in 3.3, (Huxman et al., 2004; Poulter et al., 2014; Australian Bureau of Statistics, 2006).

The global RMSE of the ext3-det-T-fit function is 3.14 μgC m^{-2} s^{-1} compared to the global 1.81 μgC m^{-2} s^{-1} of the T-fit. For the tropics the RMSE of the ext3-det-T-fit function is 4.81 μgC m^{-2} s^{-1}, while the T-fit accounts for a RMSE of 2.611 μgC m^{-2} s^{-1}. The performance decrease in terms of RMSE of around 70% compared to the T-fit is reasonable when comparing it to the computational saving costs of a factor 3. The global terrestrial CO_2 sink of the ext3-det-T-fit function is -2.87 Pg yr^{-1} compared to the T-fit function, which has a global terrestrial CO_2 sink of -2.76 Pg yr^{-1}, both for the periods from 2001-2010. This is due to the structural underestimation in the tropics (Figure 24) and the ext3-det-T-fit function will thus have a lower atmospheric CO_2 concentration over time as the difference with the T-fit function accumulates. The atmospheric sensitivity of these different T-fit functions will be further elaborated in Section 5.2.1.

4.3.2 Experiment 2: Spatial resolution experiment

The goal of this section is to decrease the spatial resolution, while maintaining a high performance of the NEE signal to save computational costs. Herewith we first test the coarsest possible global resolution, the Transcom region resolution and check whether the variability in the Transcom regions is small enough for one gridbox to represent the average NEE time series. To perform this analysis, we will check the SiBCASA's spatial NEE variability first and perform this analysis for a small region resolution as well.

The second step is to verify whether the parameters of the T-fit function show a low variability when we apply them to all gridboxes in a Transcom and small region. A low variability indicates that the parameter values of one gridbox can be assigned to other gridboxes located in the same Transcom/small region to obtain a spatial and temporal homogeneous pattern in each region. As mentioned in Section 4.2, we select the Amazon and Boreal region, which are both Transcom regions to analyze the spatial NEE variability. And we select the ATTO and ZOTTO regions, which are both small regions surrounding the observatory gridboxes to analyze the spatial NEE variability.

We conclude with a performance check in terms of a global RMSE map, in which we apply the T-fit function over a 4° by 6 ° grid (small region resolution), which is defined as the small region T-fit function and compare it to the RMSE of the T-fit function.
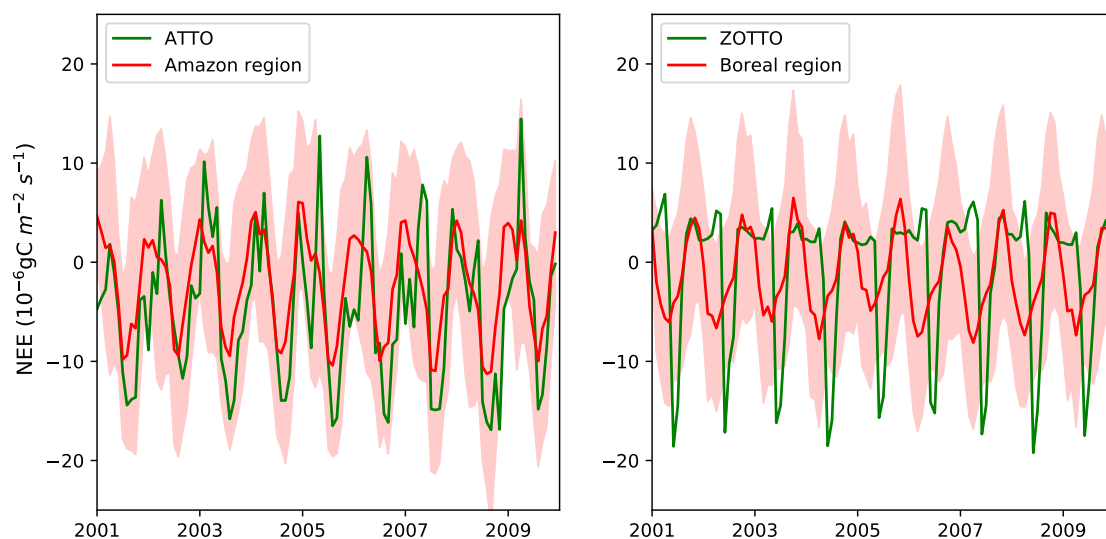
Figure 26: Averaged monthly SiBCASA NEE over the South American tropical Transcom region (left panel) and the Eurasian Boreal Transcom region (right panel). Here the red lines indicate the averaged NEE of the corresponding Transcom region with variability in terms of monthly σ, indicated by the red shading for each Transcom region. The green line represents the single NEE gridbox value for ATTO/ZOTTO as reference. In both cases the spatial NEE variability is too large to apply spatial and temporal homogeneous NEE values.

The spatial resolution can not be decreased to Transcom region scale as the variability of the Amazon and Boreal region are too large for the application of spatial and temporal homogeneous NEE values. The average Transcom region NEE time series shows large discrepancies compared to the single box NEE time series in terms of the phase and amplitude. The large NEE variability in the Amazon region is mainly caused by the different seasonality phases within the region with an average variance of 9.85 μgC m^{-2} s-1. At the equator, two wet and dry seasons seem to occur annually as the tropical rain belt passes by twice as shown in the ATTO gridbox located at 2 °S. The Northern and Southern tropics only have one wet and dry season as the tropical rain belt only passes once. Thus positive and negative NEE fluxes occur at the same time within the region.

The average Amazon region NEE signal is dominated by the Southern located tropics as the largest share of tropical landmass is located South of the equator. Therefore during the months October until March positive NEE fluxes are shown indicating the wet season in the Amazon region. The signal is moreover dominated by the single wet and dry seasonal cycle as it is more extreme in the Northern and Southern tropics, resulting in a larger seasonal NEE amplitude. The amplitude of the average Amazon NEE signal is smaller compared to single gridbox values such as ATTO. This is due to the averaging of the NEE signal while having large phase difference within the region.

The Boreal region also shows a large phase variability as the σ is 8.69 μgC m^{-2} s-1. This is mainly due to the phase difference in the spring and autumn months generally dependent on the latitude. The southerly located gridboxes have a larger summer season and earlier snow melt, which causes earlier plant growth compared to northerly located gridboxes in the Boreal Region. This causes a timing difference in NEE decrease and for autumn vice verse applies.

The effect of the phase variability in the Amazon region also applies to the Boreal region as the amplitude of the average NEE time series of the Boreal region is dampened. It also clearly shows that the phase difference compared to ZOTTO is large, even though the grid box is approximately in the middle of the Transcom region. Here from we can conclude that a single gridbox cannot represent a whole Transcom region. Moreover it is not possible to parameterize the average region's NEE signal as it does not match possible single gridbox values. To decrease the variability within the NEE signal of a region we will increase the global spatial resolution to 6°x4°(lat x lon).
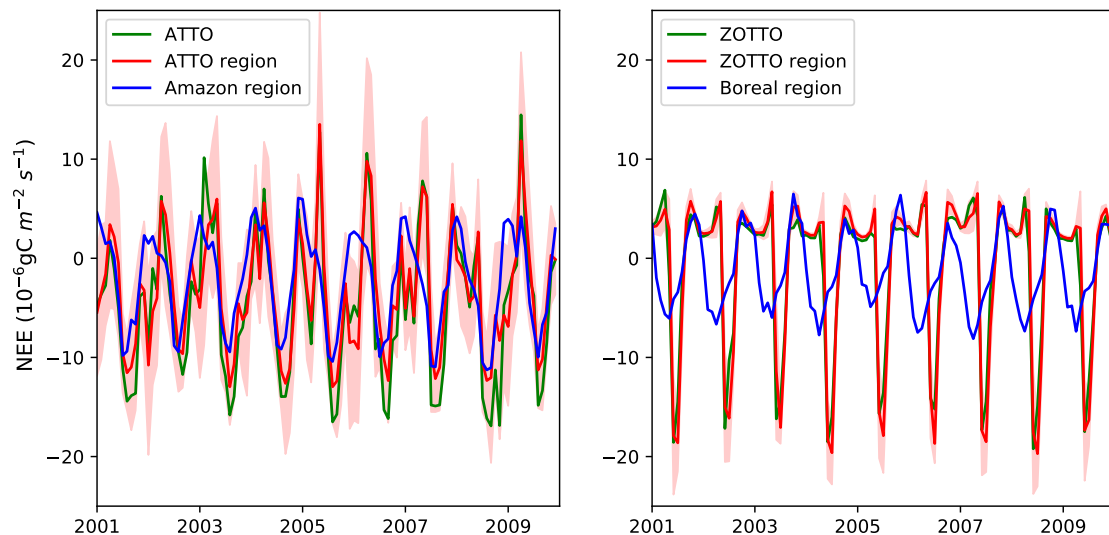

Figure 27: Averaged monthly NEE over the ATTO region (left panel) and the ZOTTO region. Monthly σ is indicated by the red shading for each small region. The ATTO and ZOTTO NEE single gridbox time series and their corresponding monthly averaged NEE Transcom regions are shown as reference. The spatial NEE variability is small as indicated by the red shading for both small regions.

Both small regions show relative small NEE variability compared to their corresponding NEE variability within each Transcom region (Figure 26 and 27). The ATTO and ZOTTO region have an average σ of 4.43 μgC m^{-2} s-1 and 1.022 μgC m^{-2} s-1 while the Amazon and Boreal region have an average σ of 9.85 μgC m^{-2} s-1 and 8.69 μgC m^{-2} s-1 respectively. Do note that both small regions have a small σ decrease compared to their computational costs increase, which increases from approximately 1/1000 to 1/24 of the original costs to run SiBCASA.

Figure 27 again shows the disability of the average Transcom region NEE signal to represent a large region. The average Boreal and $1\,\sigma$ shading of the Boreal region, and ZOTTO region NEE signal are not comparable during almost the entire year. Using the average Boreal region signal for the ZOTTO region would therefore cause large disagreement in the region's atmospheric CO_2 concentration even tough the ZOTTO region is located in the center of the Boreal region.

On a small region spatial resolution the σ between the ATTO and ZOTTO small regions differs by a factor 4.3. This can be explained by Kottek et al. (2006), which shows that in the Amazon Transcom region and especially near the equator, multiple climate zones are classified, while the Boreal region is dominated by one climate zone. To check whether this spatial resolution could be an performance and cost efficient method, we will now elaborate the variability of the parameters of the T-fit function for the previous mentioned Transcom and small regions.
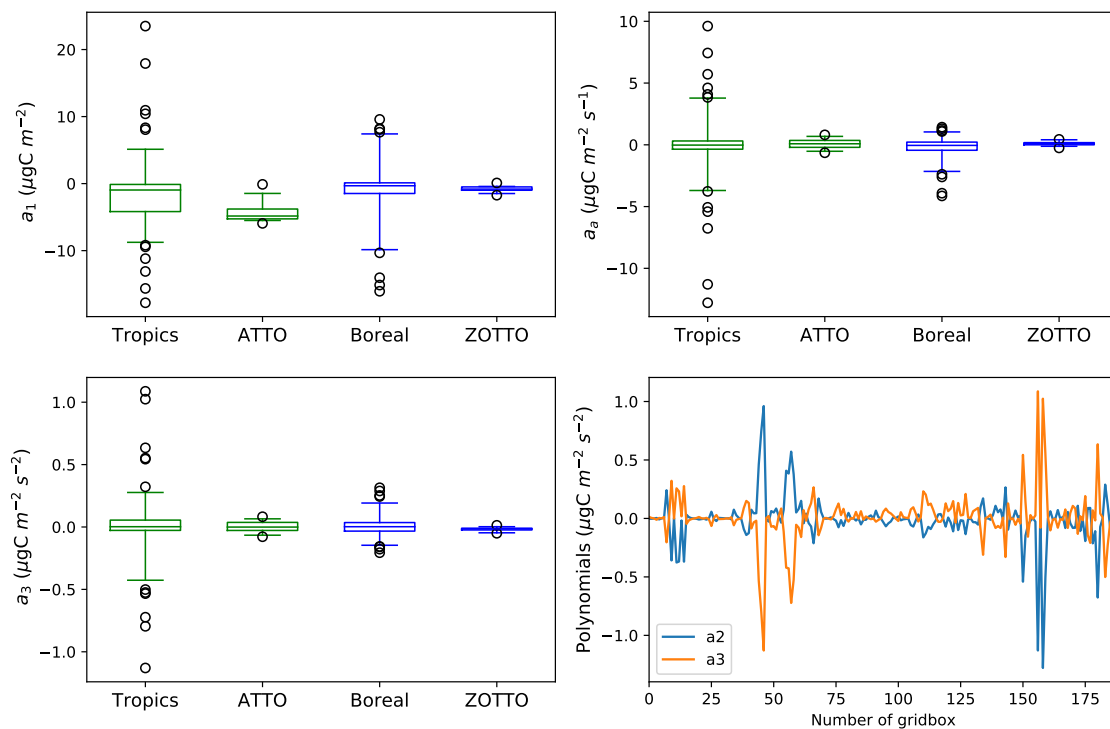
Figure 28: Top and left bottom panels: Boxplots of the Tropical and Boreal Transcom region and the small ATTO and ZOTTO ($4°$ by $6°$) region. The whiskers represent a 95% confidence interval and the white dots represent outliers. a1 represents the mean initial value, a2 represents the linear polynomial term and a3 represents the quadratic polynomial term of the fit function. Bottom right panel: Values of 1/10 of a2 (linear term) plotted versus a3 (quadratic term). Here a value of 1/10 of a2 is chosen for visualization purposes. The variability of the polynomials of the small regions is small and 1 gridbox can represent a small region.

