University of Exeter
Department of Mathematics

# Calibration of plant functional type parameters using the adJULES system

Nina Raoult

June 2017

Supervised by Dr Tim Jupp

Submitted by Nina Raoult, to the University of Exeter as a thesis for the degree of Doctor of Philosophy in Mathematics , June 2017.



I certify that all material in this thesis which is not my own work has been identified and that no material has previously been submitted and approved for the award of a degree by this or any other University.

(signature) .....

# Declaration

Parts of this thesis, most notably the description of adJULES found in Chapter 2 and the results given in the first half of Chapter 4, have been published in Raoult et al. [2016].

Nina Raoult

# Abstract

Land-surface models (LSMs) are crucial components of the Earth system models (ESMs) that are used to make coupled climate-carbon cycle projections for the 21st century. The Joint UK Land Environment Simulator (JULES) is the land-surface model used in the climate and weather forecast models of the UK Met Office. JULES is also extensively used offline as a land-surface impacts tool, forced with climatologies into the future. In this study, JULES is automatically differentiated with respect to JULES parameters using commercial software from FastOpt, resulting in an analytical gradient, or adjoint, of the model. Using this adjoint, the adJULES parameter estimation system has been developed to search for locally optimum parameters by calibrating against observations. This thesis describes adJULES in a data assimilation framework and demonstrates its ability to improve the model-data fit using eddy-covariance measurements of gross primary productivity (GPP) and latent heat (LE) fluxes. The adJULES system is extended to have the ability to calibrate over multiple sites simultaneously. This feature is used to define new optimised parameter values for the five plant functional types (PFTs) in JULES. The optimised PFT-specific parameters improve the performance of JULES at over 85% of the sites used in the study, at both the calibration and evaluation stages. The new improved parameters for JULES are presented along with the associated uncertainties for each parameter. The results of the calibrations are compared to structural changes and used in a cluster analysis in order to challenge the PFT definitions in JULES. This thesis concludes with simple sensitivity studies which assess how the calibration of JULES has affected the sensitivity of the model to $CO_2$-induced climate change.

# Acknowledgements

I would like to thank my supervisor Tim Jupp, whose maddening attention to detail drove me to finally learn about split infinitives. I am enormously grateful for his patience, motivation, and immense knowledge, and for all the guidance he has provided me over the years.

I would also like to thank Peter Cox for sharing the his depths of knowledge and his great enthusiasm, and for helping me see the big picture.

Big thanks go to Margriet for providing me with the FluxNet data needed for this project and to Anna for showing me the ropes with JULES.

I have had the pleasure of sharing an office with some wonderful people; indeed I would like to thank Nathan, Nick, Matt, Fanny, and Arwen for bring much needed life back to the office (if not too much life at times) and for providing fierce debates at lunchtime.

I would like to thank my housemates, the many iterations who have made Exeter one of my favourite places to live in. Thanks to Nic for reigniting my love of films, to Laura who has provided me with many coffee breaks and date nights, to Tessa who has always been a housemate, long before she got an official key, and finally, to James who has been there from the start.

There have also been many people who have helped keep me sane during this time even from afar. To Faye and Emily, for their many tea break phone calls and encouraging words, and to Krisz and Sami for their numerous visits. To Michael who will take me skydiving. And to the rest of Oxpeeps and the Reading lot who never fail me, I am always touched by these friendships.

Indeed, these lists of names are not exhaustive and there are many other special people that I am grateful to; if you are reading this you are likely one of them.

The most of special thanks go to my family. I am so fortunate to have a family who understands what I do and believes in me. To Mum for her continuous love and support. To Dad and Justine, who have provided a sounding board for my research, between them they cover many aspects of the PhD. To Max, with whom I revert back to a competitive child, and to Camille, who even from a different country has been there with champagne and Skype calls whenever I needed them most.

Finally to Matt for completing the perfect Hanabi firework with me.

# Contents

# List of Tables

# List of Figures

# 1. Introduction

## 1.1. Motivation

Atmospheric carbon dioxide ($CO_2$) plays a critical role in regulating the global climate. Its role can best be understood by considering the terrestrial radiation balance which determines the Earth's climate. This is the balance between incoming solar energy and outgoing longwave radiation. Changes to the atmospheric composition or surface properties of the Earth inevitably alter this balance.

Shortwave radiation emitted by the Sun warms the Earth's surface, causing it to emit longwave radiation. Some of the longwave radiation is then absorbed and re-emitted by clouds and greenhouse gases (GHGs), warming the surface and lower atmosphere. Changing the atmospheric concentrations of long-lived GHGs, including $CO_2$, alters the Earth's energy balance by changing the amount of absorbed outgoing longwave radiation. Anthropogenic emissions of GHGs have dramatically increased atmospheric concentrations, leading to changes in the Earth's climate, most of which are still unknown and hard to predict.

The oceans and the terrestrial biosphere absorb $CO_2$, removing on average 50% of anthropogenic emissions from the atmosphere. However, the driving mechanisms and feedbacks of this prcoess are not completely understood. As such, the terrestrial carbon cycle represents a large source of uncertainty in climate projections, and so the development of accurate land-surface models is an important prerequisite for reliable climate projections.

As well as removing $CO_2$ from the atmosphere, the land is a significant store of carbon, containing about three times as much as the atmosphere [Denman et al., 2007]. The carbon store is partitioned between live vegetation, litter, and soil carbon, of which the latter is the greatest. Land carbon depends on the balance of inputs and outputs of carbon from and to the atmosphere. Currently, the land is a carbon sink, meaning there is net uptake of carbon. However, changes to the land carbon sink are likely because land carbon fluxes are sensitive to changes in the climate [Cox et al., 2006]. Changes to the land carbon sink could have signification ramifications for the climate system.

Photosynthesis provides the main input of carbon to the terrestrial biosphere. This is the process by which plants use water from the soil, $CO_2$ from the atmosphere and solar energy to convert carbon dioxide to stores of chemical energy. Outputs of carbon from the terrestrial biosphere to the atmosphere include decomposition, respiration, fire, and land use changes.

Increased atmospheric $CO_2$ is expected to enhance photosynthesis, increasing carbon uptake, while warming is expected to accelerate respiration rates, decreasing carbon storage [Cox et al., 2006]. Accordingly, therefore, understanding the effects of $CO_2$ and temperature on plant growth and respiration is vital for predictions of the future land carbon balance.

## 1.2. Modelling the land-surface

### 1.2.1. The fluxes between the biosphere and the atmosphere

The movement of any material from one place to another is called a flux. The main fluxes that make up the terrestrial carbon cycle describe the exchange of carbon, water and energy between the biosphere and the atmosphere.

#### Carbon flux

Plants use water from the soil, atmospheric $CO_2$, and energy from sunlight to make carbohydrates via the photosynthesis reaction:

$$6CO_2 + 6H_2O + \text{energy} \longrightarrow C_6H_{12}O_6 + 6O_2 \tag{1.1}$$

The gross primary productivity (GPP) is the gross uptake of $CO_2$ to plants associated with this reaction. Current estimates suggest photosynthesis removes 120 gigatonnes of carbon per year ($\mathrm{GtC\,yr^{-1}}$) from the atmosphere.

The carbohydrates synthesised in this process are used in the building of tissues that make up the leaves, branches, roots, and trunk of plants. In this way, part of the $CO_2$ removed from the atmosphere is stored in the structure of plants; the residence time of plant carbon is about one year in the leaves, up to more than 100 years in certain types of tree trunks.

Plants release $CO_2$ back into the atmosphere through the process of respiration. Respiration occurs as plant cells use the carbohydrates made during photosynthesis and oxygen to release energy (Eq. 1.1 in reverse). This energy is then used to grow and maintain the plant's tissues. Plant respiration is called autotrophic respiration ($R_a$) and represents approximately half (60 $\mathrm{GtC\,yr^{-1}}$) of the $CO_2$ that is returned to the atmosphere in the terrestrial portion of the carbon cycle.

The amount of photosynthetic carbon that is not used for respiration and available for other processes is called the net primary productivity (NPP) and is equal to GPP minus $R_a$. This represents the actual sequestration of carbon by the plant biomass.

In addition to the death of whole plants, living plants lose biomass by shedding a portion of their leaves, roots and branches each year. This acts as a transfer of carbon from the plant to the soil. Dead plant material is then decomposed by microbial fauna and bacteria. The

activity of these organisms releases carbon to the atmosphere via heterotrophic respiration ($R_h$), while the main nutrients, such as nitrate and phosphate, are rapidly recycled by plants.

The net exchange of carbon between the ecosystem and the atmosphere (NEE) is the difference between the sum of the respiration fluxes ($R_{\text{eco}} = R_a + R_h$) and the photosynthesis flux:

$$\text{NEE} = R_{\text{eco}} - \text{GPP}. \tag{1.2}$$

NEE is negative when an ecosystem is a net carbon sink and positive for a net source.

### Water and Energy Flux

Two major fluxes determine the Earth's energy balance; the 'radiative' flux and the 'turbulent' flux. Radiative fluxes are associated with the shortwave radiation from the sun and reflected by the Earth's surface, and the longwave radiation emitted by Earth's surface and radiated toward the surface by the atmosphere. Turbulent fluxes are associated with heating of the atmosphere by the Earth's surface (sensible heat) and phase changes of water (latent heat), so named because both processes are driven by wind.

$$R_{\text{net}} - G = H + \text{LE} + \Delta S \tag{1.3}$$

The left-hand side of the equation denotes radiative and conductive fluxes; $R_{\text{net}}$ is the net radiation and $G$ the soil heat flux. The right-hand side contains the turbulent fluxes; $H$ is the sensible heat flux and LE is the latent heat flux. Positive $R_{\text{net}}$ supplies energy to the surface and positive $G$, $H$ and LE remove energy from the surface. Finally, $\Delta S$ denoted the change in storage of energy in the soil, vegetation, and air within the canopy.

The latent heat flux (LE) represents the summed contribution of the evapotranspiration processes: evaporation from the soil, evaporation of the rain water intercepted by the leaves, and transpiration of water by the plant through the leaves. Transpiration consists of the vaporisation of liquid water contained in plant tissues and its removal to the atmosphere. Transpiration takes place at the level of the leaf stomata; small openings on the plant leaf through which gases and water vapour pass through. The vapour exchange with the atmosphere is therefore controlled by the stomatal aperture.

$CO_2$ used for the photosynthesis reaction is also transmitted through the leaf stomata (in the opposite direction). This results in a coupling between photosynthesis and transpiration via the concept of stomatal conductance. This coupling is crucial because the plant regulates its stomatal conductance in order to minimise water loss as a function of soil moisture availability and to maximise $CO_2$ absorption. Therefore, in order to constrain the carbon cycle, it is also important to consider the LE flux.

Evapotranspiration ($E$) is also a key part of the water balance where

$$P = R + E + \Delta S \tag{1.4}$$

with $P$ representing precipitation, $R$ representing runoff and $\Delta S$ is the change in storage (in soil or the bedrock/ground water).

### 1.2.2. A brief history of land-surface models

Land-surface models (LSMs) have formed an important component of climate models for many decades [Pitman, 2003]. They have evolved greatly over time due to advances in scientific understanding and increased motivation for modelling the climate. The increasing availability of satellite and *in situ* measurements has facilitated the development of more complex models by providing data against which to validate. Although there have been great advances made over the last fifty years, even the most sophisticated LSM remains a gross simplification of the full climate system [McGuffie and Henderson-Sellers, 2001].

First generation LSMs focussed on providing the lower boundary condition for atmospheric models by calculating the land-atmosphere fluxes of heat, moisture, and momentum, and updating the surface state variables on which these fluxes depend (e.g. soil temperature, soil moisture, snow cover). These models date back to the late 1960s and early 1970s (e.g. Manabe [1969]) and were developed for use in numerical weather prediction models. As such, vegetation was treated simply as a passive structure separating, but not interacting with, the soil and the atmosphere.

In the mid to late 1990s some land-surface modelling groups began to introduce additional aspects of biology into their schemes, most notably the dynamic control of transpiration by leaf stomata and the connected rates of leaf photosynthesis [Sellers et al., 1997; Cox et al., 1999]. In the early 2000s, climate modelling groups began to use the carbon fluxes simulated by LSMs within first generation climate carbon cycle models [Cox et al., 2000; Friedlingstein et al., 2001]. These early results, and a subsequent model inter-comparison [Friedlingstein et al., 2006], highlighted the uncertainties associated with land carbon climate feedbacks. The 5th Assessment Report of the Intergovernmental Panel on Climate Change (IPCC AR5; Stocker et al. [2013]) for the first time routinely included climate models with an interactive carbon cycle (now called Earth System models or ESMs), confirming that land responses to climate and $CO_2$ are amongst the largest of the uncertainties in future climate change projections [Brovkin et al., 2013; Jones et al., 2013; Friedlingstein et al., 2013]. Any future decreased ability of the land surface to draw down atmospheric $CO_2$ could imply smaller "compatible emissions" in order to stay below key warming thresholds such as 2°C.

Uncertainties in LSMs arise from three major sources: uncertainty due to initial and boundary conditions, process uncertainty, and parameter uncertainty. Taking these one by one, uncertainty due to initial and boundary conditions include uncertainties in the forcing data and initial state of the model [Kavetski et al., 2006a,b; Ajami et al., 2007]. Process uncertainty includes the misrepresentation of land-surface processes and also the neglect of important processes such as nitrogen limitations on plant growth (see for example Thornton et al., 2007; Zaehle et al., 2010) or canopy light interception [Mercado et al.,

2009]. The drive to reduce process uncertainty almost invariably leads to increases in LSM complexity, which typically lead to the introduction of additional internal model parameters. Parameter uncertainty arises from uncertainty in these internal model parameters. The evolution of LSMs has therefore involved an attempt to reduce process uncertainty by increasing model realism and complexity, but at the cost of increasing parameter uncertainty. This thesis concerns the development and application of a technique to reduce parameter uncertainty in the widely used Joint UK Land Environment Simulator (JULES) LSM [Best et al., 2011; Clark et al., 2011].

## 1.3. Using land-surface models in climate predictions

### 1.3.1. Land-surface models as components of Earth System models

The land-surface is an integral component of the Earth System and a fundamental part of the carbon cycle. Understanding land-atmosphere exchanges of $CO_2$ is of great relevance when calculating emission reductions. In addition, since the land-surface provides food and influences the water supply, understanding the potential impact of climate change on agricultural yield is of vital importance. Land-carbon feedbacks remain uncertain and there are many questions associated with future land surface functioning in a changing climate as well as the impact of different land-use changes [Huntingford et al., 2010]. Therefore, the land-surface component is an integral part of any Earth System Model used to predict the effect of climate change.

Earth System Models (ESMs) are the main tool of climate change research. They provide remarkable and important insights into the functioning of the climate system, and are used in the Intergovernmental Panel on Climate Change (IPCC) reports to set targets for mitigation (IPCC AR5; Stocker et al. [2013]).

Until the last decade, the majority of experiments neglected the feedback between climate and the carbon cycle. Instead, to predict future climate changes, studies used general circulation models with prescribed $CO_2$ concentrations and fixed vegetation distribution. These $CO_2$ concentrations were derived from emission scenarios using relatively simple offline carbon cycle models (emission scenarios are further discussed in Sect. 1.3.3). However, in order to understand fully the large feedbacks between the climate and carbon cycle, it is essential to model the two simultaneously in coupled climate-carbon cycle models [Cox et al., 2000; Friedlingstein et al., 2006].

#### Met Office Hadley Centre models

The JULES land-surface model is one of the components of the UK Met Office Unified Model. The UK Met Office Hadley Centre has developed different configurations of the Unified Model to use for climate predictions at varying time-scales: seasonal, decadal, and centennial. With a prefix Had- (for Met Office Hadley Centre), multiple iterations exist

such as the HadCM3 (coupled-model version 3) and more recently the HadGEM3 (Global Environment Model version 3).

These climate models are usually at a lower resolution than the models used for day to day weather forecasting due to computational cost. By default they include ocean and sea-ice components coupled to the atmosphere model in order to represent the full coupled climate system. The "Earth System" configurations also add processes associated with atmospheric chemistry and the terrestrial ecosystem. The JULES land-surface model provides the land-surface element found in the latter configurations.

### Dynamic global vegetation models

In order to understand fully the role of climate-vegetation feedbacks on large timescales, the land cover needs to be treated as an interactive element . This is done by incorporating dynamic global vegetation models (DGVMs) directly within climate models. DGVMs update the plant distribution and soil carbon based on climate-sensitive $CO_2$ fluxes at the land-atmosphere interface.

When JULES is running as part of the UK Met Office Unified Model, or as the terrestrial carbon cycle component of the Met Office Hadley Centre's coupled climate-carbon cycle model, it is run with a dynamic global vegetation model called TRIFFID (Cox [2001]: Top-down Representation of Interactive Foliage and Flora Including Dynamics).

Each plant functional type (PFT) covers a different fraction of each gridbox in JULES. The area covered by each PFT is updated by TRIFFID, typically every ten days, based on the net carbon available to it and on the competition with other vegetation types, which is modelled using a Lotka-Volterra approach. Competition between the PFTs in the model is based on a dominance hierarchy ordered tree-shrub-grass, with dominant types limiting the expansion of sub-dominant types.

Each time TRIFFID is called, land-surface parameters, such as albedo and roughness length, are updated based on the new vegetation state. These land-surface parameters depend on the type, height and leaf area index of the vegetation. The new vegetation state allows for the changes in the biophysical properties of the land surface and in the terrestrial carbon storage to feed back into the atmosphere.

TRIFFID has historically used five PFTs [Clark et al., 2011]. These are the same five PFTs discussed throughout this thesis, chosen as a minimal set to represent the variation in vegetation structure (e.g canopy height, root depth) and function (e.g. C3 versus C4 photosynthesis). This ensures the inclusion of both biophysical and biogeochemical vegetation feedbacks in ESMs [Clark et al., 2011], although latter versions of TRIFFID now represent nine PFTs to improve the simulation of the global distribution of vegetation types [Harper et al., 2016].

The number of PFTs in other DGVMs will differ depending on the parameterisations used in the model. The availability of field data for defining and validating parameter values is

also an important factor in deciding the number of PFTs. When running a LSM within an ESM, detailed ecophysiological and physical parameters are needed [Clark et al., 2011]. Uncertainty in these parameter values will feed through as a source of uncertainty in the ESM.

### 1.3.2. Uncertainty in climate change predictions

Understanding and quantifying uncertainty in climate change projections is of increasing importance. It is a fundamental part of climate research, and plays an important role in mitigation planning and advice to policymakers.

Uncertainties in LSMs feed-through into uncertainties in climate change projections. Briefly touched on in Sect. 1.2.2, the uncertainties from LSMs can be grouped into initial condition, boundary condition, parameter, and structural uncertainties. Initial condition and boundary condition uncertainties, sometimes grouped as forcing uncertainty, are introduced if datasets are used to replace what in reality is an interactive part of the system [Tebaldi and Knutti, 2007]. In running JULES for this thesis, this is the uncertainty associated with the driving data which are used to represent the atmosphere and find the initial state of the model. When running climate models, the LSMs are coupled with atmospheric and ocean models, removing the need for driving data. Similarly, the uncertainty associated with the initial state of the system is less important when running long-term projections such as the multi-decadal runs used in climate predictions [Tebaldi and Knutti, 2007]. Therefore, the uncertainties from LSMs which will affect climate change projections will be the parameter and structural uncertainties associated with the model.

The parameter and structural uncertainties in the LSMs will differ between different models due to differences in physical and numerical formulations. Due to the highly complex climate system, it is impossible to describe all the processes accurately, no matter how complex the model [Tebaldi and Knutti, 2007]. Therefore, different models will choose to capture different processes and parameterise them in different ways. These choices in model design and resolution, most notably the ones which cannot be captured by changing parameter values, contribute to what is called structural (or process) uncertainty. This error is difficult to quantify since it is not always clear which processes are missing.

Parameter uncertainties in a model can be explored and quantified by perturbed physics ensembles (PPEs). This is where the model is run with an ensemble of different parameter values. Many of the attempts to quantify climate change or climate model parameters in a probabilistic sense have taken this approach [Tebaldi and Knutti, 2007]. The PPE method has the advantage of being relatively simple and easy to implement though it will only capture the uncertainty of a given model's representation of climate, i.e. it will not be able to capture uncertainty beyond the parameterisation. It is still able to provide valuable insight into the model even if just one parameter is perturbed.

The work leading up to this chapter has involved reducing parameter uncertainty in JULES. Therefore, in this chapter, the main emphasis will be on the effects of param-

eter change on the model's response. Reduction in structural uncertainty will also be briefly considered in order to contrast between parameter and structural uncertainty.

In addition to the model uncertainty found in the different components of the climate model, the full model will also suffer from forcing uncertainty. This arises mainly from incomplete knowledge of external factors influencing the climate system [Deser et al., 2012]. Climate models are run using prescribed emissions of greenhouse gasses (GHGs). These trajectories are highly uncertain due to unknowns in future world economic and social development. For example, changes in land-use, population size, emissions, and the developing of new technologies are all important factors.

### 1.3.3. Emission scenarios

To tackle the forcing uncertainty, a standard set of emission scenarios is used in most climate research. These scenarios do not reduce the uncertainty but help to understand it better in order to reach decisions that are robust under a wide range of possible futures [Schwartz, 1996].

A standard set of socio-economic and emission scenarios is used for climate research to be complementary and comparable across the scientific community. These ensure that starting conditions, historical data and projections are employed consistently across different studies. Due to the high computational cost of climate models, the scenarios can also provide a standardised starting framework for different types of experiments.

Scenarios provide plausible descriptions of how the future may evolve with respect to a range of variables [Van Vuuren et al., 2011]. In addition to future greenhouse gas and emissions of other air pollutants, the other variables considered cover technological changes, changes in energy generation and land use, global and regional economic circumstances, and population growth.

Scenarios are an integral part of the IPCC reports. They are used in the assessment of possible climate impacts, mitigation options and associated costs. It is important to note that the scenarios are not actual future predictions or policy recommendations, but a tool to explore both the scientific and real-world implications of different plausible futures.

#### Previous scenarios used in IPCC

Several sets of scenarios have been published by the IPCC. The first set of climate change scenarios published in 1992 was called IS92 [Leggett et al., 1992]. These were followed in 2000, by a second generation of projections, collectively referred to as the Special Report on Emissions Scenarios (SRES; Nakicenovic et al. [2000]). The SRES were used in two subsequent reports; the Third Assessment Report (TAR) and Assessment Report Four (AR4).

The SRES scenarios were designed to improve upon some aspects of the IS92 scenarios. Generated through an open process involving many different modelling teams, the SRES scenarios are 'baseline' scenarios; they do not take into account potential mitigation. For example, it is possible that emissions may change less than these scenarios imply through policy actions.

The SRES scenarios investigate the uncertainty of future greenhouse gas and short-lived pollutant emissions given a wide range of driving forces. Some of the cases explored the implications of economic convergence between developed and developing countries [Moss et al., 2010]. The quantitative SRES projections are complemented by storylines of the future, which facilitate the interpretation of the scenarios [Moss et al., 2010].

Both IS92 and SRES assumed there were no policy actions to mitigate climate change. Motivated by the need to explore mitigation options and evaluate adaptation strategies, SRES was superseded by Representative Concentration Pathways (RCPs) in 2014.

**Representative Concentration Pathways**

The Representative Concentration Pathways (RCPs) are the scenarios presented in the IPCC Fifth Assessment Report (AR5). The pathways are defined by their total radiative forcing pathway and level by 2100. Radiative forcing is a cumulative measure of human emissions of greenhouse gases (GHGs) from all sources expressed in Watts per square metre. There are four pathways: RCP8.5, RCP6.0, RCP4.5 and RCP2.6 (also referred to as RCP3-PD), where the numbers refer to the radiative forcing ($Wm^{-2}$) for each RCP and PD stands for Peak and Decline. Details of the pathways are found in Table. 1.3.3.

The name was chosen specifically to highlight the selection process involved in selecting the scenarios. The scenarios are 'representative' of the existing literature, and were developed independently by different modelling groups and chosen to represent a broad range of climate outcomes. The term 'pathway' is used to emphasize the trajectory that is taken over time to reach that outcome is of interest, as well the specific long-term concentration.

For this thesis, the RCPs are used to provide a link between change in atmospheric $CO_2$ and temperature rise. Figure. 1.1 shows the temperature time series generated when using the RCPs in the HadGEM2ES model.

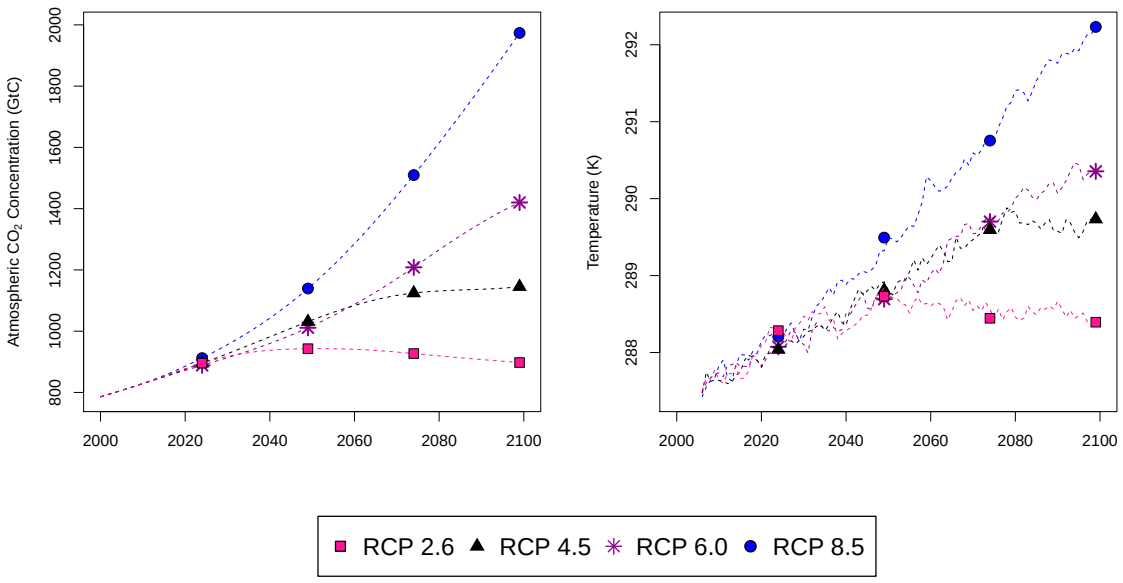



Figure 1.1.: Atmospheric carbon dioxide concentration (left panel) and simulated global mean temperature from the HadGEM2ES model (right panel), for the four RCP scenarios. Markers placed every 25 years.

| | Description | Primary Characteristics | Publication | IA Model |
|---|---|---|---|---|
| RCP 8.5 | Rising radiative forcing pathway leading to 8.5 $W\,m^{-2}$ (∼1370 ppm $CO_2$ eq) by 2100 | High population and relatively slow income growth with modest rates of technological change and energy intensity improvements, leading in the long term to high energy demand and GHG emissions in absence of climate change policies [Riahi et al., 2011]. Increasing greenhouse emission, and use of cropland and grasslands, driven by an increasing global population. | Riahi et al. [2007] | MESSAGE |
| RCP 6.0 | Stabilisation without overshoot pathway to 6 $W\,m^{-2}$ (∼850 ppm $CO_2$ eq) at stabilisation after 2100 | Medium baseline GHG emission with high mitigation. Medium air pollution. Limited move to renewables and nuclear. | Fujino et al. [2006] Hijioka et al. [2008] | AIM |
| RCP 4.5 | Stabilisation without overshoot pathway to 4.5 $W\,m^{-2}$ (∼650 ppm $CO_2$ eq) at stabilisation after 2100 | Very low baseline GHG emission with medium-low mitigation. Assumes that climate policies are invoked to achieve the goal of limiting emissions, concentrations and radiative forcing. Reforestation programs, the use of cropland and grassland decreases following considerable yield increases and dietary changes. | Clarke et al. [2007] Smith and Wigley [2006] Wise et al. [2009] | GCAM |
| RCP 2.6 | Peak in radiative forcing at ∼3 $W\,m^{-2}$ (∼490 ppm $CO_2$ eq) before 2100 and then decline (the selected pathway declines to 2.6 $W\,m^{-2}$ by 2100) | Low GHG emissions, the use of bio-energy and carbon capture and storage technologies resulting in negative emission. Increasing use of cropland due to bio-energy production and constant use of grassland as increase in production of animal products is met through a shift from extensive to more intensive animal husbandry. | Van Vuuren et al. [2007] | IMAGE |

Table 1.1.: Adapted from Table. 2 in Van Vuuren et al. [2011]. IA Model stands for Integrated Assessment Model (IAM), which are used to generate these scenarios.

## 1.4. The importance of calibration

The Earth System is highly complex and diverse, with many different physical, ecological, and biochemical processes governing it. As such, accurately reflecting all these processes with equations for a computer model is impossible. Even the most complex and sophisticated climate model is still a simplification of reality. It is impossible to model everything, and so different models choose to capture different processes, and equations are often parameterised in different ways. The optimal values of the parameters that make up these equations are often unknown and may vary over different models.

The representation of the vegetation in terrestrial models is an example of such a simplification. Since it is impossible to model every individual type of vegetation found around the globe, vegetation is grouped. The number of groupings tends to be small but still aims to cover variations in structure (such as canopy height), climate and function (e.g. C3 and C4 photosynthesis). In the JULES land-surface model, the terrestrial model used in this study, vegetation is collected into 5 such groupings called plant functional types (PFTs). Each PFT has a vector of approximately ten parameters describing it.

Even when parameters have a physical meaning that can be determined by experimental measurements, these measurements are often carried out at small scales, for example on leaves or on individual plants. These then need to be extrapolated to a value representative of an entire ecosystem. This can be difficult given the spatial variability and non-linearity of ecophysiological processes [Jarvis, 1995].

Calibration is a tool that can be used to improve these parameter values, determining the best estimates. It is the act of confronting a model with observations and changing the internal parameters so that the best match between model and observation is achieved. This is a very powerful technique, which can improve the model output and refine knowledge of the parameters. In addition, if the calibration is unable to find a satisfactory fit of the model to the observations, it is often possible to deduce structural problems in the model. These could include poorly described or missing processes in the model.

With respect to vegetation models, local flux measurements using the eddy-correlation method constitute the observational data of choice, as they provide direct, continuous, and high frequency observations of carbon and water vapour (discussed Sect. 2.2). These measurements are found at individual sites typically covering a few hectares.

## 1.5. Key questions

The mathematical framework of data assimilation makes it possible to adjust the internal parameters of a model and to quantify the contribution of this adjustment on the accuracy of the simulations. Optimising a model thus allows better parameterisation if the calibration is conclusive, or suggests new conceptual development. This thesis proposes to adopt this approach with the JULES land-surface model, using measurements of water

and carbon fluxes. The internal parameters varied in this study correspond to the PFTs of JULES and are saved in a vector.

In light of the above, this thesis aims to answer the following key questions:

KQ1: **Can a (locally) optimal vector of generic parameters for each of the JULES PFT classes be found in a robust and repeatable manner?**

Is it possible to create a robust framework in which repeatable and objective calibrations take place? How complicated does the framework need to be? Is it possible to calibrate over individual sites and pick one to be representative of the whole PFT?

KQ2: **Are the PFT definitions in JULES robust or do the observations suggest a different partitioning of the vegetation?**

Do vegetation groupings in the JULES make sense? Is it possible to calibrate over the whole PFT without creating too many outliers? And if calibrations are performed over individual sites, do these results support the current PFT grouping? Is it possible to look at the optimised parameter values for a specific site and know what type of vegetation is found at that site?

KQ3: **How do the parameter changes in JULES affect the model's response to $CO_2$ driven climate change?**

Once the model has been calibrated what does this mean for climate predictions? Does calibration affect how the model responds to changes in atmospheric $CO_2$ and temperature?

## 1.6. Thesis structure

In Chapter 2, the different components that make up the adJULES system are introduced. Beginning with the JULES land-surface model, some of the key equations are introduced in Sect. 2.1.2, putting the critical parameters in context. These equations relate mainly to photosynthesis, a vital process of the carbon cycle discussed above. In Sect 2.2, the data used to constrain the model are presented. These are *in situ* eddy-covariance data taken from the FluxNet database.

In Chapter 3, the cost function, the focal point of the optimisation scheme, is explored in greater depth. In order to answer KQ1, a multi-site cost function is developed and its robustness at different timescales is tested. In Sect. 3.3.2, the sensitivity of the system to different initial conditions is discussed.

Chapter 4 is split into two parts, both with the aim of answering KQ2. In the first half, the results of the optimisation are considered. In Sect. 4.1, the calibrations at site level are explored, i.e. using *in situ* data at specific FluxNet sites. In Sect. 4.3, new generic PFT parameter values are found by calibrating over multiple locations simultaneously.

The second half of Chapter 4 considers the ways in which the PFT representations in JULES could be redefined. The results of Harper et al. [2016] are discussed, which redefine the PFT through scientific changes. These are compared to the calibration results in Sect. 4.4.2. The PFT definitions are also challenged through a clustering experiment using the parameter values found in the site level experiments.

In Chapter 5, the main focus is KQ3. In this chapter, the calibrated and uncalibrated versions of the JULES model are run with different atmospheric $CO_2$ and temperature perturbations in order to see if the response of the model has changed. This is done first by focusing on photosynthesis in Sect. 5.1.1 and then by looking at Water Use Efficiency (the ratio of carbon uptake through photosynthesis to loss water by transpiration) in Sect. 5.1.2.

Finally, the main conclusions of the thesis are summarised in Chapter 6, with recommendations for the development of the adJULES system and a brief discussion of possible future work.

# 2. Background and Methodology

In this chapter, the different components of the adJULES system are introduced. In Sect. 2.1, a land-surface model to be optimised is considered. The observations against which to calibrate the model are considered in Sect. 2.2.

In Sect. 2.3, a data assimilation method within a Bayesian framework is discussed. Due to the mathematical nature of this thesis, this discussion of methodology is the main focus. Initially, in Sect. 2.3.2, data assimilation as a whole is discussed. This is followed in Sect. 2.3.4 by a specific look at the adjoint method used in this thesis. Automatic differentiation, which is used to generate the adjoint used in this project, is also covered in Sect. 2.3.4.

Some of the diagnostic tools used to analyse and quantify the improvements made are covered in Sect. 2.4. The chapter concludes in Sect. 2.5 by introducing the adJULES system, which is the parameter estimation system developed in this study.

## 2.1. The JULES land-surface model

The JULES land-surface model [Best et al., 2011; Clark et al., 2011] simulates the interactions between the land and the atmosphere. Originally developed from the Met Office Surface Exchange Scheme (MOSES, Cox et al. [1999]), JULES can be used in a stand-alone mode with observed atmospheric forcing data, or can be coupled into a general circulation model (GCM). JULES is currently the land surface model used in the UK Met Office Unified Model.

JULES is a mechanistic land-surface model including physical, biophysical, and biochemical processes. These control the radiation, heat, water, and carbon fluxes between land and atmosphere in response to time-series of the state of the overlying atmosphere [Best et al., 2011; Clark et al., 2011]. Processes such as photosynthesis, evaporation, plant growth, and soil microbial activity are all linked through mathematical equations. These equations quantify how soil moisture and temperature govern evapotranspiration, heat balance, respiration, photosynthesis, and carbon assimilation [Best et al., 2011; Clark et al., 2011]. JULES runs at a given sub-daily step (typically 30 minutes). Meteorological drivers such as rainfall, incoming radiation, temperature, humidity, and wind speed are used as inputs.