The polynomials suggest that a one gridbox can represent the ATTO and ZOTTO small regions (Figure 28. The mean initial value (a1) shows the largest variability for ATTO, but does still fall within the 95 % confidence interval of the tropical Transcom region. For the ZOTTO region, parameter variability is low throughout and indicates that one gridbox thus can represent a small region. From here we can also again conclude that a single gridbox cannot represent the NEE time series of a Transcom region.

We have seen in earlier time series of ATTO and ZOTTO that almost no trend can be observed in a short time series of nine years. However, the parameter values of a2 and a3 are both not close to 0 μgC m^{-2} s^{-1}. It therefore seems that a2 and a3 compensate each other over time. The right bottom panel of figure 28 confirms our hypothesis as in this case the a2 and a3 parameters of the Amazon Transcom region cancel each other out approximately. Do note that this type of visualization method does not show the exact pattern of a2 and a3, which time series have indicated (not shown). The regularity of this pattern does suggest that the terms cancel each other out over time as well.

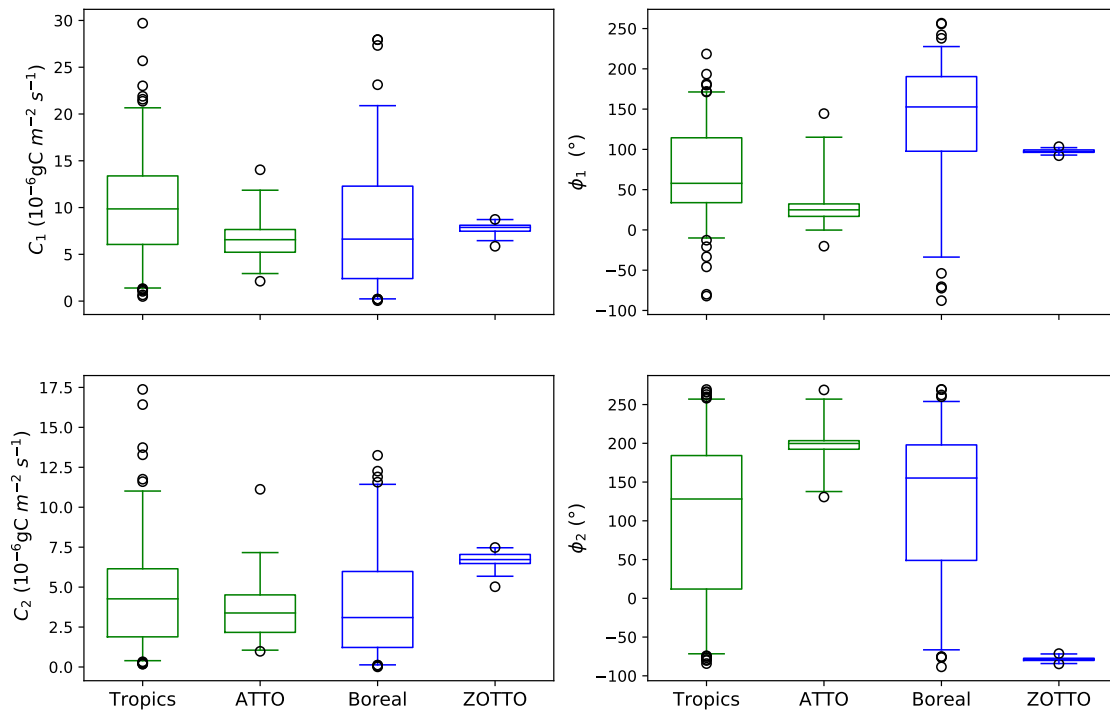
Figure 29: Top panels: Boxplot of the amplitude (C_1) and phase(ϕ_1) of the T-fit function. Bottom panels: Boxplot of the amplitude (C_2) and phase(ϕ_2) of the T-fit function. The tropics resembles the tropical Transcom region, ATTO resemble the ATTO small region, Boreal resembles the Boreal Transcom region and ZOTTO resembles the ZOTTO small region.

The amplitude and phase of the Transcom regions and small regions is highly variable for ATTO, while for ZOTTO shows a small variability. Within the ZOTTO small region, variability between climate drivers is small as the small region falls within one gridbox (Kottek et al., 2006). The Boreal Transcom region shows large variability of the phase and amplitude of the parameters of the T-fit function. This can be explained by the approximately five different climate zones in this region (Kottek et al., 2006). Note that the phase values of the Transcom regions near 250° and -100° can be interpreted as the same value.

At the ATTO small region, the phase of the dry and wet season is subjected to interannual variability, due to droughts and interannual fluctuations of the position of the ITCZ (van Schaik et al., 2018; Phillips et al., 2009; Gruber, 1972; Adam et al., 2016). Next to that, the amount of precipitation in the Amazon is spatial heterogeneous and causes a variable intensity on droughts (van Schaik et al., 2018). And finally the location of a gridbox determines when the ITCZ passes and therefore influences the start and end of the dry season (Fu and Li, 2004), which causes the phase variability in the Amazon Transcom region and ATTO region. The differences in the ATTO region are smaller compared to the tropical Transcom region as the tropical Transcom region is simply larger. The results of the third and fourth harmonic are shown in Appendix 8.1. With this information in mind, we would now like to check the global performance when selecting one gridbox for each small region.
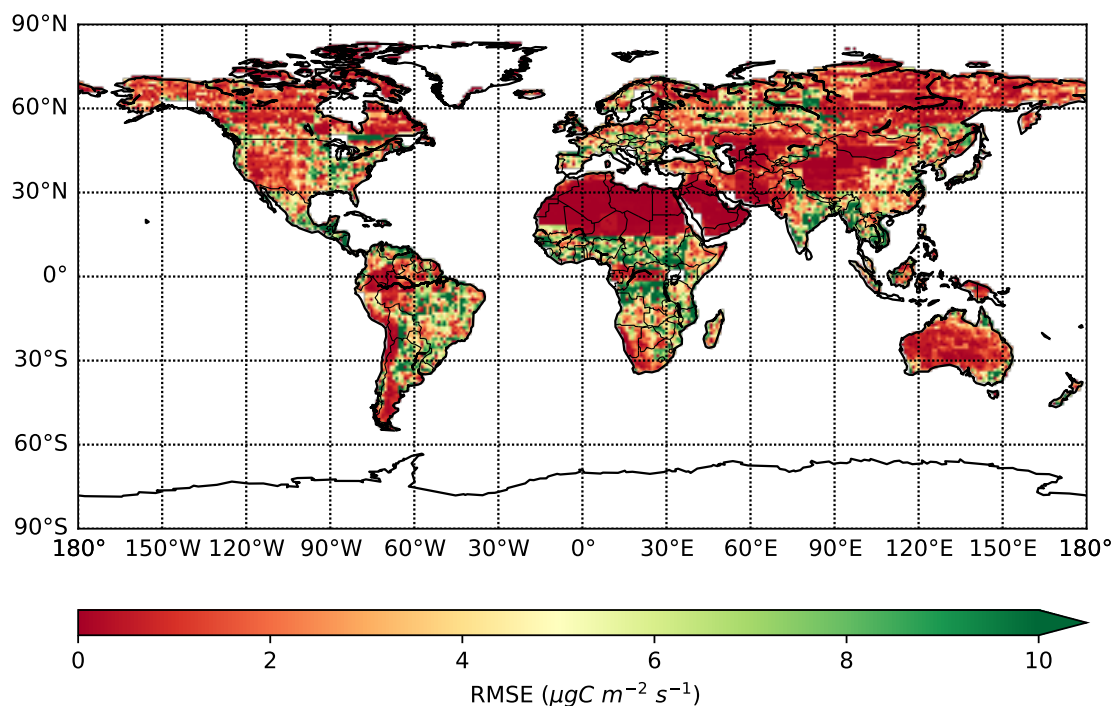

Figure 30: RMSE of the small region T-fit function - T-fit function, which indicates the performance loss of the small region T-fit function calculated from 2001-2010. The RMSE is fairly evenly distributed among the Amazon Basin and extratropics, while it shows difficulties at outer ranges of the African tropics.

The strong point about this method is the better representation of the tropics and especially of the Amazon Basin and Indonesian tropics relative to the RMSE of for example Europe. The larger and more homogeneous a climate zone, such as the northwestern Amazon Basin and the Australian desert, the lower the RMSE. Therefore the western equatorial strip of Africa is shows a low RMSE increase of 0-2 μ gC m^{-2} s^{-1}. While north of the African tropics until the Sahara desert, multiple climate zones are located in a relative small region and therefore one gridbox cannot represent a small region (Kottek et al., 2006).

The 4x6 region seems to perform well compared to its computational savings. The global RMSE increase of the 4x6 T fit compared to the T-fit function is 3.19 μgC m^{-2} and for the tropics the RMSE increase is 4.78 μgC m^{-2}, which are large as the global T-fit function only has a global RMSE of 1.81 μgC m^{-2}. However, if we use the global averaged monthly NEE values as indicator for its performance results improve. This is due to the fact that some gridboxes in each small region are underestimated and others are overestimated, which leaves us with a reasonable represented global land sink of -2.95 Pg yr^{-1} for the small region T-fit function compared to -2.76 Pg yr^{-1} for the T-fit function.The RMSE error is thus not able to show if the errors in the signal smoothen out spatially, which from a TM5 perspective will occur as longitudinal atmospheric mixing is in the order of weeks (Bolin and Keeling, 1963).

4.3.3   Experiment 3: Extended Transcom fit function experiment

We have seen in the previous section that one gridbox cannot resemble the average time series of a Transcom region. This time we will check if the total relative sum of the region (section 4.2) corresponds to the total sum of a single gridbox. If so, it should constrain the eventual CO_2 mole fraction difference with SiBCASA when we run it with our forward atmospheric transport model TM5. Hereafter we will show you the the cumulative error experiment, which will (1) indicate how much NEE monthly data we need to represent the whole time series and therefore save computational costs.(2) Check whether it is more important to have a well presented NEE sum in the first years or a probably decent presented NEE sum after the whole time series, which is from importance when running the TM5 model from a cumulative error propagation perspective. We will finally show the performance of the Transcom region proxy with the use of the global RMSE, which from here we call the Transcom T-fit.

All gridboxes are able to represent the relative sum of each Transcom region very well for the total time period it is optimized for, which is indicated by the left column of Table 1 This is due to the fact that we do not take into account the amplitude and phase of the seasonal pattern of the box, but only the 9 year total sum in this case. This implies that for implementing each optimized gridbox for all gridboxes in each Transcom region, the total amount of NEE is equal when we calculate the total NEE of each individual gridbox for a 9 year period. With this method we checked all gridboxes to select the best gridbox and therefore do not save computational costs. Therefore we will next check if the selected optimized gridboxes

after 1,2,4,8 years can resemble the 9 year total NEE flux of the Transcom region, when we run the gridbox thus for the full 9 years.

	1yr opt		2yr opt		4yr opt		8yr opt	
	1yr	9yr	2yr	9yr	4yr	9yr	8yr	9yr
TR1	1.00	0.18	1.00	2.28	1.00	1.02	1.00	0.85
TR2	1.01	0.67	1.00	0.20	1.00	0.67	1.00	0.88
TR3	1.02	1.14	1.02	0.97	1.02	0.97	0.99	0.99
TR4	1.01	2.44	1.24	0.58	1.03	0.17	1.02	0.68
TR5	0.99	1.30	1.01	0.51	1.00	0.86	1.00	1.55
TR6	1.03	1.22	0.98	0.77	1.02	0.77	1.00	1.04
TR7	1.00	0.49	1.00	1.55	1.00	0.38	1.00	1.17
TR8	1.01	1.29	1.00	0.52	1.00	0.64	1.00	1.21
TR9	1.01	1.93	1.00	0.92	1.00	1.13	1.01	1.01
TR10	1.00	3.46	1.01	79.78	1.00	22.11	1.01	0.37
TR11	1.00	1.72	1.00	0.32	1.00	1.28	1.00	1.15
Average	1.01	0.81	1.00	8.04	1.00	2.72	1.00	0.92

Table 1: Cumulative error experiment of the Transcom T-fit function from 2001-2010. Here 1 yr opt is thus the box optimization method for 2001. TR1 stands for Transcom region 1 etcetera, in which section 4.2.1 shows the corresponding name and location of each Transcom region. The calculation of the performance is explained in section 4.2.

The cumulative error experiment can decrease computational costs by about a factor 1/9 as optimizing for 8 years gives an representation of 0.92 of the sum of for a period of 9 years (Table 1). However, this time we do not take into account phase and amplitude of the seasons. TR10 shows large discrepancy of the performance compared to the other Transcom regions. TR10 is the Australian continent and is known for its large IAV in the NEE signal and therefore more difficult to get a well representative sum. Moreover, the northern part of Australia is located in the tropical climate zone according to Kottek et al. (2006), which indicates that selecting one gridbox to represent a desert and tropical climate zone seems not feasible. To check whether a performance of 0.92 after a selection period of 8 years (Table 1) corresponds to a good performance in terms of RMSE, we now show the global RMSE. This indicates the representativeness of the seasonal cycle, which is not optimized in this method. We namely optimized for a good representation of the total NEE fluxes for each period disregarding the yearly seasonal cycle. Thus a gridbox with a large seasonal amplitude could perform equally to a gridbox with a small seasonal amplitude and is one of the weaknesses of this method.
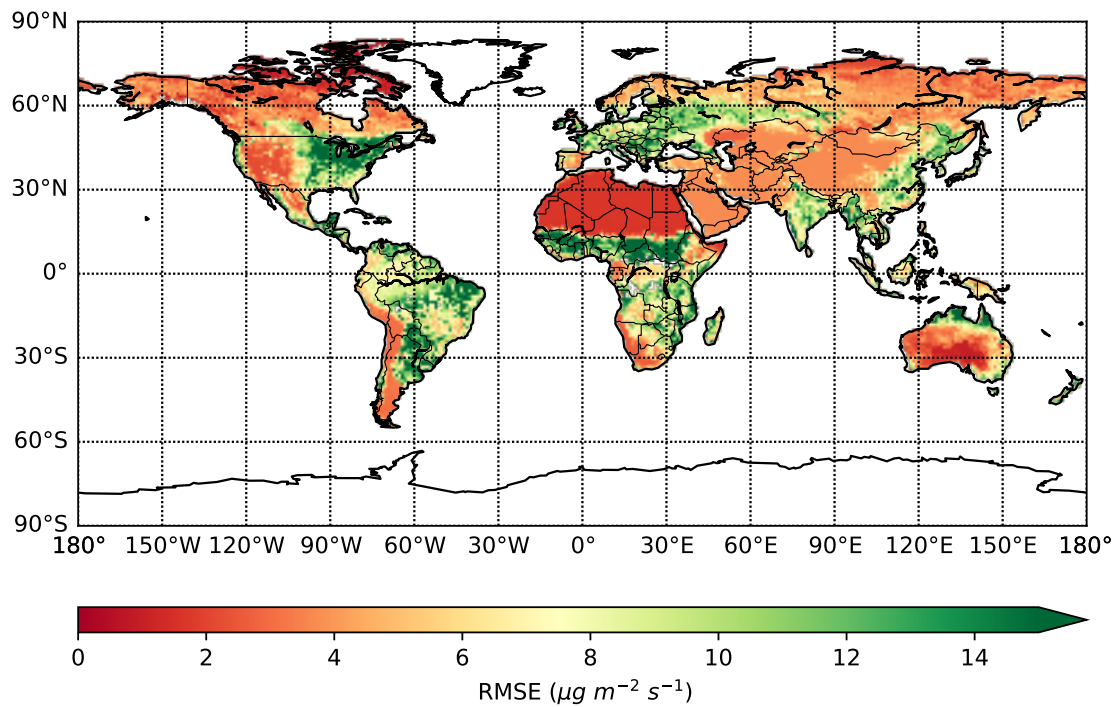
Figure 31: Performance of the Transcom T-fit function of the optimized gridboxes for each Transcom region after an optimization period of 8 years for an 9 year time series in terms of RMSE.

The Transcom T-fit function is not efficient in terms of computational costs savings versus performance decrease as the RMSE error is very high. Globally we have a RMSE of 6.37 μ gC m^{-2} and in the tropics we have a RMSE of 8.62 μ gC m^{-2}. Do note that the Amazon Basin does not have the largest RMSE as one gridbox in northern South America is able to somewhat represent the NEE signal of all gridboxes, due to the fact that the Transcom region is dominated by the tropical climate zone. For Australia, we can confirm our earlier findings, in which the northern tropical region of Australia is not well represented. Therefore the ability to represent the seasonal well temporally and spatially seems of great importance as the 0.92 performance indicator after 8 out of 9 years in the previous cumulative error experiment does not translate to a good performance in terms of RMSE.

4.4 Conclusion

We have seen that saving computational costs by a factor 3 temporally leads to a well represented RMSE, however does show a large structural bias in terms of monthly averaged NEE. This is mainly caused by IAV in the first 3 years of the NEE signal, leading to parameterization difficulties. Decreasing the spatial resolution, the 4x6 T-fit showed high global RMSE, but global and tropical monthly NEE average are well presented indicating a potential good performance in terms of atmospheric sensitivity as spatial NEE differences are smoothened. The Transcom T-fit function shows a low performance compared to computational savings as the seasonal cycle of different climate zone within a Transcom region are not well presented.

5 Atmospheric sensitivity analysis of the fit functions with TM5

5.1 Introduction

The previous sections showed the possibilities of decreasing the computational costs from a spatial and temporal perspective of the T-fit function. Here the monthly averaged NEE and RMSE gave a first indication of the performance of these simplified NEE fluxes. The T/P-fit function showed a good monthly average estimate, however does have difficulty capturing the IAV of the signal for which an inversion can be applied. Before CarbonTracker can optimize these simplified biosphere surface fluxes, they first need to be transformed to CO_2 mole fractions. Therefore the SiBCASA NEE output and its simplified NEE fluxes will be used as input for the global chemistry transport model TM5 (Peters et al., 2005; Van der Laan-Luijkx et al., 2017), (see Section 2). The TM5 model will run with simplified NEE fluxes as input and will thus deviate from the TM5 run with SiBCASA. The atmospheric sensitivity of the simplified NEE fluxes, with SiBCASA as reference, show whether they are suitable as input for CarbonTracker.

The aim of this Section is therefore to capture the atmospheric sensitivity in the TM5 forward model run with simplified different NEE input fluxes. This gains a deeper understanding of the possibilities to implement the simplified NEE fluxes as input for TM5 and eventually as prior for CarbonTracker. To reach our aim, we first need to answer the following research question:

1. Which processes contribute to the atmospheric CO_2 signal in terms of growth rate, seasonal amplitude, and background concentration at different locations?

The atmospheric CO_2 concentration will be shown at four different atmospheric observatory locations across the whole latitude of the Earth. This should show the latitudinal gradient as the largest share, $\sim 90\%$ of fossil fuel emissions occur in the Northern hemisphere while taking into account the inter-hemispheric mixing of 1-1.5 years (Conway and Tans, 1999). These locations are also selected in view of follow-up studies using an inversion run with CarbonTracker to determine the quality of the prior (TM5 output). Therefore this Section will take the CO_2 time series with SiBCASA input as reference for the other runs. Hereafter we can answer the main research question:

2. What is the sensitivity of atmospheric CO_2 in forward atmospheric simulations using TM5 with simplified NEE fluxes for different locations?

5.2 Methods

This section will first explain more about TM5 and introduce the four observation stations and explain how monthly and global mean CO_2 concentration are determined. Hereafter the set-up of the TM5 runs will be elaborated with its different inputs to determine the atmospheric sensitivity.

5.2.1 TM5

TM5 is a two-way nested chemistry transport model which was originally set up to model regional atmospheric composition (Krol et al., 2005). The entire Earth's atmosphere is taken into account as sources and sinks are not regionally limited. This is especially true for CO_2, which is a long-lived trace gas and therefore local sources are affected by long range transport from anthropogenic emissions, which are not uniformly distributed around the Earth (Peters et al., 2004).

However from a computational perspective, running TM5 on a high resolution global grid is not possible. Therefore Krol et al. (2005) used the 'nesting' approach, in which a regional high resolution model transfers information into a global coarse resolution model. Thus the global domain provides the boundary conditions for the nested region. This is the main new feature of TM5 which originates from first tracer model set up by Heimann et al. (1988) and builds upon the TM3 model, in which many of the parameterization and concepts are used. TM5 has also been used to independent determine and verify the national and regional CO_2 uncertainty of the United States to establish fossil fuel emission reduction targets (Basu et al., 2016).

5.2.2 Observation stations and measurements of CO_2 mole fractions

Throughout this Section, four CO_2 observatory stations will be compared which comprises the whole range of latitudes. The observations will validate the processes (e.g. atmospheric mixing) taken into account for the reference TM5 run with SiBCASA input. The selected stations are (1) Barrow, Alaska (71 °N, 157 °W), (2) Mauna Loa, Hawaii, (20 °N, 156 °W) (3) American Samoa (14.2 °S, 171 °W), and (4) Antarctica, South Pole (90.0 °S, 59 °E) (Figure 32. These observatory stations are taken as baseline and represent the background air for a large region from the Cooperative Global Air Sampling Network (Conway et al., 1994). The observation stations are remotely located and therefore do not show significant CO_2 'spikes' caused by local anthropogenic sources or vegetation (NOAA a, 2016). The observational data can be found at NOAA c (2018)
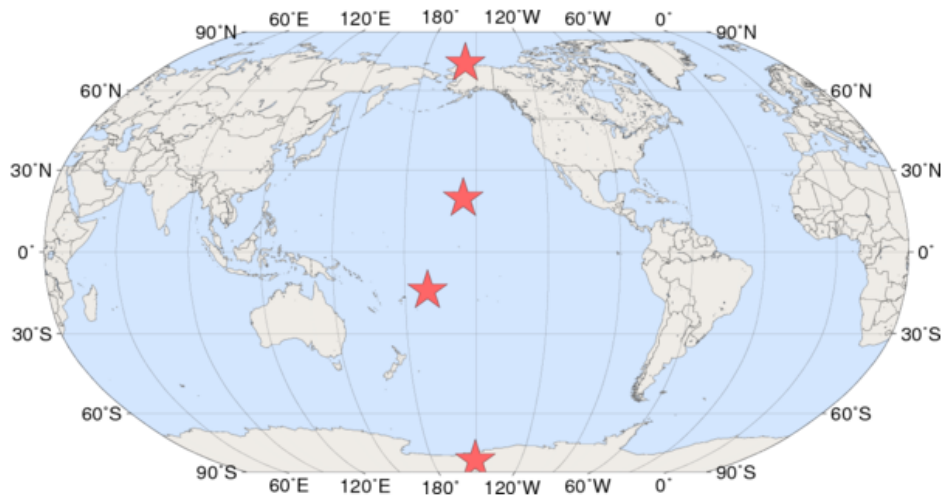
Figure 32: Locations of the four CO_2 observation stations. From North to South: Barrow, Mauna Loa, American Samoa and the South Pole (NOAA e, 2018).

For the determination of the monthly averaged CO_2 mole fraction, daily averages from the National Oceanic and Atmospheric Administration are in situ measured and fitted through the curve fitting method of Thoning et al. (1989) as explained in NOAA f (2016) and Section 3.2.2. The daily averages are calculated from a selection of hourly averages which satisfy the background concentration and therefore represent regional air. The selection process filters out influences of nearby anthropogenic CO_2 emissions and local vegetation. More detailed information about the selection criteria is given by NOAA b (2018). The observational data is available at (NOAA c, 2018).

The global CO_2 mole fraction estimates are based on measurements from a subset of observation stations (NOAA d, 2016). Herewith only stations with mainly a well-mixed boundary layer are selected, typically at remote locations at sea level such as the observation station at American Samoa. Hereafter the noise of weekly measurements due to synoptic-scale atmospheric variability is reduced with the smooth curve method of Thoning et al. (1989). The global mean surface values are determined by a weekly synchronization period.

5.2.3 Settings of TM5 runs

TM5 can be run on global 4°x6°, 2°x3° and 1°x1° horizontal resolution (longitude × latitude) and can perform with a vertical resolution of 25, 34 or 60 layers. The runs in this section make use of a 4°x6° and 25 vertical layer resolution without the use of the nesting approach. The coarse resolution decreases computational costs and the results are only used to test the atmospheric sensitivity of the fit functions and therefore only used as indicators. The output is in CO_2 mole fractions and the temporal resolution is in months which is consistent with the temporal resolution in Sections 3 and 4.

The global mean surface values of the TM5 runs are obtained by averaging the TM5 output CO_2 mole fraction at every gridbox at surface level. The growth rate and latitudinal gradient are calculated with the use of Thoning et al. (1989). The latitudinal gradient is obtained by the difference of the background concentration of Barrow minus the background concentration at the South Pole.

Name	Input	Resolution (lat x lon)	Temporal (yrs)
SiBCASA	SiBCASA	180 × 360	9
Fit function	Fit function	180 × 360	9
T-fit Function	T-fit function	180 × 360	9
4x6 fit function	Fit function	180 × 360	9
Transcom fit function	Fit function	180 ×360	9

Table 2: Overview with the settings of the TM5 runs.

The results section for each location will show the sensitivity using different settings as input for TM5 (2. The SiBCASA run will be used as reference rather than observations of the stations themselves as the SiBCASA NEE fluxes have not been subjected to the ensemble data assimilation of CarbonTracker just like the other simplified NEE fluxes.

For the fit function and T-fit function the SiBCASA signal is parameterized as elaborated in Section 3. Note that the P-fit function is not used as input for a TM5 run, because the flux analysis results in Section 4 show large resemblance and therefore will not show noteworthy results for this atmospheric sensitivity analysis.

The 4x6 fit function and Transcom fit function will indicate the efficiency of saving computational costs versus simplifying the NEE fluxes. Herewith it would be ideal to use the 4x6 T-fit function, ext3-det-T-fit function and the Transcom T-fit function for the atmospheric sensitivity analysis, but is not possible due to time constrains. Do note Section 3 showed only a small performance difference between the fit and T-fit function and therefore the fit function can be used as indication of the atmospheric sensitivity of the T-fit function for the simplified NEE fluxes.

5.3   Results and discussion

This section will first address the processes which influence the CO_2 signal of the stations and which are not taken into account compared to observations. Herewith we will explain (1) background concentration, (2) the seasonal cycle and (3) growth rate. This will also include a comparison of the processes between the stations. Hereafter the differences in CO_2 for each station will be compared and the sensitivity of TM5 with the different inputs (see section 5.2) will be determined.
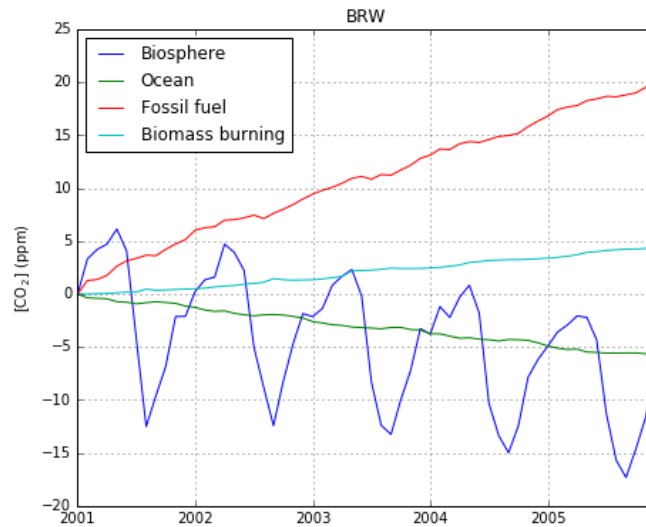
Figure 33: Accumulation of the CO_2 components into the atmosphere over time for Mauna Loa from 2001 until 2006.