The vegetation in the JULES model is categorised into five plant functional types (PFTs); broadleaf trees (BT), needleleaf trees (NT), C3 grasses (C3G), C4 grasses (C4G), and

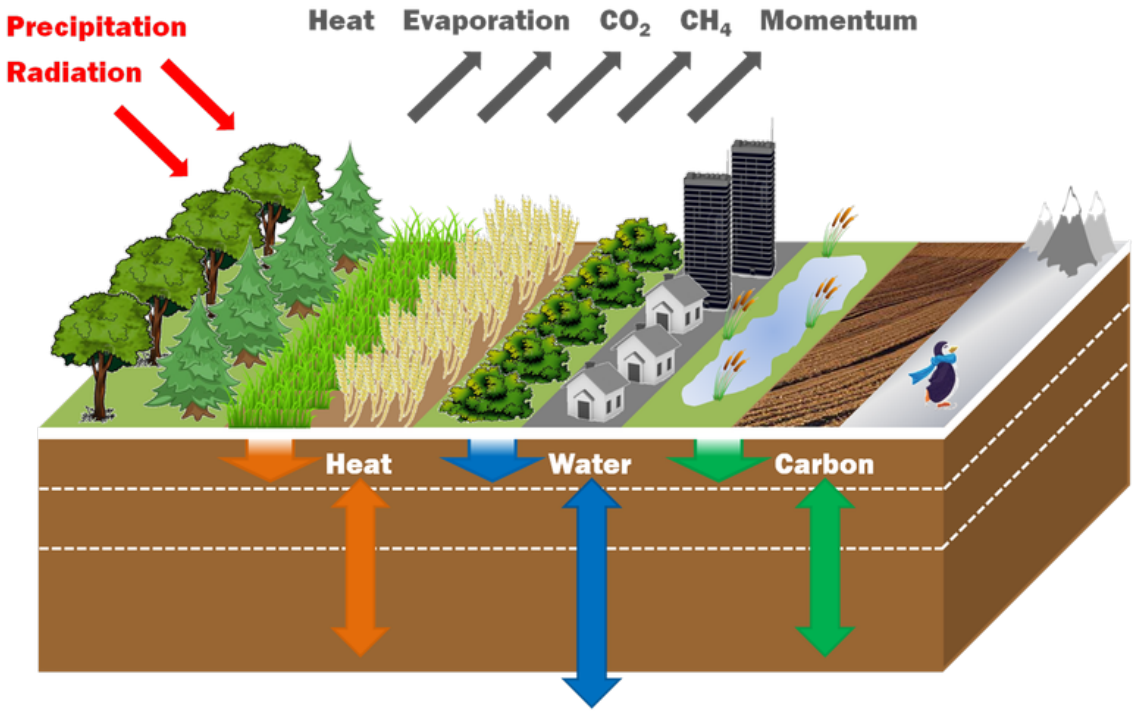



Figure 2.1.: Schematic diagram of the JULES land-surface model retrieved from http://jules.jchmr.org/content/about.

shrubs (Sh). Each PFT class has a different set of input parameters, the default values for which are taken from a previous study [Blyth et al., 2010]. The leaf-level carbon assimilation is calculated differently depending on whether a plant is modelled with a C3 or a C4 photosynthetic pathway.

The five PFTs, along with four non-vegetation surface types, exist on separate tiles within each gridbox in JULES [Best, 2005]. The fractional area of each surface type can be prescribed for each gridbox. A separate energy and carbon balance is calculated for each tile. The gridbox average for each flux is taken by weighting the values from each tile.

In JULES, soil processes are modelled in several layers. However, all the tiles lie over and interact with the same soil column [Best, 2005]. Meteorological driving variables are needed for each gridbox along with variables that describe the soil properties at that location.

JULES can be run for any number of gridboxes. It can also be run at a point where the inputs are taken to represent conditions at that point. This latter configuration is used in this project.

### 2.1.1. The JULES version used in this study

The JULES land-surface model is a community built model and under continual development. The most up to date version of JULES in circulation at the start of this project was version 3.4. The adJULES system utilises the adjoint of the JULES model - a complex piece of code which requires time (and money when using commercial programs) to

develop. This has meant that it has not been possible to keep up to date with the most recent version of JULES and still be able to conduct scientific studies. The most recent version of JULES to have an adjoint is 2.2. Therefore this is the version used in the thesis. The most pertinent changes to JULES between versions 2.2 and 3.4 (summarised from JULES version release notes found at http://jules.jchmr.org/content/about) are as follows:

- **JULES v3.0**: integrates the IMOGEN impacts tool into the JULES release. IMOGEN is an emulation of climate change using pattern-scaling calibrated against the Met-Office Hadley Centre GCM. A version of IMOGEN compatible with version 2.2 does exist and is briefly discussed in Chapter 5.
- **JULES v3.1:** restructuring of the code and interface.
- **JULES v3.2:** includes new output variables for isoprene. emissions
- **JULES v3.3:** set up to run in parallel and some improvements to the numerics in the soil hydrology.
- **JULES v3.4:** more changes to the structure of the code.

Note that there have been no major scientific changes between versions 2.2 and 3.4, only code structure changes. The scientific changes that have been made exist as switches which are switched off in default JULES runs.

Despite the above, there have been a few significant scientific changes since the beginning of this thesis. First, is the inclusion of a crop module in JULES (version 4.0). As a result, when considering which FluxNet sites to use for this study, the crop sites were purposefully omitted. The second major change has been to the plant physiology and PFT partitioning described in Harper et al. [2016]. This work is further discussed in Chapter 4.

Since the trunk of the JULES code has not changed significantly between versions 2.2 and 3.4, the results found using this version will still be informative, with structural problems identified, as well as parameter interactions and sensitivities, being easily be transferrable to the newest versions of the model. Future work includes updating the adJULES system to include the most current version of JULES. The framework and tools created during this study will be applicable to later versions, provided the adjoint is generated.

Several high impact publications have been produced using version 2.2 of the model (e.g. Gedney et al. [2006], Sitch et al. [2007], Cox et al. [2008], Booth et al. [2012]). Most importantly for this study perhaps is Blyth et al. [2011], which lays out benchmarking tests for JULES and acts as a precursor to this project. Without calibrating the model, Blyth et al. [2011] found, for example, that JULES systematically underestimated photosynthesis at temperate sites, while overestimating evaporation. These are the type of issues that the adJULES system aims to tackle.

### 2.1.2. Key parameters and their equations

Even a relatively simplistic land-surface representation such as JULES has over a hundred internal parameters representing the environmental sensitivities of the various land-surface types and PFTs within the model. In general these parameters are chosen to represent measurable quantities within the real world (e.g. aerodynamic roughness length, surface albedo, plant root-depth). This allows observationally-based estimates of these parameters to be made in the early stages of the model development process.

| Symbol | Name in code | Description | Units |
|---|---|---|---|
| $n_0$ | `nl0` | Top leaf nitrogen concentration | kgN $(\text{kgC})^{-1}$ |
| $f_0$ | `f0` | Maximum ratio of internal to external $CO_2$ | - |
| $d_r$ | `rootd_ft` | Root depth | m |
| $\alpha$ | `alpha` | Quantum efficiency | mol $CO_2$ $(\text{mol PAR})^{-1}$ |
| $\frac{\delta c}{\delta L}$ | `dcatch_dlai` | Rate of change of canopy interception capacity with leaf area index (LAI) | kg $m^{-2}$ |
| $T_{\text{low}}$ | `tlow` | Lower temperature for photosynthesis | °C |
| $T_{\text{upp}}$ | `tupp` | Upper temperature for photosynthesis | °C |
| $dq_c$ | `dqcrit` | Humidity deficit at which stomata close | kg $kg^{-1}$ |

Table 2.1.: The key JULES parameters used in this thesis.

The eight parameters calibrated within this study (see Table 2.1) relate predominantly to leaf-level stomatal conductance ($g$) and photosynthesis ($A$). Four of the parameters control the responses of $g$ and $A$ to environmental conditions, such as surface temperature ($T_{\text{upp}}$, $T_{\text{low}}$), solar radiation ($\alpha$), and atmospheric humidity deficit ($dq_c$). The calibration parameters $f_0$ and $n_0$ essentially control the maximum values of leaf-level stomatal conductance and photosynthesis. The remaining two calibration parameters influence the hydrological partitioning at the land surface and relate to the amount of rainfall intercepted by the plant canopy ($\delta c/\delta L$), and the root depth ($d_{\text{r}}$) from which each PFT can access soil water for transpiration. The simulated latent heat flux and gross primary productivity have been found to be especially sensitive to these parameters in previous studies [Blyth et al., 2010].

The full set of equations within the JULES model is documented in the literature by Best et al. [2011] and Clark et al. [2011], but the key equations for C3 vegetation are highlighted below. The C4 vegetation uses slightly different equations, but these only apply to two of the sites in this study.

In JULES, leaf-level photosynthesis and stomatal conductance are treated with a coupled model [Cox et al., 1998]. Based on the models of Collatz et al. [1991, 1992], leaf-level photosynthesis $A$ is controlled by the carboxylation rate (which depends on $n_0$, $T_{\text{low}}$, $T_{\text{upp}}$) and light-limited photosynthesis (which depends on $\alpha$).

When unstressed by water availability, the potential leaf-level photosynthesis $A_p$ is the

smoothed minimum of the following three limiting rates:

- Rubisco-limited rate

$$W_c = V_{\text{cmax}}(n_0, T_{\text{upp}}, T_{\text{low}}) \left[ \frac{c_i(f_0, dq_c) - c_*}{c_i(f_0, dq_c) + K_c(1 + O_a/K_o)} \right] \tag{2.1}$$

- Light-limited rate

$$W_l = \alpha(1 - \omega) I_r \left[ \frac{c_i(f_0, dq_c) - c_*}{c_i(f_0, dq_c) + 2c_*} \right] \tag{2.2}$$

- Rate of transport of photosynthetic products

$$W_e = 0.5 V_{\text{cmax}}(n_0, T_{\text{upp}}, T_{\text{low}}) \tag{2.3}$$

where $\mathrm{V_{cmax}}$ ($\mathrm{molCO_2\ m^{-2}\,s^{-1}}$; Eq. 2.5) is the maximum rate of carboxylation of Rubisco, $c_i$ (Pa; Eq. 2.4) is the leaf internal $CO_2$ partial pressure, $O_a$ (Pa) is the partial pressure of atmospheric oxygen, $K_c$ and $K_o$ (Pa) are Michaelis-Menten constants for $CO_2$ and $O_2$ respectively, $c_*$ (Pa) is the $CO_2$ compensation point in the absense of mitochondrial respiration, $\omega$ is the leaf scattering coefficient for PAR, and $I_r$ is the incident photosynthetically active radiation (PAR, $\mathrm{mol\,m^{-2}\,s^{-1}}$).

The internal $CO_2$ concentration $c_\mathrm{i}$ is assumed to be dependent on the external $CO_2$ concentration $c_\mathrm{a}$ and the atmospheric humidity deficit $dq$ [Cox et al., 1998] via the equation

$$\frac{c_i - c_*}{c_a - c_*} = f_0 \left( 1 - \frac{dq}{dq_c} \right), \tag{2.4}$$

where $c_*$ is the $CO_2$ compensation point, and $f_0$ and $dq_c$ are parameters that are calibrated in this study. Rearranging this equation, $c_i$ can be written as a function of these two parameters.

$\mathrm{V_{cmax}}$ is given by the following equation dependent on temperature $T$:

$$V_{\text{cmax}} = \frac{n_0 n_e 2^{(T-25)/10}}{\left[1 + e^{0.3(T - T_{\text{upp}})}\right] \left[1 + e^{0.3(T_{\text{low}} - T)}\right]}. \tag{2.5}$$

where $n_e = 8 \times 10^{-4}$ $\mathrm{molCO_2\,m^{-2}\,s^{-1}\ kg\,C\ (kg\,N)^{-1}}$. To account for soil moisture stress, the potential leaf photosynthesis $A_p$ is multiplied by a soil water factor $\beta$ to get the leaf photosynthesis $A$ [Cox et al., 1998],

$$A = A_p \beta \tag{2.6}$$

$\beta$ is the dimensionless moisture stress factor, which is related to the mean soil moisture concentration in the root zone ($\theta$), and the critical and wilting point concentrations ($\theta_c$

and $\theta_w$) as follows:

$$\beta = \begin{cases} 1 & \text{for } \theta > \theta_c \\ \frac{\theta-\theta_w}{\theta_c-\theta_w} & \text{for } \theta_w < \theta \leq \theta_c \\ 0 & \text{for } \theta \leq \theta_w \end{cases}. \tag{2.7}$$

$\beta$ is also used to calculate the transpiration $E'$. Each soil layer $k$, the flux extracted is $e_k^0 E$ where

$$e_k^0 = \frac{r_k \beta_k}{\sum_k r_k \beta_k} \tag{2.8}$$

with $r_k$ denoting the fraction of roots in soil layer $k$ extending from depth $z_{k1}$ to $z_k$

$$r_k = \frac{\exp(-2\frac{z_{k-1}}{d_r}) - \exp(-2\frac{z_k}{d_r})}{1 - \exp(-2\frac{z_k}{d_r})}. \tag{2.9}$$

The plant root depth $d_r$ is a parameter optimised in this thesis.

The stomatal conductance for water vapour $g$ is diagnosed in JULES from the leaf-level photosynthesis $A$ and the internal and external $CO_2$ concentrations:

$$g = 1.6\frac{A}{c_a - c_i}. \tag{2.10}$$

The factor of 1.6 converts the stomatal conductance for $CO_2$ into a stomatal conductance for water vapour.

The photosynthesis model used in JULES is based on scaling up observed processes at the leaf scale to represent the canopy. The scaling to canopy level can be done in several ways. In this study the simple 'big leaf' approach was adopted [Clark et al., 2011], although optimisations can also be carried out for more complex canopy radiation options [Mercado et al., 2009].

In the big leaf approach, incident radiation decreases through the canopy following Beer's law [Monsi and Saeki, 1953]:

$$I_c = I_0 e^{-kL_c} \tag{2.11}$$

where $I_c$ is irradiance beneath the canopy, $I_0$ irradiance at the top of the canopy, $k$ is a light extinction coefficient and $L_c$ is the canopy leaf area index. Leaf-level photosynthesis, which is assumed to vary proportionally with the vertical distribution of irradiance [Sellers, 1985], can be expressed in a similar manner

$$A_1 = A_0 e^{-kL_c} \tag{2.12}$$

where $A_0$ denotes the photosynthesis at the top of the canopy. The canopy photosynthesis is then calculated as the integral of leaf-level photosynthesis over the entire canopy leaf area index

$$A_c = \int_0^{L_c} A_1 dL = \frac{A_0}{k}\left(1 - e^{-kL_c}\right) \tag{2.13}$$

The canopy-level conductance is expressed in a similar manner.

Finally, for rainfall rate $R$, the change in canopy water content is calculated as

$$\frac{\partial C}{\partial t} = R - T_F(C_m) \tag{2.14}$$

where $T_F$ is the through-fall, which depends on $C_m$, the maximum canopy water that can be held by vegetation. This in turn is calculated by

$$C_m = C_0 + \frac{\delta c}{\delta L} L \tag{2.15}$$

where $C_0$ is the interception by leaf-less vegetation, $L$ is leaf area index (LAI) and $\frac{\delta c}{\delta L}$ is the rate of change of water holding capacity with LAI.

The default values and prescribed ranges for each of the parameters used in this study are shown in Table 2.2. The lower and upper bounds set for each parameter are elicited from expert opinion.

| | BT | NT | C3 | C4 | Sh | Lower bound | Upper bound |
|---|---|---|---|---|---|---|---|
| $n_0$ | 0.046 | 0.033 | 0.073 | 0.06 | 0.06 | 0.001 | 0.2 |
| $\alpha$ | 0.08 | 0.08 | 0.12 | 0.06 | 0.08 | 0.001 | 0.999 |
| $f_0$ | 0.875 | 0.875 | 0.9 | 0.8 | 0.9 | 0.5 | 0.99 |
| $T_{\text{low}}$ | 0 | −10 | 0 | 13 | 0 | −50 | 40 |
| $T_{\text{upp}}$ | 36 | 26 | 36 | 45 | 36 | 25 | 50 |
| $d_{\text{r}}$ | 3 | 1 | 0.5 | 0.5 | 0.5 | 0.1 | 4 |
| $\frac{\delta c}{\delta L}$ | 0.05 | 0.05 | 0.05 | 0.05 | 0.05 | 0.001 | 0.1 |
| $dq_c$ | 0.09 | 0.06 | 0.1 | 0.075 | 0.1 | 0.001 | 0.2 |

Table 2.2.: PFT-specific JULES parameters optimised in this study (Table 2.1). The prior values and ranges for each PFT are given.

The eight parameters optimised were selected prior to this study and are part of a fix subset integrated in the adJULES system. Increasing the subset of possible parameters to choose from involves regenerating the adjoint code. As discussed in Sect. 2.1.1, generating a new adjoint is beyond the scope of this project. Therefore no new parameters could be optimised at this stage. With more time, a sensitivity test would help pick out the best parameters to optimise.

### 2.1.3. Driving data

In order to model the behaviour of vegetation evolving against time, JULES uses the following local driving data at each time-step:

- `Pstar:` Air pressure (Pa)
- `T:` Air temperature (K)
- `q:` Specific humidity ($\mathrm{kg\,kg^{-1}}$)
- `wind:` Wind speed ($\mathrm{m\,s^{-1}}$)
- `Precip:` Precipitation ($\mathrm{kg\,m^{-2}\,s^{-1}}$)
- `SW_down:` Downward shortwave radiation ($\mathrm{W\,m^{-2}}$)
- `LW_down:` Downward longwave radiation ($\mathrm{W\,m^{-2}}$)

Data for each of these variables are obtained from half-hourly gap-filled meteorological data measured at *in situ* field sites. The driving data are also used to spin up the model from an arbitrary starting point to a steady soil moisture and temperature state.

## 2.2. FluxNet data

Eddy-covariance flux data are used to drive and calibrate the model. These data are part of FluxNet [Baldocchi et al., 2001]. At each FluxNet site there is a tall 'flux' tower that measures the net carbon and water fluxes between the surrounding atmosphere and vegetated canopies. The data are collected at a high temporal resolution (half-hourly) continuously day and night, in some cases over multiple years [Papale, 2012]. The FluxNet database contains more than 500 locations worldwide, and all of the data are processed in a consistent manner using standard methodologies including correction and gap-filling [Papale et al., 2006].

### 2.2.1. The eddy-covariance technique

#### The mathematics of eddy-covariance

Air flow can be imagined as a horizontal flow of numerous rotating eddies, all having horizontal and vertical components [Burba and Anderson, 2010]. The vertical movement of the components can be measured from the tower. As the wind passes a tower, the tower measures the covariance between the concentration of $CO_2$ (or water vapour concentration) and the vertical wind component of each eddy [Baldocchi et al., 2001].

In turbulent flow, vertical flux is equal to the mean time of the product of air density $\rho_a$ ($\mathrm{mol\,m^{-3}}$), vertical wind speed $w$ ($\mathrm{m\,s^{-1}}$) and the mixing ratio $\chi$ of the gas of interest

$$F = \overline{\rho_a \chi w}. \tag{2.16}$$

When calculating carbon fluxes, the gas of interest is $CO_2$. When calculating latent heat, the gas of interest is water vapour.

Reynolds decomposition [Reynolds, 1895] is a mathematical technique used to separate a scalar $\phi$ into a sum of its average ($\overline{\phi}$) and fluctuating ($\phi'$) parts. The fluctuations, or 'perturbations', are defined such that their time average equals zero ($\overline{\phi'}$=0).

Using Reynolds decomposition on $F$, each component can be broken down into its mean and fluctuation. By assuming that the air density fluctuations are negligible ($\rho_a' \approx 0$) and that vertical flow is negligible for horizontal homogeneous terrain ($\overline{w} \approx 0$), i.e. there is no divergence or convergence, $F$ can be rewritten as:

$$F = \overline{\rho_a}\,\underbrace{\overline{\chi' w'}}_{eddy-covariance} \tag{2.17}$$

**Limitations**

Before discussing the limitations of the eddy-covariance data, it is important to clarify the difference between *error* and *uncertainty*. Error is the difference between the measurement and truth, whereas uncertainty is the confidence placed on the measurement.

The varying footprints, i.e. the area 'seen' by the flux tower, can be a source of errors and uncertainties. This can affect the data quality particularly if the area is inhomogeneous and patchy [Göckede et al., 2006]. Several errors can also occur due to instrumentation limits such as collection frequency. Most of these problems can be solved by applying correction procedures accordingly [Papale et al., 2006].

Despite the sophisticated instruments and data processing methods, technical and human faults are inevitable. These faults create gaps in the data. Common failures during the collection phase include power breaks due to solar panels and damaged instruments due to animals or lightning [Papale, 2012]. Gaps are also created in the data during the data quality control phase; measurements not acquired during the ideal conditions are filtered out. Falge et al. [2001] estimated that 35% of data are missing or rejected (based on a 19 site experiment) whilst Papale et al. [2006] suggested between 20-60% of the data were rejected based on quality filters alone.

The presence of gaps in the time-series is not necessarily a problem in itself. Due to the high temporal resolution of the measurements, there can exist many similar observations found in similar conditions. The problem occurs when the gaps do not happen randomly but systematically, e.g. over large timescales such as solar panels powering down in winter. Gap filling algorithms exist to cover such periods. Extensive information on these can be

found in Falge et al. [2001], Moffat et al. [2007], and Papale [2012].

There are three other known sources of errors in eddy-covariance. First, the underestimation of night time fluxes is a recurrent problem identified by Goulden et al. [1996]. Eddy flux measurements can underestimate the net ecosystem exchange during periods with low turbulence and therefore limited air mixing. This error occurs at night when there is a net emission of $CO_2$. Therefore, the ecosystem respiration is often underestimated and the carbon assimilation overestimated [Moncrieff et al., 1996]. Second, is the problem of energy closure. The incoming radiant heat does not equal the sum of the outgoing latent and sensible heat fluxes. These two errors are called systemic errors. They are constant but unknown.

The last source of error is due to random measurement errors in the flux data. These include errors due to the stochastic nature of turbulence, the variations in the direction of the wind, and instrumental error. There have been different approaches used to assess these errors. Hollinger and Richardson [2005] compare the measurements from two towers with the same footprint, and Richardson et al. [2006] compare measurements made on successive days from the same tower under the same environmental conditions. The most common method is the model residual approach which uses the difference between the data and the outputs from a highly tuned empirical model [Richardson et al., 2008; Stauch et al., 2008; Lasslop et al., 2008]. The model error is assumed to be negligible and therefore the model residual can be attributed almost entirely to random measurement error [Moffat et al., 2007].

### Partitioning NEE into GPP and $R_{eco}$

The carbon flux between the canopy and the atmosphere (NEE) is defined as the difference between the the respiration of the ecosystem ($R_{\text{eco}}$) and the carbon assimilated by photosynthesis (GPP). The measured NEE therefore does not reveal the actual value of either flux. Algorithms have been developed to partition NEE into these two components.

The data used in this thesis are partitioned using the algorithm described in Reichstein et al. (2005). This proposed methodology uses two key ideas. First, it exploits the fact that photosynthesis does not occur at night, i.e. GPP is 0 and therefore NEE = $R_{\text{eco}}$. Second, it assumes that the temperature sensitivity of $R_{\text{eco}}$ follows the exponential regression model laid out in Lloyd & Taylor (1994):

$$R_{\text{eco}} = R_{\text{ref}} \cdot \exp\left[E_0 \left(\frac{1}{T_{\text{ref}} - T_0} - \frac{1}{T - T_0}\right)\right], \tag{2.18}$$

where $T_{\text{ref}} = 10°\text{C}$ and $T_0 = -46.02°\text{C}$, and $E_0$ determines the temperature sensitivity. To calculate $R_{\text{ref}}$ and $E_0$, $R_{\text{eco}}$ is calibrated using night-time NEE over a 15 day window, and then extrapolated during the day-time. Finally, GPP is calculate from the estimated $R_{\text{eco}}$ and measured NEE.

It is important to remember that $R_{eco}$ and GPP are calculated by a model and therefore subject to its assumptions. The values are derived from NEE measurements and hence cannot be treated as independent.

### 2.2.2. FluxNet sites used in this thesis

Data from 160 sites were made available for this study by M. Groenendijk. These sites were selected based on data availability; sites with missing input variables or data gaps of more than 50 % during the growing season were omitted.

Since the FluxNet database does not distinguish between the different types of grasslands, using Met Office ancillary files, the grasslands were partitioned into C3 grasses and C4 grasses according to fractional cover. In the case of C3 grasses, sites were picked only when the fractional cover was over 60 %. Since the C4 grasses are under-represented in the FluxNet database, this boundary was lowered to include all sites where C4 grass was the dominant PFT. Crops were not included in either grass class, as discussed in section 2.1.1.

In all, one year of FluxNet data is used for each site considered in this study at the calibration stage. Where multiple years are available, the most complete year was chosen. For each site the model is spun up to a steady soil moisture and temperature state. Where possible, the 2 years of data preceding the year of comparison were applied repeatedly in the spin up. Where this was not possible, the first year of data was repeatedly applied. Only sites with at least 2 years of data are used in this study, so that the spin-up year is different from the experiment year. In each case, the model was spun up for at least 50 years. For deciduous sites and crop sites, leaf area index values are taken from MODIS data for the appropriate year. Where possible, a second year of FluxNet data was spun up to be used at the evaluation stage of this study. This second year was chosen to be the second most complete year when more than 1 year was available.

After filtering out sites without a dominant PFT or without at least 2 years of data, 81 where left. The sites used in each of the PFT classes are described in Table A.1 and their locations shown in Fig. 2.2. From Fig. 2.2 it is clear that the FluxNet sites used are concentrated in the northern hemisphere, notably in Europe and North America.

To constrain photosynthetic parameters described in section 2.1.2, this study uses the gross primary productivity (GPP) and latent heat (LE) fluxes calculated at each of these sites. As described in section 2.2.1, GPP is one of the products partitioned from the NEE carbon flux, the other being $R_{eco}$ [Reichstein et al., 2005]. The $R_{eco}$ flux is sensitive to soil parameters which are not part of the parameters used to define the PFTs. Since one of the aims of this thesis is to find new generic parameters for each PFT, the soil parameters were not chosen as part of the study and, therefore, the respiration flux was not used for calibration. Even though GPP data are model-derived estimates, which could introduce an additional uncertainty into the results, they are separate from the respiration flux. This is why the data stream is chosen instead of the NEE flux.

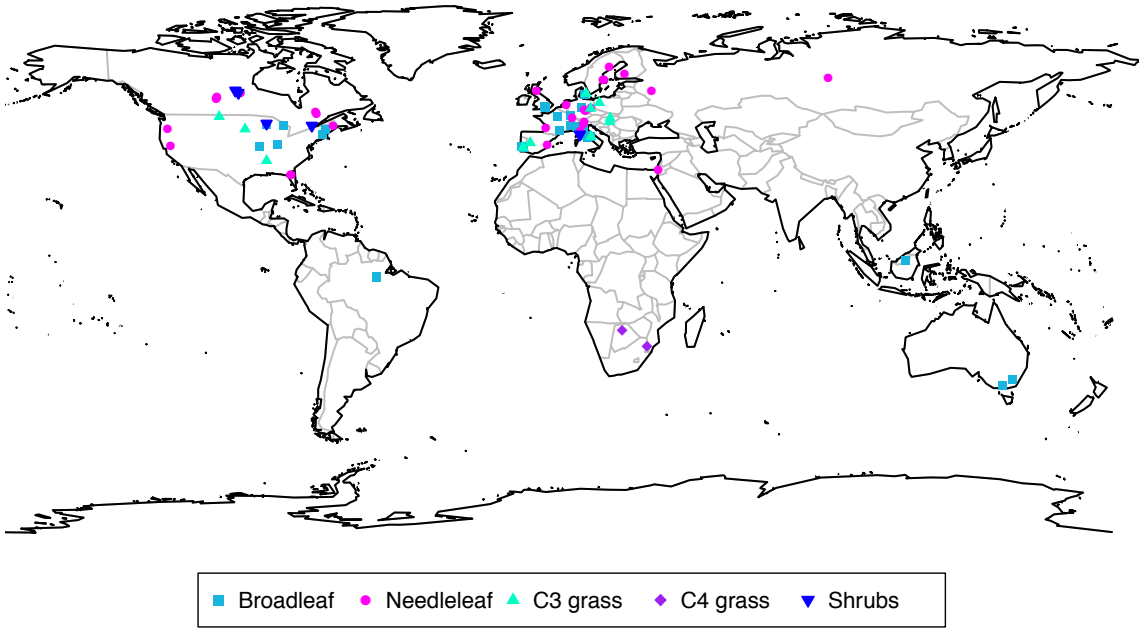



Figure 2.2.: The distribution of the FluxNet sites used in this thesis, broken down by PFT.

The choice of fluxes against which to calibrate will affect the results. Due to model structural errors, calibration against two particular observables could cause model simulations of other fluxes (not used in the tuning) to become worse [Gupta et al., 1999].

### 2.2.3. Other site data used

In an attempt to run the experiments as closely as possible to a standard JULES run, input fields of vegetation structure and soil type were drawn from the UK Met Office ancillary files used in the HadGEM2 configurations. The LAI seasonal cycle used is derived from a MODIS product [Myneni et al., 2002] from Boston University. The values taken for each of the experiment sites correspond to the closest grid point at which data are available. This could lead to inconsistencies between the actual vegetation at a given site and the vegetation structure and soil type used in the model.

The decision to prescribe LAI dated from Luke [2011] when adJULES was set up the calibrate the FluxNet sites from Blyth et al. [2011]. However, since LAI can also be directly calculated by JULES, this would be more desirable for future experiments. The use of prognostic LAI would solve the mismatch in spatial scale between flux towers and MODIS.

### 2.2.4. Example of a JULES run

Figure 2.3 shows the JULES modelled output for gross primary productivity (GPP), the carbon flux used in this project. This run uses driving data from a broadleaf site in Denmark (DK-Sor). As mentioned above, this run is conducted at a single point.

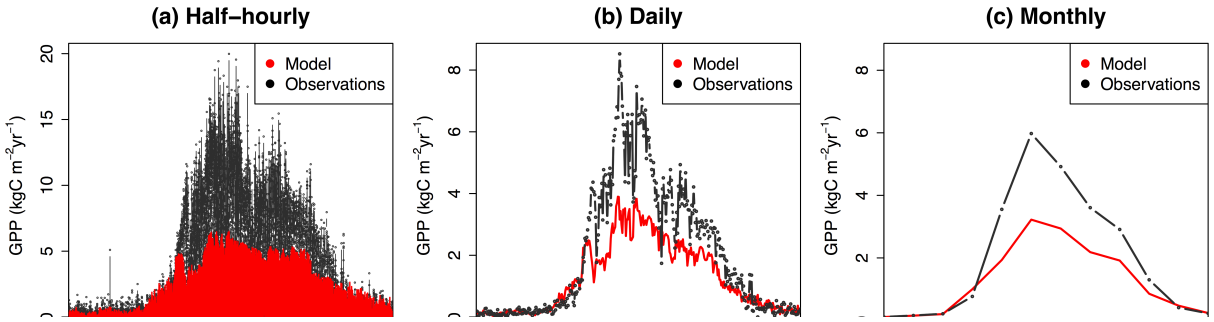



Figure 2.3.: A default JULES run for GPP at the DK-Sor site. One year is shown at different frequencies: half-hourly (a), daily (b) and monthly (c). The modelled time-series (red) is plotted against observations from that site (black).

The JULES model outputs half-hourly data (Fig. 2.3a). Figures 2.3b,c show this first run averaged over different time-scales. The observations shown in each plot are averaged in the same manner. These observations are eddy-covariance flux data taken from the DK-Sor site (see section 2.2). In all cases, the modelled GPP is seen to be underestimated compared to the observations. Due to known structural problems in the photosynthesis model, this thesis does not try to fit subdaily cycles but will focus on fitting the seasonal cycle.

For $m$ time points, 365 for daily data and 12 for monthly data, the modelled time-series can be saved in a vector $\mathbf{m}_t$ of length $m$. A similar observation vector $\mathbf{o}_t$ can also be constructed. This notation is used for the rest of this thesis.

## 2.3. Data Assimilation

So far, the land-surface model has been introduced, along with the parameters to be optimised, and the observational data against which to calibrate. This next section concerns different optimisation techniques and applications.

### 2.3.1. Terminology

To improve the performance of a model, two components can be optimised. Either the values of unknown parameters (*parameter estimation*) or the predictions of the model according to a given data set (*state estimation*). This is achieved by trying to find an 'optimal match' between the model and the observations by varying the properties of the model [Peng et al., 2011]. This process is also sometimes referred to as *calibration*.

The term *Data Assimilation* is commonly used to describe the process of using observations to refine the initial state within a numerical representation of a system [Bouttier and Courtier, 2002]. This is most obviously the case for weather forecasting, in which the temperature, humidity and wind fields define the initial state. However, data assimilation techniques have also been used for parameter estimation, for example in hydrological

models (Madsen [2003], Liu and Gupta [2007]), and carbon cycle data assimilation systems (CCDAS; Rayner et al. [2005], Kaminski et al. [2013]). In parameter optimisation by data assimilation, the internal parameters of a model take on the role of the dynamical state variables in initial state estimation by data assimilation. Nevertheless, the underlying techniques (e.g. of defining a model adjoint and minimising the error in the fit to data), are very similar in these two applications of data assimilation. This study is certainly not the first to define parameter estimation of this form as data assimilation (Braswell et al. [2005], Stöckli et al. [2008], Verbeeck et al. [2011], Kuppel et al. [2012], Hararuk et al. [2014]).

Another term in the literature is *Model-data fusion*. It encompasses both data assimilation (in its classic definition) and inversion techniques (i.e. using observed properties to constrain scientific processes) [Peng et al., 2011].

Finally, *optimisation* refers to the branch of mathematics concerned with minimising (or maximising) an objective function. In this project, the objective function is a cost function based on the differences between observed and modelled fluxes.

### 2.3.2. Data assimilation methods

Data assimilation allows for integrating multiple types of data, while making allowance for associated uncertainties and including prior knowledge. The optimisation techniques used in data assimilation fall into two categories: batched and sequential.

#### Notation

Taking the common notation from Bouttier and Courtier [2002], the dimension of the state space of the system is denoted by $n$ and the dimension of the observation space by $m$:

- $\mathbf{x}$ state vector of size $n$ that describes the state of the forecast model
- $\mathbf{x}_t$ the true state of the system
- $\mathbf{x}_b$ the background state, an initial guess
- $\mathbf{x}_a$ the analysis which is an estimate of $\mathbf{x_t}$
- $\mathbf{x}_f$ the forcast
- $\mathbf{y}$ observation vector of size $m$ (denoted $\mathbf{o}$ in the later parts of this work)
- $\mathbf{H}$ observation operator which maps $\mathbf{x}$ from model space to observation space (denoted $\mathbf{m}$ in the later parts of this work)
- $\mathbf{B}$ the background error covariance matrix
- $\mathbf{R}$ the observation error covariance matrix

In parameter estimation, the state vector $\mathbf{x}$ becomes a vector of parameters, with $n$ denoting the number of parameters used in the optimisation.

### Sequential methods

Sequential methods process the data one by one. Initialised by the background state $\mathbf{x}_b$, these methods work as a two-part cycle. First the model is used to evolve a previous state estimate (analysis) to find a forecast. Then, once an observation becomes available, the observation is used to update the forecast state to give an improved analysis.

The most famous example of sequential data assimilation is the Kalman filter (Williams et al., 2009). The Kalman gain $\mathbf{K}$ is an evolving weight matrix dependent on model state errors and observation errors. It is used to adjust the model forecast $\mathbf{x}_f$ to an analysis $\mathbf{x}_a$:

$$\mathbf{x}_a = \mathbf{x}_f + \mathbf{K}(\mathbf{y} - \mathbf{H}(\mathbf{x}_f)) \tag{2.19}$$

The Ensemble Kalman filter is an extension of the Kalman filter used for non-linear problems [Raupach et al., 2005]. The idea behind it is to use a statistical sample (ensemble) of state estimates instead of a single estimate. For example, Quaife et al. (2008) uses the Ensemble Kalman filter to estimate model parameters.

### Batch methods

Unlike sequential methods, batch methods process all the data at once. The model is run forwards with the background state $\mathbf{x}_b$ for the whole of the assimilation window, before being confronted by observations.

In batched optimisation, a cost function $J$, also known as an objective function, is minimised. In most examples, the cost calculates the difference between model outputs $H(x)$ and observations $\mathbf{y}$ as well as the mismatch between background $\mathbf{x}_b$ and optimal state $\mathbf{x}$ [Williams et al., 2009]:

$$J(\mathbf{x}) = \frac{1}{2}\left[(\mathbf{y} - \mathbf{H}(\mathbf{x}))^T\mathbf{R}^{-1}(\mathbf{y} - \mathbf{H}(\mathbf{x})) + (\mathbf{x} - \mathbf{x}_b)^T\mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}_b)\right]. \tag{2.20}$$

Batch methods can be further broken down into two categories; gradient-based and 'global search' methods. These categories refer to the approach used to minimise the cost function.

Gradient descent algorithms find a direction in parameter space along which to minimise the cost function. These methods are deterministic and highly efficient, often converging with relatively few iterations. However, they may discover local rather that global minima [Williams et al., 2009]. Local minima refer to turning points found in some neighbourhood that need not be a global minimum (Fig. 2.4). A local minimum is found either when the gradient is zero or at the boundaries of the domain. These are both termination criteria for gradient descent algorithms, however, they do not guarantee the lowest point. A global minimum is the lowest value the function $J(\mathbf{x})$ can take over the whole $\mathbf{x}$ domain.

Posterior uncertainties for gradient descent algorithms are calculated using model output sensitivity or in some cases the second derivative of the cost function with respect to

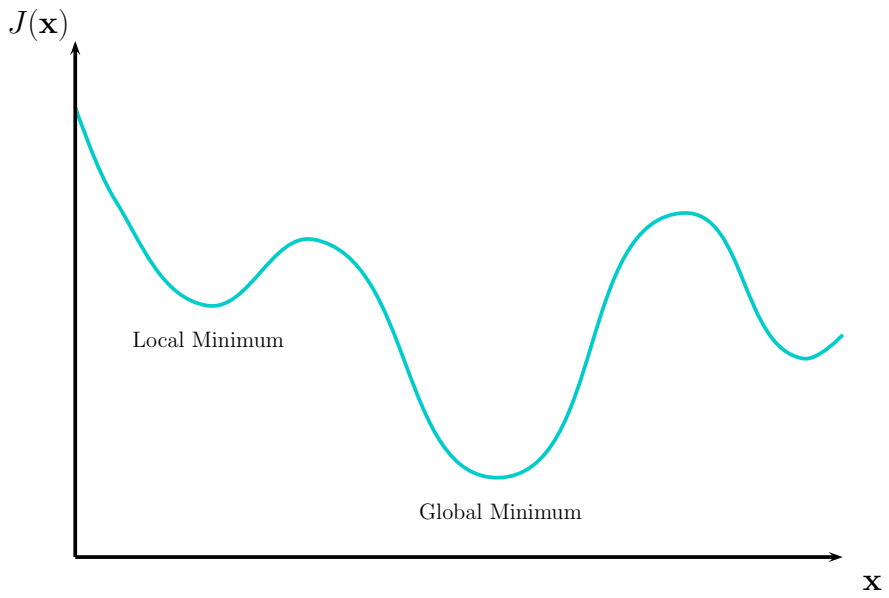



Figure 2.4.: Visualisation of global minimum vs local minima in one-dimension.

the parameters (Hessian). For an example of a gradient descent algorithm used for parameter optimisation, see Santaren et al. [2007], where the cost function is differentiated analytically in order to find the descent direction (see Sect. 2.3.4).

Global search algorithms are often based on a random generator e.g. genetic algorithms, or Markov Chain Monte Carlo, as used in a study by Braswell et al. [2005]. They are more likely to find a global minimum and are well adapted to highly non-linear models. However, they tend to have a high computational cost [Peng et al., 2011].

In this study, the adjoint model is used to find the descent direction (further described in Sect. 2.3.4).

A combination of batch methods is used in Vrugt et al. [2005]. The study starts with a global search to find the right area of parameter space, followed by a gradient-based method to find the minimum.

### 2.3.3. Data assimilation with LSMs

There have been many studies using a large variety of data assimilation methods to improve land-surface models. The optimisation techniques used range from simple *ad hoc* parameter tuning to rigorous data assimilation frameworks, and cover various LSMs, to derive vectors of parameters that improve model–data fit significantly (e.g. Wang et al., 2001, 2007; Reichstein et al., 2003; Knorr and Kattge, 2005; Raupach et al., 2005; Santaren et al., 2007; Thum et al., 2008; Williams et al., 2009; Peng et al., 2011). However, very few studies have quantified the errors in the observations, model parameters, and model structure systematically and consistently [Wang et al., 2009].

Peng et al. [2011], Raupach et al. [2005], Wang et al. [2009] and MacBean et al. [2016] all provide comprehensive summaries of these studies outlining the advantages and disadvantages of the methods used. Peng et al. [2011] clearly lays out different methodologies with their key features and applications to LSMs in a concise table. Williams et al. [2009] also give an overview, describing a clear framework for using FluxNet data to improve land-surface models, and a breakdown of data assimilation methods. They also give an explanation of FluxNet sites along with the characteristics and limitations of the data available.

**Parameter estimation with LSMs**

In numerical weather prediction, data assimilation has predominately been used to optimise the state whilst keeping the parameters fixed. This is because the physics are mostly known and well understood. In terrestrial carbon cycle models however, where most of the equations are unknown, finding the correct set of parameters is more pertinent [Luo et al., 2015]. Due to their easy implementation, Monte Carlo Markov Chain (MCMC) methods have dominated the field. Smaller ecosystem models are much less computationally expensive than numerical weather prediction models and so many ensemble runs can be performed. For example, Rosolem et al. [2012] uses a multi-operator genetic algorithm on the Simple Biosphere 3 model and data assimilation experiments on the Data Assimilation Linked Ecosystem Carbon (DALEC) model [Williams et al., 2005] have predominately been sequential and MCMC.

For larger scale global models, variational methods are much more efficient. Key examples of such implementations are the ORganizing Carbon and Hydrology In Dynamic EcosystEms model (ORCHIDEE) [Krinner et al., 2005] and the Biosphere Energy Transfer HYdrology scheme (BETHY) as part of a Carbon Cycle Data Assimilation System (CCDAS) (Rayner et al. [2005]; Kaminski et al. [2013]).

The majority of LSMs group vegetation into a small number of PFTs. Model parameters are assumed to be generic over each PFT. Through different optimisation techniques, some studies have tried to assess the robustness of PFT-specific parameters (e.g. Kuppel et al., 2014). Medvigy et al. [2009] and Verbeeck et al. [2011] both showed that parameters derived at one site can perform well on a similar site and over the surrounding region (Medvigy and Moorcroft, 2011). However, a contradictory study by Groenendijk et al. [2011] found that there was cross-site parameter variability after optimisation within the PFT groupings.

Many of these studies calibrate the model at individual measurement sites. Given the small spatial footprint of each flux tower, this can often result in over-tuning. This over-tuning may occur when a single site does not represent the full range of a plant functional type (PFT), given different tree types, tree ages, and above-ground biomass found at each site. There may be some anomalous plants in the flux tower footprint that are not representative of the PFTs over a broader area. The optimised model parameters are site

specific and often struggle to perform as well when generalised over other sites [Xiao et al., 2011].

In the last few years, there has been a move towards deriving PFT-specific parameters using data from multiple sites, the results of which have been generally positive (e.g. Xiao et al., 2011 and Kuppel et al., 2012). Both of these studies used data from multiple sites in their optimisation (calling it multi-site optimisation) and have commented on the robustness of this technique, showing that the choice of the initial parameter vector had little effect on the optimised values.

Kuppel et al. [2012] compared different approaches for finding generic PFT-specific parameters, such as averaging optimised parameter vectors over PFTs and directly optimising over multiple sites. They found that the latter method was best for finding PFT-specific parameters. The multi-site optimisation procedure was refined in Kuppel et al. [2014], extended to other PFTs, and evaluated at a global scale.

For global modelling, there is a clear need to find generic parameters and associated uncertainties for each PFT by optimising against observations in a reproducible way.

#### Data assimilation with JULES

There have been very few studies of data assimilation on JULES, most of which have used ensemble methods to improve the state of the model. Ghent et al. [2010] use an ensemble Kalman filter (described in section 2.3.2) to improve the simulations of land surface temperature and some preliminary work by Quaife et al. [2014] looks at assimilating satellite data into JULES via a particle filter (also known as sequential Monte Carlo). These methods are relatively easy to implement, but costly when large ensembles are used. Due to the random element of these methods, the results are not always reproducible.

The JULES study most resembling the work performed in this thesis is the unpublished work of Pearson et al. [2009]. Also aiming to optimise carbon cycle parameters, their presentation states that due to the complexity of JULES, for parameter estimation, variational data assimilation wins over monte carlo methods.

Due to the lack of a rigorous and established data assimilation framework, the default parameters in JULES have remained unchanged for many years. When calibrating the model, rather *ad hoc* methods are used with the modeller varying the parameters that he/she believes are most relevant to the model performance. Such model tuning is by its very nature subjective, lacks reproducibility, and is often sub-optimal because the modeller is unable to explore the full feasible parameter space through such a manual technique.

The adJULES system provides this much-needed framework. Using the adjoint method and a gradient-descent method, it has the ability to find minima rapidly across multiple parameters via matrix inversion and has the advantage of reproducibility.

### 2.3.4. The adjoint method

#### Descent algorithm

Most of the theory in this section in taken from Press et al. [2007]. This thesis uses a gradient-based algorithm, a method which uses the derivative of a function to pick a direction in parameter space along which to minimise (or maximise) the function.

One example is *gradient-decent*, also known as *method of steepest descent*. As the name suggests, it uses the gradient to find the direction of steepest descent. The negative gradient acts as a compass pointing downhill. Starting at point $\mathbf{x}_0$, this methods works by iteratively minimising a function along this downhill direction for a step size $\gamma$. In order to use this method, the function needs to be defined and differentiable in a neighbourhood around each $\mathbf{x}_i$.

$$\mathbf{x}_{i+1} = \mathbf{x}_i - \gamma_i \nabla f(\mathbf{x}_i),\ i \geq 0. \tag{2.21}$$

For small enough $\gamma_i$, a sequence $\{\mathbf{x}_i\}$ is constructed with $f(\mathbf{x}_i) > f(\mathbf{x}_{i+1})$, which hopefully converges to a local minimum. The step size $\gamma_i$ is allowed to change between iterations. The choice of $\gamma_i$ is vital to this method. If $\gamma_i$ is too small, the method will be too slow and will need many iterations to converge. If $\gamma_i$ is too big, it might miss the minimum.

Similarly, Newtonian algorithms of descent attempt to construct a sequence $\mathbf{x}_i$ converging towards some value $\mathbf{x}^*$, but here, they also use the second derivative of $f$. In one dimension, the algorithm is derived by taking the second order Taylor expansion of $f$ around $x_i$ where $\delta x$ denotes a small change in $x$.

$$f(x_i + \delta x) \approx f(x_i) + f'(x_i)\delta x + \frac{1}{2} f''(x_i)\delta x^2. \tag{2.22}$$

This expression is then differentiated with respect to $\delta x$ and set to zero. This yields the following relationship

$$\delta x = -f'(x_i)/f''(x_i). \tag{2.23}$$

This in turn gives an iterative sequence that converges to the stationary point $x^*$

$$x_{i+1} = x_i + \delta x = x_i - f'(x_i)/f''(x_i). \tag{2.24}$$

This methods works by fitting a quadratic function around $x_i$ and stepping towards the minimum of that quadratic. By using this curvature information, the method takes a more direct route than the gradient-descent method.

Generalising this iterative scheme to higher dimensions, the gradient, called the Jacobian matrix, is denoted by $\nabla f$ and the second derivative, called the Hessian matrix (or curvature matrix) is denoted by $\mathbf{C}$,

$$\mathbf{x}_{i+1} = \mathbf{x}_i - \mathbf{C}^{-1} \cdot [\nabla f(\mathbf{x}_i)],\ i \geq 0. \tag{2.25}$$

The Hessian is the matrix composed of second partial derivatives, i.e.

$$\mathbf{C}_{jk} = \frac{\partial^2 f}{\partial x_j \partial x_k}, \tag{2.26}$$

where $x_j$ represents the different elements of $\mathbf{x}$.

Finding the inverse of the Hessian, $\mathbf{C}^{-1}$, in high dimensions can be costly. Quasi-Newton methods approximate the Hessian used in the optimisation algorithm at every iterative step $i$. This approximation helps cut computational costs. There are several ways the Hessian can be approximated, of which the BFGS method is the most used (suggested independently by Broyden, Fletcher, Goldfarb, and Shanno, in 1970).

In this thesis, a cost function similar to Eq. 2.20 is minimised iteratively using a gradient-based algorithm called L_BFGS_B (Byrd et al. [1995], optim: R Development Core Team [2015]). This is based on the BFGS quasi-Newton method but is modified to use limited memory, for computational affordability, and box constraints, so each parameter is given an upper and lower bound based on an educated opinion or physical reasoning [Byrd et al., 1995].

### Theory of automatic differentiation

Knowledge of the gradient of the cost function with respect to the model variables is needed as part of the descent algorithm. An efficient way to calculate the gradient of the cost function with respect to many parameters is through an adjoint model [Giering and Kaminski, 2003].

Automatic differentiation is software used to find the derivative of a function specified by a computer program [Naumann, 2011]. Since every computer program can be broken down into a sequence of elementary arithmetic operations (e.g. $+$, -, $\times, \div \dots$) and elementary functions (e.g. $\exp, \log, \sin, \cos \dots$), the chain rule is applied repeatedly to these operations to compute the derivative accurately.

Consider a model as a mapping of input variables onto output variables,

$$f : \mathbb{R}^n \to \mathbb{R}^m.$$

The first derivative of the model (assuming the mapping is differentiable) is the Jacobian matrix. Automatic differentiation allows us to evaluate the Jacobian matrix numerically. There are two 'modes' of automatic differentiation: the forward mode which generates the tangent linear model (similar to a person using the chain rule), and the reverse mode which generates the adjoint model. Both modes generate identical derivatives [Errico, 1997], the choice of forward or reverse mode refers to the order in which the derivatives are computed. The Jacobian matrix can be generated with $n$ sweeps in forwards mode, or $m$ sweeps in the reverse mode.

### Simple example

To illustrate the difference between the two modes, following Berland [2006]'s example, consider the following function:

$$f(x_1, x_2) = x_1 x_2 + \sin(x_1) \tag{2.27}$$

This can be broken down into a sequence of operations (namely $\times, +, \sin$) on work variables $w_i$ as follows:

$$w_1 = x_1 \tag{2.28}$$
$$w_2 = x_2 \tag{2.29}$$
$$w_3 = \sin(w_1) \tag{2.30}$$
$$w_4 = w_1 w_2 = x_1 x_2 \tag{2.31}$$
$$w_5 = w_3 + w_4 = w_1 w_2 + \sin(w_1) = x_1 x_2 + \sin(x_1) \tag{2.32}$$

When differentiating in the forward mode, as shown in Fig. 2.5, the work variables are differentiated sequentially, moving forwards through the sequence of operations and down the figure. To give the different derivatives, the computation is seeded by $\dot{w}_1, \dot{w}_2 \in \{0, 1\}$ with $\dot{w}_1 \neq \dot{w}_2$. When $\dot{w}_1 = 1$, the function is differentiated with respected to $x_1$. Similarly, when $\dot{w}_2 = 1$, the function is differentiated with respected to $x_2$.

The differentiation procedure is as follows:

$$\dot{w}_1, \dot{w}_2 \in \{0, 1\} \tag{2.33}$$
$$\dot{w}_3 = \frac{\partial w_3}{\partial w_1}\dot{w}_1 = \cos(w_1)\dot{w}_1 \tag{2.34}$$
$$\dot{w}_4 = \frac{\partial w_4}{\partial w_1}\dot{w}_1 + \frac{\partial w_4}{\partial w_2}\dot{w}_2 = \dot{w}_1 w_2 + w_1 \dot{w}_2 \tag{2.35}$$
$$\dot{w}_5 = \dot{w}_3 + \dot{w}_4 = \dot{w}_1 w_2 + w_1 \dot{w}_2 + \cos(w_1)\dot{w}_1 \tag{2.36}$$

It is easy to see that seeding with $(\dot{w}_1, \dot{w}_2) = (1, 0)$ gives

$$\frac{\partial f}{\partial x_1} = x_2 + \cos(x_1) \tag{2.37}$$

and similarly seeding with $(\dot{w}_1, \dot{w}_2) = (0, 1)$ gives

$$\frac{\partial f}{\partial x_2} = x_1. \tag{2.38}$$

The reverse mode is shown in Fig. 2.6. This time the operation starts with the complete function and propagates backwards through all dependencies. This reverse pass starts at the end, i.e. with

$$\bar{f} = \frac{df}{df} = 1. \tag{2.39}$$

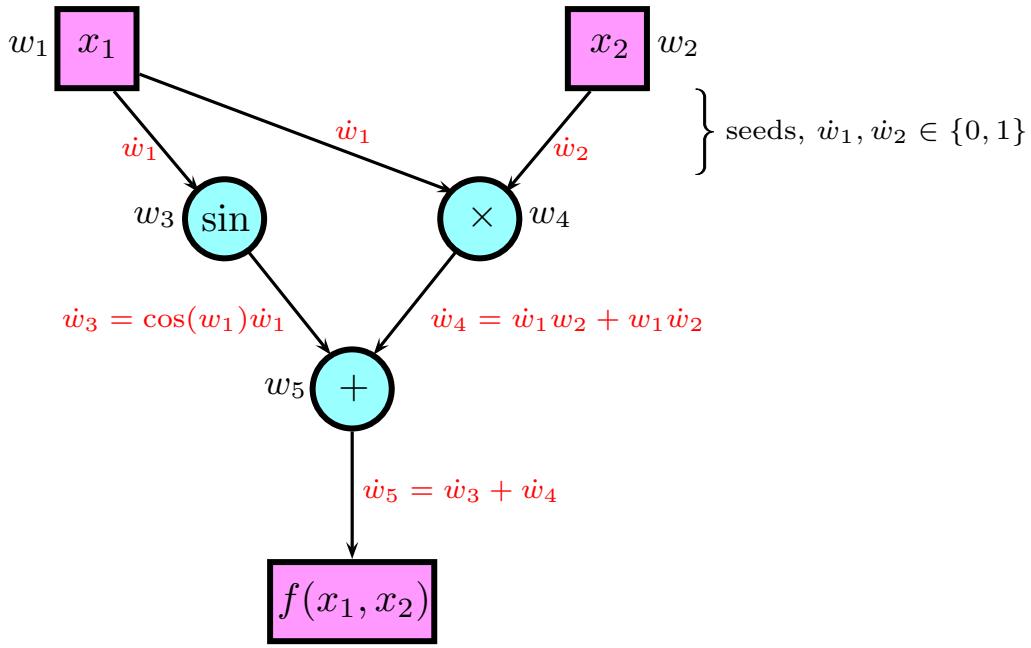



Figure 2.5.: Automatic differentiation in forward mode on $f(x_1, x_2)$, adapted from Berland [2006].

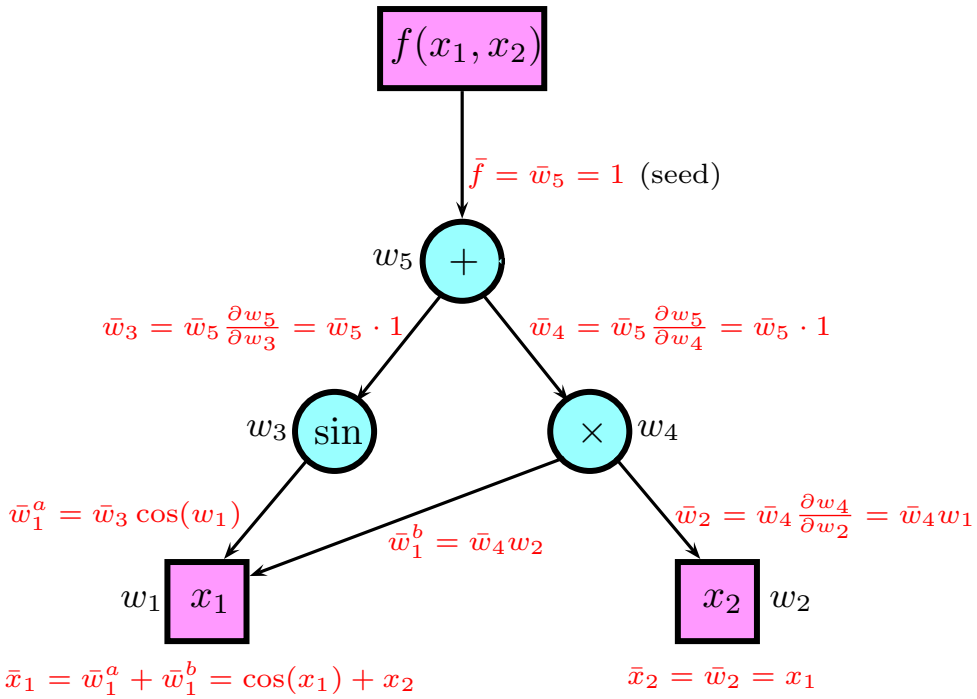



Figure 2.6.: Automatic differentiation in reverse mode on $f(x_1, x_2)$, adapted from Berland [2006].

This expression states that a change in $f$ results in exactly the same change in $f$. This is equivalent to seeding the computation with $\bar{f} = \bar{w}_5 = 1$. Since there is only one output, only one derivation computation is required.

Using the chain rule and the work variables $w_i$ as defined above, the differentiation procedure in reverse mode is as follows:

First, the computation is seeded with 1

$$\bar{w}_5 = \frac{\partial f}{\partial w_5} = 1 \tag{2.40}$$

To calculate the next work variables down the tree, the chain rule is utilised to get $\bar{w}_4 = \frac{\partial f}{\partial w_5} \times \frac{\partial w_5}{\partial w_4}$ and $\bar{w}_3 = \frac{\partial f}{\partial w_5} \times \frac{\partial w_5}{\partial w_2}$. Since $w_5$ is linear depend on $w_3$ and $w_4$

$$\bar{w}_4 = \bar{w}_5 \frac{\partial w_5}{\partial w_4} = \bar{w}_5 \frac{\partial}{\partial w_4}(w_3 + w_4) = \bar{w}_5 \times 1 = \bar{w}_5 = 1 \tag{2.41}$$

$$\bar{w}_3 = \bar{w}_5 \frac{\partial w_5}{\partial w_3} = \bar{w}_5 \frac{\partial}{\partial w_3}(w_3 + w_4) = \bar{w}_5 \times 1 = \bar{w}_5 = 1 \tag{2.42}$$

Using the chain again, $\bar{w}_2 = \frac{\partial f}{\partial w_4} \times \frac{\partial w_4}{\partial w_2}$. Therefore, from the definition of $w_4$ and rules of partial derivatives, $\bar{w}_2$ is calculated as follows

$$\bar{w}_2 = \bar{w}_4 \frac{\partial w_4}{\partial w_2} = \bar{w}_4 \frac{\partial}{\partial w_2}(w_1 w_2) = \bar{w}_4 w_1 = w_1 \tag{2.43}$$

To move down to $w_1$, there are two paths possible since $w_1$ contributes to $f$ via $w_3$ and $w_4$. Both are possibilities are calculated using the chain rule and summed together to get $\bar{w}_1$

$$\bar{w}_1^a = \bar{w}_3 \frac{\partial w_3}{\partial w_1} = \bar{w}_3 \frac{\partial}{\partial w_2}(\sin(w_1)) = \bar{w}_3 \cos(w_1) = \cos(w_1) \tag{2.44}$$

$$\bar{w}_1^b = \bar{w}_4 \frac{\partial w_4}{\partial w_1} = \bar{w}_4 \frac{\partial}{\partial w_2}(w_1 w_2) = \bar{w}_4 w_2 = w_2 \tag{2.45}$$

$$\bar{w}_1 = \bar{w}_1^a + \bar{w}_1^b = \cos(w_1) + w_2 \tag{2.46}$$

This calculation yields

$$\frac{\partial f}{\partial x_1} = \bar{x}_1 = \bar{w}_1^a + \bar{w}_1^b = \cos(x_1) + x_2 \tag{2.47}$$

and

$$\frac{\partial f}{\partial x_1} = \bar{x}_2 = x_1, \tag{2.48}$$

which are identical to the partial derivatives found when differentiating in forwards mode.

In this example, since $f : \mathbb{R}^2 \to \mathbb{R}$, $n = 2$ and $m = 1$. The forward mode took two sweeps, whereas the reverse mode only required one sweep. One sweep in reverse mode is more involved compared to one sweep in forwards mode, requiring six operations compared to three. However, since the forwards mode here requires two sweeps, it is easy to see that given a high enough number of variables $x_i$, i.e. situations when $m \ll n$, the reverse mode

is more efficient.

This study uses automatic differentiation software from FastOpt [FastOpt, 2010]. In the example, the software processes the FORTRAN code for the function:

```
subroutine f(x, y)
real x(2),y
y = x(1) * x(2) + sin(x(1))
end
```

and produces code for the Jacobian:

Forward Mode (Tangent Linear)

```
subroutine f_tl( x, x_tl, y, y_tl )
real x(2),x_tl(2),y,y_tl
y = x(1)*x(2)+sin(x(1))
y_tl = x_tl(2)*x(1)+x_tl(1)*(x(2)+cos(x(1)))
end
```

Reverse Mode (Adjoint)

```
subroutine f_ad( x, x_ad, y, y_ad )
real x(2), x_ad(2), y, y_ad
y = x(1)*x(2)+sin(x(1))
x_ad(2) = x_ad(2)+y_ad*x(1)
x_ad(1) = x_ad(1)+y_ad*(x(2)+cos(x(1)))
y_ad = 0.
end
```

Both algorithms can be used to calculate the derivative of `f(x, y)`. The algorithm generated in forward mode is known as the *tangent linear*, whereas the code generated in reverse mode is called the *adjoint*.

### Creating the adjoint model

For the purposes of parameter estimation, the function in question is the cost function $J : \mathbb{R}^n \to \mathbb{R}$, where $n$ is the number of parameters to optimise.

Calculating $\nabla J$ is most efficient in reverse mode as only one sweep is needed to generate the derivative with respect to all parameters. As discussed in the previous section, the complexity of one sweep in forward mode is proportional to the complexity of the original code. The complexity of one sweep in reverse mode is greater than this, but not as great as $n$-times the complexity of the original code. Therefore, for large enough $n$, the reverse mode is more efficient for calculating $\nabla J$ [Bartholomew-Biggs et al., 2000].

JULES has been differentiated line by line using commercial software from FastOpt [FastOpt, 2010] to create the adjoint model adJULES. In order to differentiate the model, discontinuities in the model caused by step changes in the code (i.e. `IF`, `MAX` and `MIN` statements) needed to be smoothed out [FastOpt, 2010]. It is possible to replace such statements with

smoother alternatives. For example, `IF` statements of the form:

```
IF x > a
      f = b
ELSE
      f = c
```

can be replaced with the following function:

$$f(x) = \frac{1}{1 + e^{-2k(x-a)}} b + \frac{1}{1 + e^{2k(x-a)}} c \tag{2.49}$$

where $k$ is a 'sharpness' constant. It is easy to see that the left-hand term dominates when $x > a$ and similarly, the right-hand term dominates when $x < a$.

For `MAX` and `MIN` functions, also using a sharpness constant $k$, a smooth alternative for the function `g=max(x,y)` is

$$g(x, y; k) = \frac{\ln(\exp(kx) + \exp(ky))}{k} \tag{2.50}$$

and for the function `h=min(x,y)`, a smooth alternative is

$$h(x, y; k) = \frac{-\ln(\exp(-kx) + \exp(-ky))}{k}. \tag{2.51}$$

Smoothing parts of the model code is allowed because generally the world operates in a smooth manner. For example, a tree could be modelled in such a way that when the temperature is below a threshold, it loses all its leaves at once. In reality, the tree would be losing its leaves gradually with decreasing temperature. In practice, the `MAX` and `MIN` statements have been the main causes of discontinuities in JULES [Luke, 2011].

Generating the adjoint to use in the adJULES system is complicated and very costly. As such, the adJULES system has struggled to keep up with the new releases of JULES which occur approximately every six months. In comparison, the ORCHIDEE land-surface model developed in France uses the tangent linear model [Verbeeck et al., 2011]. The developers of ORCHIDEE have had trouble generating the adjoint due to complexity of derivation process. [Chevallier, 2016].

Instead of using an out-dated version of the model, it is possible to do these calculation using finite differences which doesn't require the adjoint code. However, the adjoint method, once implement, does reduce computational time and increases accuracy. The speed of finite differences will be similar forward mode calculation whereas as discussed previously, the reverse mode can be faster, especially for parameter estimation schemes. In terms of accuracy, the adjoint method computes the derivatives exactly (up to machine precision) while finite differences incur truncation errors [Homescu, 2011]. Also the size of the step $h$ needed for finite difference varies with the current value of input parameters, making the

problem of choosing $h$, such that it balances accuracy and stability, a challenging one. The adjoint method on the other hand, is automatic and does not require time spent choosing step-size parameters.

## 2.4. Tools for analysis

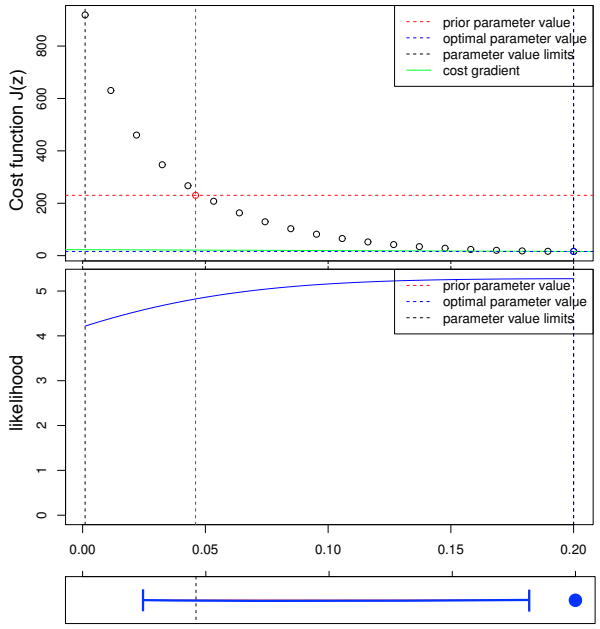
### 2.4.1. Parameter uncertainty

In addition to generating optimal parameter values, it is possible to estimate the uncertainty associated with each parameter. This is done by using the Hessian, described in Eq. (2.26). The Hessian is the second derivative of the cost function calculated using the adjoint code. It is evaluated at the optimal parameter value, yields information about the curvature of the cost function at the local minimum. A "sharp" cost function, where the cost function is steep either side of the optimal parameter value, indicates lower parameter uncertainty. This can also be interpreted as meaning that a small deviation from the optimal parameter value yields a large increase in cost. Conversely, a "flat" cost function indicates higher parameter uncertainty, or little change in cost caused by deviation from the optimal parameter value.

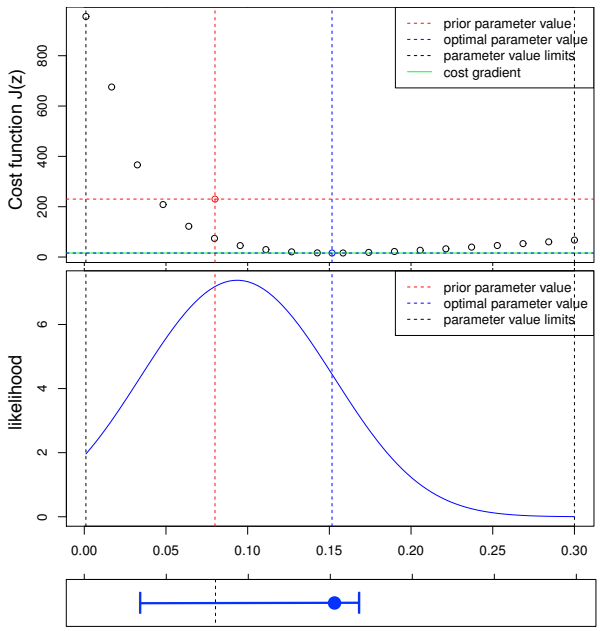
In order to generate statistics associated with the curvature of the cost function, the Hessian is used to generate samples from the posterior distribution. This is a truncated multivariate normal distribution [Genz et al., 2015] because of the box constraints placed on the prior. These box constraints come from the upper and lower bounds allowed for each parameter as specified in Table 2.2. Using Gibbs sampling [Geman and Geman, 1984], an ensemble of plausible parameter vectors is generated from this distribution, for a statistically satisfactory match between observations and modelled time series. The multivariate normal parameter distribution allows marginal density plots to be generated for each parameter. To visualise the parameter uncertainties, error bars are used to represent the 80 % quantile range (10th to 90th percentile) for each optimal parameter.

Fig. 2.7 illustrates this method. Calculated in a preliminary experiment, this figure shows four of the optimised parameters at the Harvard Forest FluxNet site (US-Ha1). In each case, the top plot is a 'slice' through parameter space, the red line showing the prior value and the blue the optimal value. To generate this 'slice', the cost function is evaluated at different increments of parameter range. The zero derivative at the optimal parameter confirms a local minimum.