The biosphere, ocean, fossil fuel and biomass burning fluxes are the four components, which cause atmospheric CO_2 concentration changes over time (Keeling et al., 1976) (Figure 33). Here fossil fuel and biomass burning are CO_2 sources and the biosphere and ocean are CO_2 sinks. The seasonal cycle of Mauna Loa is mainly influenced by the biosphere due to the seasonal cycle of GPP, R and therefore NEE (33).The amplitude of the seasonal cycle at Mauna Loa is approximately 15 ppm from 2001 until 2006 and shows the largest IAV in terms of amplitude and phase of the four components, due to climate variations.
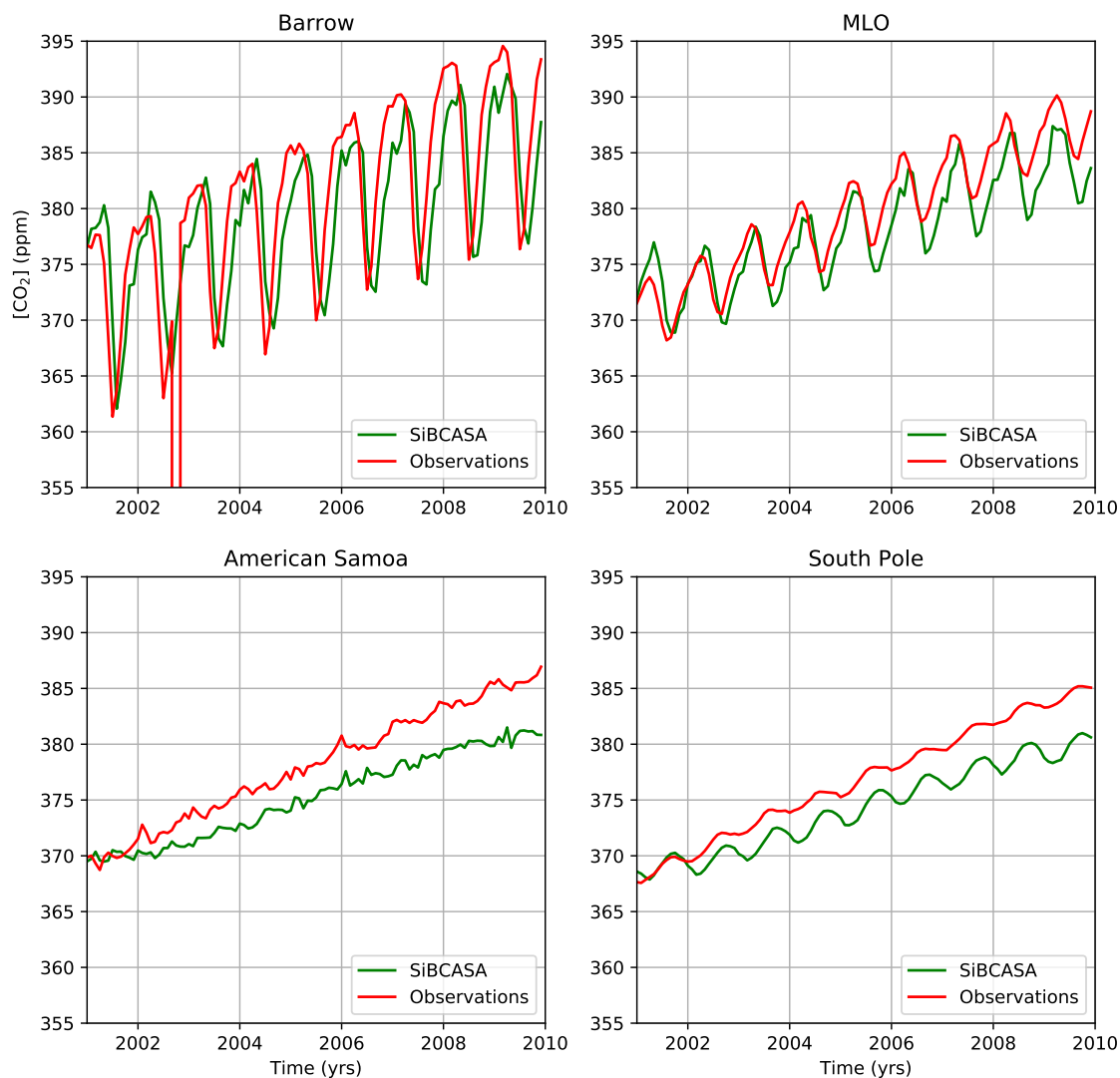
Figure 34: CO_2 concentration at four stations for the time period of 2001 - 2010.

The long-term growth rate of the SiBCASA run and the observations show large discrepancy (Figure 34), which cause the difference of about 0.5 ppm for the stations and globally (not shown). The average growth of the four stations combined is 0.63 ppm yr^{-1} lower for the SiBCASA run compared to observations which is line with global difference. The global average observed growth rate is 1.97 ppm yr^{-1} and for the SiBCASA run only 1.36 ppm yr^{-1} (Figure 34).

The global terrestrial CO_2 sink of SiBCASA is reasonably high compared to the findings of Le Quéré et al. (2016) and could explain the lower atmospheric CO_2 growth rate of SiBCASA compared to observations. During the 2001-2010 period, SiBCASA has a global terrestrial CO_2 sink of -2.76 PgC yr^{-1}, which is at the maximum range of Le Quéré et al. (2016) with a global terrestrial carbon sink of -1.9 $\pm$ 0.9 PgC yr^{-1} between 2006 and 2015.

As expected, the background concentration shows a development of the North-South CO_2 latitude gradient (Figure 34), (Conway and Tans, 1999). Barrow shows correspondingly the highest background concentration and the South Pole the lowest as anthropogenic emissions are substantially higher in the Northern hemisphere and interhemispheric mixing is slow. The difference between the two stations over the time period is 5.2 ppm on average and is substantially higher than the 4.0 ppm difference of Conway et al. (1994); Fraser et al. (1983), which is concluded from the period of 1981 -1992. This trend can be explained by the relative higher increase in anthropogenic emissions on the Northern hemisphere for the following years (Boden et al., 2013). Current observations also show a 5.2 ppm Barrow-South Pole latitudinal gradient and the long-term trend of the gradient shows minimal variability in both cases (not shown). Therefore the interhemispheric mixing time of TM5 seems to be comparable with observations.

The amplitude of the observations correspond with SiBCASA, whereby the Northern hemisphere located stations show large seasonality in the CO_2 mixing ratio with a gradual decay towards the equator as explained by the biosphere component in the total CO_2 time series (Fraser et al., 1983; Matsueda et al., 2002; Denning et al., 1995). Together with the strongly seasonal vegetation, meridional transport, and large scale convective mixing attribute to the large seasonal cycle across the Northern hemisphere (Denning et al., 1995).

A relative small seasonal cycle can be observed in the Southern hemisphere compared to the Northern hemisphere as indicated at the South Pole. The amplitude of the CO_2 mixing ratio is lower as land and thus the the terrestrial biosphere is smaller. The South Pole is still able to represent this seasonal cycle due to interhemispheric mixing as mentioned in previous paragraph.

At American Samoa, the observations are not dominated by the large seasonal cycle of the biosphere fluxes. The location of the South Pacific Convergence Zone (SPCZ),a monsoon through in the ITCZ, prevents easy mixing between the hemispheres(Murayama et al., 2003; Matsueda et al., 2002). However, in the Australian summer the ITCZ is poorly defined as it shifts northwards, which causes penetration of the Northern hemispherical air. This air mixes with the Southern hemispherical air which has an opposite seasonal cycle and therefore neutralizes the seasonal cycle (Waterman et al., 1989). Now we gained a better understanding of the processes, which account for spatial and temporal patterns of CO_2 mole fractions, we now increase our understanding of the atmospheric sensitivity of the simplified NEE fluxes.
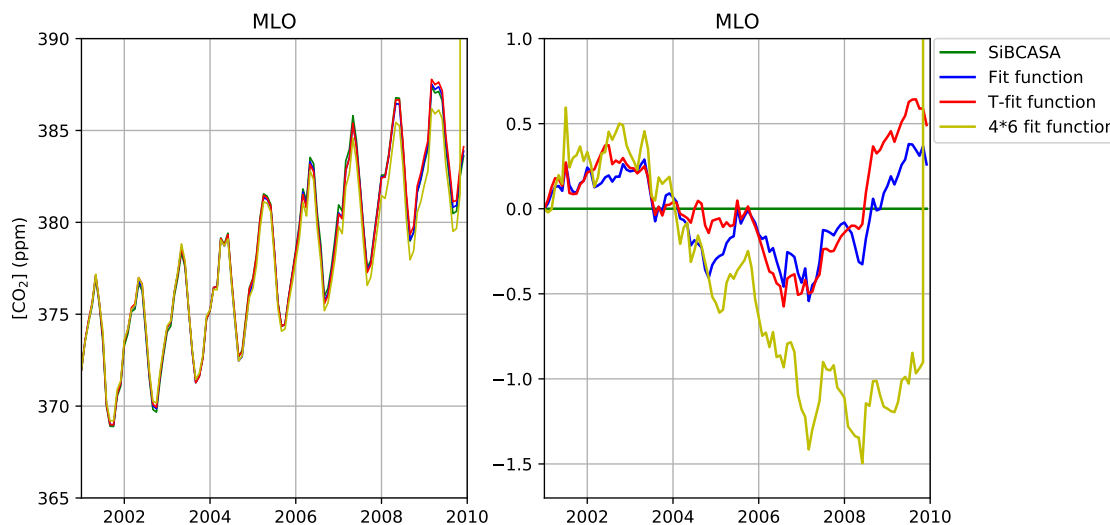
Figure 35: Left panel: time series of the atmospheric CO_2 mole fractions with SiBCASA and its simplified NEE flux methods as input for TM5 for Mauna Loa from 2001 until 2010. Right panel: time series of the simplified NEE flux methods with SiBCASA as reference for Mauna Loa for 2001 until 2010.

The three simplified NEE flux methods show great similarity with SiBCASA in terms of global atmospheric CO_2 mole fraction (Figure 35). The fit and T-fit function both show approximately the same atmospheric sensitivity of a CO_2 mole fraction difference of 0.5 ppm compared to SiBCASA (Figure 35).

In 2004 the T-fit function outperforms the fit function, while in 2009 the opposite is depicted. Checking the RMSE's, we have seen that the T-fit function outperforms the fit function with an RMSE of 1.812 μgC m^{-2} s^{-1} compared to 2.163 μgC m^{-2} s^{-1} respectively, in terms of NEE fluxes. The T-fit seems to have a larger structural bias in fitting the NEE fluxes of SiBCASA compared to fit function. This means that the least square regression method when calculating the T-fit function has a tendency to over or underestimate the NEE fluxes over a time. This results in an equal performance compared to the fit function even though the RMSE is lower. This shows the disadvantage of using the RMSE as performance metric.

The 4x6 fit function shows lowest performance as expected, due the largest RMSE and average monthly NEE difference as concluded in section 4. Over time the error of the 4x6 fit function increases due to the cumulative error propagation. The computational decrease of a factor 24 spatially has led to approximately a 1.5 ppm difference of CO_2 mole fraction compared to SiBCASA. This is quite a large difference as in CarbonTracker these errors are extrapolated in its calculations (Koren and Peters, 2018).

Globally and for the other three observatory stations, we obtain similar results with a 0.5 ppm difference for the T-fit and fit function, and a within 1.5 ppm difference for the 4x6 fit function (See Appendix 8.2). The largest differences are expressed in the monthly erratic behaviour in terms of atmospheric sensitivity. Barrow is highly erratic from month to month while the south pole shows low intermonthly variability at these small CO_2 mole fraction differences. At the south pole, the effect of the difference in biosphere fluxes between SiBCASA and the other simplified NEE fluxes is smoothened as atmospheric mixing has taken place before it reached the observatory station. On the other hand, Barrow is attached to the mainland of Alaska, and surrounded by boreal forest in the south. Therefore the changes between SiBCASA and the simplified NEE fluxes is more pronounced in the amplitude of the atmospheric sensitivity at Barrow when the wind direction is favorable .

The ext3-det-T-fit function, is most likely not applicable for decreasing the computational costs. The global terrestrial CO_2 sink of the ext3-det-T-fit function compared to the 4x6 T-fit function is substantially larger with a -2.95 PgC yr^{-1} and -2.87 PgC yr^{-1} respectively. Note that SiBCASA and the T/P-fit function show a global terrestrial CO_2 sink of -2.76 PgC yr^{-1}, all calculated over a time period from 2001-2010. This indicates that the CO_2 mole fraction difference will be larger than 1.5 ppm and therefore the performance decrease outweighs the computational savings of a factor 3 from a temporal perspective.
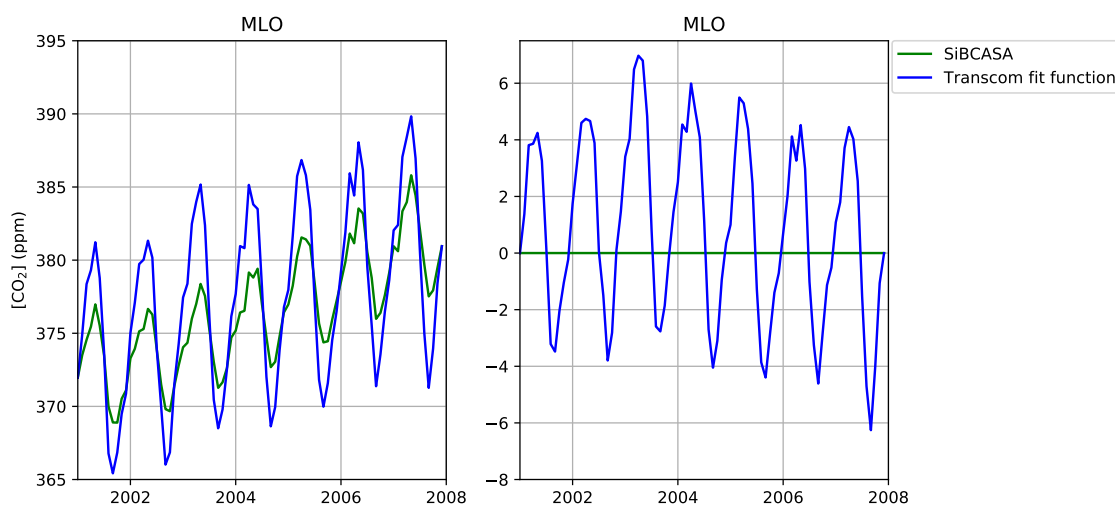

Figure 36: Left panel: time series of the atmospheric CO_2 mole fractions with SiBCASA and the Transcom fit function as input for TM5 for Mauna Loa from 2001 until 2010. Right panel: time series of the Transcom fit function with SiBCASA as reference for Mauna Loa for 2001 until 2010.