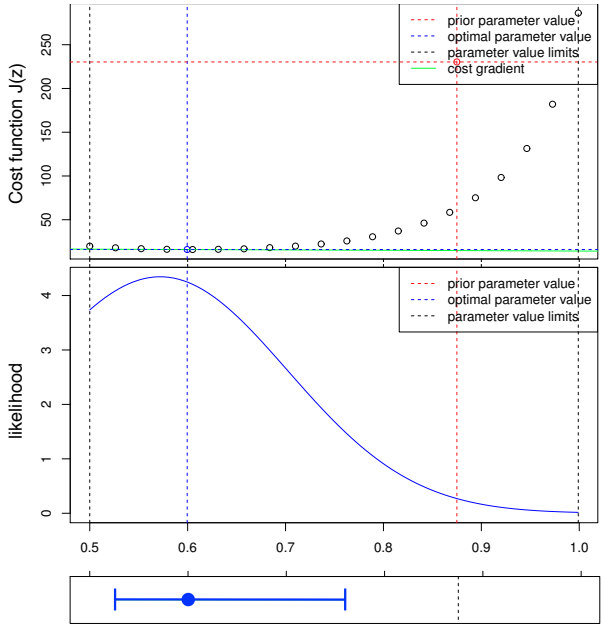
The plot below the slice is a marginal density plot taken from the multivariate truncated normal distribution. Note that the curvature of the marginal density plot does not always follow the slice. This is because the second derivative is calculated locally at the optimal parameter value instead of over the whole range shown by the slice. The marginal density plot shows one dimension of an $n$-dimensional multivariate normal distribution, and as such, the optimal values, shown in blue, are not always found at the peak of the distributions.



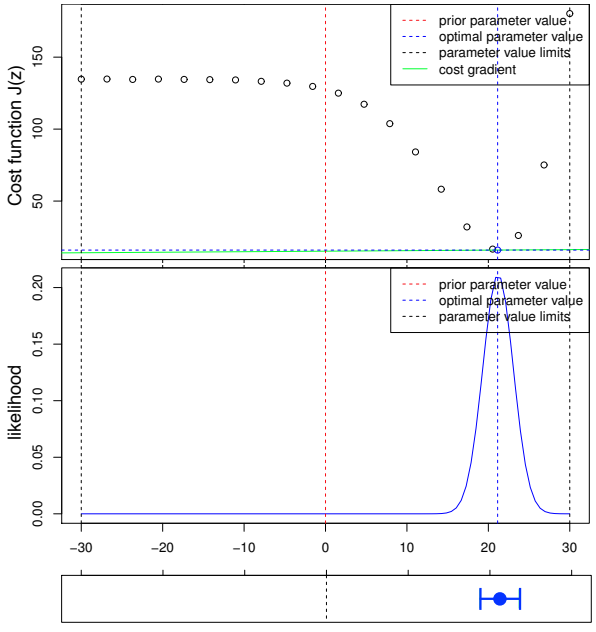

$n_0$




$\alpha$




$f_0$




$T_{\text{low}}$

Figure 2.7.: Parameter distributions at Harvard Forest. The first row for each parameter shows a 'slice' through parameter space. Prior value (red) and the optimised value (blue) are shown. The second row shows the posterior distribution for that parameter and the third rows shows the 80 % quantile range

The error bars below the plots show the 80% quantile range for each the optimal values. The prior parameter is shown as a dashed line, and the box spans the prescribed range over which the parameter is allowed to vary in the adJULES system.

In this example, the $n_0$ optimal value is outside the 80% quantile range. The cost function is 'flat' for this parameter, indicating high parameter uncertainty. Changing this optimal parameter value will have little effect on the cost; $n_0$ is unconstrained and can take any value within our bounds. In this case, it has hit the upper limit. The $\alpha$ and $f_0$ parameters in this example are better constrained. The $T_{low}$ parameter is the most constrained parameter. It has a 'sharp' cost function and therefore a much lower uncertainty associated with it.

The plausible parameter ensemble can also be plotted in parameter space to show the correlation between parameter values, as shown in the two-dimension marginal density plot in Fig. 2.8. The densely populated areas on the plot show likely or plausible parameter combinations. The sparsely populated or unpopulated areas on the plot show unlikely or implausible parameter combinations. In this figure, it is also possible to identity a positive correlation structure between the two parameters.

When considering these two-dimensional marginal density plots, as in one-dimension, it is important to remember that they represent only two dimensions of a high-dimensional multivariate normal distribution which is truncated. Consequently, the optimal parameter values (which are modes of the full high-dimensional distribution) may not coincide with modes of the one- and two-dimensional marginal distributions.

### 2.4.2. Metrics of model-data fit

#### Fractional error

To measure the performance of the model run using different parameter vectors, the fraction of variance unexplained $\epsilon^2$ is used to define the fractional error $\epsilon$. This metric was chosen to quantify not only the model-data fit using different parameter vectors at each site, but also to show how each site performed relative to others.

Given a parameter vector, a modelled time-series $\mathbf{m}_{i,t}$ with $m$ data points is generated using JULES, where $i$ denotes one of the observable data streams (in this case LE and GPP). This is compared to an observed time-series $\mathbf{o}_{i,t}$. For each data stream $i$, the fraction of variance unexplained by the model is

$$\epsilon_i^2 = \frac{\sum_{t=1}^{m}(\mathbf{o}_{i,t} - \mathbf{m}_{i,t})^2}{\sum_{t=1}^{m}(\mathbf{o}_{i,t} - \bar{\mathbf{o}}_i)^2}, \text{ where } \bar{\mathbf{o}}_i = \frac{1}{m}\sum_{t=1}^{m}\mathbf{o}_{i,t}. \tag{2.52}$$

It follows that the mean fraction of variance unexplained across two data streams,

$$\epsilon^2 = \frac{\epsilon_1^2 + \epsilon_2^2}{2}, \tag{2.53}$$

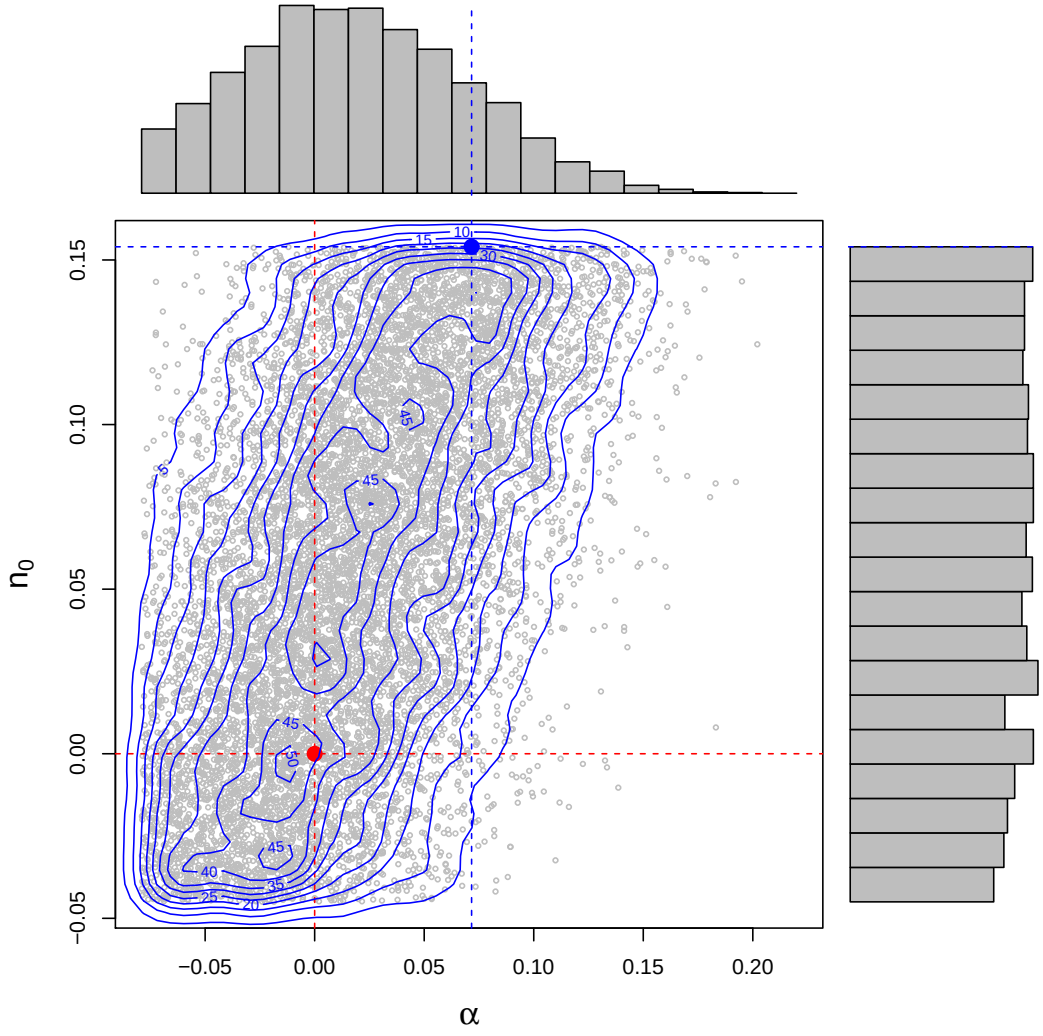



Figure 2.8.: Two-dimensional marginal density plot between $\alpha$ and $n_0$ at US-Ha1. The 1-D marginal distributions for each parameter is found on the edges. The dimensions of the plot represent the prior range of each parameter. Red points/dashed lines represent initial parameter values. Blue points/dashed lines represent optimised parameter values. Blue contours illustrate the posterior distribution.

is a single dimensionless measure of model misfit. The fractional error $\epsilon$ can then be interpreted as the typical (root-mean-square) error expressed as a fraction of the (root-mean-square) magnitude of the observed seasonal cycle. Thus, $\epsilon = 0$ represents a perfect match to the observations, while $\epsilon = 1$ corresponds to the error in a null model whose prediction $\mathbf{m}_{i,t}$ always equals the observational mean $\bar{\mathbf{o}}_i$.

In hydrology, this is related to a metric known as the Nash–Sutcliffe efficiency [Nash and Sutcliffe, 1970] equivalent to $1-\epsilon^2$, and has been used by many studies to perform cross-site comparisons.

### Taylor diagrams

The fractional error is a good tool for cross-site comparison but it does not give much information about model performance at each site. Taylor diagrams (Taylor, 2001) can provide more insight into the fit by considering the relationship between observed variance var($\mathbf{o}_t$), modelled variance var($\mathbf{m}_t$), error variance var($\mathbf{o}_t - \mathbf{m}_t$) and model-observation correlation cor($\mathbf{o}_t$, $\mathbf{m}_t$).

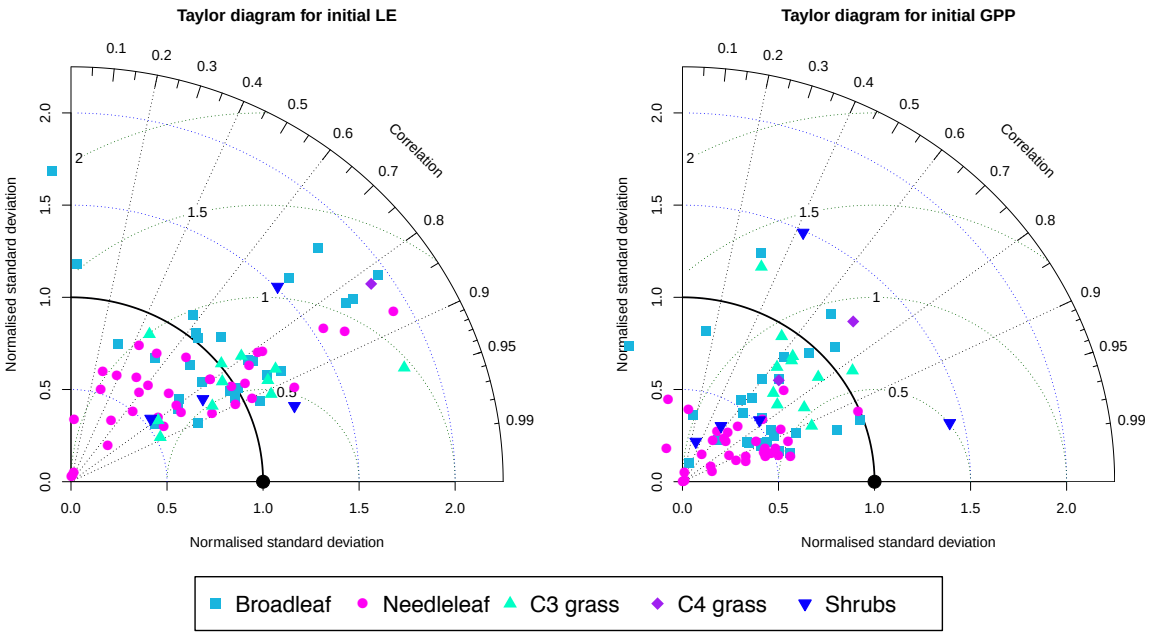



Figure 2.9.: Initial model-data fit calculated over monthly time-series represented by Taylor diagrams. Observed time series (black dot) can be compared with modelled time series generated with the default JULES parameters at all of the FluxNet sites used in this study. Radial distance from the origin (dotted lines) represents normalised standard deviation $\sqrt{var(\mathbf{m}_t)/var(\mathbf{o}_t)}$, and so a modelled time series with the correct variance lies on the thick black line. Angular position represents the correlation between modelled and observed time series. The distance from the black dot (dotted green lines) represents the normalised standard deviation in the errors $\sqrt{var(\mathbf{o}_t - \mathbf{m}_t)/var(\mathbf{o}_t)}$. Three sites are omitted in the LE plot due to extremely high variances: BR-Sa1 (BT), ZA-Kru (C4G), US-Los (Sh).

The Taylor diagrams in Fig.2.9 show the initial error in each of the data streams at each of the FluxNet sites focused on in this project.

For the initial latent heat fit (left), needleleaf sites tend to underestimate the seasonal cycle, whereas the broadleaf sites tend to overestimate it. The C3 grass sites tend to have a latent heat seasonal cycle of the right magnitude. The correlation between the modelled time series and observation time series is relatively high for the majority of the sites, most lying between 0.6 and 0.9.

For the initial GPP fit (right), the seasonal cycle is typically underestimated for all PFT types, especially needleleaf sites. The correlation between model and observed time series is also above 0.6 for the majority of the sites.

#### Bias

Since Taylor diagrams are based on a decomposition of the variance of the errors, they are insensitive to any systematic offset in the model. It therefore makes sense to consider in addition the normalised bias

$$b = \frac{|\mu_m - \mu_o|}{\sigma_o} \tag{2.54}$$

where $\mu_m$ and $\mu_o$ are the means of the modelled and observed time series respectively, and $\sigma_o$ denotes the standard derivation of the observations.

## 2.5. The adJULES system

To conclude this chapter, all of the components described above are combined to make the adJULES system.

### 2.5.1. The theory of adJULES

The JULES land-surface model generates a modelled time-series for a given vector of internal parameters, $\mathbf{z}$. The internal parameters $\mathbf{z}$ are the parameters chosen in section 2.1.2, and are a subset of all available parameters denoted by $\mathbf{x}$. The cost function, $J(\mathbf{z})$, consists of a weighted sum of squares of the difference between $\mathbf{m}_t$ (the vector of model outputs at time $t$), and $\mathbf{o}_t$ (the vector of observations at time $t$), combined with a term quadratic in the difference between parameter values $\mathbf{z}$ and initial parameter values $\mathbf{z}_0$:

$$J\left(\mathbf{z};\hat{\mathbf{z}},\mathbf{z}_0\right) = \frac{1}{2}\left[\sum_t \left(\mathbf{m}_t(\mathbf{z}) - \mathbf{o}_t\right)^T \mathbf{R}\left(\hat{\mathbf{z}}\right)^{-1}\left(\mathbf{m}_t(\mathbf{z}) - \mathbf{o}_t\right) + \lambda(\mathbf{z}-\mathbf{z}_0)^T\mathbf{B}^{-1}(\mathbf{z}-\mathbf{z}_0)\right]. \tag{2.55}$$

Here, $\mathbf{R}(\hat{\mathbf{z}}) = \frac{1}{m}\sum_{t=1}^{m}(\mathbf{m}(\hat{\mathbf{z}})_t - \mathbf{o}_t)(\mathbf{m}(\hat{\mathbf{z}})_t - \mathbf{o}_t)^T$ denotes the error cross product matrix produced by a JULES run with parameter value $\hat{\mathbf{z}}$. In an optimisation, $\mathbf{z}$ and $\hat{\mathbf{z}}$ are updated separately in nested loops, having both been initialised to the default JULES parameter value $\mathbf{z}_0$. In the inner loop, $\mathbf{z}$ is varied to minimise the cost function (termination criterion:

$\nabla J \approx 0$) for the current value of $\hat{\mathbf{z}}$. In the outer loop, $\hat{\mathbf{z}}$ is reset to the new value of $\mathbf{z}$ from the inner loop (termination criterion: change in $\hat{\mathbf{z}}$ negligible). At the end of an optimisation, therefore, the matrix $\mathbf{R}$ conveys information about the error correlation structure in a JULES run with optimal parameter values.

The matrix $\mathbf{B}$ describes the prior covariances assigned to the parameters, and is here chosen to be a diagonal matrix proportional to the inverse square of the ranges allowed for each parameter. The prior uncertainties are therefore assumed to be uncorrelated between the parameters. The $\lambda$ parameter controls the relative importance of the background (i.e. the right-hand term in Eq. 2.55) and the error term (i.e. the left-hand term in Eq. 2.55). This is further discussed in Sect. 3.2.3. In most experiments, $\lambda$ acts as a switch; when set to 1, the background term is included, when set to 0, the prior is non-informative and takes on a top-hat distribution.

The optimal vector of parameters is the vector $\mathbf{z}$ that minimises the cost function (Eq. 2.55). The aim of adJULES is to find this vector. The adJULES system minimises the cost function iteratively using the gradient descent algorithm L-BFGS-B described in section 2.3.4. At each iteration, the gradient $\nabla J(\mathbf{z})$ of the cost function $J(\mathbf{z})$ is computed with respect to all parameters, using the adjoint model of JULES (see section 2.3.4). The adjoint is generated with the automatic differentiator tool TAF (Transformation of Algorithms in Fortran; see Giering et al. [2005]).

Once the cost function reaches the minimum, a locally optimal parameter vector is returned. This process is then repeated, the locally optimised parameters are fed back through JULES, generating a new modelled time-series and hence a new cost function. The loop terminates when the modelled time series no longer improves or when the difference between consecutive $z$ vectors is below a tolerance of $1 \times 10^{-5}$ (Fig. 2.10). At the end, $\mathbf{z}_1$ denotes the locally optimal parameter vector and the second derivative of the cost function with respect to the parameters can be used to calculate posterior uncertainties (as described in section 2.4.1).

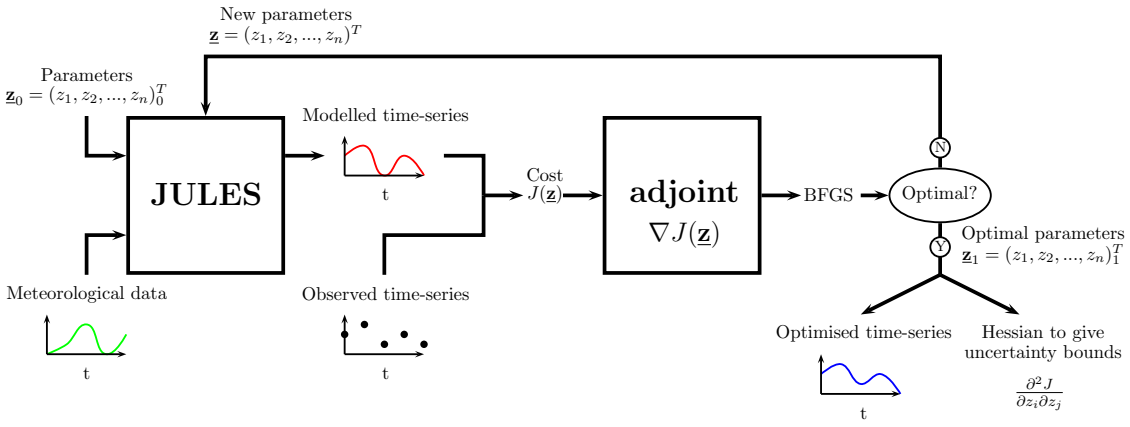



Figure 2.10.: Schematic of the adJULES parameter estimation system starting with the initial parameter vector $\mathbf{z}_0$. This is usually based on default JULES parameter values [Blyth et al., 2010]. The optimised parameter vector is denoted $\mathbf{z}_1$.

The tools in section 2.4.2 can then be used to quantity the model-data fit given by the optimal parameter vector.

### 2.5.2. Inherited version of the adJULES system

The adJULES system was originally developed by Tim Jupp and FastOpt at the University of Exeter. Set up to calibrate a subset of 50 soil and vegetation of parameters against three fluxes: net ecosystem exchange (NEE), sensible heat (H) and latent heat (LE), adJULES could be run 'at a point', i.e. could calibrate at a single location.

Since its creation circa 2008 (five years prior to the start of this project), the adJULES system as been through some changes, most notably the addition of new data streams against which to calibrate in Luke [2011].

As it stands, the adJULES system inherited for this project was set up to optimise:

- 94 of the physical JULES parameters covering five PFTs and four soil layers
- individually over the 9 FluxNet sites described in Blyth et al. [2011]
- simultaneously over a subset of the 6 different data streams (NEE, H, LE, T, GPP, Resp)
- following the cost function:

$$J\left(\mathbf{z};\hat{\mathbf{z}}\right) = \sum_{t} \left(\mathbf{m}_t(\mathbf{z}) - \mathbf{o}_t\right)^T \mathbf{R}\left(\hat{\mathbf{z}}\right)^{-1} \left(\mathbf{m}_t(\mathbf{z}) - \mathbf{o}_t\right) \tag{2.56}$$

- by iteratively looping 3 times (the main loop in the schematic shown in Fig.2.10)
- using the adjoint generated from JULES version 2.2
- using the BFGS optimisation scheme

In the following chapter, the cost function used in the adJULES system is more closely examined. It is extended from the definition in Eq. 2.56 to include a background term and to include the ability to calibrate over multiple measurement sites simultaneously.

# 3. Improvements to the adJULES system

In this chapter the cost function is considered more closely. This is done in two ways, first by defining it within a Bayesian framework (Sect. 3.1.1), and then by looking specifically at the error covariance matrices **R** and **B** (Sect. 3.1.2 and 3.1.3 respectively). The mathematical theory used to define these matrices is covered, and in Sect. 3.1.4, alternative formulations of the cost function are discussed.

In order to calibrate over multiple sites simultaneously, the cost function is extended to a 'multi-site' framework in Sect. 3.2 and its robustness is tested at different timescales (Sect. 3.2.2). In Sect. 3.2.3, the background term **B** is reconsidered. In a multi-site framework both the **B** matrix and the additional sites in the calibration were found to place strong constraints on the optimisation.

This chapter concludes in Sect. 3.3 by looking at the system's sensitivity to initial conditions, both in terms of fit and optimal parameter vectors.

## 3.1. Building the cost function

### 3.1.1. Bayesian framework

In order to understand the statistical assumptions made in the adJULES scheme, the cost function is built up within a Bayesian framework following the example of Tarantola [1987].

Bayes' theorem states [de Laplace, 1820]:

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)} \tag{3.1}$$

where $A$ and $B$ are events and $P(A) \neq 0$. This formulation allows for the combination of the prior distribution $P(B)$, with the probability of $A$ given $B$, to give a posterior distribution $P(B|A)$ of $B$ given $A$.

Changing the notation to that of this study, this theorem can be used to obtain the cost function described by Eq. 2.55. Let the observations be denoted by the vector $\mathbf{o} \in \mathbb{R}^m$, where $m$ represents the number of observations. The model outputs are represented by vector $\mathbf{m}(\mathbf{z}) \in \mathbb{R}^m$, where $\mathbf{z} \in \mathbb{R}^n$ is the model parameter vector with $n$ denoting the number of parameters to be optimised.

Starting from a known parameter value called the background $\mathbf{z}_0$, knowledge of the parameters $\mathbf{z}$ is described by the probability density function $p(\mathbf{z})$. Define $p(\mathbf{o}|\mathbf{z})$ to be the likelihood of the observations $\mathbf{o}$ given the parameter vector $\mathbf{z}$. This distribution takes into account the uncertainties associated with the prediction of $\mathbf{o}$. The probability sought after is $p(\mathbf{z}|\mathbf{o})$, which describes the distribution of the parameters given the observations $\mathbf{o}$. Using Bayes' theorem (Eq. 3.1), these probabilities can be combined to give:

$$p(\mathbf{z}|\mathbf{o}) \propto p(\mathbf{o}|\mathbf{z})p(\mathbf{z}). \tag{3.2}$$

In a Bayesian setting, normal (Gaussian) distributions are commonly used to represent the different terms of the optimisation [Tarantola, 1987]. The choice comes from the central limit theorem which states that the sum of a sequence of random variables with finite variances converges towards a normal distribution [de Laplace, 1820].

The normal probability distribution function (PDF) for variable $\mathbf{x} \in \mathbb{R}^q$ is completely defined by the first two moments, mean and variance:

$$p(\mathbf{x}) = \frac{1}{(2\pi)^{q/2}\Sigma^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{x}-\bar{\mathbf{x}})^T\Sigma^{-1}(\mathbf{x}-\bar{\mathbf{x}})\right] \tag{3.3}$$

where $\bar{\mathbf{x}}$ and $\Sigma$ are the mean of $\mathbf{x}$ and the covariance matrix associated with $\mathbf{x}$ respectively.

By setting up a prior centred on $\mathbf{z}_0$, the PDF of $p(\mathbf{z})$ is given by

$$p(\mathbf{z}) \propto \exp\left[-\frac{1}{2}(\mathbf{z}-\mathbf{z}_0)^T\mathbf{B}^{-1}(\mathbf{z}-\mathbf{z}_0)\right] \tag{3.4}$$

where the $\mathbf{B}$ matrix is the background covariance matrix. Similarly, the likelihood function of $p(\mathbf{o}|\mathbf{z})$ is given by

$$p(\mathbf{o}|\mathbf{z}) \propto \exp\left[-\frac{1}{2}(\mathbf{o}-\mathbf{m}(\mathbf{z}))^T\mathbf{R}^{-1}(\mathbf{o}-\mathbf{m}(\mathbf{z}))\right] \tag{3.5}$$

where the $\mathbf{R}$ matrix is the covariance matrix of the observation errors.

Using Bayes' theorem to combine these analytical expressions, the posterior can be calculated as follows:

$$p(\mathbf{z}|\mathbf{o}) \propto \exp\left[-\frac{1}{2}\left[(\mathbf{o}-\mathbf{m}(\mathbf{z}))^T\mathbf{R}^{-1}(\mathbf{o}-\mathbf{m}(\mathbf{z})) + (\mathbf{z}-\mathbf{z}_0)^T\mathbf{B}^{-1}(\mathbf{z}-\mathbf{z}_0)\right]\right]. \tag{3.6}$$

Therefore, in order to maximise $p(\mathbf{z}|\mathbf{o})$, the function

$$J(\mathbf{z}) = \frac{1}{2}\left[(\mathbf{o}-\mathbf{m}(\mathbf{z}))^T\mathbf{R}^{-1}(\mathbf{o}-\mathbf{m}(\mathbf{z})) + (\mathbf{z}-\mathbf{z}_0)^T\mathbf{B}^{-1}(\mathbf{z}-\mathbf{z}_0)\right] \tag{3.7}$$

needs to be minimised.

This gives the formulation of the cost function used in this thesis and reflects the assumption of Gaussian probability distributions for the observed values and the a priori information about the parameters [Kaminski et al., 2002]. In the case of a linear model,

Tarantola [1987] shows that the posterior probability distribution is also Gaussian and that the uncertainties in the optimal parameters are quantified by a covariance matrix which is the inverse of the Hessian at the minimum. In the case of a non-linear model, uncertainties derived in this way are merely a local approximation [Kaminski et al., 2002].

It is clear that there are two components to the cost function in Eq. 3.7. First, a cost function based on model-data fit,

$$J_o(\mathbf{z}) = \frac{1}{2}(\mathbf{o} - \mathbf{m}(\mathbf{z}))^T \mathbf{R}^{-1}(\mathbf{o} - \mathbf{m}(\mathbf{z})) \tag{3.8}$$

and second, a cost function which penalises when the optimal parameters are too far from the prior values

$$J_b(\mathbf{z}) = \frac{1}{2}(\mathbf{z} - \mathbf{z}_0)^T \mathbf{B}^{-1}(\mathbf{z} - \mathbf{z}_0). \tag{3.9}$$

When discussing the set up of the **R** and **B** covariances in the next few sections, each component is considered separately.

### 3.1.2. Observations covariance matrix R

As described above, **R** is the observation error covariance matrix. This matrix should include both the model errors and measurement errors.

Taking them one by one, the model errors represent the errors in model process representation and structure. This error is very hard to characterise since it is not directly observable. In general model error is likely to dominate **R** though some studies choose to ignore it, assuming the error is small or is compensated for by the cost. In this study, the model error is assumed to be absorbed in the output error residual.

The measurement errors cover the errors from the data themselves. These errors can be characterised as random or systematic. In eddy-covariance data, random errors arise from the measurement instruments, the stochastic nature of turbulence and the varying FluxNet footprints [Lasslop et al., 2008]. The systematic errors can be constant, e.g. from inaccurately calibrated instruments, or appear only at certain times, e.g. $CO_2$ errors at night (see Sect. 2.2.1 for more details). For later theory to apply, the measurement errors are assumed to be normally distributed.

Most studies opt for a diagonal **R** matrix assuming uncorrelated errors. Indeed, including correlations between observations is a recent addition to numerical weather prediction studies [Stewart et al., 2013]. Accounting for some correlation structure in **R** has been shown to improve data assimilation results and forecast accuracy [Weston et al., 2014].

In carbon model data assimilation, the inclusion of correlated observation error is even more recent due to the relative infancy of the field. There are nonetheless a couple of relevant studies worth noting.

The first is the work of Pinnington et al. [2016]. Using a 4D-Var framework, Pinnington

et al. [2016] use time correlations between observation errors to update the **R** covariance matrix, which is shown to reduce the root mean square error in the 14 year forecast of daily NEE by 44%.

The second is the work conducted by the ORCHIDEE team covered in MacBean et al. [2016] and Peylin et al. [2016]. Both these investigate stepwise assimilation of multiple data streams into carbon models, MacBean et al. [2016] using simple toy models and Peylin et al. [2016] using the ORCHIDEE model itself. Peylin et al. [2016]'s study keeps the **R** matrix uncorrelated throughout, but does comment on the possibility of propagating correlations through the different steps. MacBean et al. [2016] considers the impact of a correlated **R** matrix when simultaneously assimilating the data streams. Both studies argue that given a robust framework and an adequate description of the error covariance matrices, simultaneous and step-wise assimilation of data streams should be equivalent. MacBean et al. [2016] found that the inclusion of correlation between data streams was of increasing importance if the information content of the observations was too low, i.e. highly uncertain or sparse.

#### Setting up R

In adJULES, **R** is set up to denote the error cross-product matrix produced by a JULES run with a parameter value **z**. This choice of **R** allows for non-independent time-series. The adJULES system has the ability to calibrate against different subsets of the multiple observables integrated in the system by masking the observables not used in the optimisation. Further to this, **R** is set up to scale the cost function to $\nu$ degrees of freedom, where $\nu$ denotes the difference between the number of data points $m$ and the number of parameters optimised $n$. This choice of scaling is justified at the end of this section.

Given a parameter vector **z**, the errors between the modelled time series and the observed time series, also known as residuals, can be generated as follows

$$\mathbf{e}_t(\mathbf{z}) = \mathbf{m}_t(\mathbf{z}) - \mathbf{o}_t.$$

The errors in each data stream make up the columns of **e**, and the rows represent the errors at each time point. The Gram matrix

$$\mathbf{E} = \mathbf{e}^T\mathbf{e}$$

can be used to represent the non-central second moment of errors. The central second moment of errors would give the covariance centred at the mean. However, in this case ellipses of cost around the origin are more informative since the best possible fit is when the error is zero.

The errors between the different data streams can be correlated and of different magnitudes. The left-hand plot in Fig. 3.1 illustrates this. It considers the initial daily errors in modelled LE against modelled GPP at the Harvard site (US-Ha1). The errors are clearly

correlated and there is a notable difference in the magnitude of the scales of both data streams.

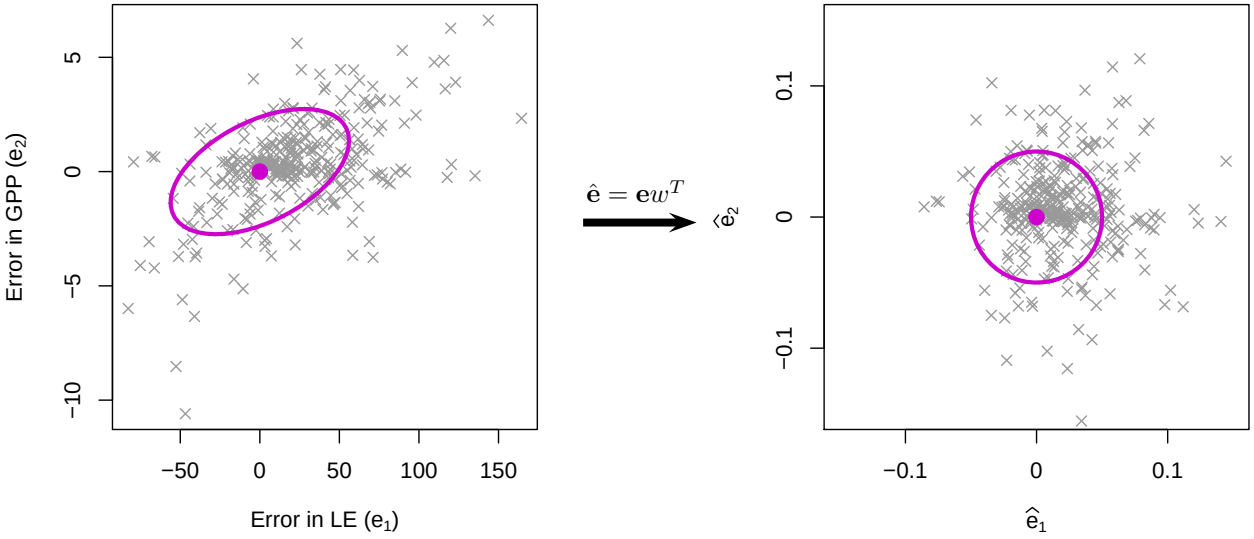



Figure 3.1.: Transformation of the errors at US-Ha1 in observation space. The crosses represent the different errors in LE and GPP at daily time points and the ellipses represent the covariance matrix. The original observation space (LHS) is transformed through a rotation and scaling to an uncorrelated space (RHS).

By rotating and scaling the axes through a transformation, the errors can be decorrelated, as shown in Fig. 3.1. In order to do this, it is necessary to find a basis $w$ such that $\hat{\mathbf{e}} = \mathbf{e}w^T$ has uncorrelated columns and therefore the resulting covariance matrix is diagonal. This is done using the Cholesky decomposition of $\mathbf{E}^{-1}$.