The atmospheric sensitivity analysis of the Transcom fit shows that the large potential computational savings lead to a too large performance decrease with a maximum of 7 ppm and an average of 5 ppm. This was also already concluded in section 4 and therefore the total sum of the optimal selected gridbox can not resemble the relative sum of one Transcom. The method is able to represent the total NEE fluxes for the nine years time period, however is not able to represent the global seasonal cycle. On a global scale, the Transcom fit shows a maximum atmospheric sensitivity of 4.8 ppm and an average of 3.2 ppm (Appendix 8.2).

5.4 Conclusion

The seasonal cycle of the atmospheric CO_2 mole fraction is due to the seasonal cycle of biosphere in especially the Northern hemisphere and well represented in TM5. The latitudinal gradient is also well presented as it corresponds to Conway and Tans (1999). SiBCASA shows a structural underestimation of the CO_2 mole fraction of approximately 0.5 ppm. The fit function and T-fit function show an atmospheric sensitivity of 0.5 ppm and therefore comparable with SiBCASA, which indicate that both can be used to determine seasonal, interannual and long-term trends. Further simplified NEE fluxes such as the 4x6 fit function, lead to an atmospheric sensitivity of at least 1.5 ppm compared to SiBCASA, which is not suitable for inversions.

6   Discussion and recommendation

To save computational costs, with around a maximum of 0.5 ppm difference compared to SiBCASA, we can apply two strategies: (1) increase the performance of the T/P-fit function and (2) decrease the computational costs less drastically.

6.1   Performance increase of the T/P-fit function

The seasonal cycle of NEE is captured well for the extratropics and for a large extent in the tropics. The large IAV in the NEE signal in the tropics can only partly be resolved by introducing the climate sensitivity parameter γ, which makes use of linear regression between NEE and T or P anomalies. Rödenbeck et al. (2018) was able to link natural processes to γ and robustly explain the effect of ENSO on the tropics and to a smaller extent for the tropics. However, in our case, we were not able to biophysically interpret γ_{NEE-T} and γ_{NEE-P} as it seems that the drought response of SiBCASA in the tropics is lacking (Section 3.3.3). Laan-Luijkx et al. (2015) already showed that SiBCASA has difficulty capturing the drought response in SiBCASA. In their case both the GPP and R lowered, causing a non-significant effect on the NEE. A drought response lack was also captured by Van der Velde (2015) when studying the effects of droughts on water-use efficiency.

One of the limiting factors of GPP in SIBCASA is the Rubisco enzymatic conversion (ω_c), which is a catalyst for photosynthesis. The limited rate of ω_c (mol m$^{-2}s^{-1}$) is calculated with:

$$\omega_c = f_c(V_m, c_i, O_2) \tag{30}$$

Here f_c is the full function as described in (Sellers et al., 1996b), V_m is the catalytic capacity of the Rubisco enzym (mol m$^{-2}s^{-1}$), c_i is the partial pressure of CO_2 in the the leaf (Pa) and ω_c is also dependent on O_2, which is the partial pressure of O_2 in the leaf (Pa). And V$_m$ is given by:

$$V_m = V_{max}f_T(T_c)f_w(W_i) \tag{31}$$

Here V_{max} is the maximum catalytic capacity of the Rubisco enzym (mol m^2 s^{-}1), T_c the temperature at the top of the canopy (K), W_i is the soil moisture wetness and f_T and f_w are the full functions as described by Sellers et al. (1996a). Therefore a limiting factor of GPP, ω_c, is dependent on W_i/ van Schaik et al. (2018) shows that the implementation of the surface hydrology model PCR-GLOBWB with the precipitation product MSWEP improves the surface hydrology of the Amazon Basin and supply SiBCASA with the soil moisture fraction. PCR-GLOBWB does has an improved infiltration scheme and accounts for run-off unlike SiBCASA, which has a high soil infiltration capacity, which leads to an under representation of the runoff precipitation ratio. This has led to an improvement of the SiBCASA drought response in which van Schaik et al. (2018) was able to distinguish the drought responses in different parts of the Amazon and could therefore increase the interpretability of especially γ_{NEE-P}.

The unknown deep rooting depth of the Amazon vegetation is a second uncertainty of the drought response of SiBCASA in the Amazon (van Schaik et al., 2018) The rooting depth is based on observations at one location in the Amazon, which may not be suitable for the whole Amazon Basin (Maeght et al., 2013; Laan-Luijkx et al., 2015). Running SiBCASA with an implementation of PCR-GLOBWB with MSWEP increases the drought response and should improve the performance and interpretability of γ_{NEE-P} and also of γ_{NEE-T} as they are closely linked (Berg et al., 2015). This also accounts for the implementation of a more advanced deep rooting depth when more observations are available across the Amazon Basin. Taller and older Amazon forests for instance are found to have a deeper rooting depth, which makes them more resilient to precipitation variability as the forest has access to deep soil moisture (Giardina et al., 2018). Giardina et al. (2018) therefore suggest that age and height are another control factor of GPP in response to precipitation anomalies. Currently only the Normalized Difference Vegetation Index (NDVI) is taken into account in SiBCASA, which is derived from satellite data. This gives us an estimate about the green leaf biomass and is used as input to estimate the GPP with the use of a photosynthesis model (Schaefer et al., 2008; Ichii et al., 2001).

The findings of Giardina et al. (2018) are based on new remote sensing observations of sun-induced fluorescence (SIF), which is a proxy of GPP. A satellite instrument at the top of the atmosphere directly detects the wavelength of re-emitted light of chloroplasts during photosynthesis. SIF is therefore a direct observation of GPP and improves the signal especially in the tropics as noise is reduced compared to other retrieval products (Koren et al., 2018). SIF increases the spatio-temporal response of GPP and is able to detect a different response between the eastern and western part of the Amazon. A comparison with Yang et al. (2018); Jiménez-Muñoz et al. (2016); Gloor et al. (2018) showed that the largest anomalies of precipitation, temperature, terrestrial water storage and evapotranspiration are in line with SIF anomalies. This would bypass the need of NDVI satellite data to estimate and constrain the GPP fluxes, which could lead to an increased performance of the γ_{NEE-P} and γ_{NEE-T} and their biophysical interpretation.

All recommendations could thus increase the T/P-fit function's performance to capture the IAV in the NEE time series. However, the results also suggest that adjusting the T/P-fit function could improve its performance. The phase and amplitude of the T/P-fit function should for example be dependent on climate anomalies as well, which should increase the performance of the T/P-fit function in the tropical and semi-arid climate zones.

To increase the performance and interpretability of γ_{NEE-T} and γ_{NEE-P}, we could make use of a larger data set. Herewith we obtain a larger amount of NEE and T or P residuals to estimate the monthly climate sensitivity parameter. For this estimation we have currently used a linear regression dependent on only 9 years of residuals to estimate γ_{NEE-T} and γ_{NEE-P}, in which a larger data set could decrease its uncertainty as well (Raudys and Jain, 1991). Moreover, a climate sensitivity parameter that would interpret both temperature and precipitation would explain a larger portion of the IAV in the NEE signal, however would lead to a non-physical interpretation. In the case of $\gamma_{NEE-T-P}$, we would first use linear regression on the NEE IAV and T IAV to estimate the γ_{NEE-T}. Hereafter we would use a second linear regression with the residuals of the T-fit function compared to SiBCASA and apply it to the linear regression with P IAV, which is not possible as we would take the residuals of the IAV. Another possible subsequent step can be running an T/P-fit function inversion with CarbonTracker, which could optimize the climate sensitivity parameters γ_{NEE-T} and γ_{NEE-P}.

6.2 Performance increase of the computational saving methods

For the second strategy we recommend to increase the spatial resolution to a 2°x3° (lat/lon) grid. The 4°x6° fit function showed differences up to 1.5 ppm, in which one NEE time series represents an approximate 450 km by 650 km area. When increasing the resolution, the climate zone overlap in a 2x3 local region will significantly lower according to the climate zone classification of Køppen-Geiger (Kottek et al., 2006). The higher resolution should lead to a better performance as one gridbox will represent a 2x3 local region better. Do note that it will possible resemble a better atmospheric resemblance of the NEE fluxes compared to a saving costs temporally.

When decreasing the temporal costs by a factor three, the global RMSE increases by a factor 1.7 compared to the T-fit function. While decreasing the spatial resolution to 6°x4° small regions, we obtain a global RMSE increase by a factor 2.8. The downside of the temporal cost savings is the large structural bias as we parameterize the NEE time series in the three year. If for example droughts occur in these years, an overestimation of IAV in the NEE signal is parameterized, which leaves us with a over or under estimation of especially the seasonal cycle. Next to the seasonal cycle, the IAV can only be estimated from three data points as the we estimate monthly γ values. Herewith we observe a large structural bias as the 9 year monthly average NEE fluxes do not correspond to those of SiBCASA. To make this method more robust, a larger sample of the original dataset is needed to parameterize a T-fit function. This should lead to a lower RMSE and more importantly a better representation of the monthly NEE fluxes, which seem a better indication for the atmospheric sensitivity analysis.

7 Conclusion

Anthropogenic carbon emissions accelerate climate change and impact the global carbon cycle. About 25% of CO_2 is currently taken up by the terrestrial biosphere (Quéré et al., 2015). The carbon fluxes are subjected to seasonal, interannual and long-term trends and substantial variability, which are for a large share a response to climate variations and trends. The 2010 drought in the Amazon for example, turned the Amazon from carbon sink into a carbon source and herewith increased the atmospheric growth rate in 2010. The effect of GPP and R on in particular droughts contributes to largest land sink uncertainties.

To constrain the climate carbon cycle responses, it is essential to employ observational records. However, commonly used Eddy Covariance measurements can only be used for a very local scale and large scale observations are scarce (Baldocchi et al., 2001). Therefore biogeochemical models are set up to quantify the large scale-carbon fluxes such as SiBCASA (Schaefer et al., 2008). To decrease the uncertainty of the model, inverse modelling can applied which optimizes the carbon fluxes with the use of CO_2 mole fraction observations (Peters et al., 2005). The data assimilation system CarbonTracker, can optimize these fluxes, however is computational costly as weekly calculations have to be performed.

To quantify carbon fluxes on a longer time scale to capture the interannual and long-term trend, (Cox et al., 2013) identified an emergent linear relationship between the sensitivity of the annual atmospheric CO_2 growth rate and sensitivity of tropical land carbon storage to warming. (Rödenbeck et al., 2018) based his statistical model on the findings of (Cox et al., 2013) to obtain a climate sensitivity parameter (γ_{NEE-T}, with the use of linear regression of the NEE and T anomalies.

Together with the least square regression model of (Thoning et al., 1989), we build are own statistical model, which captures the seasonal and NEE IAV with SiBCASA as input. We based our statistical model on both γ_{NEE-T} and γ_{NEE-P} (precipitation) to capture the response of droughts in especially the Amazon, which we called the T or P-fit function. Both fit functions show similar results performance wise as the seasonal NEE cycle is captured well globally.In regions with highly pronounced IAV in the NEE time series, such as the tropics, the performance of the T/P fit functions decreases with a factor 1.5 approximately in terms of RMSE. The biophysical interpretation of γ_{NEE-T} does not correspond with the findings of (Rödenbeck et al., 2018) as SiBCASA does not show a good drought response due to a lack of detail in the surface hydrology. Moreover, an inversion run would increase the representation of the IAV in the NEE time series and biophysical interpretation of the γ's would increase. The fit and T-fit function both showed an atmospheric sensitivity of 0.5 ppm when using these simplified NEE fluxes as input for the atmospheric transport model TM5. This atmospheric sensitivity is suitable as potential prior for CarbonTracker as SiBCASA also shows an atmospheric sensitivity of 0.5 ppm compared to observations over a 9 year time period.

To further save computational costs we have (1) used less monthly NEE data to parameterize the T-fit function with temporal savings of a factor 3. The global RMSE square error was reasonable low, however the structural bias was high when IAV in the parameterization period was pronounced. This will very likely lead to an atmospheric sensitivity of >1.5 ppm compared to SiBCASA over a 9 year time period. Therefore a larger portion of the monthly NEE data need to be taken into account the parameterize the SiBCASA NEE time series to decrease the structural bias and decrease its atmospheric sensitivity.

We also (2) decreased the spatial resolution of the T-fit function, in which one gridbox was selected to represent a 4° x 6° (latitude x longitude) small region. This method showed a high RMSE, but a relatively small global monthly NEE difference compared to the original T-fit function, due to the absence of a structural bias. The atmospheric sensitivity of the 4x6 fit function is within a range of 1.5 ppm over a 9 year time period, and is therefore not suitable as prior for CarbonTracker. To decrease the atmospheric sensitivity further research with the use of higher resolution is needed. We (3) finally checked the spatio-temporal resemblance of one gridbox for a whole Transcom region of the T-fit function, which showed a high RMSE as the NEE fluxes were too strongly simplified lead to an global atmospheric sensitivity of 3.2 ppm approximately.