In the adJULES system, $\mathbf{E}$ is multiplied by the number of observables $n_o$ before taking the Cholesky decomposition of its inverse.

$$[n_o\mathbf{E}]^{-1} = U^T U$$

This upper triangular $U$ matrix becomes the basis for $w$. Now uncorrelated, the errors can be scaled to represent the desired number of degrees of freedom. By further scaling $U$ by $\sqrt{\nu}$, so that $w = \sqrt{\nu}U$, $J_o(\mathbf{z})$ can be written as a $\chi^2$ objective function with $\nu$ degrees of freedom

$$J_o(\mathbf{z}) = \frac{1}{2}\sum_{t=1}^{n} \mathbf{e}_t(\mathbf{z})^T \mathbf{R}^{-1} \mathbf{e}_t(\mathbf{z}) \tag{3.10}$$

$$= \frac{1}{2}\sum_{t=1}^{n} \left(\frac{\hat{\mathbf{e}}_t(\mathbf{z})}{\sigma_t}\right)^2 \equiv \chi^2 \tag{3.11}$$

where $\sigma_t$ denotes the standard deviations for each data point $\hat{\mathbf{e}}_t(\mathbf{z})$. Therefore, the observation correlation matrix $\mathbf{R}$ is set to be

$$\mathbf{R}^{-1} = w^T w = \nu U^T U. \tag{3.12}$$

### Updating R

The choice of $\mathbf{R}$ in this optimisation scheme is dependent on $\mathbf{z}$. After minimising the cost function once, the model can be run again with the new parameters to give new errors and therefore a new covariance matrix $\mathbf{R}$.

This updating of $\mathbf{R}$ can be considered to be a separate optimisation loop. Let $\mathbf{R}(\hat{\mathbf{z}})$ denote the error covariance matrix for parameter vector $\hat{\mathbf{z}}$. Initialised by the JULES default parameter vector $\mathbf{z}_0$, the $J_o$ part of the cost function $J$ can be written as

$$J_o\left(\mathbf{z};\hat{\mathbf{z}}\right) = \frac{1}{2}\sum_t \left(\mathbf{m}_t(\mathbf{z}) - \mathbf{o}_t\right)^T \mathbf{R}\left(\hat{\mathbf{z}}\right)^{-1}\left(\mathbf{m}_t(\mathbf{z}) - \mathbf{o}_t\right) \tag{3.13}$$

The full cost function is optimised using the BFGS algorithm (described in Sect. 2.3.4) to give an optimal parameter vector $\mathbf{z}_{\text{temp}}$. Next, $\hat{\mathbf{z}}$ is reset to the new parameter vector $\mathbf{z}_{\text{temp}}$ and this new cost function is optimised. In the early versions of adJULES, this process was repeated three times. In the newest version of adJULES, the number of iterations allowed has been increased to let the process converge. The new termination criterion is that the change in $\hat{\mathbf{z}}$ should be negligible (i.e. $1 \times 10^{-5}$). Typically, five iterations are needed. The final output parameter vector is saved as $\mathbf{z}_1$.

Figure 3.2 shows sequential optimisations performed on a site in Denmark (DK-Sor). Successive parameter vectors were able to converge to $\mathbf{z}_1$ after eight optimisations. The first optimisation can be seen to reduce the errors the most, especially for the summer months which start with the largest errors. Some of the other months can be seen to get slightly worse in order to allow more reduction in error at the summer months.

Each optimisation of the cost $J$ takes approximately 15 iterations. The cost function is then updated and a new optimisation takes place. This is repeated on average five times, giving overall just under 100 iterations. In comparison, in Santaren et al. [2007]'s ORCHIDEE parameter optimisation experiments, the model also took approximately 100 iterations to converge to a minimum of the cost function. Running an MCMC algorithm on a similar experiment took over 1000 iterations to converge.

### Chi-square fitting

The cost function $J_o$ has been constructed in such a way that it can be written as a $\chi^2$ objective function on $\nu$ degrees of freedom. Theory taken from Press et al. [2007] (Chapter 15: Modelling of Data) is used in this section to explain this choice.

Consider the case with uncorrelated observations. Suppose that each data point $\mathbf{o}_t$ has a measurement error that is independently random and distributed as a normal (Gaussian) distribution around the 'true' model $\mathbf{m}$. If each point has its own, known standard deviation $\sigma_t$, then the probability of the data set is the product of the probabilities of each

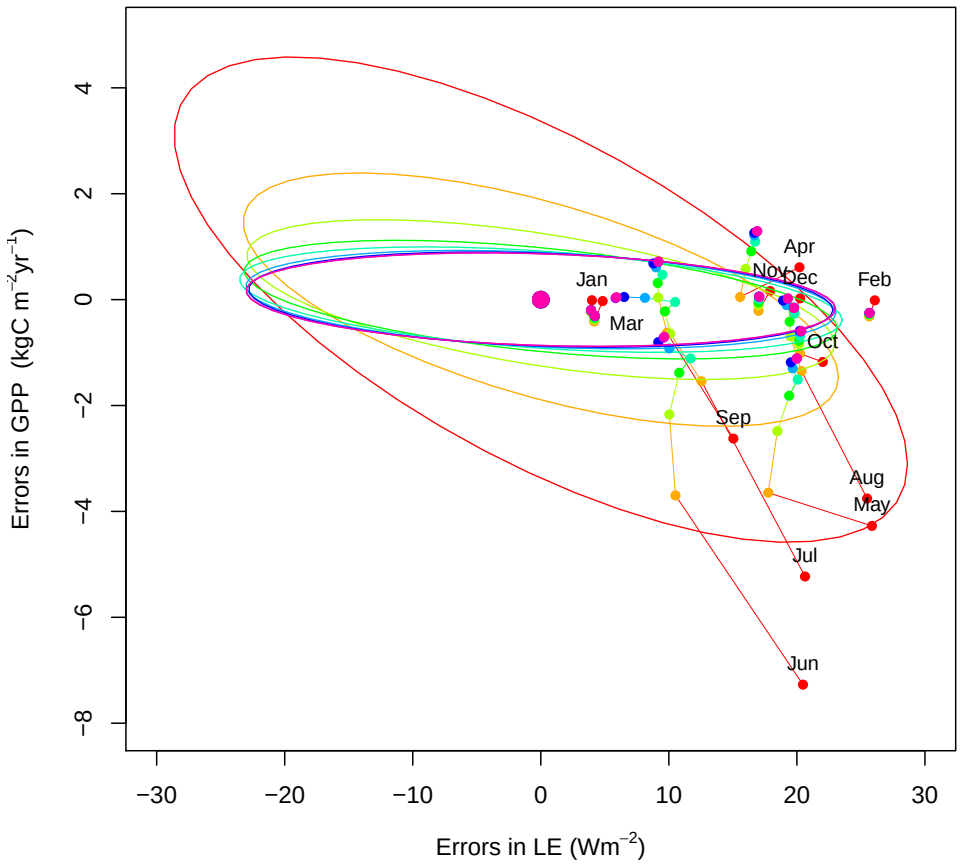



Figure 3.2.: The monthly errors at the DK-Sor site shown at each step of the successive optimisations performed. The ellipses illustrate the non-central second moment of errors used in defining **R**. The different colours represent each iterative step and the points represent the error in modelled GPP and LE of the given month at each step.

point.

$$P \propto \prod_{t=1}^{m} \left\{ \exp\left[ -\frac{1}{2} \left( \frac{\mathbf{o}_t - \mathbf{m}_t(\mathbf{z})}{\sigma_t} \right)^2 \right] \right\} \tag{3.14}$$

Maximising this is equivalent to minimising the negative of its logarithm

$$\left[ \sum_{t=1}^{m} \frac{[\mathbf{o}_t - \mathbf{m}_t(\mathbf{z})]^2}{2\sigma_t} \right]. \tag{3.15}$$

This is the same as minimising the quantity

$$\chi^2 \equiv \sum_{t=1}^{m} \left( \frac{\mathbf{o}_t - \mathbf{m}_t(\mathbf{z})}{\sigma_t^2} \right)^2. \tag{3.16}$$

This is called 'chi-square' fitting. The quantity $\chi^2$ is a sum of $m$ squares of normally distributed quantities, each normalised to unit variance. Once the parameter vector $\mathbf{z} = \{z_1, \ldots, z_n\}$ has been adjusted to minimise the value of $\chi^2$, the terms in the sum are not all statistically independent. For models linear in the parameters, the probability

distribution for different values of $\chi^2$ at the optimum is a $\chi^2$ distribution on $\nu = m - n$ degrees of freedom. The $\chi^2$ statistic can also be used to give a quantitative measure for the goodness-of-fit of the model. The contours of constant $\Delta\chi^2$ can be used as boundaries for different confidence regions.

To summarise the above by highlighting the assumption made: if i) the measurement errors are normally distributed and ii) the model is (locally) linear in its parameters, then $\chi^2$ at the optimum is drawn from a $\chi^2_\nu$ distribution.

In cases where the uncertainties associated with the set of measurements are not known in advance, considerations related to $\chi^2$ fitting are used to derive a value for $\sigma$. Assuming that all measurements have the same standard deviation, $\sigma_t = \sigma$, and that the model does fit well, then an arbitrary value for $\sigma$ can be assigned to $\sigma$ for the minimisation procedure. The model parameters found by minimising $\chi^2$ with this arbitrary $\sigma$, denoted $\mathbf{z}_{\text{opt}}$, can then be used to calculate the value of $\sigma$:

$$\sigma^2 = \sum_{t=1}^{m} \frac{[\mathbf{o}_t - \mathbf{m}_t(\mathbf{z}_{\text{opt}})]^2}{\nu} \tag{3.17}$$

The optimal parameter values are insensitive to $\sigma$. However, it is important to assign the correct value of $\sigma$ since it controls the curvature at the optimum. The curvature at the optimum is used to generate the posterior uncertainties associated with the optimal parameter vector.

In this study, since the standard deviation of each error is not known, this method is used. In the adJULES procedure, once an optimum has been reached, $\mathbf{R}$ is updated and the optimisation is run again (as discussed above). At an optimum, $\chi^2$ is known to be distributed on $\nu$ degree of freedom. At the initial step, the distribution is unknown, however the function can be scaled arbitrarily through the first choice of $\sigma$. The choice is made to scale this initial objective function to also be $\chi^2$ on $\nu$ degrees of freedom. This choice of scaling and therefore definition of $\mathbf{R}$ was made by T. Jupp when first creating the adJULES system.

### 3.1.3. Background covariance matrix B

The background covariance matrix $\mathbf{B}$ contains statistical information about the prior distribution of the parameters (or of the state variables if used). In most studies, this is a diagonal matrix where the diagonal elements of $\mathbf{B}$ are the standard deviation of each parameter.

The off-diagonal elements of $\mathbf{B}$ quantify the correlations between errors in the parameters. Including correlations in the $\mathbf{B}$ matrix has been shown to improve data assimilation results significantly in both numerical weather predictions [Bannister, 2008] and more recently in carbon cycle assimilation [Pinnington et al., 2016]. In Pinnington et al. [2016], correlations in the background error covariance matrix are included by running an ensemble through a

set of ecological dynamical constraints based on expert judgement. Including off-diagonal elements in the **B** matrix is beyond the scope of this study.

A background term is vital in numerical weather prediction studies where an initally state is needed [Bouttier and Courtier, 2002]. In contrast, for parameter estimation studies, the background term is not always implemented. Berger et al. [2012] argue that inclusion of initial conditions or boundary conditions is what differentiates data assimilation studies from parameter estimation ones. Though the background term is omitted in Berger et al. [2012], the study performs a preliminary sensitivity test, starting at different initial conditions, to see if the same optimum is reached. In Kaminski et al. [2002], the main pre-cursor study to the CCDAS work, the background term is omitted since the knowledge it would add was believed to be already embodied in the model. Other studies choose to leave out $J_b$ for simplicity or when unsure how to define the matrix.

There are many benefits to including a background term. Most importantly, it ensures that the problem is well-posed, even when few observations are present [Bannister, 2004]. A mathematical problem is said to be well-posed if there exists a unique solution and if the solution's behaviour changes continuously with the initial conditions [Hadamard, 1902]. This term is therefore necessary whenever the information contained in the observations is insufficient to guarantee a unique optimal solution.

In addition the background covariance matrix **B** serves as a penalisation term. This means the optimisation never deviates too far from the initial guess. This is useful when the expert guesses are thought to be close to the truth (e.g. when parameters can be measured physically). The correct minimum of $J$ is expected to be close the prior values.

Including the background term, however, means that its error characteristics need to be determined. Defining the background error covariance matrix is not a trivial task and it is important to determine these background errors correctly.

#### Setting up B

In adJULES, the definition of the background term $J_b$ (Eq. 3.9) has been extended to include a factor $\lambda$, called the constant of proportionality,

$$J_b(\mathbf{z}, \lambda) = \frac{\lambda^2}{2}(\mathbf{z} - \mathbf{z}_0)^T \mathbf{B}^{-1}(\mathbf{z} - \mathbf{z}_0). \tag{3.18}$$

This can simply be set to zero in order to omit the background term. Otherwise, this controls the relative importance of the term.

The role of $\lambda$ can be best understood by reverting back to the PDF definition of the cost function (Eq. 3.2),

$$p(\mathbf{z}|\mathbf{o}) \propto \exp\left[-\left(J_o(\mathbf{z}) + J_b(\mathbf{z}, \lambda)\right)\right] = \exp[-J_o(\mathbf{z})] \cdot \exp[-J_b(\mathbf{z}, \lambda)]. \tag{3.19}$$

By excluding the background term from the cost function, the prior is taken to be a top

hat distribution, i.e. every value in the allowed range is equally likely. Adding the term to the cost function is the same as multiplying by $e^{-J_b(\mathbf{z},\lambda)}$ where $J_b(\mathbf{z},\lambda)$ is a quadratic term centred at $\mathbf{z}_0$, the initial value of the parameter, and $\lambda$ controls the shape of the distribution. This is illustrated in Fig. 3.3.

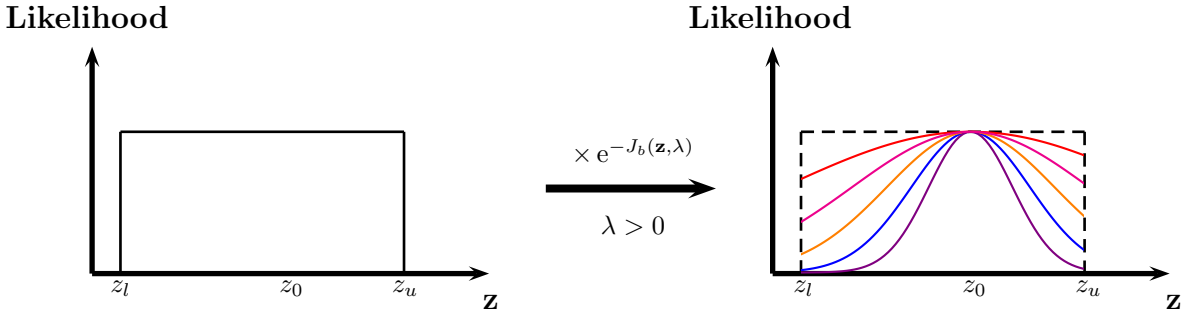



Figure 3.3.: The shape of the prior distribution of $\mathbf{z}$ for changing values of $\lambda$.

Increasing values of $\lambda$ give a more pronounced bell curve. Note that the peak of this curve is centred at $\mathbf{z}_0$, and the curves are cut off at the bounds given by the lowest value allowed for $\mathbf{z}$ ($\mathbf{z}_l$) and the highest value allowed for $\mathbf{z}$ ($\mathbf{z}_u$). Larger values of $\lambda$ help condition the problem and force parameter values to be close to the initial value $\mathbf{z}_0$.

The $\mathbf{B}$ matrix is set to be a diagonal matrix whose diagonal elements are proportional to the inverse square of the ranges allowed for each parameter. Therefore the larger the bounds, the larger the error. This is also scaled by $\nu$ to ensure the $J_b$ term is of equal weight to the $J_o$ term in the main cost function when $\lambda = 1$,

$$\mathbf{B} = \nu \cdot \mathrm{diag}\left(\frac{1}{\mathbf{z}_\mathrm{u} - \mathbf{z}_\mathrm{l}}\right)^2. \tag{3.20}$$

In previous multi-site studies, e.g. Kuppel et al. [2012, 2014], the prior range was also used to define the background covariance matrix $\mathbf{B}$. The range was further multiplied by a factor of 40% [Kuppel et al., 2012] and one-sixth [Kuppel et al., 2014] respectively. This factor is equivalent to the constant of proportionality $\lambda$ discussed in this thesis.

In Sect. 3.2.3, the effect of $\lambda$ on the optimisation is considered in greater depth. However, without knowing the optimal value for this factor, $\lambda$ is simply used as a switch taking values of either 0 or 1 for the main experiments of this thesis. The addition of the background term is a new feature in this thesis, previously it was simply omitted.

### 3.1.4. Alternative formulations

The cost function is the central focus of the data assimilation system. Its definition determines what is to be minimised, and therefore which parameters are to be found. For example, if the aim were to find the parameters that captured the peak of the seasonal cycle, the cost function could be constructed to focus on that alonw.

In a study called OptIC (Optimisation InterComparison; Trudinger et al. [2007]), five dif-

ferent parameter estimation methods were compared. The methods covered were adjoint, Kalman filer, MCMC, Levenberg-Marquardt and genetic algorithm. The model tested was a simple representation of carbon dynamics in a terrestrial biosphere model. The choice of cost function was shown to be more important than the method in finding the optimal parameters. In fact, the methods were equally successful, and variations arose from different cost functions used [Trudinger et al., 2007]. This study highlights the fact that careful consideration of the cost function is vital to any optimisation scheme.

The cost function chosen in this study is the one most commonly used in similar optimisations. However, it is worth considering the other possible formulations. Apart from changing the choice of $\mathbf{R}$ and $\mathbf{B}$, there are two other ways in which the cost function could be changed.

#### Metric minimised

The first method is to change the metric minimised. The cost function in this study minimises a weighted sum of squares. Other metrics that could be used as the cost function include the correlation coefficient, root-mean square error (RMSE), bias, and FVU ($\epsilon^2$ as described in Sect. 2.4.2). RMSE most closely resembles the metric used in the current cost function, taking the form:

$$\text{RMSE} = \sqrt{\frac{\sum_{t=1}^{m}(\mathbf{o}_t - \mathbf{m}_t(\mathbf{z}))^2}{m}} \tag{3.21}$$

Each of these metrics have their own advantages and disadvantages. For example, the correlation coefficient indicates the precision of the model but is not robust and is insensitive to additive differences. RMSE is sensitive to large errors and gives strong emphasis on fitting peak values. Fischer et al. [2013] found that changing metrics for different timescales was beneficial. Though not used in the optimisation, these metrics are considered when assessing the improved model-data fit in this study.

#### Multi-objective optimisation

The second method is to use a multi-objective cost function instead of a single-objective function as used in this thesis. This means minimising multiple metrics simultaneously, where each metric defines a separate objective function. If the objective functions within the multi-objective framework are conflicting, no single solution can simultaneously minimise all the objectives. There is, therefore, a set of solutions instead of one unique one. Each solution is said to be Pareto optimal. Without additional subjective preference information, all Pareto optimal solutions are considered equally good [Dehuri et al., 2015].

By removing the need to find one unique optimal solution, Gupta et al. [1998] argue that a multi-objective cost function bypasses some of the statistical assumptions made in setting up the cost function, namely assumptions made about model error. In some

studies the model error component of the **R** covariance matrix can be assumed either to be small, or to be 'absorbed' into the output error residual. Taking these assumption in turn, the magnitude of model error for some portions of the model response may, in fact, be substantially larger than the output measurement error [Gupta et al., 1998]. In the other case, the model error component does not necessarily behave statistically in the same manner as the output measurement error. In fact, model errors do not necessarily have any inherent probabilistic properties that can be used to construct an objective function [Gupta et al., 1998]. Therefore Gupta et al. [1998] conclude that there is no 'statistically correct' choice for the objective function, and, in turn, no statistically correct 'optimal' choice for the parameters.

However, multi-objective methods tend to be complex and based on approaches using random generation of parameter sets (e.g. Yapo et al. [1998]). These tend to be computationally expensive and very few studies use gradient based methods. Those that do, convert multi-objective optimisation problems into a single objective function using different weights. In Izui et al. [2015], these weighting coefficients are then adaptively determined by solving a linear programming problem. These techniques are used in engineering and design optimisation problems which tend to be less complex than terrestrial parameter estimation problems.

Though beyond the scope of this work, it is interesting to keep these ideas in mind. Multi-objective functions can be a good way of bypassing some of the statistical assumptions made and the problems encountered in optimisation problems.

## 3.2. Extending cost function to multiple sites

In its simplest form, adJULES runs at a single grid-point location and so the derived optimal parameter vector is site-specific. On the other hand, multi-site optimisation aims to find values for a common set of parameters using data from multiple locations. One of the key accomplishments of this thesis has been to implement multi-site optimisations into the adJULES system. The definition of the cost function (Eq. 2.55) has been extended to include the observations from all $S$ sites and its derivative found in order to use the L-BFGS-B algorithm again. The extended cost function is the sum of the individual cost functions for each site $s$. Similarly, the first and second derivatives of this new cost function can be defined using the sum of the derivatives at the individual sites.

$$f(\mathbf{z};\hat{\mathbf{z}},\mathbf{z}_0) = \frac{1}{2}\left[\sum_s\sum_t(\mathbf{m}_{t,s}(\mathbf{z})-\mathbf{o}_{t,s})^T\mathbf{R}_s(\hat{\mathbf{z}})^{-1}(\mathbf{m}_{t,s}(\mathbf{z})-\mathbf{o}_{t,s}) + S\lambda(\mathbf{z}-\mathbf{z}_0)^T\mathbf{B}^{-1}(\mathbf{z}-\mathbf{z}_0)\right] \tag{3.22}$$

An additive cost function, where the optimisation criterion is to minimise the total cost, was chosen over a cost function where all individual cost functions are required to improve.

All of the sites were used in finding the optimal parameter vector for each PFT, so that sites that do not improve with the rest of the PFT suggest incorrect classification of the site or issues with the PFT definitions.

### 3.2.1. Further extensions to the cost function

In moving to a multi-site framework, a few different ideas where implemented. These still exist within the adJULES system as further extensions to the cost function, however they are not throughly explored in this thesis.

A simple extra feature is the ability to apply different weights to each of the sites in the cost function. Deciding how important each site should be in calculating the cost adds a level of subjectivity, so in this study each site is equally weighted.

The other extensions implemented provide the multi-site optimisations with the ability to vary extra parameters locally. This feature allows for the calibration of a common set of parameters over multiple sites, while also allowing other parameters at the sites to improve individually. Two methods were considered, though only the latter works in achieving this properly.

The first idea involves an additive cost function where all parameters could vary with a penalisation term forcing the parameters deemed common over sites to be as close together as possible. The main issue with this implementation was that the common parameters were never equal, the penalisation weighting needed to be too high. This showed that the optimal parameters between sites were very different.

The second method involves creating a large parameter vector $\mathbf{y}$ where the first few elements are the common parameters and the rest of the local parameters are appended in some site order.

Consider the parameter vector $\mathbf{z}_i$ at site $i \in \{1, S\}$ where $S$ is the number of sites. This vector can be broken down into two subsets, subset $\mathbf{c}_i$ which contains the parameters to be common over all sites, and subset $\mathbf{r}_i$, the rest of the parameters to be optimised at that site. The cost function $g(\mathbf{y})$ can be written as a sum of the individual cost functions at each site:

$$g(\mathbf{y}) = \sum_{i}^{S} f(\mathbf{z}_i).$$

This is achieved by splitting the $\mathbf{y}$ back into its components to calculate the cost

$$\mathbf{y} = \begin{pmatrix} \mathbf{c} \\ \mathbf{r}_1 \\ \vdots \\ \mathbf{r}_i \\ \vdots \\ \mathbf{r}_m \end{pmatrix} \longmapsto \begin{cases} \mathbf{z_1} = \begin{pmatrix} \mathbf{c} \\ \mathbf{r}_1 \end{pmatrix} & \longrightarrow \quad f(\mathbf{z_1}) \\ \vdots & \\ \mathbf{z_i} = \begin{pmatrix} \mathbf{c} \\ \mathbf{r}_i \end{pmatrix} & \longrightarrow \quad f(\mathbf{z_i}) \\ \vdots & \\ \mathbf{z_S} = \begin{pmatrix} \mathbf{c} \\ \mathbf{r}_S \end{pmatrix} & \longrightarrow \quad f(\mathbf{z_S}) \end{cases} .$$

In turn the gradient of $g(\mathbf{y})$ is calculated:

$$\nabla g(\mathbf{y}) = \begin{pmatrix} \sum_i^s \nabla f(\mathbf{c}) \\ \nabla f(\mathbf{r}_1) \\ \vdots \\ \nabla f(\mathbf{r}_i) \\ \vdots \\ \nabla f(\mathbf{r}_S) \end{pmatrix} .$$

Therefore the adJULES system has the ability not only to calibrate a common set of parameters over different FluxNet sites but also to let the parameters vary locally. Due to time constraints, and since the main aim of the project is to find a common set of parameters, this extension was not explored further.

The results discussed in the rest of this thesis are derived from the additive function described in Eq. 3.22. The local parameters, such as soil moisture, were read from ancillary files and are not optimised. The optimisation concentrates on only one set of parameters, all chosen to be common.

### 3.2.2. Testing the robustness of the multi-site technique

To test the robustness of the multi-site technique, random subsets of the broadleaf sites were optimised. The optimal parameters were then tested at the remaining sites. Broadleaf sites were used because this PFT is the best represented in the FluxNet network. Due to time and computational constraints, only the deciduous broadleaf subset was considered.

Initially, five randomly selected sets of five sites were used in the optimisation stage. These sets will be referred to as training sets. The optimal parameter vectors were then evaluated at the rest of the sites, the validation sets. The training sets selected are shown in Table 3.1.

Four different setups were tested. First, two different data frequencies were used: daily data and monthly data. Second, in both cases, the background term was either included

| | | | | | |
|---|---|---|---|---|---|
| **Set 1** | US-Bar | US-MOz | IT-Non | IT-Ro1 | US-MMS |
| **Set 2** | US-Ha1 | FR-Hes | DE-Hai | IT-Ro2 | UK-PL3 |
| **Set 3** | US-MMS | UK-PL3 | US-Ha1 | IT-Non | IT-LMa |
| **Set 4** | IT-LMa | FR-Hes | DE-Hai | IT-Non | US-MOz |
| **Set 5** | IT-Ro1 | US-MMS | FR-Fon | IT-PT1 | US-MOz |

Table 3.1.: The five subsets of deciduous FluxNet sites used to test the robustness of the multi-site technique.

(i.e. prior distribution assumed to be proportional to one over the prescribed range for each parameter) or ignored (i.e. use of a non-informative top-hat prior).

Let $\mathbf{z}_s^{\text{off}}$ denote the locally optimised $\mathbf{z}$ vector found at sites when optimising without a background term. Similarly let $\mathbf{z}_s^{\text{on}}$ denote the case when the background term was included. The default JULES parameters are denoted by $\mathbf{z}_0$. The parameter vectors found by optimising over each subset of 5 sites is denoted by $\mathbf{z}_m$, with 'on' and 'off' superscript to indicate the use of a background term.

The first thing to highlight is the fact that each optimised $\mathbf{z}_m$ vector is different from the vector found by averaging all the $\mathbf{z}_s$ from its set. It is also different from any of the $\mathbf{z}_s$ vectors in the set. This means that the multi-site optimisation has found a new and different optimal parameter vector which cannot be found by a more straightforward method.

The results for fitting over monthly data are shown in Fig. 3.4 and Fig. 3.5. Each plot shows a different metric used to quantify improvement. Figure 3.4 uses fractional error as described in section 2.4.2 and Fig. 3.5 uses reduction in average RMSE.

For each site, $\mathbf{z}_s^{\text{off}}$ and $\mathbf{z}_s^{\text{on}}$ give similar reductions in fractional error (Fig. 3.4), with $\mathbf{z}_s^{\text{off}}$ performing very slight better. The exception to this are IT-Non and UK-Ham, for which $\mathbf{z}_s^{\text{on}}$ performs noticeably worse than $\mathbf{z}_s^{\text{off}}$, and in the case of the IT-Non site, worse than the default JULES parameters. In the case of IT-Non, this apparent decline in fit is due to the choice of metric. The fractional error is different to the metric minimised in the cost function. Figure 3.5 shows a different metric, the average reduction in RMSE at each site. Using this metric, $\mathbf{z}_s$ improves of all the sites.

The optimised parameter vectors $\mathbf{z}_m$ generally perform well, both on the sites used in the training sets and the sites used in the validations sets. For all sites, excluding UK-PL3, at least three of the five $\mathbf{z}_m$ parameter vectors improve the model-data fit. For two-thirds of the sites, the model-data fit improves no matter what parameter vector is used to generate the modelled time-series. For the UK-PL3 site, even when the site is included in the training set, the new parameter vector does not improve the fit. This points to UK-PL3's incompatibility in this PFT.

Even though the fit can be seen to deteriorate the most for US-Bar and US-Ha1 in Fig. 3.5, Fig. 3.4 shows that these are sites that start off with relatively low errors.

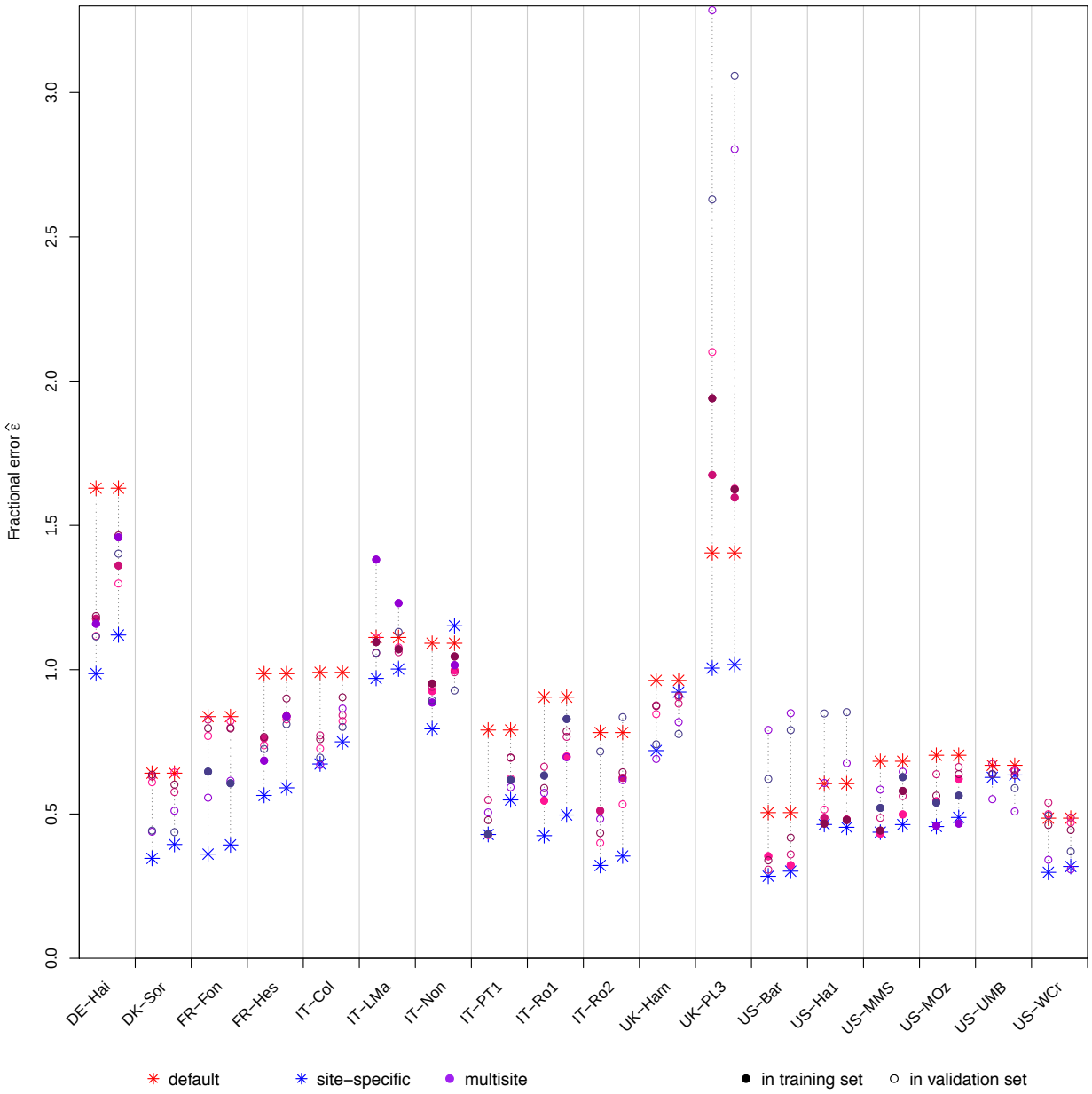



Figure 3.4.: The effect of parameter vectors $\mathbf{z}$ on the overall model-data fit at each of the sites tested, using the metric described in section 2.4.2. The left-hand side in each site panel contains runs without background term, the right-hand side contains runs with background term included. Original default JULES parameters (∗), site-specific optimal parameters (∗), and the multi-site parameters found by optimising over each set of five sites (●○,●○,●○,●○,●○), denoted set 1, set 2, set 3, set 4, set 5 respectively. Sites in the training set (filled circles), sites in validation set (open circles).

In some cases, the 5-site optimised parameter vector $\mathbf{z}_m$ outperforms the locally optimised set $\mathbf{z}_s$, even for sites not uses in that particular training set. For example, the US-UMB site shows this phenomena in both metrics. This shows the multi-site optimisation has found a different local minimum that minimises the error to a greater extent (further discussed in Sect. 3.2.4). Note that starting with a different first guess would also lead to different results.