The T/P fit function thus show possibilities to decrease the computational costs of Carbon-Tracker and resolve seasonal, interannual and long-term NEE patterns. Current research to new climate and carbon flux proxies could further enhance our understanding of the impact climate variability to the terrestrial carbon cycle.
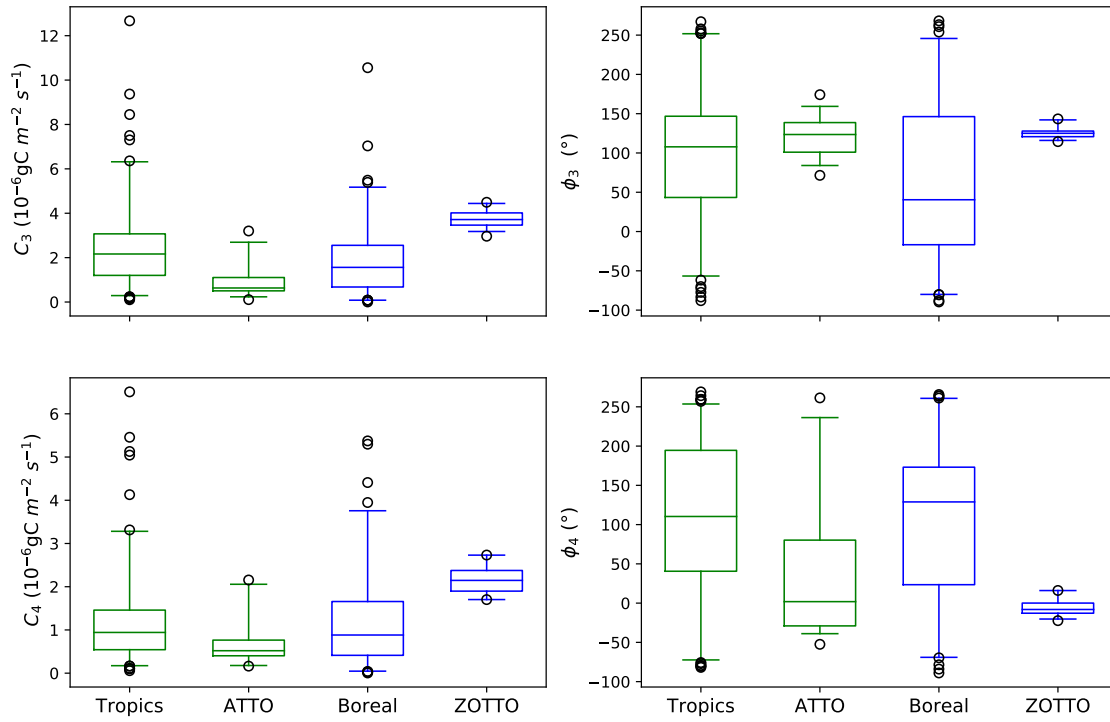
8 Appendix

8.1 Appendix A

Figure 37: Top panels: Boxplot of the amplitude (C_3) and phase(ϕ_3) of the T-fit function. Bottom panels: Boxplot of the amplitude (C_4) and phase(ϕ_4) of the T-fit function. The tropics resembles the tropical Transcom region, ATTO resemble the ATTO small region, Boreal resembles the Boreal Transcom region and ZOTTO resembles the ZOTTO small region.
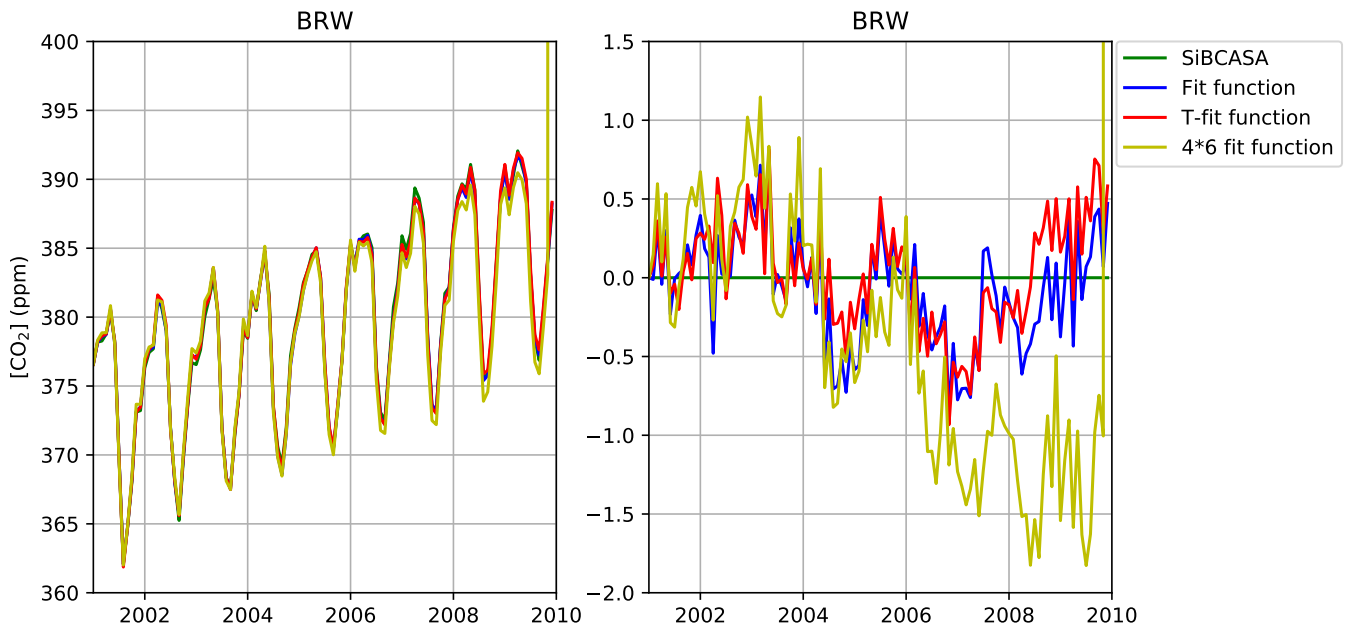
8.2 Appendix B

Figure 38: Left panel: time series of CO_2 mole fraction of the simplified NEE fluxes. Right panel: time series of atmospheric sensitivity of simplified NEE fluxes compared to SiBCASA. Both are time series of Barrow observatory station, United States .
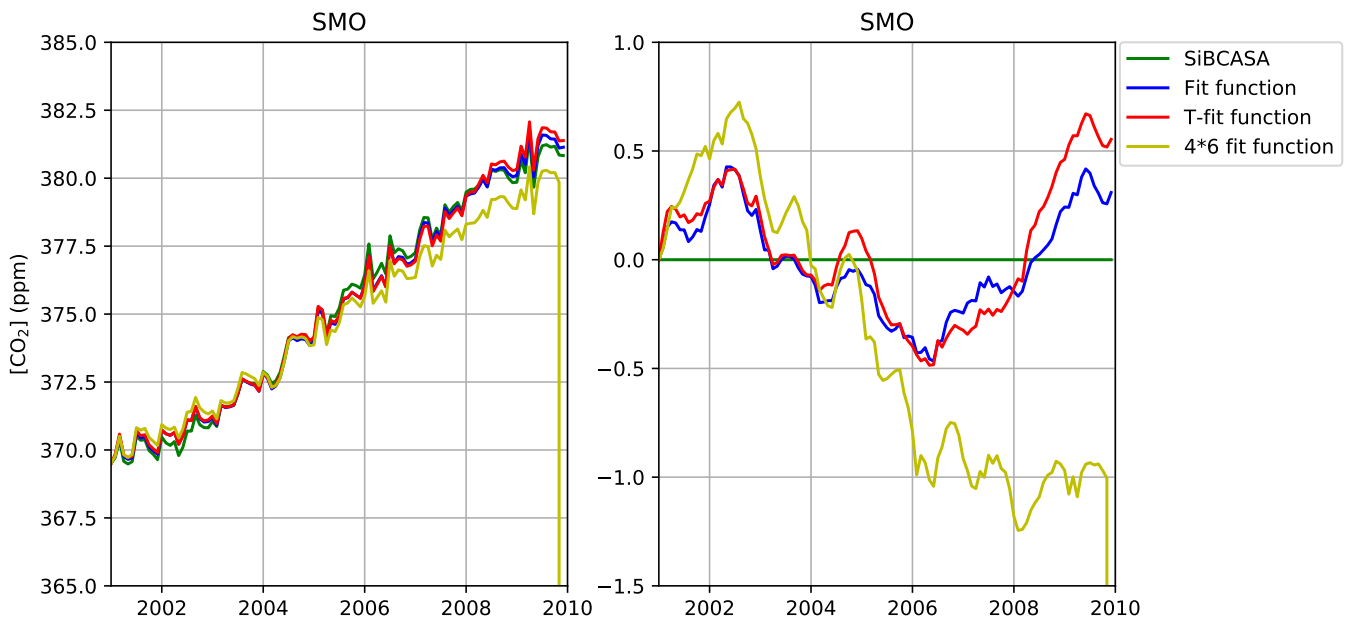

Figure 39: Left panel: time series of CO_2 mole fraction of the simplified NEE fluxes. Right panel: time series of atmospheric sensitivity of simplified NEE fluxes compared to SiBCASA. Both are time series of American Samoa (SMO)
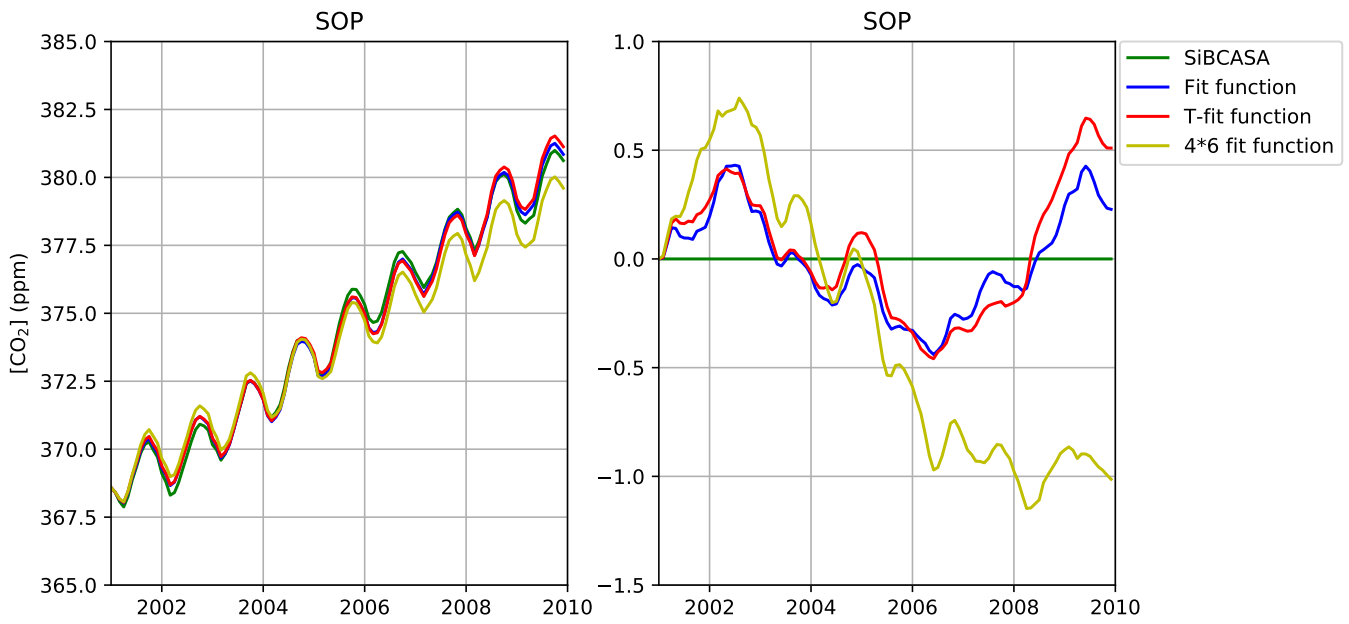
Figure 40: Left panel: time series of CO_2 mole fraction of the simplified NEE fluxes. Right panel: time series of atmospheric sensitivity of simplified NEE fluxes compared to SiBCASA. Both are time series of the South Pole (SOP)

Figure 41: Left panel: time series of CO_2 mole fraction of the simplified NEE fluxes. Right panel: time series of atmospheric sensitivity of simplified NEE fluxes compared to SiBCASA. Both are global time series.)