When optimising over the daily data, the results are even more robust. The optimised parameter vectors $\mathbf{z}_m$ improve 15/18 of the sites regardless of which parameter vector is

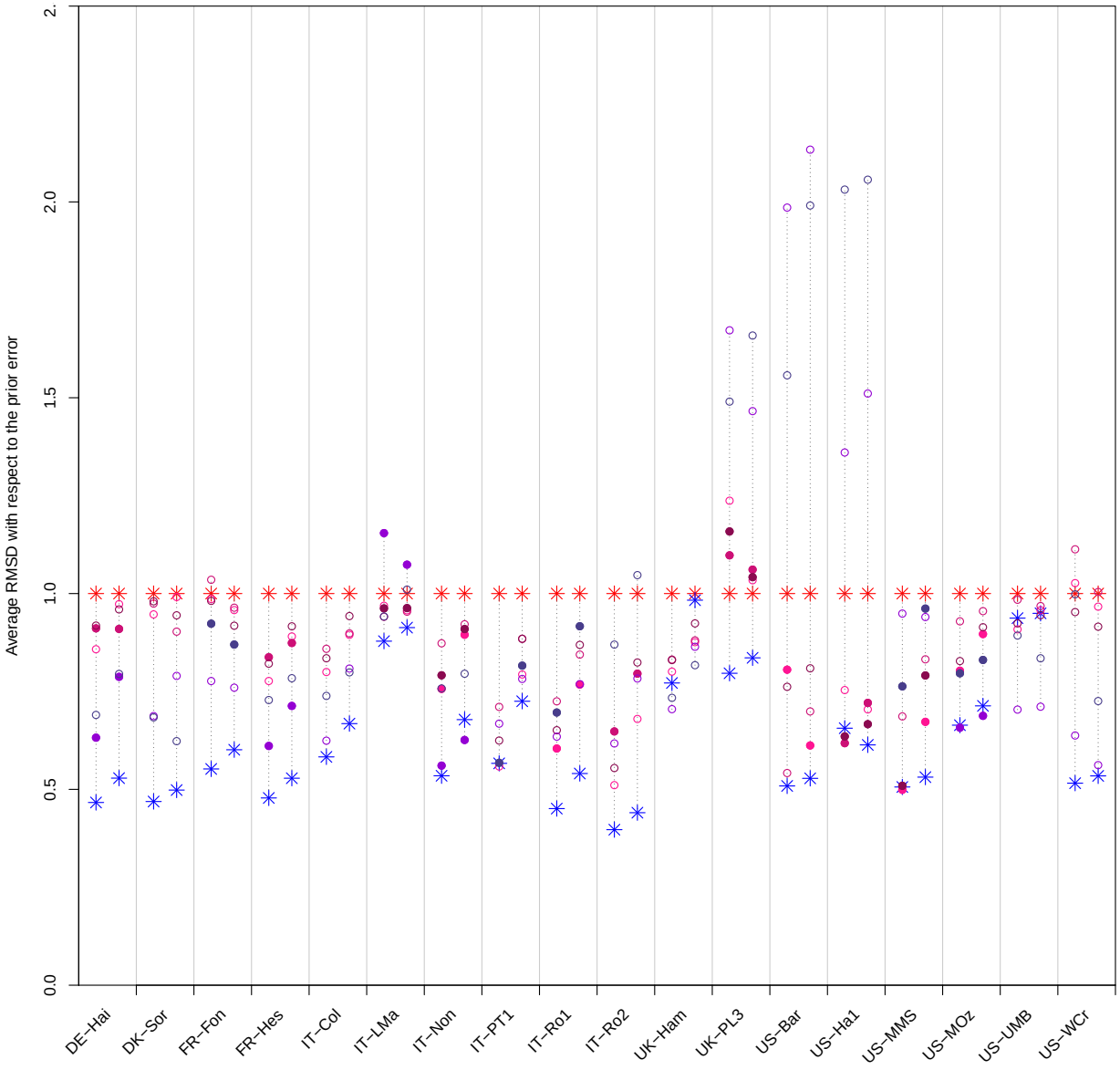



Figure 3.5.: Same as Fig.3.4 but using average RMSE reduction as metric.

used. The UK-PL3 site still performs badly regardless of the $\mathbf{z}_m$ vector used, however the deterioration of fit is not so pronounced. US-Bar and US-Ha1 are the other two sites which do not improve with any the $\mathbf{z}_m$. They do however improve for 3/5 vectors tested, the other two parameter vectors giving a similar or slightly worse fit than the default JULES parameters.

The reduction in error for the daily data is less than for the monthly data. This is due to the fact the daily data have more points to fit. The $\mathbf{z}_m$ vectors are also less likely to outperform the locally optimised parameter vectors. Due to more data points constraining the optimisation, it is possible that the problem is better posed and therefore the solutions more likely to be unique.

Table 3.2 investigates a parameter vector's ability to improve the model-data fit at a different observation frequency than that over which it was calibrated. In all cases, $\mathbf{z}_m$ improves the majority of the sites regardless of the observation frequency used in either

| Parameter vector found by calibrating over | | Fit tested at data frequency | Fraction of sites improving using the vector found at set 1 | 2 | 3 | 4 | 5 | Average error reduction using $\mathbf{z}_s$ at each site |
|---|---|---|---|---|---|---|---|---|
| Daily | No background | Daily | 0.94 | 0.94 | 0.94 | 0.78 | 0.83 | 23.6% |
| | | Monthly | 0.89 | 0.83 | 0.89 | 0.78 | 0.78 | 34.4% |
| | With background | Daily | 0.94 | 0.94 | 0.94 | 0.83 | 0.83 | 18.5% |
| | | Monthly | 0.94 | 0.89 | 0.94 | 0.83 | 0.83 | 27.0% |
| Monthly | No background | Monthly | 0.89 | 0.83 | 0.94 | 0.78 | 0.78 | 35.6% |
| | | Daily | 0.89 | 0.94 | 0.89 | 0.78 | 0.89 | 20.6% |
| | With background | Monthly | 0.89 | 0.89 | 0.94 | 0.78 | 0.72 | 27.8% |
| | | Daily | 0.89 | 0.94 | 0.94 | 0.78 | 0.72 | 16.0% |

Table 3.2.: Optimal parameter vectors calibrated over different observation frequencies are tested at different timescales. Model-fit data quantified in each case using the $\epsilon$ metric described in Sect. 2.4.2.

the calibration or validation. In each case, the level of improvement is comparable with average error reductions in the range 5-15%.

In all cases, the UK-PL3 site does not improve. On closer inspection, this site has a very different seasonality to the rest of the sites for this PFT. Interestingly, when this site is included in the set (sets 2 and 3), the optimal parameter vector found is one of the best performing in this experiment. This shows that the multi-site optimisation does have the ability to find a best-fit set of the parameters even when there is an outlier included in the set. This may be due to the outlier stopping overfitting occurring when optimising over a subset of sites, helping keep the parameter vector more general.

The right-hand side of Table 3.2 considers the locally optimised parameter vectors. The vectors found by calibrating over daily data perform even better when transposed to the monthly data. The reduction in error is similar between the experiments.

Overall the results are promising, showing that the optimised parameters, even when calibrated from a small subset of sites, can be generalised over the rest of the set. The technique has also been found to be robust with respect to outliers and when optimising over different timescales.

### 3.2.3. Exploring the background term in a multi-site framework

As described previously, the matrix $\mathbf{B}$ describes the prior covariances assigned to the parameters, and is here chosen to be a diagonal matrix proportional to the inverse square of the ranges allowed for each parameter. The prior uncertainties are assumed to be uncorrelated between the parameters, and the constant of proportionality $\lambda$ controls the relative importance of $J_b$ and $J_o$.

Moving to a multi-site framework, it became apparent that including the background term

in some of the experiments returned very narrow posterior uncertainties, as illustrated in Fig. 3.6. In these cases, the background term was found to dominate the cost function. For $J_b$ to be as small as possible, the optimal parameter vector needs to be as close as possible to the initial one. Hence the optimisation never moves too far from the original value and the posterior bounds are extremely narrow.

In Fig. 3.6, the optimisation performed without $J_b$ returns mostly uncertain parameter values, with over half of them having posterior uncertainties spanning the whole box. These are all seen to collapse to a very narrow cloud when $J_b$ is included in the optimisation. Note that there are some cases where this very narrow posterior range does not include the prior value within its bounds. This is apparent for the $f_0$ parameter.

This collapse was found to be especially apparent when calibrating over daily data. The optimisation is already constrained by the increased number of data points i.e. 365 data points for the eight parameters. Adding the extra $J_b$ constraint only limits the optimisation further. It follows that adding more sites to the optimisation will also restrain it.

A possible explanation for the difference between the results found with/without the Bayesian term could be due to the choice of $\mathbf{R}$ matrix. Since the choice of $\mathbf{R}$ is based on the prior model-data RMS, very large misfit in this prior model-data RMSE could deweight the $J_o$ term so that the $J_b$ becomes too dominant. The prior misfit in the daily data will be much larger than monthly data with more outliers.

#### Condition number

In order to understand the role of $\lambda$ on the posterior distributions, the Hessian matrix at the optimum, $\mathbf{C}$, is investigated. One of the arguments for including a background, and hence having $\lambda > 0$, was to ensure that the problem was well-posed (see Sect. 3.1.3). A problem is said to be well-posed if there exists a unique solution and if the solution's behaviour changes continuously with the initial conditions.

Another argument for including a background term is that it helps condition the problem. While solutions may be continuous with respect to the initial conditions, when solved with finite precision, or with errors in the data, they may suffer from numerical instability. A small perturbation in the initial data can result in much larger errors in the answer [Trefethen and Bau, 1997]. Such problems are said to be ill-conditioned. Even if a problem is well-posed, it may still be ill-conditioned. An ill-conditioned problem is indicated by a large condition number.

The condition number associated with the linear system

$$\mathbf{Ax} = \mathbf{b} \tag{3.23}$$

gives a bound on how inaccurate the solution $\mathbf{x}$ will be after approximation. It is defined as the maximum ratio of the relative error in $\mathbf{x}$ divided by the relative error in $\mathbf{b}$. Given

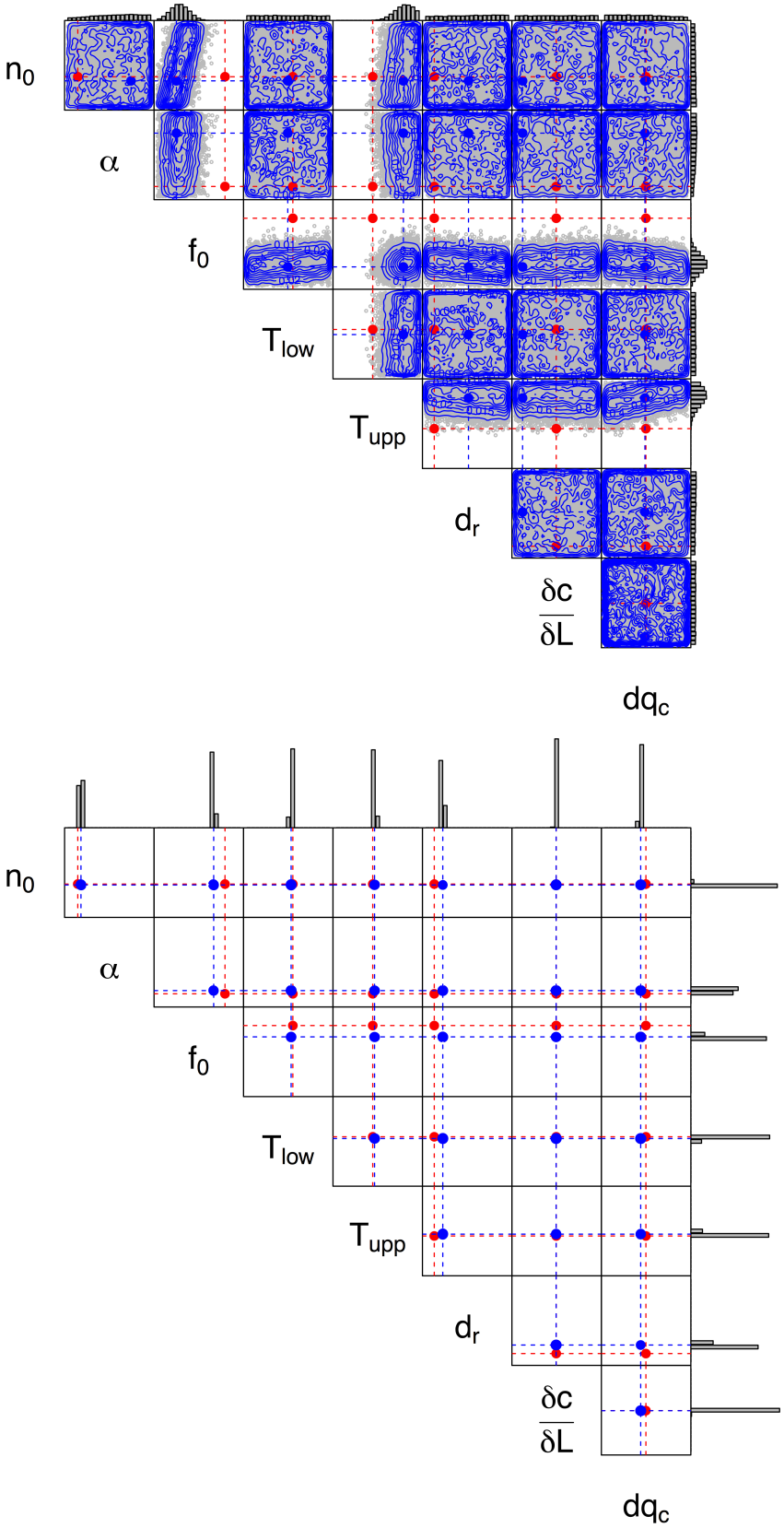



Figure 3.6.: The correlations between parameters calibrated over daily data at all the C3 grass sites. The parameter vector found in experiments excluding the background term (top) and including the background term (bottom). Each subfigure shows a 2-D correlation map, within which each box is a 2-D marginal plot. Bar graphs show 1-D marginal distributions for individual parameters. The dimensions of the boxes represent the prior range of each parameter. Red points/dashed lines represent initial parameter values. Blue points/dashed lines represent optimised parameter values. Blue contours illustrate the posterior distribution.

a perturbation $\delta$, the equation becomes

$$\mathbf{A}(\mathbf{x} + \delta\mathbf{x}) = \mathbf{b} + \delta\mathbf{b}. \tag{3.24}$$

Using matrix norms and the submultiplicative property they satisfy, the two following expressions can be generated:

$$\|\delta\mathbf{x}\| \leq \|\mathbf{A}^{-1}\|\|\delta\mathbf{b}\|$$

$$\|\mathbf{b}\| \leq \|\mathbf{A}\|\|\mathbf{x}\|.$$

In turn these can be used to generate the relative error of the solution:

$$\frac{\|\delta\mathbf{x}\|}{\|\mathbf{x}\|} \leq \|\mathbf{A}\|\|\mathbf{A}^{-1}\|\frac{\|\delta\mathbf{b}\|}{\|\mathbf{b}\|} \tag{3.25}$$

which shows that the error is determined by the constant $\kappa(\mathbf{A}) = \|\mathbf{A}\|\|\mathbf{A}^{-1}\|$. This constant is the condition number of the system (matrix).

Note that

$$\kappa(\mathbf{A}) = \left\|\mathbf{A}^{-1}\right\| \cdot \|\mathbf{A}\| \geq \left\|\mathbf{A}^{-1} \cdot \mathbf{A}\right\| = 1. \tag{3.26}$$

If the condition number is $\mathcal{O}(1)$, the matrix is well conditioned which means its inverse can be computed with good accuracy. If the condition number is very large, then the matrix is said to be ill-conditioned. Such a matrix is almost singular, and the computation of its inverse, or solution of a linear system of equations is prone to large numerical errors. Generally, if the condition number $\kappa(\mathbf{A}) = \mathcal{O}(10^k)$, then in addition to accuracy lost to numerical precision, an extra $k$ digits of accuracy could also be lost [Cheney and Kincaid, 2012].

The condition number may also be infinite. This implies however that the problem is ill-posed; it does not have a unique, well-defined solution. In such cases, the matrix is not invertible.

Since the Hessian is symmetric and positive definite, the condition number of $\mathbf{C}$ can be calculated using its eigenvalues,

$$\kappa(\mathbf{C}) = \frac{|\mu_{\max}(\mathbf{C})|}{|\mu_{\min}(\mathbf{C})|}, \tag{3.27}$$

where $\mu_{\max}(\mathbf{C})$ and $\mu_{\min}(\mathbf{C})$ are maximal and minimal eigenvalues of $\mathbf{C}$ respectively (see Lewis et al. [2006] for proof). As well as the sensitivity of the solution to perturbations, the speed of convergence of the optimisation scheme is also dependent on $\kappa(\mathbf{C})$.

One solution in addressing ill-conditioned problems is to change the system to one with a lower condition number and solve that equivalent problem. This is called preconditioning. For example, Haben [2011] preconditions a system similar to Eq. 3.23 with a symmetric positive definite matrix $\mathbf{P}$ to give the system

$$\hat{\mathbf{A}}\mathbf{w} = \hat{\mathbf{b}}, \tag{3.28}$$

where $\hat{\mathbf{A}} = \mathbf{P}^{\frac{1}{2}}\mathbf{S}\mathbf{P}^{\frac{1}{2}}$ is the preconditioned matrix, $\mathbf{w} = \mathbf{P}^{\frac{1}{2}}\mathbf{x}$ and $\hat{\mathbf{b}} = \mathbf{P}^{\frac{1}{2}}$. Preconditioning is beyond the scope of this thesis, however it could provide an interesting avenue for further work.

**Tuning $\lambda$**

To understand better the role of $\lambda$ on optimisation procedure, multi-site calibrations were conducted over the daily data for different values of $\lambda$. Daily data were used since this is where the collapse of posterior distribution was most obvious. The posterior uncertainties were found to be highly unconstrained for $\lambda = 0$ and extremely tight for $\lambda = 1$, therefore investigations were carried out over this range.

In a perfect optimisation, at the optimum, the gradient is zero and for a non-zero Hessian, the curvature matrix is positive definite (if the Hessian is zero, the optimum is ensured if the least non-zero derivative is even-numbered and positive). However, since the optimisations are performed on a computer with limited numerical precision, the termination criterion for the optimisation is when the gradient reaches a very low threshold and so the Hessian is not guaranteed to be positive definite.

The Hessian describes the curvature of a locally fitted quadratic at the optimum. If the Hessian is not positive definite at the optimum, it is likely that one or more slices of the cost function through parameter space will be flat or have a slight negative curvature. These are the parameters that are ill-constrained. When generating the posterior uncertainties using this Hessian, these directions of non-positive curvature through parameter space are made slightly positive, in order to make the Hessian positive definite. This does not change the uncertainty associated with these parameters very much since they will still be seen to be highly unconstrained.

For low values of $\lambda$, the majority of parameters were found to be hitting the boundaries of the prescribed ranges and the Hessian was found not to be positive definite at the optimum. When $\lambda$ was increased, the optimal parameters would move away from the bounds, and for high enough $\lambda$, the numerical Hessian would be positive definite. A non positive definite Hessian could also be made to become positive definite by removing the offending parameters from the matrix. The number of parameters that needed to be removed in order to achieve this also decreased as $\lambda$ increased. In both cases, the relationship was linear; once the Hessian became positive definite at the optimum it remained so.

When considering the condition number of the Hessian at the optimum, for low values of $\lambda$, $\kappa(\mathbf{C}) = \mathcal{O}(10^{13})$. The lowest value of $\lambda$ for which Hessian was found the positive definite gave $\kappa(\mathbf{C}) = \mathcal{O}(10^{5})$. In the range tested, once the condition number had dropped down, increasing $\lambda$ did not lower this order of magnitude any further. Though the condition number is improved, it remains high.

For each of the daily multi-site experiments, the lowest value of $\lambda$ such that the Hessian is positive definite at the optimal parameter value was selected, denoted $\lambda_{\text{opt}}$. This allows

uncertainties to be generated around each parameter (albeit some relatively tight) but prevents the gradient descent algorithm from reaching the boundaries of the prescribed prior range and also ensures a (relatively) low condition number. The values of $\lambda_{\text{opt}}$ were found to be lower for multi-site optimisations with many sites than those performed over fewer sites. This means that only a low weighting of the background term was needed to condition the problem. Given more time, this would have been an interesting relationship to investigate further.

In Raoult et al. [2016], $\lambda$ was manually tuned in this manner for each of the multi-site optimisation in order to prevent the strong collapse in posterior distributions discussed above. Due to improvements and corrections made to the code since, these results have been updated for this thesis. These are presented in Chapter 4. The main change is from bi-monthly data used in the paper to monthly data presented in this thesis. The updated calibrations are performed over fewer data points, and as such the posterior collapse is less pronounced. Due to time constraints and a lack of understanding of how best to weight the background term, Chapter 4 only considers $\lambda$ values 0 and 1.

### 3.2.4. Extra comments on the multi-site implementation

The two experiments described above point to a 'smoothing' hypothesis. In Sect. 3.2.2, the parameter vectors found over multiple sites were sometimes found to outperform the locally optimised parameter vectors. In these cases, a different and better minimum had been discovered. In Sect. 3.2.3, the addition of more sites to the optimisation meant that only a low-weighted background term was needed to help condition the problem. The additional data provided by the extra sites replaced the extra constraints given by the background term.

The idea of 'smoothing' was first suggested in Kuppel et al. [2014]. This is the idea that the added constraints placed on the parameters by increasing the number of sites causes the cost function to become 'smoother'. As a result, this may render the optimisation scheme less likely to become trapped in local minima. One of the motivations for the next section is to test this hypothesis.

## 3.3. Testing the sensitivity to initial conditions

In this section, adJULES is initialised with different parameter vectors to test the sensitivity of the system. In these experiments, Latin hypercube sampling (LHS) is used to generate different starting parameter vectors. LHS is a statistical method for generating near-random samples of parameter values from a multidimensional distribution. It aims to spread the sample of points as evenly as it can across all possible values. This is achieved by partitioning each input distribution into a given number of intervals of equal probability and selecting one sample from each interval. This ensures that the whole of parameter space is sampled.

### 3.3.1. Initial cost

Using LHS, 100 different starting parameter vectors were selected. The initial cost using each of these vectors was calculated at each of the broadleaf sites and compared to the cost at the local optimum.

For 22/28 of the sites tested, each locally optimised parameter vector performed better than any of the 100 randomly sampled parameter vectors. For a further two sites, only 1/100 randomly sampled parameter vectors was found to give a lower error. For site PT-Esp, this number was increased to 2/100. In each of these cases, the random parameter vector was different.

For the last three sites, a larger number of random parameter vectors outperforms the locally optimised ones. Two of the these sites, BR-Sa1 and ID-Pag, have extremely high and unrealistic initial fractional errors $\epsilon$. These sites are seen to change very little when parameter vectors are changed. These sites do not truly improve when confronted by new parameter vectors, therefore the locally optimised vector and the randomly generated set of parameter vectors are comparable. For these sites to improve, structural changes need to be made to the model.

The final site to discuss in this experiment is US-UMB, for which 11% of the randomly generated parameter vectors outperform the locally optimised one. It is not clear why this site is anomalous. These 11 parameter vectors were found to give similar reduction in errors to that of the optimised one, on average adding an extra reduction of 3% to the 15% reduction achieved by the locally optimised parameter vector.

This result gives confidence in the adJULES system. If it were possible to sample randomly in parameter space and perform better than with an optimisation scheme, the system would quickly become obsolete.

### 3.3.2. Sensitivity tests

For this experiment, LHS was used to select 25 different starting parameter vectors. These were used to generate the initial monthly JULES run at each of the 18 deciduous broadleaf tree sites. The difference between these runs and the observations at each site was used as the basis of the cost functions minimised.

Calibrations using each of these 25 different starting parameter vectors, along with the default JULES starting parameter vector, were performed both locally at each of the sites and over the whole of the 18 site subset. This resulted in 26 optimal parameter vectors at each of the 18 sites. These calibrations were performed twice, once without a background term in the cost function and once with.

Figure 3.7 shows the overall RMSE reduction for which all of the optimised parameter vectors are responsible. The spread of reductions is much larger for the experiments conducted with the background term than without. This may seem counterintuitive since

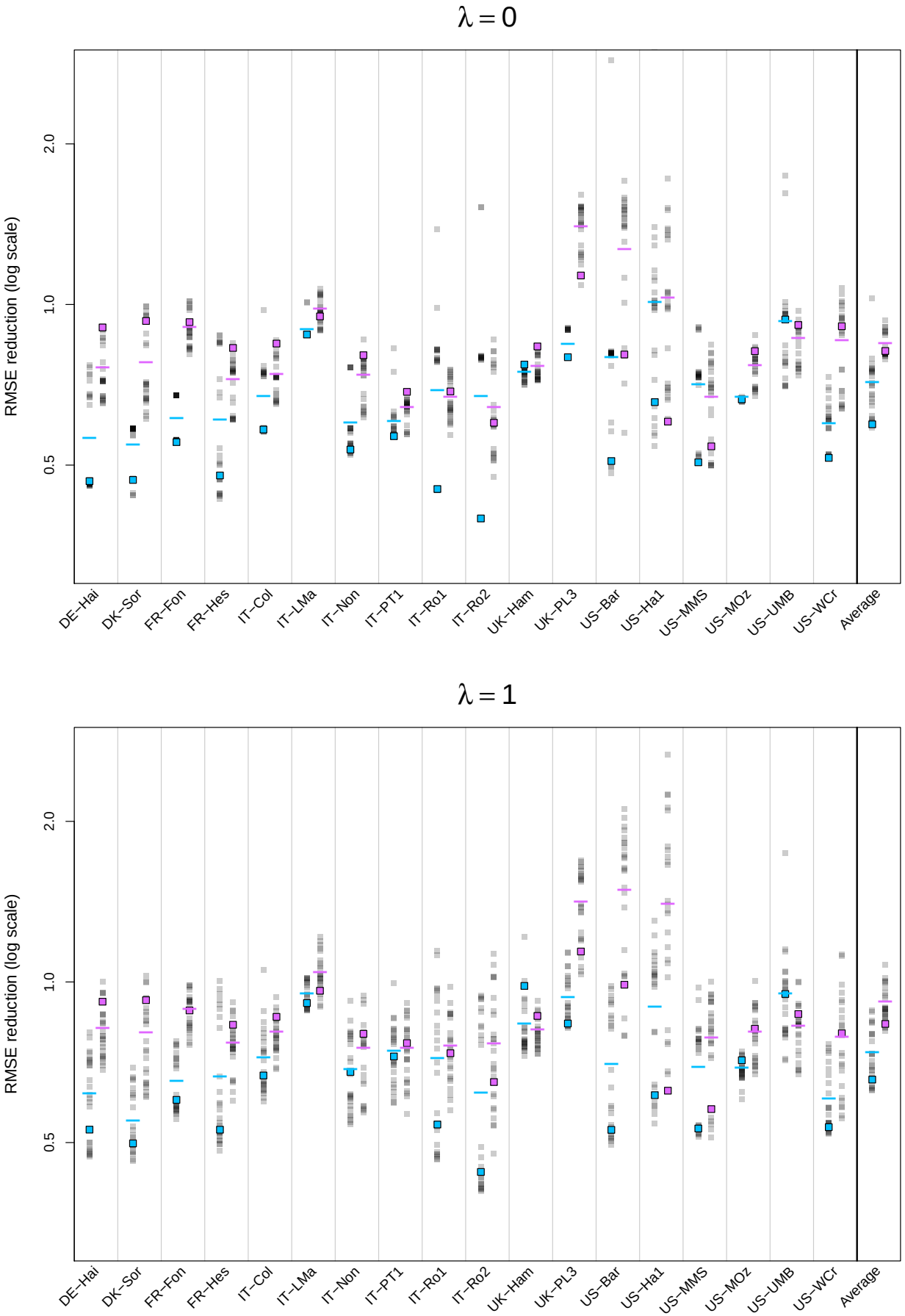



Figure 3.7.: Averaged model-data RMSE reduction of the GPP and LE fluxes. Calibrations ran both without a background term (upper panel) and with it (lower panel). Results are shown both for the locally optimised parameter vectors (on the left) and for the multi-site parameter vectors (on the right). The parameter vectors optimised starting from the default JULES parameter vectors are highlighted (coloured box; blue-single, purple-multi) as is the mean reduction at each site (horizontal bar; blue-single, purple-multi). Average column corresponds to the total average reduction for which each starting vector is responsible.

including the background term is thought to ensure uniqueness. However, since the background term also acts as a penalisation term, it stops the optimisations from straying too far from the initial parameter values, which in this experiment are all different.

For the $\lambda = 0$ experiment, the locally optimised parameter vectors starting from the default JULES parameter vector perform the best for the majority of the sites. This is not the case for the $\lambda = 1$ experiment. For the majority of the sites however, the parameter vector optimised by starting from the original JULES parameters does give RMSE reductions below the average.

The multi-site parameter vector optimised from the default JULES starting values does not appear to be performing the best of the multi-site parameter vectors. Indeed, for most sites this value is higher than the average of all reductions. However, since the multi-site parameter vector was found over multiple sites, it makes more sense to consider the overall RMSE reduction across all the sites. This is shown in the average column of the figure. In both experiments, this has a small spread.

If the addition of more sites tended to smooth the cost function, the spread in the possible multi-site RMSE values would be smaller than the spread from the single-site RMSE values. When considering the individual sites, this is true for 10/18 of the sites shown for $\lambda = 0$, and 8/18 of the sites when $\lambda = 1$. When considering the overall average, the multi-site spread is smaller than the local one.

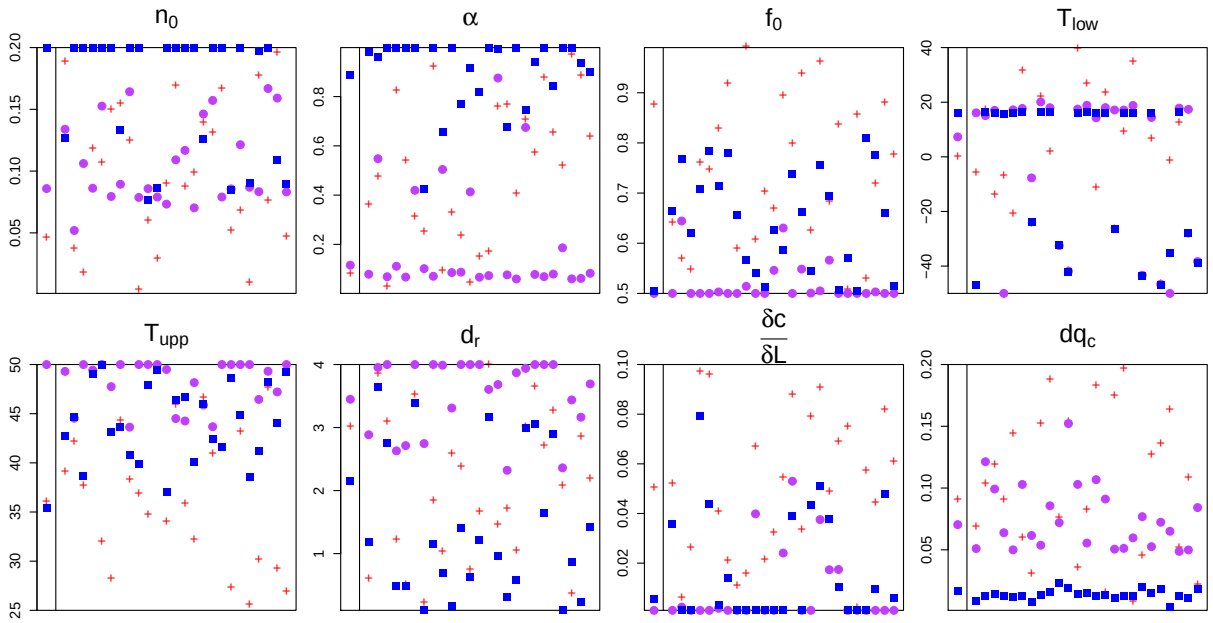
Figure 3.8 considers the optimal parameter values themselves. When the background term is excluded, many of the parameters hit the bounds. This happens consistently regardless of the initial parameter vector. The $\alpha$ parameter can be seen to hit the upper bound when locally optimised, whereas when part of the multi-site optimisation, this parameter is found near the lower bound.

For both experiments, the $T_{\text{low}}$ parameter seems to opt for a value just under 20 for the majority of starting vectors. The $\frac{\delta c}{\delta L}$ parameter hits the lower bound when the background term is excluded. When the background term is added to the cost function, this parameter does not move far from its starting value.

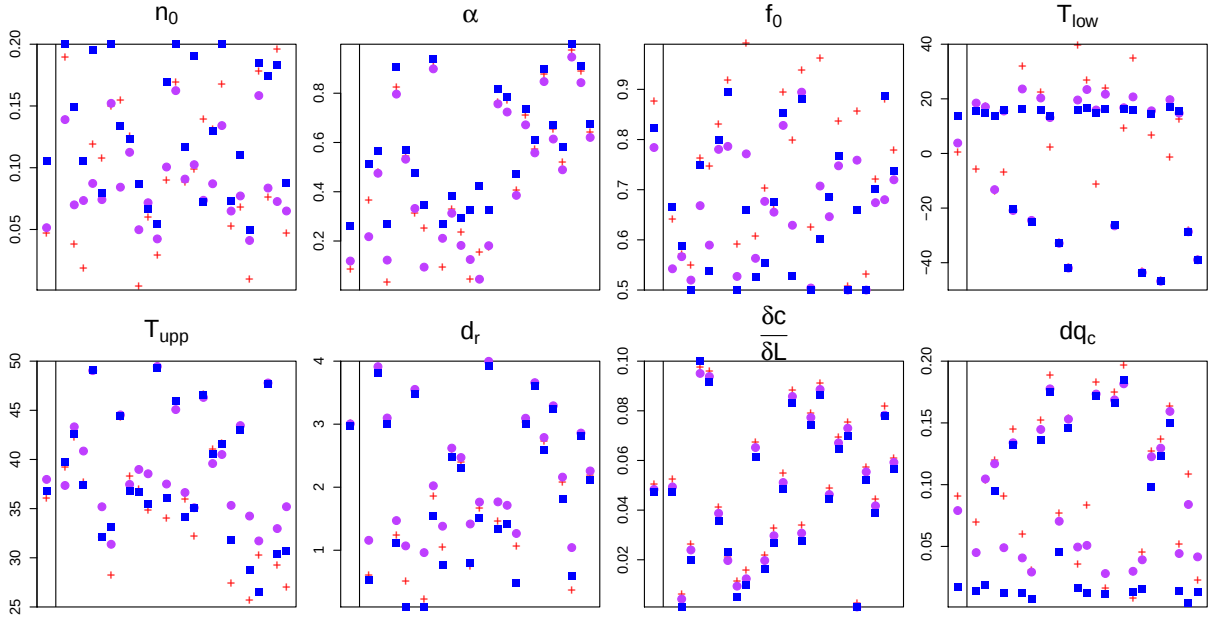
Due to the distribution of possible optimal parameter vectors, especially when $\lambda = 1$, the system is seen to be sensitive to initial conditions. Though the parameter values shown were taken at one specific site, the results are similar across the different sites. In this experiment, it is not possible to conclude whether the addition of the multiple sites to the optimisation scheme is 'smoothing' the cost function as the distribution of the multi-site parameters are as varied as the single-site parameter values.

## 3.4. Closing remarks

The multi-site framework developed in this chapter is shown to be successful and robust. It has the ability to calibrate over multiple sites simultaneously, and a parameter vector




(a) Optimisation performed without a background term ($\lambda = 0$)




(b) Optimisation performed with a background term ($\lambda = 1$)

Figure 3.8.: The optimal parameter values found after performing optimisations over German site DE-Hai, starting with the default JULES parameters (left-hand side separated by the vertical line) and the 25 randomly generated vectors in order along the $x$-axis. Results for single-site optimisations (blue) and multi-site optimisations are shown (purple) vertically lined up with the initial value for each parameter (red crosses). The $y$-axis span the prescribed parameter ranges.

found over a small subset of sites can be generalised to bigger sets.

The development of the multi-site framework has, however, opened up more research questions around the weighting and definition of the background matrix **B**. The addition of more sites to the cost function is seen to add more constraints, on top of the ones controlled by the number of data points in the optimisation and the relative strength of this background term. More work is needed to understand the relationship between these three factors and the constraints they place on the cost function.

In Chapter 4, the multi-site framework developed in this chapter is used to find new PFT-generic parameter vectors for the JULES model. The locally-optimised parameter vectors generated at each site are also used to re-examine the PFT groupings and to consider other ways in which vegetation could be partitioned.

# 4. Improving the Plant Functional Types in JULES

In this chapter the main results of the adJULES optimisations are considered. As described in Chapter 2, the parameters optimised in this thesis relate to the plant functional types (PFTs) in JULES. The parameters are optimised with the aim of improving the PFTs' representations within the model.

In Sect. 4.1, the ability of the adJULES system to calibrate at specific FluxNet sites is investigated. Referred to as single-site optimisations, these are shown to perform well. The main improvement trends for each PFT are discussed. The potential to generalise the parameter vectors found at each single-site over each of the PFT is explored in Sect. 4.2. One way this is achieved is by averaging the single-site results, the other is by choosing one representative vector to describe the whole PFT.