Figure 42: Left panel: time series of CO_2 mole fraction of the 4x6 simplified NEE fluxes. Right panel: time series of atmospheric sensitivity of simplified NEE fluxes compared to SiBCASA. Both are time series of Mauna Loa (MLO).
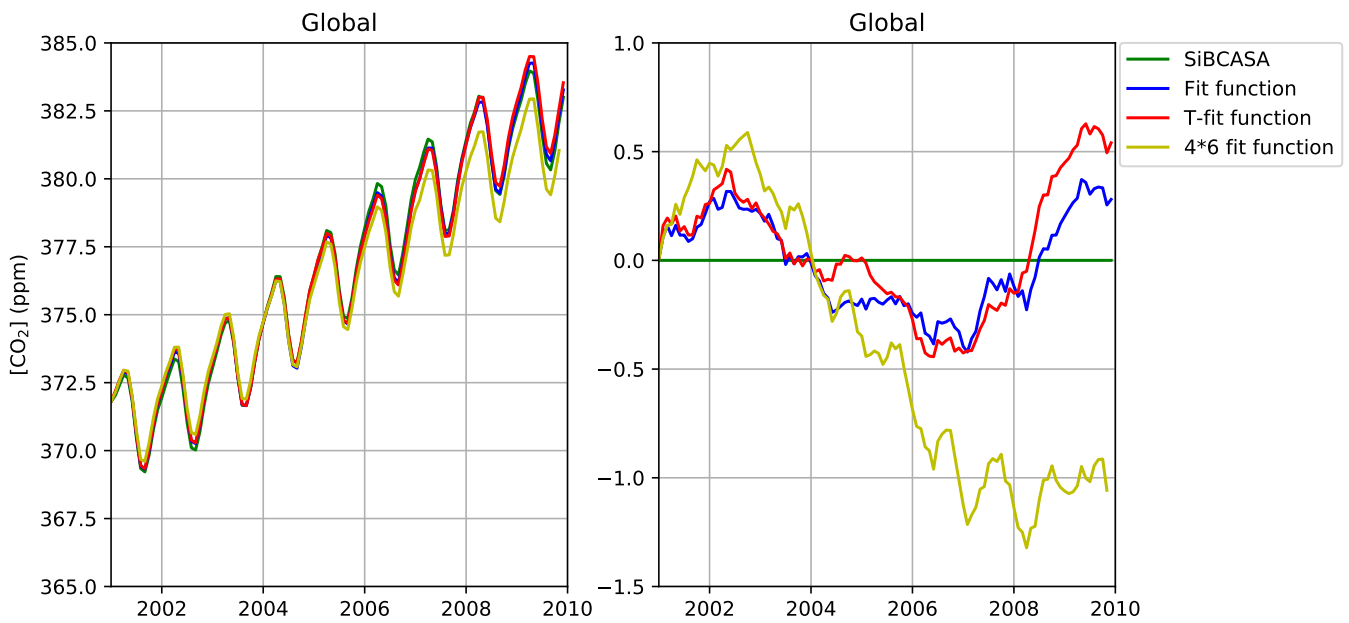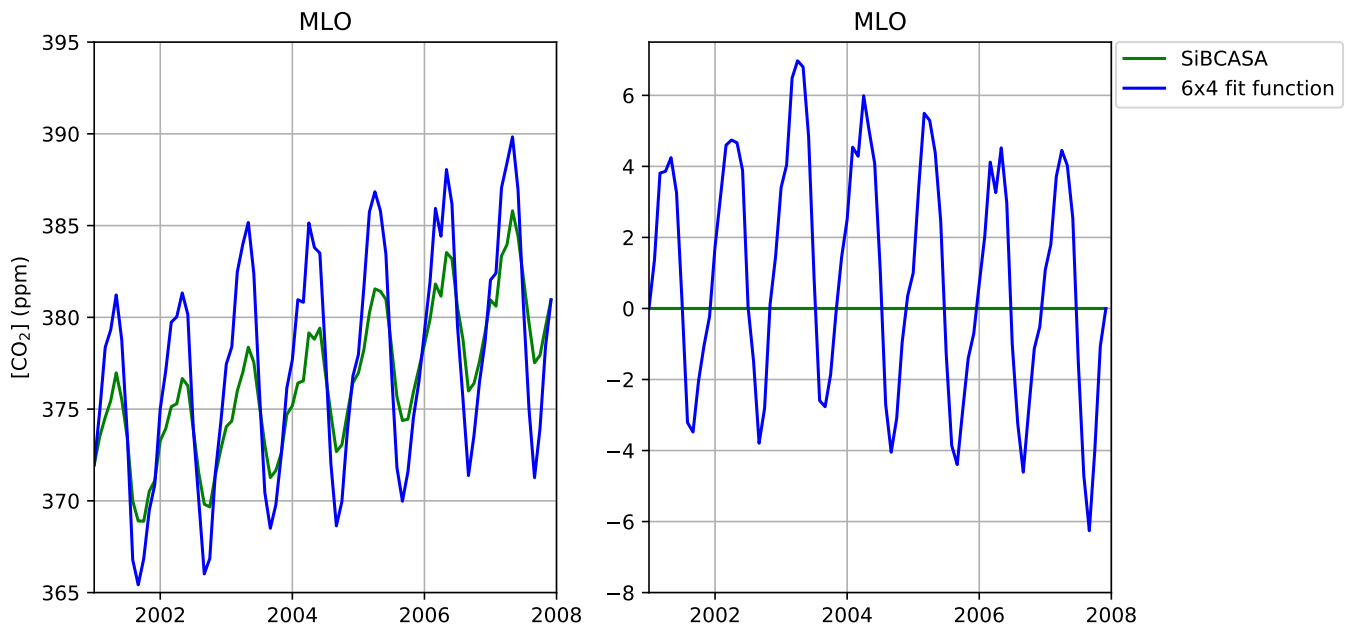
References

Adam, O., Bischoff, T., and Schneider, T. (2016). Seasonal and interannual variations of the energy flux equator and itcz. part i: Zonally averaged itcz position. *Journal of Climate*, 29(9):3219–3230.

Andreae, M., Acevedo, O., Araùjo, A., Artaxo, P., Barbosa, C., Barbosa, H., Brito, J., Carbone, S., Chi, X., Cintra, B., et al. (2015). The amazon tall tower observatory (atto): overview of pilot measurements on ecosystem ecology, meteorology, trace gases, and aerosols. *Atmospheric Chemistry and Physics*, 15(18):10723–10776.

Arpe, K., Leroy, S., Lahijani, H., and Khan, V. (2012). Impact of the european russia drought in 2010 on the caspian sea level. *Hydrology and earth system science*, 16:19–27.

Australian Bureau of Statistics (2006). Australian deserts, climatic aspects of australia's deserts.

Australian Government - Bureau of Meteorology (2016). Average annual, seasonal and monthly rainfall.

Australian Government - Bureau of Meteorology (2019). The recent climate – news, reports and summaries. Accessed: 2019-01-10.

Baldocchi, D., Falge, E., Gu, L., Olson, R., Hollinger, D., Running, S., Anthoni, P., Bernhofer, C., Davis, K., Evans, R., et al. (2001). Fluxnet: A new tool to study the temporal and spatial variability of ecosystem-scale carbon dioxide, water vapor, and energy flux densities. *Bulletin of the American Meteorological Society*, 82(11):2415–2434.

Ballantyne, A., Alden, C., Miller, J., Tans, P., and White, J. (2012). Increase in observed net carbon dioxide uptake by land and oceans during the past 50 years. *Nature*, 488(7409):70.

Basu, S., Miller, J. B., and Lehman, S. (2016). Separation of biospheric and fossil fuel fluxes of co2 by atmospheric inversion of co2 and 14co2 measurements: Observation system simulations. *Atmospheric Chemistry and Physics*, 16(9).

Berg, A., Lintner, B. R., Findell, K., Seneviratne, S. I., van den Hurk, B., Ducharne, A., Chéruy, F., Hagemann, S., Lawrence, D. M., Malyshev, S., et al. (2015). Interannual coupling between summertime surface temperature and precipitation over land: Processes and implications for climate change. *Journal of Climate*, 28(3):1308–1328.

Boden, T. A., Andres, R. J., and Marland, G. (2013). Global, regional, and national fossil-fuel co2 emissions (1751-2010)(v. 2013). Technical report, Carbon Dioxide Information Analysis Center (CDIAC), Oak Ridge National . . . .

Bolin, B. and Keeling, C. (1963). Large-scale atmospheric mixing as deduced from the seasonal and meridional variations of carbon dioxide. *Journal of Geophysical Research*, 68(13):3899–3920.

Bonan, G. (2015). *Ecological climatology: concepts and applications*. Cambridge University Press.

Borodina, A., Fischer, E. M., and Knutti, R. (2017). Emergent constraints in climate projections: a case study of changes in high-latitude temperature variability. *Journal of Climate*, 30(10):3655–3670.

Brienen, R. J., Phillips, O. L., Feldpausch, T. R., Gloor, E., Baker, T. R., Lloyd, J., Lopez-Gonzalez, G., Monteagudo-Mendoza, A., Malhi, Y., Lewis, S. L., et al. (2015). Long-term decline of the amazon carbon sink. *Nature*, 519(7543):344.

CarbonTracker Europe (2017). Biosphere module.

CarbonTracker Team (2017). Carbontracker documentation ct2016 release.

Chang, P. and Zebiak, S. (2015). {TROPICAL} {METEOROLOGY} & amp; {CLIMATE} | el niño and the southern oscillation: Theory. In North, G. R., Pyle, J., and Zhang, F., editors, *Encyclopedia of Atmospheric Sciences (Second Edition)*, pages 97 – 101. Academic Press, Oxford, second edition edition.

Chiang, J. C. and Sobel, A. H. (2002). Tropical tropospheric temperature variations caused by enso and their influence on the remote tropical climate. *Journal of climate*, 15(18):2616–2631.

Cleverly, J., Boulain, N., Villalobos-Vega, R., Grant, N., Faux, R., Wood, C., Cook, P. G., Yu, Q., Leigh, A., and Eamus, D. (2013). Dynamics of component carbon fluxes in a semi-arid acacia woodland, central australia. *Journal of Geophysical Research: Biogeosciences*, 118(3):1168–1185.

Conway, T. J. and Tans, P. P. (1999). Development of the co2 latitude gradient in recent decades. *Global Biogeochemical Cycles*, 13(4):821–826.

Conway, T. J., Tans, P. P., Waterman, L. S., Thoning, K. W., Kitzis, D. R., Masarie, K. A., and Zhang, N. (1994). Evidence for interannual variability of the carbon cycle from the national oceanic and atmospheric administration/climate monitoring and diagnostics laboratory global air sampling network. *Journal of Geophysical Research: Atmospheres*, 99(D11):22831–22855.

Cox, P. M., Pearson, D., Booth, B. B., Friedlingstein, P., Huntingford, C., Jones, C. D., and Luke, C. M. (2013). Sensitivity of tropical carbon to climate change constrained by carbon dioxide variability. *Nature*, 494(7437):341.

Davey, M., Brookshaw, A., and Ineson, S. (2014). The probability of the impact of enso on precipitation and near-surface temperature. *Climate Risk Management*, 1:5–24.

DeFries, R. and Townshend, J. (1994). Ndvi-derived land cover classifications at a global scale. *International Journal of Remote Sensing*, 15(17):3567–3586.

Denning, A. S., Fung, I. Y., and Randall, D. (1995). Latitudinal gradient of atmospheric co2 due to seasonal exchange with land biota. *Nature*, 376(6537):240.

Doughty, C. E., Metcalfe, D., Girardin, C., Amézquita, F. F., Cabrera, D. G., Huasco, W. H., Silva-Espejo, J., Araujo-Murakami, A., Da Costa, M., Rocha, W., et al. (2015). Drought impact on forest carbon dynamics and fluxes in amazonia. *Nature*, 519(7541):78.

Frank, D., Reichstein, M., Bahn, M., Thonicke, K., Frank, D., Mahecha, M. D., Smith, P., Velde, M., Vicca, S., Babst, F., et al. (2015). Effects of climate extremes on the terrestrial carbon cycle: concepts, processes and potential future impacts. *Global Change Biology*, 21(8):2861–2880.

Fraser, P., Pearman, G., and Hyson, P. (1983). The global distribution of atmospheric carbon dioxide: 2. a review of provisional background observations, 1978–1980. *Journal of Geophysical Research: Oceans*, 88(C6):3591–3598.

Friedlingstein, P., Cox, P., Betts, R., Bopp, L., von Bloh, W., Brovkin, V., Cadule, P., Doney, S., Eby, M., Fung, I., et al. (2006). Climate–carbon cycle feedback analysis: results from the c4mip model intercomparison. *Journal of climate*, 19(14):3337–3353.

Fu, R. and Li, W. (2004). The influence of the land surface on the transition from dry to wet season in amazonia. *Theoretical and applied climatology*, 78(1-3):97–110.

Gatti, L., Gloor, M., Miller, J., Doughty, C., Malhi, Y., Domingues, L., Basso, L., Martinewski, A., Correia, C., Borges, V., et al. (2014). Drought sensitivity of amazonian carbon balance revealed by atmospheric measurements. *Nature*, 506(7486):76.

Giardina, F., Konings, A. G., Kennedy, D., Alemohammad, S. H., Oliveira, R. S., Uriarte, M., and Gentine, P. (2018). Tall amazonian forests are less sensitive to precipitation variability. *Nature Geoscience*, page 1.

Gloor, E., Wilson, C., Chipperfield, M. P., Chevallier, F., Buermann, W., Boesch, H., Parker, R., Somkuti, P., Gatti, L. V., Correia, C., et al. (2018). Tropical land carbon cycle responses to 2015/16 el niño as recorded by atmospheric greenhouse gas and remote sensing data. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 373(1760):20170302.

Gruber, A. (1972). Fluctuations in the position of the itcz in the atlantic and pacific oceans. *Journal of the Atmospheric Sciences*, 29(1):193–197.

Gurney, K. R., Law, R. M., Denning, A. S., Rayner, P. J., Baker, D., Bousquet, P., Bruhwiler, L., Chen, Y.-H., Ciais, P., Fan, S., et al. (2002). Towards robust regional estimates of co2 sources and sinks using atmospheric transport models. *Nature*, 415(6872):626.

Haverd, V., Ahlström, A., Smith, B., and Canadell, J. G. (2017). Carbon cycle responses of semi-arid ecosystems to positive asymmetry in rainfall. *Global change biology*, 23(2):793–800.

Hawkins, E., Ortega, P., Suckling, E., Schurer, A., Hegerl, G., Jones, P., Joshi, M., Osborn, T. J., Masson-Delmotte, V., Mignot, J., et al. (2017). Estimating changes in global temperature since the preindustrial period. *Bulletin of the American Meteorological Society*, 98(9):1841–1856.

Heimann, M., Monfray, P., and Polian, G. (1988). Long range transport of rn-222: A test for 3d tracer models. *Chemical Geology*, 70:98.