In Sect. 4.3, multi-site optimisations are performed over each PFT. This gives a more robust and objective way of finding a new parameter vector for a given PFT. The uncertainties associated with each parameter and the correlations between parameter pairs are discussed, as well as assessing the improvement in model-data fit at each site (Sect. 4.3.1 and 4.3.2).

The chapter concludes by considering other ways to improve the PFTs in JULES. First, this is done by looking at some of the structural changes in Sect. 4.4, namely changes in the canopy representation and the recent work by Harper et al. [2016]. Second, the PFT groupings are reconsidered in Sect. 4.5. Considering the single-site optimisations, clustering algorithms are used to see if the sites fall into natural groupings.

The experiments in the first half of this chapter are an updated version of the published results in Raoult et al. [2016]. Some of the results differ slightly due to minor improvements made to inconsistencies found in the code; the averaging window was not quite monthly and the atmospheric carbon concentrations were too low. Different background weightings ($\lambda$) are considered in this chapter compared to those in the paper for the multi-site optimisations, as discussed in Sec. 3.2.3. Unless stated otherwise, the background term is included in the following optimisations, i.e. $\lambda = 1$, in order to give equal weighting to both terms in Eq. 2.55.

## 4.1. Single-site optimisations

First, each site was optimised individually in order to find site-specific parameter vectors. Typically, this required about 150 function evaluations to find a local optimum. In these experiments, 1-year runs at the different sites were optimised against monthly averaged latent heat (LE) and GPP. A site dominated by each PFT was picked to represent the general improvements made. The main seasonal cycles of LE and GPP for the different sites are shown in Fig. 4.1. The rest of the sites can be found in B

Most broadleaf sites follow the pattern illustrated (Fig. 4.1, top row). Normally, for broadleaf sites, a standard JULES run will underestimate GPP. The optimisation does a good job in correcting this, bringing the modelled time-series closer to the observations. In contrast, LE does not improve as much.

Similarly, for the needleleaf sites (Fig. 4.1, second row) the JULES model output tends to overestimate LE and underestimate GPP. The parameter vector found in the optimisation improves the fit of both data streams, most notably for GPP. At sites in which a double peak seasonality is apparent, the optimised model captures this better than the original model.

GPP is also underestimated for the C3 grass sites (Fig. 4.1, middle row) and, for the majority of the sites, the optimisation does a good job of correcting this. The LE flux tends to have the right magnitude before optimisation, unlike the GPP flux, but adJULES does not manage to improve this output significantly. In the example shown, the JULES model using the default parameter vector already performs very well, so little improvement is possible, but this is not always the case. The new set of parameters is also good at simulating multiple peaks in the LE and GPP fluxes, when they are observed.

There are only two C4 grass sites in the set and JULES does not perform very well on these before or after optimisation (Fig. 4.1, fourth row). The original stomatal conductance–photosynthesis model within JULES was developed based on fluxes measured over C4 grass as part of the FIFE field experiment [Cox et al., 1998]. However, there are relatively few FluxNet sites over C4-dominated landscapes, and only two in the extended data set used here. As a result, the sensitivity of stomatal conductance and photosynthesis to environmental factors has been less well tested for C4 grasses. These results highlight the need to reassess JULES and other land-surface models for predominantly C4 landscapes.

The shrub sites show no general pattern (Fig. 4.1, fifth row). Some sites overestimate LE, whilst others underestimate it, and similarly for GPP. The level of improvement varies over sites. For some, the magnitude of GPP fails to get close to the magnitude of the observations, both before and after optimisation. However, it is hard to pick out a general pattern for this PFT, since there are only five sites in this set.

Overall, the adJULES system works well in finding optimal parameter vectors, which improve the performance of JULES at individual sites, regardless of PFT. The systematic underestimation of GPP in default JULES improves the most. This larger improvement

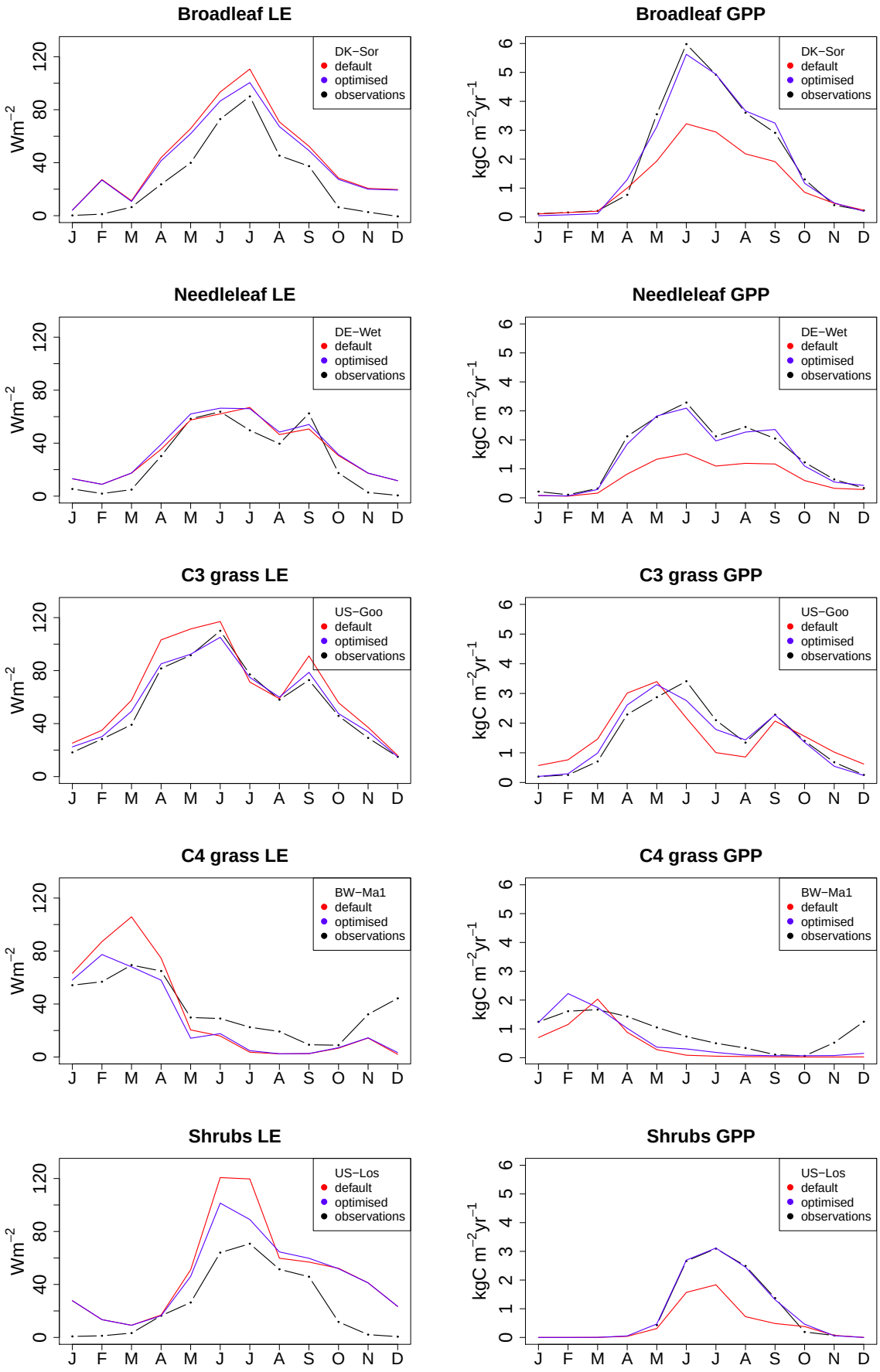



Figure 4.1.: Time-series plots for illustrative site-specific evaluations showing LE (left) and GPP (right) for each of the different PFTs. Observations (black) are compared to JULES runs using default parameters (red) and site-specific optimal parameters (blue).

in the GPP fit reflects the larger set of optimised parameters that are exclusively related to the carbon cycle. Different parameters may need to be incorporated, for example some relating to soil, in order for the LE flux to improve further. In fact, calibrating against LE and GPP fluxes without a full set of parameters controlling LE risks changing the results compared to a GPP only calibration in a non-meaningful way.

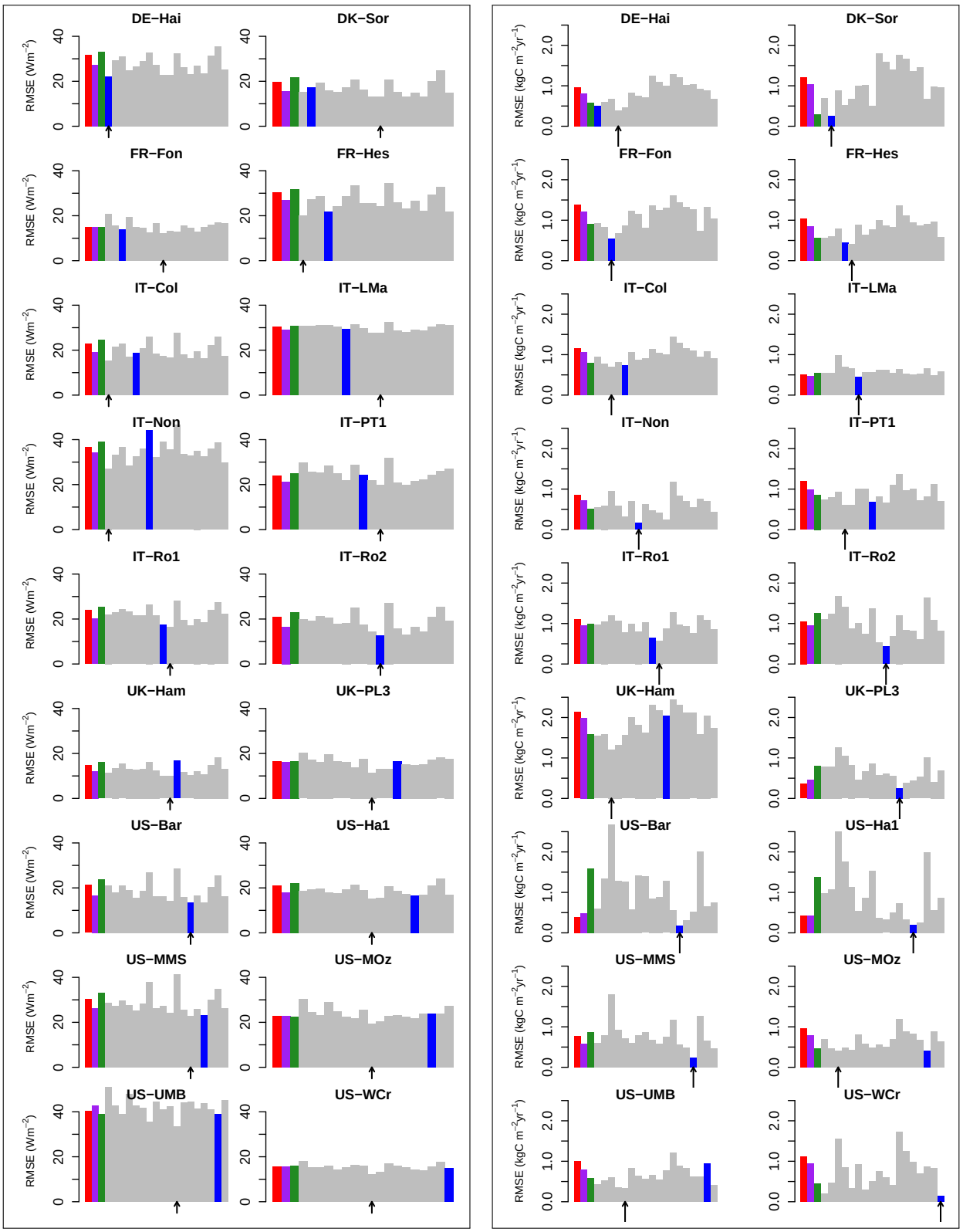
## 4.2. Moving towards a generic parameter vector to describe each PFT

One of the motivations for this work is to generate an optimal parameter vector to best describe each PFT. This improved parameter vector can be found in several ways. One approach is to transpose optimal parameter vectors found at one site to other sites with the same plant functional type. Another is to average all the optimal parameter vectors from the whole PFT. Finally, a generic parameter vector can be found by performing a multi-site optimisation for the PFT. Note that the averaged parameter vector is different to that found by performing a multi-site optimisation (as discussed in Sect. 3.2.2). A multi-site calibration of the model is more involved, and so the hope is that this parameter vector is the most efficient at reducing errors in the model-data fit. The experiments conducted in this section are similar to those found in Kuppel et al. [2012].

Due to time and computational constraints, this section focuses on the deciduous broadleaf subset. Figure 4.2 shows the model-data RMSEs for LE and GPP at each of these sites. The bars shown at each site result from applying different parameter vectors. First, the default JULES parameters are used to run the model (shown in red). This is compared to runs using the multi-site parameter vector and the average parameter vector (shown in green). The average parameter vector is calculated by taking the mean of the single-site optimisation vectors. Finally, the optimal single-site parameters are applied (shown in grey, and highlighted blue when the parameters were optimised at that particular site). Note that even though only the deciduous subset of the broadleaf sites is considered in this figure, the multi-site parameter vector was calculated over all of the broadleaf sites. Similarly, for the average parameter vector, the mean was taken over all the broadleaf sites.

The single-site parameter vector found at each site is expected to perform the best, however, this is only the case for 3/18 of the sites in Figure 4.2a. With the exception of US-UMB which has a relatively high LE RMSE, the multi-site parameter consistently improves the model fit to the LE flux. In contrast, the averaged parameter vector gives a similar or worse fit when compared to the default parameters.

For a few sites, most notably IT-Non, the single-site parameter vector optimised at that site worsens the model's fit to the LE flux but significantly improves the model's fit to the GPP flux. This, and the relatively low overall reduction seen in the LE fluxes, is due to the small number of LE-related parameters optimised in this study. When several fluxes




(a) Latent Heat

(b) GPP

Figure 4.2.: Monthly model-data RMSEs for LE and GPP. For each site, the prior model (red) is compared to runs found using the parameters from single-site optimisations performed locally (blue) and at the other sites (grey). These grey bars are ordered in the site order (as listed in Table. A.1), highlighted blue when at the corresponding site. Two further vectors are considered: the parameters found from optimising over all the broadleaf sites simultaneously (purple) and the parameter vector found by averaging all of the single sites (green). The arrows highlight the parameter vectors resulting in the largest RMSE reduction.

are being optimised simultaneously, it is possible to improve the fit to one flux whilst degrading the fit to another and still have an overall reduction in total RMSE.

For the GPP flux shown in Fig. 4.2b, over half of the sites now perform best with their locally optimised parameter vector. Transposing the single-site parameter vectors most often results in significantly higher GPP RMSE values than the site-corresponding parameter vector. In general, the single-site parameter vectors are not generic enough to be transposed to other sites. The multi-site parameter set tends to improve the fit at each site (improves the fit for 15/18 of the sites) though to a much lesser extent than the site-specific set. The averaged parameter vector performs very well for some sites, in fact it improves 11/18 of the sites to a greater extend than the multi-site vector. For a few of the sites however, e.g. US-MSS and US-MOz, the averaged parameter vector performs very poorly.

Using the FVU metric described in Sect. 2.4.2, the total improvement at each site can be quantified. These values can be used to compare the average FVU for which each parameter vector is responsible. The parameter vector calibrated at site US-MMS performs the best overall, closely followed by the multi-site parameter vector. On closer inspection, the US-MSS parameter vector gives a lower average FVU value, but improves fewer of the sites overall. Of the 21 parameter vectors tested (i.e. 18 single-site, one default, one average, and one multi-site), ten perform worse than the default parameters. The average parameter vector performs seventh best.

Therefore, with this metric the multi-site parameter vector performs better than the averaged parameter vector. Even though a single-site parameter vector is seen to perform slightly better using the average FVU metric, the multi-site parameter vector is more consistent in improving all sites. The multi-site parameter vector has been found in a robust and reproducible manner.

An argument for the averaged parameter vector could also be made since it is seen to perform reasonably well. The two sites it performs worse at warp the average FVU metric. These two sites do start off with relatively low errors compared to the other sites. The advantage of the averaged parameter vector that it is simpler to calculate, however, given the non-linearity of the Earth System, it is possibly less mathematically robust. With the averaged parameter vector, it is also harder to calculate the posterior distributions associated with each parameter. The posterior distributions calculated when using the multi-site parameter vector are found by used the second derivative of the multi-site cost function. The averaged parameter vector however does not have one total cost function - the posterior distributions are calculated at each site. Some mechanism for combining these individual posterior distributions would be needed to find the error of the averaged parameter vector. This could be an avenue for future work, however, in this thesis, the multi-site optimisation is used.

## 4.3. Multi-site optimisations

Optimisations were performed over all available sites for each of the PFTs. The optimised model parameters for each of the PFTs are presented in Fig. 4.3, both for experiments with a background term and experiments without a background term. The background term refers to the prior distribution used in the cost function: the prior distribution is assumed to be proportional to one over the prescribed range for each parameter when the term is included, and a non-informative top-hat prior is assumed when the term is excluded (discussed at length in Sect. 3.1.3).

Generally, the parameters found when optimising without the background tend to have larger uncertainties than those found in the experiments with the background term. The clearest example of this is found in the $\frac{\delta c}{\delta L}$ parameter. This parameter, which determines the efficiency of rainfall interception by the plant canopy, is completely unconstrained when $\lambda = 0$. The 80% confidence interval spans the whole box and, for the broadleaf vector, this parameter hits the bounds. However, when the background term is included, this parameter does not change significantly from its original value for any of the PFTs. The uncertainty bounds are relatively tight and symmetrical. The rest of the parameters show more variation.

As described in Sect. 2.4.1, the optimal values need not be in the centre of the uncertainty range, the probability density function can be skewed. This can be seen for most of the parameters in the needleleaf parameter vector ($\lambda = 1$); the optimal parameters tend to be at the lower end of the confidence interval.

Most of the time the background term, acting as a penalising term, causes the optimal parameters to remain closer to the prior value than in experiments without the term. However, there are some where this is not the case; for example the $dq_c$ parameter for the needleleaf trees and C4 grasses. This is due to the fact that the parameter belongs to a vector of parameters which is changing during the optimisation and therefore some parameters will end up contributing more to the penalisation term than others.

Even for the penalised experiments (i.e. with $\lambda = 1$), the PFTs display high uncertainty in at least one of the parameters optimised; for the optimised broadleaf set for example, $T_{\text{low}}$ is noticeably unconstrained. For C4 grasses, $d_r$ is so unconstrained that the optimal value found lies outside the 80 % confidence interval. Needleleaf trees show large uncertainty in $dq_c$, whereas the C3 grasses show large uncertainty in $n_0$ and $T_{\text{upp}}$. For shrubs, the parameter with the largest uncertainty is $n_0$.

Some of these uncertainties can be explained by referring back to the JULES equations found in Sect. 2.1.2. Consider $V_{\text{cmax}}$ (Eq. 2.5), one of the main components for calculating photosynthesis rates. This is controlled by $T_{\text{low}}$, $T_{\text{upp}}$ and $n_0$. $V_{\text{cmax}}$ curves for a variety of $T_{\text{low}}$ values are shown in Fig. 4.4. $T_{\text{low}}$ only affects the left hand slope of the $V_{\text{cmax}}$ functional graph. Increasing the value of $T_{\text{low}}$ causes the lower end of the curve to become steeper. In temperate and tropical regions where temperatures do not go very low, the bottom half of the curve is not sampled. Without knowing the shape of the bottom

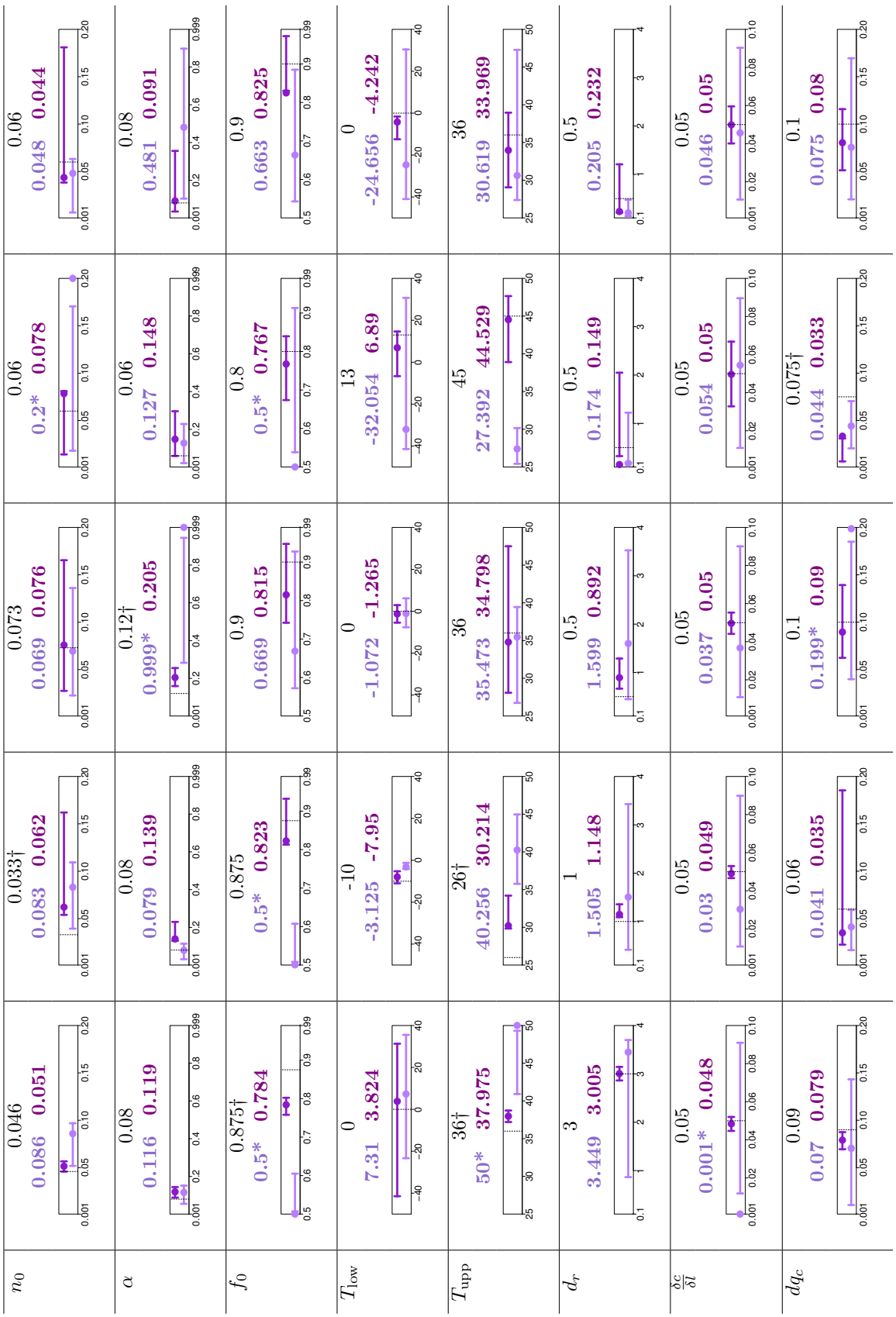



Figure 4.3.: The new parameter values found by optimising over each PFT using a cost function without background term i.e. $\lambda = 0$ (light purple), and a cost function with the manually weighted background term (dark purple). The new parameter values hitting the prescribed ranges are highlighted with an asterisk (∗). The prior value for each parameter is found on top, † denotes cases were the initial value is outside both of the new uncertainty bounds. The error bars show the uncertainty ranges given as an 80% confidence interval. The range of each box is the prescribed range the parameters were allowed to vary over and the vertical lines show the initial value for each parameter.

of the curve, the value of $T_{\text{low}}$ cannot be determined. This could explain why $T_{\text{low}}$ is unconstrained for the broadleaf trees, whereas for needleleaf sites, which exist in much colder climates, the parameter is highly constrained.

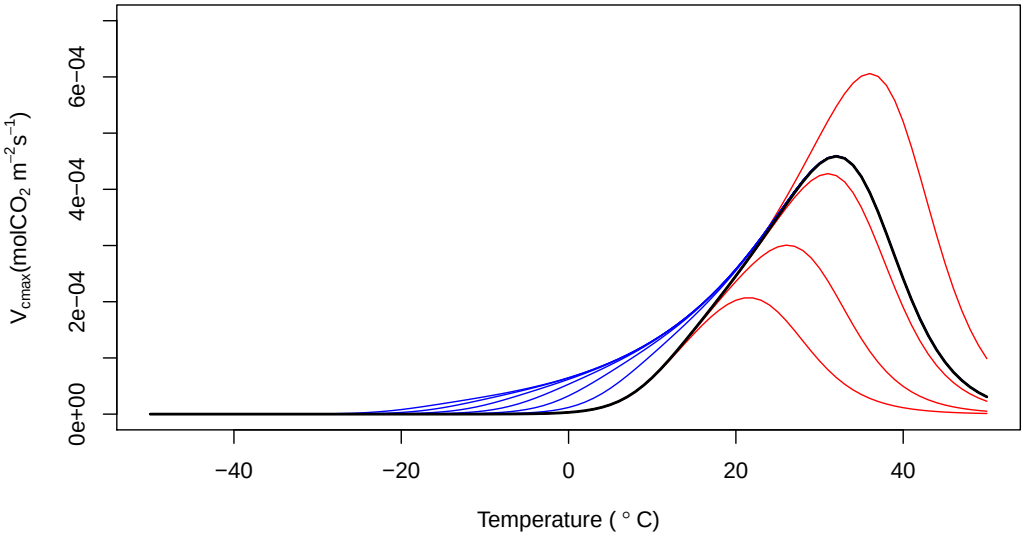
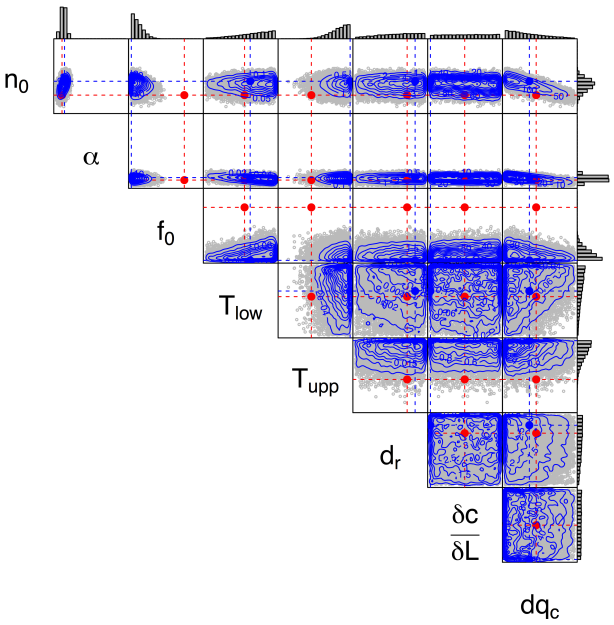



Figure 4.4.: Functional graph of $V_{\text{cmax}}$ against temperature (T). Black line shows the broadleaf default curve with $T_{\text{low}} = 10$,$T_{\text{upp}} = 36$°C and $n_0 = 0.46$. The blue lines show varying values of $T_{\text{low}}$ ranging from -20°C to 10°C increasing in increments of 5°C, and the red lines show varying values of $T_{\text{upp}}$ ranging from 25°C to 40°C. Decreasing values of $T_{\text{low}}$ flatten the LHS of the curve, increasing values of $T_{\text{upp}}$ push up the peak of the curve.

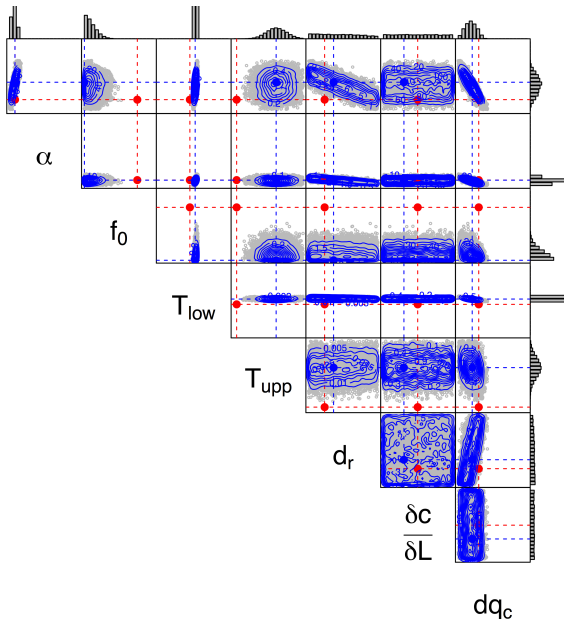
The new PFT-generic parameters are taken from the experiments with the background term. Overall, $\alpha$ is constrained for all of the PFT, opting for a low value. The root depth parameter ($d_{\text{r}}$) remains high for trees and low for shrubs and grasses. Note that for some of the parameters, the prior value lies outside the posterior uncertainty bounds.

These results can be used to improve the JULES model. The results suggest for example that $f_0$ needs to be set lower than currently in the model. The fact that the parameter is hitting the lower bound for more the PFT may also highlight the need to reexamine the equations it is found in. Similarly for $T_{\text{upp}}$ in broadleaf trees, when the experiment is unconstrained, the parameter hits the top bound. From Fig. 4.4, $T_{\text{upp}}$ can be seen to control the peak of the curve. $T_{\text{upp}}$ is increased in the optimisation the fix the underestimation in GPP. To keep $T_{\text{upp}}$ within the realistic range, moving from the big-leaf light saturated model to a multi canopy model could help constrain $T_{\text{upp}}$ by more often pushing the system into light-limited regimes.

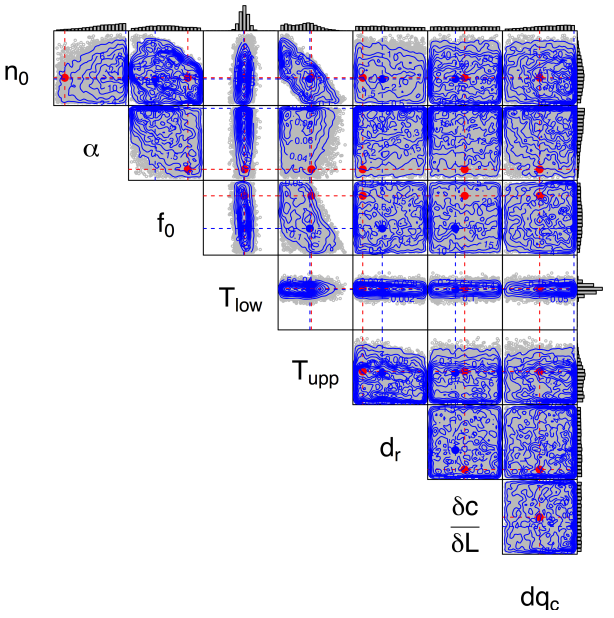
The parameters found during the constrained experiments when be recommended for future runs - the model-data fit for GPP is improved but the parameters are still within the expected ranges. However more experiments are also recommended against more fluxes and with different parameter vectors before became permanent changes in the JULES model.



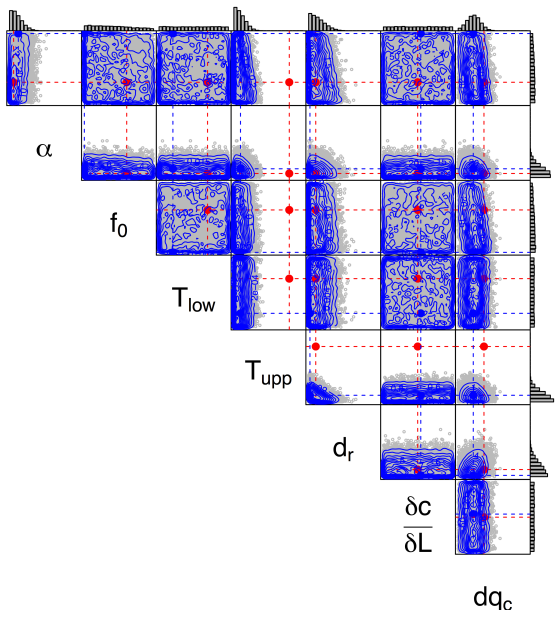

(a) BT



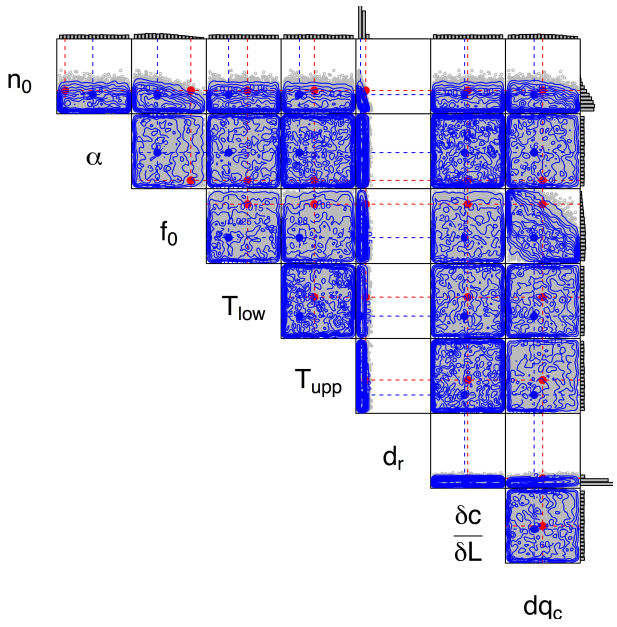

(b) NT




(c) C3

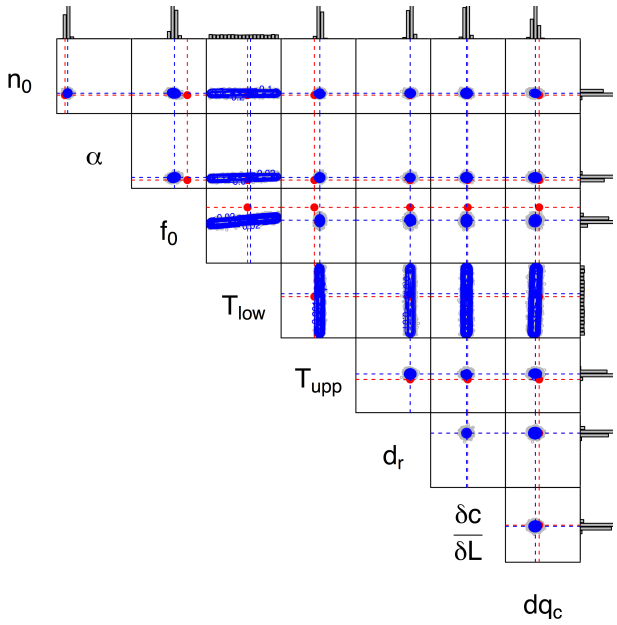



(d) C4

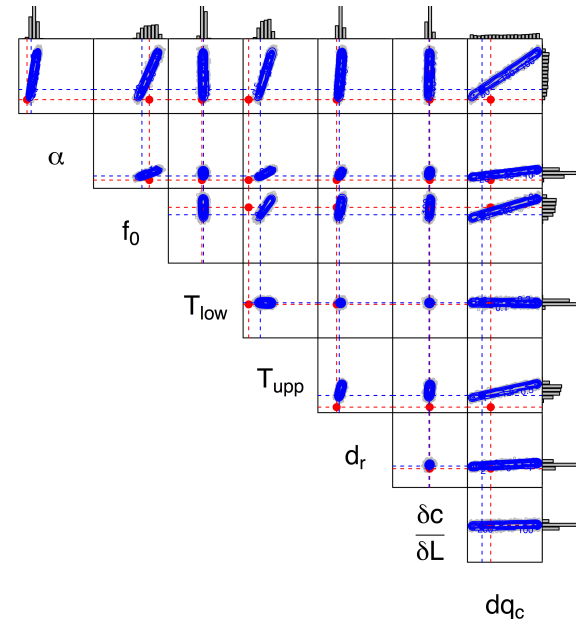



(e) Sh

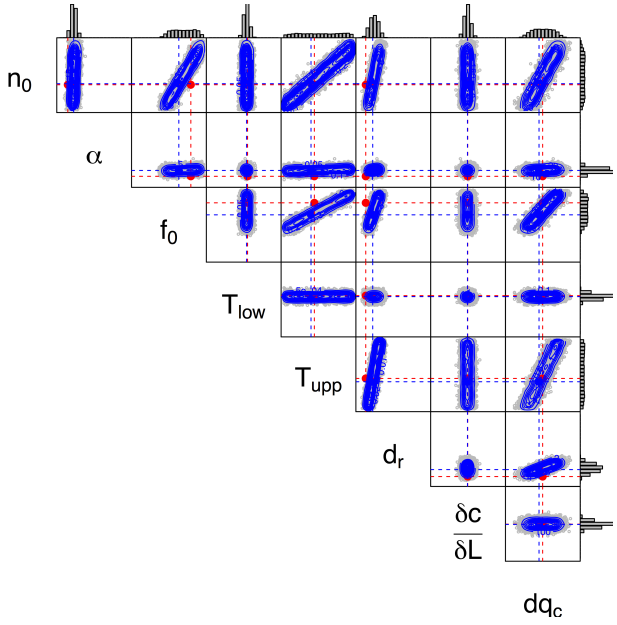
Figure 4.5.: The correlations between parameters for PFT-specific parameter optimisations found in experiments excluding the background term ($\lambda = 0$). Each subfigure shows a 2-D correlation map, within which each box is a 2-D marginal plot. Bar graphs show 1-D marginal distributions for individual parameters. The dimensions of the boxes represent the prior range of each parameter. Red points/dashed lines represent initial parameter values. Blue points/dashed lines represent optimised parameter values. Blue contours illustrate the posterior distribution.