Heimann, M., ScHULzE, E.-D., WiNDERLicH, J., Andreae, M., Chi, X., Gerbig, C., Kolle, O., Kübler, K., Lavric, J., Mikhailov, E., et al. (2014). The zotino tall tower observatory (zotto): Quantifying large scale biogeochemical changes in central siberia. *Nova Acta Leopoldina NF*, 117(399):51–64.

Houghton, R. (2000). Interannual variability in the global carbon cycle. *Journal of Geophysical Research: Atmospheres*, 105(D15):20121–20130.

Houze, R. J. (1993). *Cloud dynamics*. Academic Press.

Huxman, T. E., Snyder, K. A., Tissue, D., Leffler, A. J., Ogle, K., Pockman, W. T., Sandquist, D. R., Potts, D. L., and Schwinning, S. (2004). Precipitation pulses and carbon fluxes in semiarid and arid ecosystems. *Oecologia*, 141(2):254–268.

Ichii, K., Matsui, Y., Yamaguchi, Y., and Ogawa, K. (2001). Comparison of global net primary production trends obtained from satellite-based normalized difference vegetation index and carbon cycle model. *Global Biogeochemical Cycles*, 15(2):351–363.

Jiménez-Muñoz, J. C., Mattar, C., Barichivich, J., Santamaría-Artigas, A., Takahashi, K., Malhi, Y., Sobrino, J. A., and Van Der Schrier, G. (2016). Record-breaking warming and extreme drought in the amazon rainforest during the course of el niño 2015–2016. *Scientific reports*, 6:33130.

Jung, M., Reichstein, M., Schwalm, C. R., Huntingford, C., Sitch, S., Ahlström, A., Arneth, A., Camps-Valls, G., Ciais, P., Friedlingstein, P., et al. (2017). Compensatory water effects link yearly global land co2 sink changes to temperature. *Nature*, 541(7638):516.

Keeling, C. D., Bacastow, R. B., Bainbridge, A. E., Ekdahl Jr, C. A., Guenther, P. R., Waterman, L. S., and Chin, J. F. (1976). Atmospheric carbon dioxide variations at mauna loa observatory, hawaii. *Tellus*, 28(6):538–551.

Koren, G. and Peters, W. (2018). Personal communication.

Koren, G., van Schaik, E., Araújo, A. C., Boersma, K. F., Gärtner, A., Killaars, L., Kooreman, M. L., Kruijt, B., van der Laan-Luijkx, I. T., von Randow, C., et al. (2018). Widespread reduction in sun-induced fluorescence from the amazon during the 2015/2016 el niño. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 373(1760):20170408.

Kottek, M., Grieser, J., Beck, C., Rudolf, B., and Rubel, F. (2006). World map of the köppen-geiger climate classification updated. *Meteorologische Zeitschrift*, 15(3):259–263.

Krol, M., Houweling, S., Bregman, B., Broek, M. v., Segers, A., Velthoven, P. v., Peters, W., Dentener, F., and Bergamaschi, P. (2005). The two-way nested global chemistry-transport zoom model tm5: algorithm and applications. *Atmospheric Chemistry and Physics*, 5(2):417–432.

Laan-Luijkx, I., Velde, I., Krol, M., Gatti, L., Domingues, L., Correia, C., Miller, J., Gloor, M., Leeuwen, T., Kaiser, J., et al. (2015). Response of the amazon carbon balance to the 2010 drought derived with carbontracker south america. *Global Biogeochemical Cycles*, 29(7):1092–1108.

Le Quéré, C., Andrew, R. M., Canadell, J. G., Sitch, S., Korsbakken, J. I., Peters, G. P., Manning, A. C., Boden, T. A., Tans, P. P., Houghton, R. A., et al. (2016). Global carbon budget 2016. *Earth System Science Data (Online)*, 8(2).

Lewis, S. L., Brando, P. M., Phillips, O. L., van der Heijden, G. M., and Nepstad, D. (2011). The 2010 amazon drought. *Science*, 331(6017):554–554.

Li, W. and Fu, R. (2004). Transition of the large-scale atmospheric and land surface conditions from the dry to the wet season over amazonia as diagnosed by the ecmwf re-analysis. *Journal of Climate*, 17(13):2637–2651.

L'Heureux, M. L., Takahashi, K., Watkins, A. B., Barnston, A. G., Becker, E. J., Di Liberto, T. E., Gamble, F., Gottschalck, J., Halpert, M. S., Huang, B., et al. (2017). Observing and predicting the 2015/16 el niño. *Bulletin of the American Meteorological Society*, 98(7):1363–1382.

Maeght, J.-L., Rewald, B., and Pierret, A. (2013). How to study deep roots—and why it matters. *Frontiers in plant science*, 4:299.

man2html (2002). Ccgcrv.

Matsueda, H., Inoue, H. Y., and Ishii, M. (2002). Aircraft observation of carbon dioxide at 8–13 km altitude over the western pacific from 1993 to 1999. *Tellus B: Chemical and Physical Meteorology*, 54(1):1–21.

Murayama, S., Harada, K., Gotoh, K., Kitao, T., Watai, T., and Yamamoto, S. (2003). On large variations in atmospheric co2 concentration observed over the central and western pacific ocean. *Journal of Geophysical Research: Atmospheres*, 108(D8).

Myhre, G., Samset, B. H., Schulz, M., Balkanski, Y., Bauer, S., Berntsen, T. K., Bian, H., Bellouin, N., Chin, M., Diehl, T., et al. (2013). Radiative forcing of the direct aerosol effect from aerocom phase ii simulations. *Atmospheric Chemistry and Physics*, 13(4):1853.

Myneni, R. B., Yang, W., Nemani, R. R., Huete, A. R., Dickinson, R. E., Knyazikhin, Y., Didan, K., Fu, R., Juárez, R. I. N., Saatchi, S. S., et al. (2007). Large seasonal swings in leaf area of amazon rainforests. *Proceedings of the National Academy of Sciences*, 104(12):4820–4823.

NOAA a, E. S. R. L. G. M. D. (2016). Baseline observatories. Accessed: 2018-11-28.

NOAA b, E. S. R. L. G. M. D. (2018). Baseline observatories. Accessed: 2018-11-28.

NOAA c, E. S. R. L. G. M. D. (2018). Noaadata. Accessed: 2018-11-28.

NOAA d, E. S. R. L. G. M. D. (2016). Noaa/esrl calculation of global means. Accessed: 2018-12-11.

NOAA e, E. S. R. L. G. M. D. (2018). Noaa/esrl calculation of global means. Accessed: 2018-12-11.

NOAA f, E. S. R. L. G. M. D. (2016). Curve fitting methods applied to time series in noaa/esrl/gmd. Accessed: 2019-01-15.

Pachauri, R. K., Allen, M. R., Barros, V. R., Broome, J., Cramer, W., Christ, R., Church, J. A., Clarke, L., Dahe, Q., Dasgupta, P., et al. (2014). *Climate change 2014: synthesis report. Contribution of Working Groups I, II and III to the fifth assessment report of the Intergovernmental Panel on Climate Change*. IPCC.

Peters, W., Jacobson, A. R., Sweeney, C., Andrews, A. E., Conway, T. J., Masarie, K., Miller, J. B., Bruhwiler, L. M., Pétron, G., Hirsch, A. I., et al. (2007). An atmospheric perspective on north american carbon dioxide exchange: Carbontracker. *Proceedings of the National Academy of Sciences*, 104(48):18925–18930.

Peters, W., Krol, M., Dlugokencky, E., Dentener, F., Bergamaschi, P., Dutton, G., Velthoven, P. v., Miller, J., Bruhwiler, L., and Tans, P. (2004). Toward regional-scale modeling using the two-way nested global model tm5: Characterization of transport using sf6. *Journal of Geophysical Research: Atmospheres*, 109(D19).

Peters, W., Krol, M., Van Der Werf, G., Houweling, S., Jones, C., Hughes, J., Schaefer, K., Masarie, K., Jacobson, A., Miller, J., et al. (2010a). Seven years of recent european net terrestrial carbon dioxide exchange constrained by atmospheric observations. *Global Change Biology*, 16(4):1317–1337.

Peters, W., Miller, J., Schaefer, K., van der Velde, I., van der Werf, G., Dolman, A., Carvalhais, N., and Tans, P. (2010b). Forest carbon imbalance information improves atmosphere based carbon data assimilation systems. In *AGU Fall Meeting Abstracts*.

Peters, W., Miller, J., Whitaker, J., Denning, A., Hirsch, A., Krol, M., Zupanski, D., Bruhwiler, L., and Tans, P. (2005). An ensemble data assimilation system to estimate co2 surface fluxes from atmospheric trace gas observations. *Journal of Geophysical Research: Atmospheres*, 110(D24).

Phillips, O. L., Aragão, L. E., Lewis, S. L., Fisher, J. B., Lloyd, J., López-González, G., Malhi, Y., Monteagudo, A., Peacock, J., Quesada, C. A., et al. (2009). Drought sensitivity of the amazon rainforest. *Science*, 323(5919):1344–1347.

Pickers, P. and Manning, A. (2015). Investigating bias in the application of curve fitting programs to atmospheric time series. *Atmospheric Measurement Techniques*, 8(3):1469–1489.

Potter, C., Randerson, J., Field, C., Matson, P., Vitousek, P., Mooney, H., and Klooster, S. (1995). A process-oriented model based on global satellite and surface data. *Global Biogeochem. Cy*, 7:811–842.

Poulter, B., Frank, D., Ciais, P., Myneni, R. B., Andela, N., Bi, J., Broquet, G., Canadell, J. G., Chevallier, F., Liu, Y. Y., et al. (2014). Contribution of semi-arid ecosystems to interannual variability of the global carbon cycle. *Nature*, 509(7502):600.

Quéré, C. L., Moriarty, R., Andrew, R., Canadell, J., Sitch, S., Korsbakken, J., Friedlingstein, P., Peters, G., Andres, R., Boden, T., et al. (2015). Global carbon budget 2015. *Earth System Science Data*, 7(2):349–396.

Raudys, S. J. and Jain, A. K. (1991). Small sample size effects in statistical pattern recognition: Recommendations for practitioners. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 3:252–264.

Rödenbeck, C., Zaehle, S., Keeling, R., and Heimann, M. (2018). How does the terrestrial carbon exchange respond to inter-annual climatic variations? a quantification based on atmospheric co2 data. *Biogeosciences*, 15(8):2481–2498.

Schaefer, K., Collatz, G. J., Tans, P., Denning, A. S., Baker, I., Berry, J., Prihodko, L., Suits, N., and Philpott, A. (2008). Combined simple biosphere/carnegie-ames-stanford approach terrestrial carbon cycle model. *Journal of Geophysical Research: Biogeosciences*, 113(G3).

Sellers, P. J., Randall, D., Collatz, G., Berry, J., Field, C., Dazlich, D., Zhang, C., Collelo, G., and Bounoua, L. (1996a). A revised land surface parameterization (sib2) for atmospheric gcms. part i: Model formulation. *Journal of climate*, 9(4):676–705.

Sellers, P. J., Tucker, C. J., Collatz, G. J., Los, S. O., Justice, C. O., Dazlich, D. A., and Randall, D. A. (1996b). A revised land surface parameterization (sib2) for atmospheric gcms. part ii: The generation of global fields of terrestrial biophysical parameters from satellite data. *Journal of climate*, 9(4):706–737.

Spinoni, J., Naumann, G., Vogt, J. V., and Barbosa, P. (2015). The biggest drought events in europe from 1950 to 2012. *Journal of Hydrology: Regional Studies*, 3:509–524.

Thoning, K. W., Tans, P. P., and Komhyr, W. D. (1989). Atmospheric carbon dioxide at mauna loa observatory: 2. analysis of the noaa gmcc data, 1974–1985. *Journal of Geophysical Research: Atmospheres*, 94(D6):8549–8565.

Van der Laan-Luijkx, Ingrid, T., Van der Velde, I. R., Van der Veen, E., Tsuruta, A., Stanis-
lawska, K., Babenhauserheide, A., Zhang, H. F., Liu, Y., He, W., et al. (2017). The car-
bontracker data assimilation shell (ctdas) v1. 0: implementation and global carbon balance
2001-2015. *Geoscientific Model Development*, 10(7):2785–2800.

Van der Velde, I. (2015). *Studying biosphere-atmosphere exchange of CO2 through Carbon-13
stable isotopes*. Wageningen University.

van Schaik, E., Killaars, L., Smith, N. E., Koren, G., van Beek, L., Peters, W., and van der
Laan-Luijkx, I. T. (2018). Changes in surface hydrology, soil moisture and gross primary
production in the amazon during the 2015/2016 el nino. *Philosophical Transactions of the
Royal Society B: Biological Sciences*, 373(1760):20180084.

Waterman, L. S., Nelson, D. W., Komhyr, W. D., Harris, T. B., Thoning, K. W., and Tans,
P. P. (1989). Atmospheric carbon dioxide measurements at cape matatula, american samoa,
1976–1987. *Journal of Geophysical Research: Atmospheres*, 94(D12):14817–14829.

Welp, L. R., Keeling, R. F., Meijer, H. A., Bollenbacher, A. F., Piper, S. C., Yoshimura, K.,
Francey, R. J., Allison, C. E., and Wahlen, M. (2011). Interannual variability in the oxygen
isotopes of atmospheric co2 driven by el niño. *Nature*, 477(7366):579.

Wenzel, S., Cox, P. M., Eyring, V., and Friedlingstein, P. (2014). Emergent constraints on
climate-carbon cycle feedbacks in the cmip5 earth system models. *Journal of Geophysical
Research: Biogeosciences*, 119(5):794–807.

Yang, J., Tian, H., Pan, S., Chen, G., Zhang, B., and Dangal, S. (2018). Amazon drought
and forest response: Largely reduced forest photosynthesis but slightly increased canopy
greenness during the extreme drought of 2015/2016. *Global change biology*, 24(5):1919–
1934.