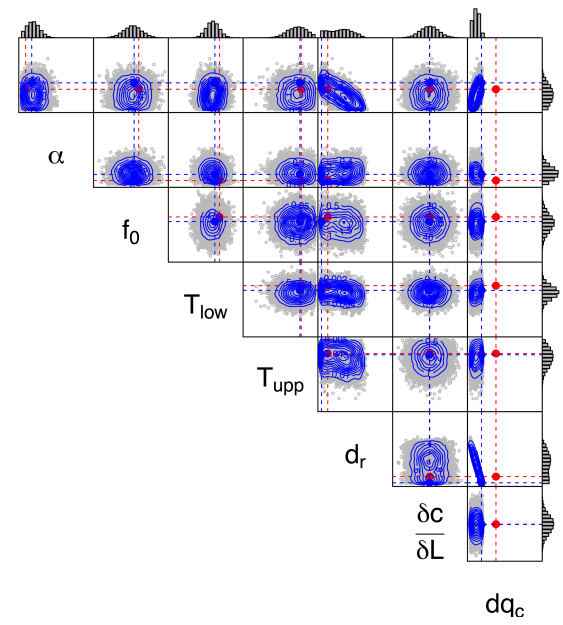

(a) BT



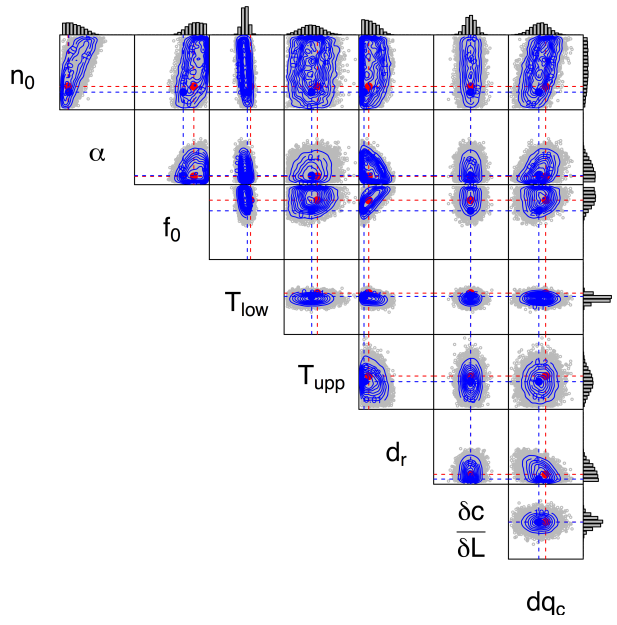

(b) NT




(c) C3




(d) C4




(e) Sh

Figure 4.6.: As in Fig. 4.5 but for experiments with the background term, i.e. $\lambda = 1$

The uncertainties shown in Fig. 4.3 are 1-D marginal distributions. To understand further how the parameters are correlated, consider the 2-D representation in Fig. 4.5 and Fig. 4.6.

In Fig. 4.5, when the background is excluded, the parameters can be seen to be unconstrained with little or no correlation between parameters. This is especially apparent in the case of the grasses and shrubs. The tree sites show more structure. As discussed in Sect. 3.2.3, this could be linked to the fact that there are more sites in these PFTs placing more constraints on the optimisation than for the non-tree PFTs.

Now consider Fig. 4.6, when the background is included. For all of the PFTs, the posterior parameter uncertainties exclude a large part of the prior ranges. The cloud of plausible points tends to be restrictive and tight for most parameters.

Figures 4.5 and 4.6 show clear correlation of some parameters, especially between $n_0$, $f_0$ and $dq_c$. Correlations are most notable for tree sites when $\lambda = 0$, and for needleleaf and C3 grass sites when $\lambda = 1$. Interestingly, the correlation between $n_0$ and $dq_c$ changes sign depending on whether the background is used in the optimisation. This is still not understood.

In the case where the background term is excluded (or with a low weighting as is the case in Raoult et al. [2016]), many of these correlations can be understood in terms of the underlying structure of the JULES model (Sect. 2.1.2). For example, the correlations between these three parameters are consistent with adJULES attempting to fit the stomatal conductance $g$, which controls the transpiration flux from taller vegetation. The stomatal conductance has the approximate form

$$g \approx 1.6 \frac{A}{c_\mathrm{a}} \left( \frac{1}{(1 - f_0) + f_0 \frac{dq}{dq_c}} \right) \tag{4.1}$$

$$= 1.6 \frac{A}{c_\mathrm{a}} \frac{1}{(1 - f_0)} \left( \frac{1}{1 + \left( \frac{f_0}{1 - f_0} \right) \frac{dq}{dq_c}} \right) \tag{4.2}$$

if it is assumed that $c_* \ll c_\mathrm{i}$ and $c_* \ll c_\mathrm{a}$ (this is a combination of Eq. 2.4 and 2.10 using the parameters defined in Table. 2.1).

The maximum rate of leaf photosynthesis ($A$) is controlled largely by the leaf nitrogen content $n_0$, especially in this big-leaf version of JULES (Cox et al., 1999). The best-fit parameters for tree PFTs seem to imply $f_0$ is close to the lower bound set at 0.5. This value eliminates the $f_0/(1-f_0)$ term in Eq. 4.2. As a result, maintaining a realistic $g$ value, and therefore a realistic LE flux, will require that $n_0$ and $f_0$ vary proportionally, and that $n_0$ and $dq_c$ values are negatively correlated. This negative correlation can be seen in Fig. 4.5 for the tree PFTs. This correlation of parameters is less obvious for the grass PFTs because evapotranspiration is controlled less by stomatal conductance and more by the smaller aerodynamic conductances associated with shorter vegetation.

### 4.3.1. Assessment of PFT-specific optimal parameters

The performance of the PFT-specific parameters is now compared to the default JULES values and to the parameters found by optimising independently at each measurement site. For each site, the fractional error in both the calibration year and the evaluation year is displayed Fig. 4.7.

For all sites, the fractional error in calibration years decreases when moving from default to site-specific optimal parameters in the calibration years (with the exception of the IT-Non and UK-Ham sites, which improve when a different metric is applied). Remarkably, the site-specific optimal parameters also improve the model–data fit in evaluation years for 54/64 (84 %) of sites. Similarly, the PFT-specific optimal parameter vector improves the fit (in both calibration and evaluation years) for 85 % of the sites; 76/79 sites for the calibration years and 55/64 sites for the evaluation years.

Consider first the broadleaf sites (Fig. 4.7, top two rows). For the majority of sites displayed in the top broadleaf panel, the reduction in fractional error in moving from default to site-specific optimal parameters is substantial and sometimes as much as a factor of 2. In the calibration year, the PFT-specific optimal parameter vector improves 26 of the 27 broadleaf sites shown, although at one of the sites, IT-Lec, the fit shows no change. The improvement is typically about half as good (on a log scale) as the improvement using the site-specific optimal parameters. In other words, the reduction in fractional error moving from default to PFT-specific optimal parameters is sometimes as much as a factor of $\sqrt{2}$. Amongst broadleaf sites, only UK-PL3 gets notably worse. Investigation shows that this site behaves differently from the rest of the sites in the set, both in the magnitude of the fluxes and seasonality. This UK site is in the Pang–Lambourn catchment, which has chalk soil with macropores that permit significant lateral subsurface flows of soil moisture. These horizontal flows cannot be captured in a model like JULES, which is essentially 1-D in the vertical below the soil surface.

Similar levels of fit and error reduction can be seen in the evaluation years in the broadleaf set. Only IT-Col shows no improvement for either vector, the PFT-specific optimal parameter vector does not worsen the fit at this location. For IT-Non, US-UMB, and IT-Cpz, the PFT-specific parameter vector outperforms the site-specific vector. This illustrates that the PFT-specific vector can be robust, whereas the locally optimised vectors might over-tune to the specific behaviour of the calibration year.

Results are similar for the needleleaf sites, the majority of the sites show noticeable improvements in both the calibration and evaluation years when using site-specific optimal parameter vectors. For some of the sites in this PFT, the improvement when using the PFT-specific parameter vector is similar to, or outperforms, that obtained with the site-specific parameter vector for the evaluation year. This illustrates that this subset of sites fits together well as a single PFT. Some sites in the needleleaf PFT remain unchanged regardless of the parameter vector used. Anomalous sites that should be noted are CA-Qcu, CA-SF3 and US-Blo. The CA-Qcu site is the only one in this PFT that does not

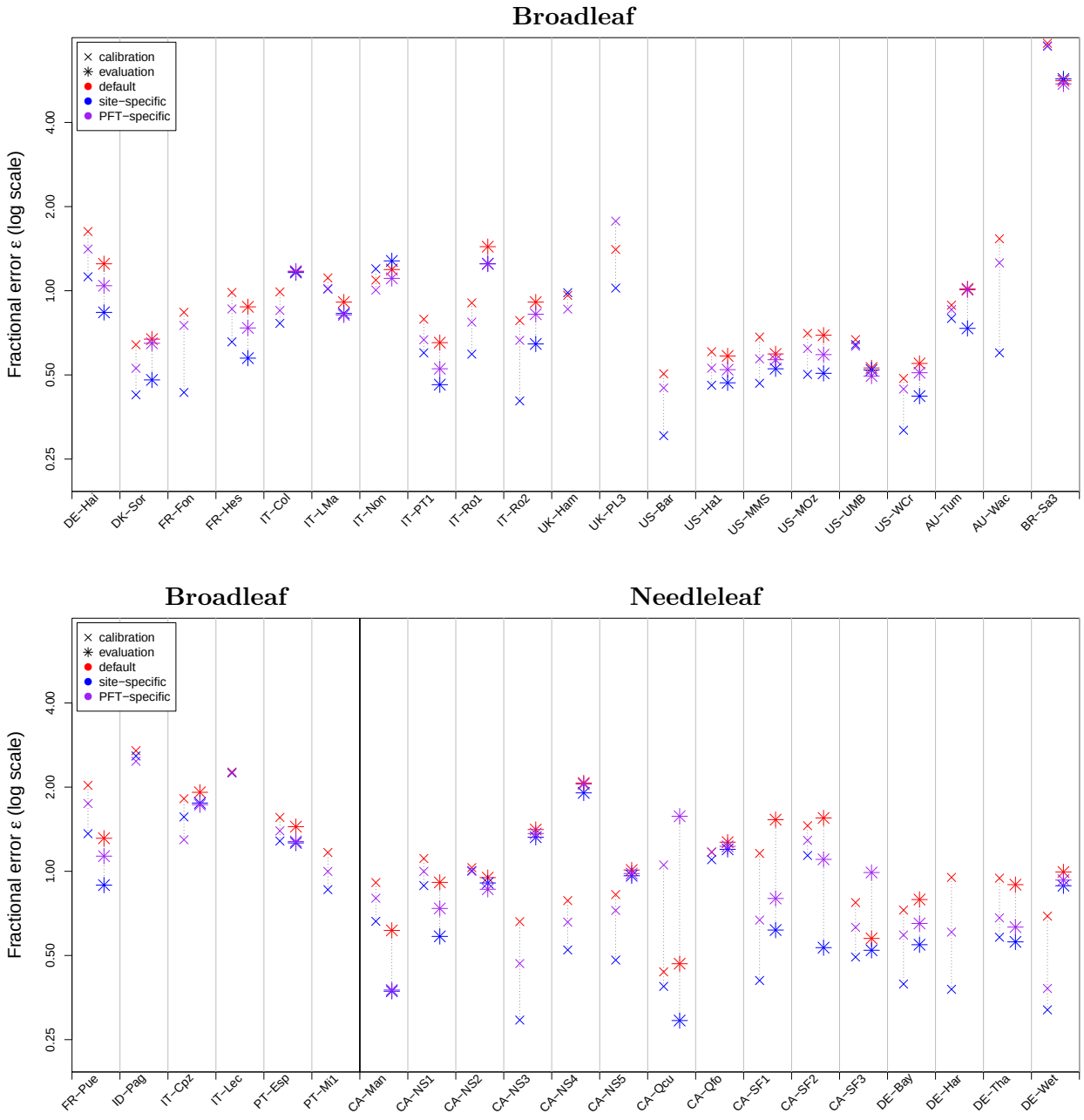



Figure 4.7.: Calibration and evaluation of site-specific and PFT-specific parameter optimisation at FLUXNET sites, using the metric described in Sect. 2.4.2. Fractional error shown for default JULES parameters (red), site-specific optimal parameters (blue), PFT-specific optimal parameters (violet). Results are shown both for the calibration year ($\times$, on left) and for the evaluation year ($*$, on right). No evaluation year was available for some sites (broadleaf: FR-Fon, UK-Ham, UK-PL3, US-Bar, ID-Pag, IT-Lec, PT-Mi1; needleleaf: SE-Sk2, UK-Gri, US-Me4, US-SP1; shrubs: DE-Gri, DK-Lva, PL-wet). Sites with very large initial errors have been removed from the plot (broadleaf: BR-Sa1; shrubs: IT-Pia).

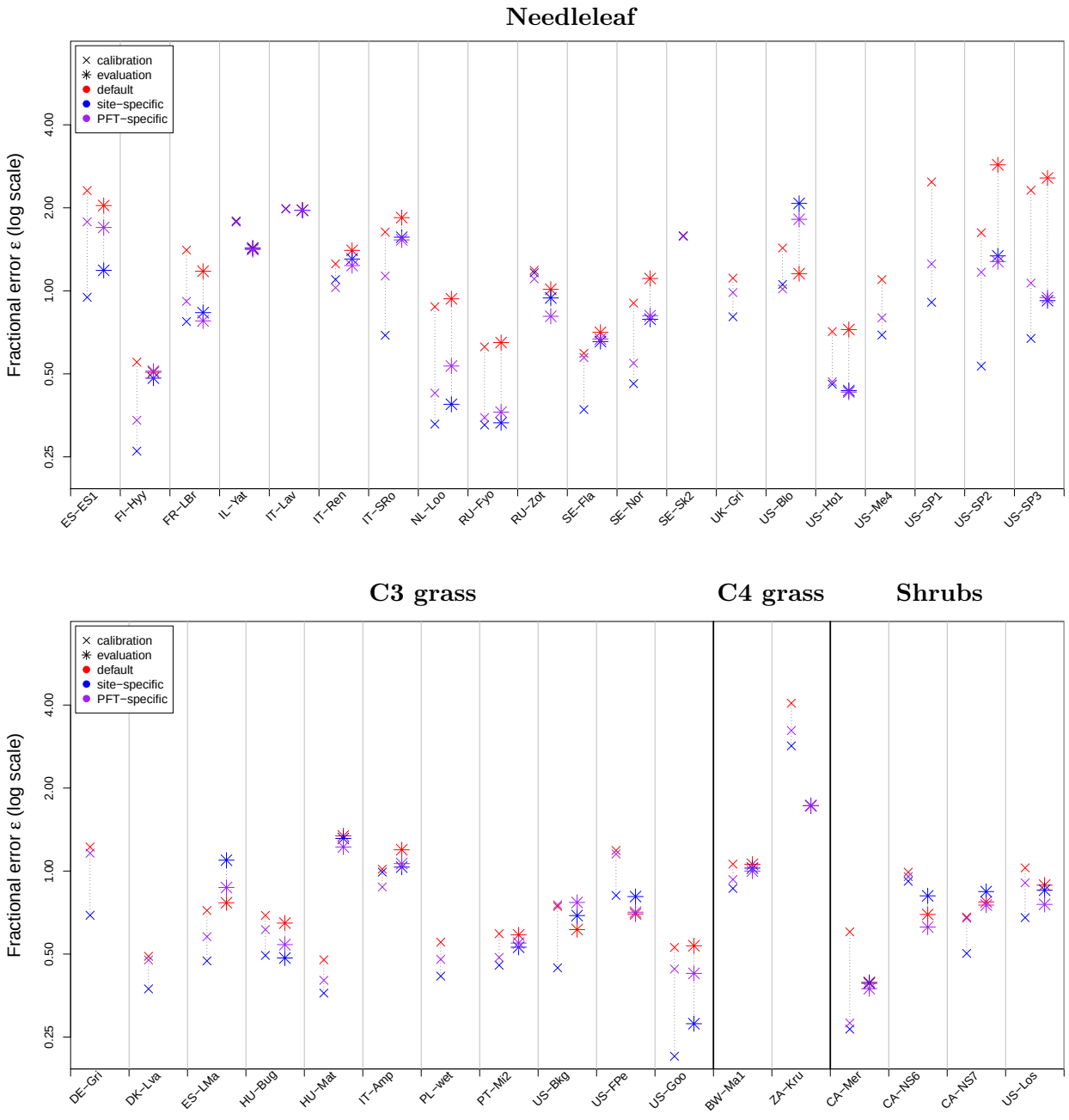



Figure 4.7.: (continued)

improve when using the PFT-specific vector, for either the calibration or evaluation years. This site has a lower annual cycle of GPP than the rest in this set. The CA-SF3 site improves when using the site-specific parameter vector in the evaluation year, but not using the PFT-specific vector. The US-Blo site improves in the calibration year, but when confronted with the evaluation year, both the site-specific vector and PFT-specific vector worsen the fit. This evaluation year has unusually high LE, which might be causing this discrepancy.

The last panel of Fig. 4.7 shows the C3 grass sites, the C4 grass sites and the shrub sites. For the C3 grass sites, half of the evaluation years either have a better fit with the PFT-specific parameter vector than with site-specific parameter vector or with the unoptimised default parameter vector. This suggests that the seasonal cycle differs over the different years at these sites. For the C4 grass sites, which started with relatively high errors, the new parameter vectors improve the sites slightly for the calibration year but hardly at all for the evaluation year. This set of two sites is too small to draw any proper conclusion about the C4 grass parameters. There is a clear need for more data from C4 grass sites. Finally, the shrubs can be seen to improve for all the sites. For the shrub sites, both the site-specific and the PFT-specific provide a better fit of the model to the observations of the calibration year. The improvement is minor for these sites, except for CA-Mer, which halves its fractional error. When confronted with observations from the evaluation years, the model also improves the fit of half of the sites for both site-specific and PFT-specific parameters. For the other two sites, the site-specific optimal vector increases error but the PFT-specific vector reduces it. In fact, for all of the sites in this PFT, the PFT-specific parameters outperform the site-specific optimal vectors over evaluation years. This is another example of the PFT-specific parameter vector being more robust.

For some sites, e.g. IT-Cpz and US-Bo1, the PFT-specific optimum outperforms the site-specific optimum in the calibration year. This is even the case when considering a different metric to fractional error $\epsilon$ used in this figure. This phenomenon was also noted by Kuppel et al. [2014], who suggest that the added constraints placed on the parameters by increasing the number of sites causes the cost function to become "smoother". This may render the optimisation scheme less likely to become trapped in local minima.

### 4.3.2. Analysis of improvement in fit

As discussed in Sect. 2.4.2, the fractional error is a good tool for cross-site comparison, however, it does not give much information about the way in which the optimised parameter vectors improve the fit at each site. Taylor diagrams (Taylor, 2001) provide more insight into how the fit has been improved by considering the relationship between observed variance $\mathrm{var}(\mathbf{o}_t)$, modelled variance $\mathrm{var}(\mathbf{m}_t)$, error variance $\mathrm{var}(\mathbf{o}_t - \mathbf{m}_t)$ and model–observation correlation $\mathrm{cor}(\mathbf{o}_t, \mathbf{m}_t)$.

The Taylor diagrams in Fig. 4.8 illustrate the improvement in performance of the optimised model for both the site-specific and PFT-generic parameters during calibration years for

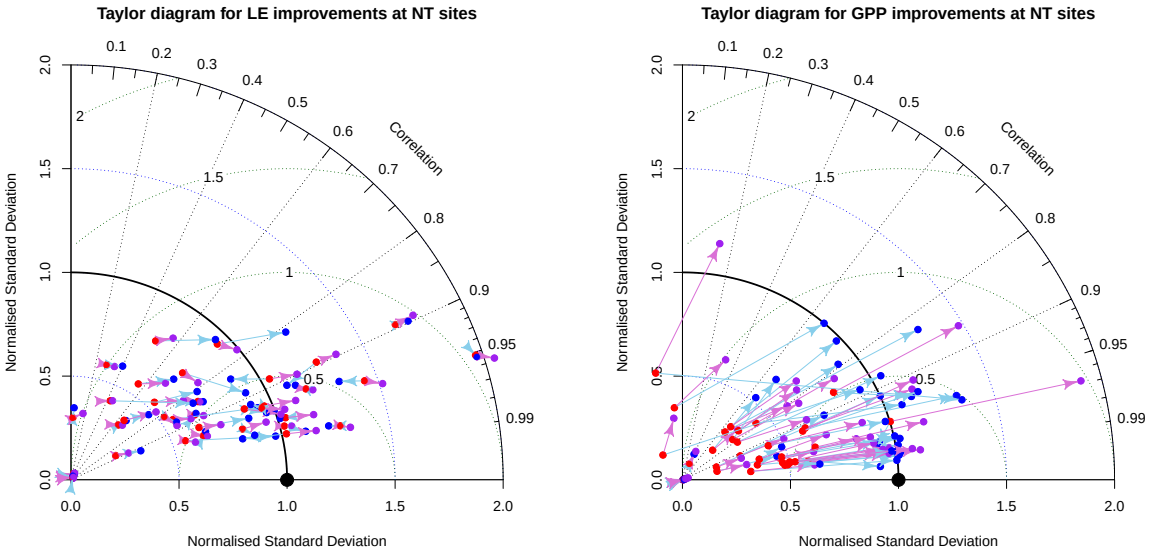



Figure 4.8.: Improvements in fit represented by "Taylor diagrams". Observed time-series (black dot) can be compared with modelled time-series for default parameters (red dots), site-specific optimal parameters (blue dots) and PFT-generic optimal parameters (purple dots). Radial distance from the origin (dotted lines) represents normalised standard deviation $\sqrt{\mathrm{var}(\mathbf{m}_t)/\mathrm{var}(\mathbf{o}_t)}$, and so a modelled time-series with the correct variance lies on the thick black line. Angular position represents the correlation between modelled and observed time-series. The distance from the black dot (dotted green lines) represents the normalised standard deviation in the errors $\sqrt{\mathrm{var}(\mathbf{o}_t - \mathbf{m}_t)/\mathrm{var}(\mathbf{o}_t)}$.

the needleleaf sites (plots for evaluation years are very similar).

For latent heat at needleleaf sites (left), the improvement is minimal. The underestimated seasonal cycle is seen to improve very slightly. The correlation between the modelled time-series and observation time-series does not improve much but for the majority of the sites this starts off relatively high (over 0.6). Other PFTs also show small improvements for latent heat.

For GPP at needleleaf sites (right), the seasonal cycle is typically underestimated and improves noticeably for both the single-site parameter vectors and the PFT-generic parameter vectors. The correlation between model and observed time-series does not change greatly. The Taylor diagram for GPP at broadleaf sites is very similar. For grasses and shrubs, the change is less drastic, though some of the sites have a more notable increase in correlation.

In addition to Taylor diagrams, the normalised bias $b$ (Eq. 2.54) can be used to assess the fit. This metric has the ability to highlight any systematic offset in the model. Calculating this statistic separately shows a reduction in bias in either the latent heat or GPP flux for 97.5% of the sites and both for two-thirds of the sites. The bias reduction in GPP at 90% of the sites was most notable. Sites where the LE bias was not reduced, tended to have larger reductions in GPP bias. This again highlights the fact that the parameters optimised are mainly related to this flux.

Taken together, these measures show that the observed improvements in model fit are due mainly to the adjustment of the magnitude of the annual cycle and a reduction in bias.

## 4.4. Structural changes to the model

The adJULES system produces the (locally) best possible fit to observations, given the existing model physics and prescribed driving data. If the fit is still inadequate, this may be due to the model and data themselves, rather than the parameter values. This section considers a couple of structurally different configurations of the JULES model to that used in the calibrating experiments.

### 4.4.1. Change in canopy model

As described in Sect. 2.1.2, these experiments use the 'big-leaf' canopy model. In JULES, there are several more options available for the treatment of radiation interception and scaling up to the canopy scale. The big-leaf canopy model is the simplest, and the most complicated includes a multi-layer canopy with sunlit and shaded leaves in each layer, two-stream radiation with sunflecks penetrating below the top layer, and light-inhibition of leaf respiration [Clark et al., 2011].

JULES was evaluated using the multi-layer approach for a temperature coniferous forest site in the Netherlands (Loobos, Jogireddy et al. [2006]) and a tropical broadleaf rainforest site in the Brazilian Amazon (BR-Ma2, Mercado et al. [2007]). Two of the German sites, DE-Hai and DE-Wet, a broadleaf and needleleaf site respectively, were also covered in Clark et al. [2011]. These studies all focused on the diurnal cycle of GPP, which was found to improve with the more complex canopy representation.

When using the same experimental set up used through-out this thesis, the change to the multi-layered canopy slightly improved LE but did not improve the GPP seasonal cycle.

### 4.4.2. Alternative PFT definitions

Conducted in parallel to the majority of the work found in this thesis, Harper et al. [2016] also tackles the task of improving the PFT representations in JULES. The study suggests a move from a five-PFT representation to a nine-PFT representation. This is achieved by first splitting the broadleaf and shrubs sites into deciduous and evergreen subsets, and then by further partitioning the evergreen broadleaf sites into tropical and temperature.

The first split of broadleaf tree sites and shrub sites into deciduous and evergreen serves to represent better the range of leaf life spans and metabolic capacities that exist in nature [Harper et al., 2016]. The difference between deciduous and evergreen plants is modelled mainly by the way the plants use nitrogen. Nitrogen is not only used in photosynthesis, but also in the growth and maintenance of the leaf structures. Since evergreen species have

less access to nutrients, they allocate a lower fraction of leaf nitrogen to photosynthesis and more to structure, defence and tolerance mechanisms than deciduous species [Harper et al., 2016]. As a result, evergreen plants tend to have longer life spans and a higher leaf mass per unit area [Takashima et al., 2004; Poorter et al., 2009].

In order to model these differences, Harper et al. [2016] introduce two new parameters into the JULES model: leaf mass per unit area (LMA, $\mathrm{kg\,m^{-2}}$) and leaf nitrogen per unit mass ($N_m$, $\mathrm{kgN\ kg^{-1}}$). The latter parameter replaces the $n_0$ parameter used in this thesis. These new parameters are used to update the $\mathrm{V_{cmax}}$ equation (Eq. 2.5) and the equations calculating the nitrogen concentration of the roots, stem, and leaves.

The second separation of evergreen broadleaf sites into tropical and temperate sites also effects $\mathrm{V_{cmax}}$. Tropical evergreen trees tend to have a lower measured $\mathrm{V_{cmax}}$ per leaf N per unit area ($N_A = N_m \cdot \mathrm{LMA}$) than temperate evergreen trees, resulting in maximum assimilation rates [Kattge et al., 2011].

The land cover at each FluxNet site is coded by one of the IGBP classes (IGBP: International Geosphere-Biosphere Programme). Excluding the non-vegetated land classes, these include; five types of forests: evergreen needleleaf (ENF), deciduous needleleaf (ENF), deciduous broadleaf (DBF), evergreen broadleaf (EBF), and mixed (MF); grasslands (GRA); two types of cropland: crops (CRO) and mosaicked cover (CVM); wetlands (WET); two types of shrubland: open (OSH) and closed (CSH); and two types of savanna: normal (SAV) and woody (WSA). The classification of a FluxNet site depends on the criteria found in Table A.2.

Note that while Harper et al. [2016] suggest increasing the number PFTs in JULES, these do not follow the extra categories found in the IGBP classes. The nine new PFT in JULES are denoted: tropical broadleaf evergreen trees (BET-Tr), temperate broadleaf evergreen trees (BET-Te), broadleaf deciduous (BDT), needleleaf evergreen trees (NET), needleleaf deciduous trees (NDT), C3 grasses, C4 grasses, evergreen shrubs (ESh), and deciduous shrubs (DSh). The original parameters from Clark et al. [2011] (Table. 2.2) have been adjusted to account for the new PFTs and new parameter values have been taken from the TRY database [Kattge et al., 2011].

Generated with the help of A. Harper, Fig. 4.9 and Fig. 4.10 compare the improvement made to model-data fit by adJULES and Harper et al. [2016]'s structural changes. The changes made in Harper et al. [2016] are not purely structural - it also includes some parameter adjustment through trail and error (different to an adJULES optimisation). The runs are generated using slightly different data and an updated version of JULES (version 4.2), into which the changes discussed in Harper et al. [2016] have been implemented. As discussed in Sect. 2.1.1, there have been very few changes to the main science between versions. This is also a good way to check whether the changes suggested by adJULES work on later versions of the model.

The sites displayed in these figures are the ones shown in Harper et al. [2016]'s study. These consist of the nine sites found in Blyth et al. [2011]'s benchmarking study, and five

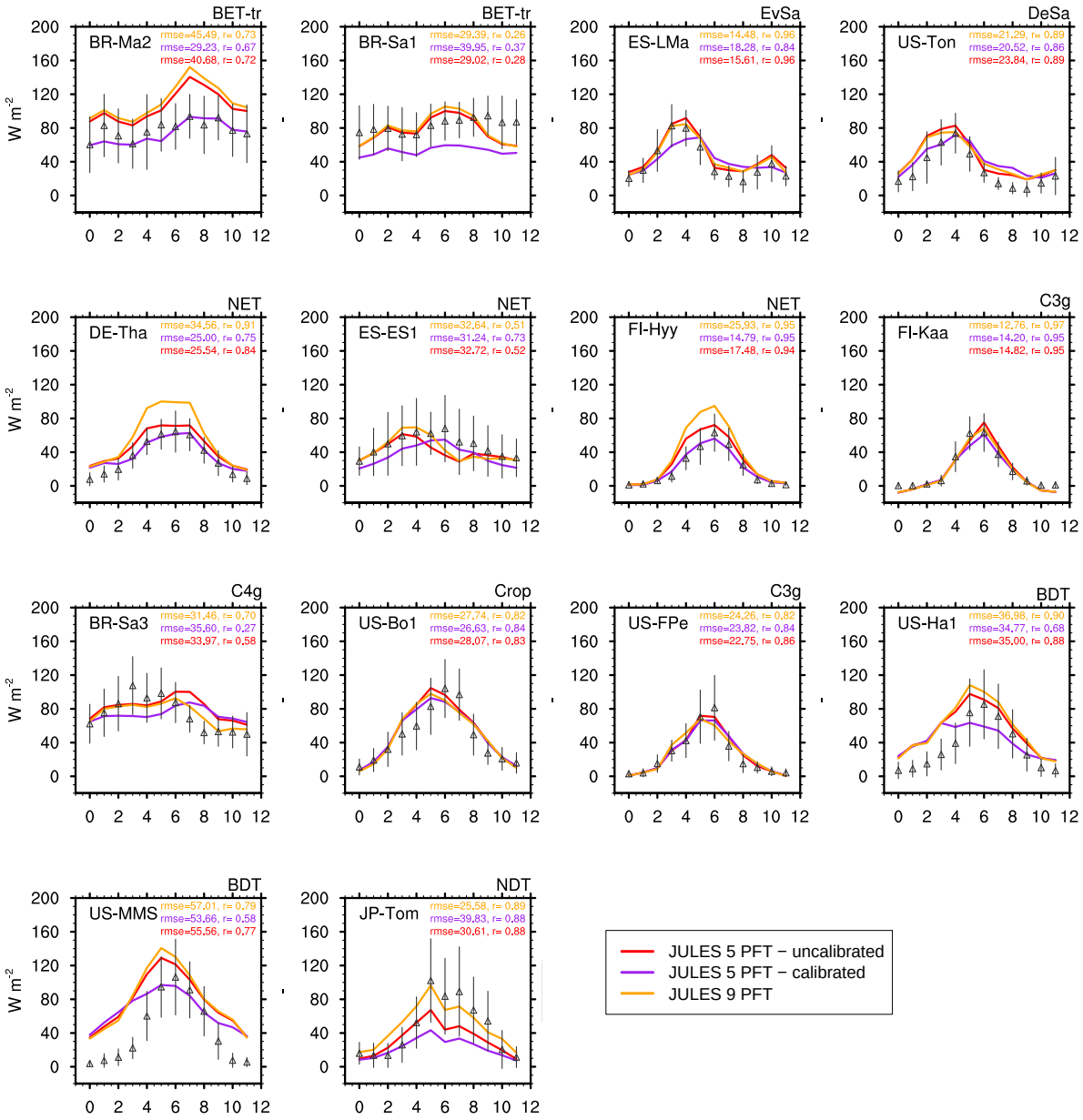



Figure 4.9.: Monthly mean fluxes of latent heat. Observations ± standard deviation from FluxNet are shown with triangles and vertical lines. The three JULES simulations are JULES with five PFTs and default parameters (red); JULES with five PFTs and the PFT-generic parameter vectors found in this study using adJULES (purple); and JULES with nine PFTs and the changes described in Harper et al. [2016] (orange).

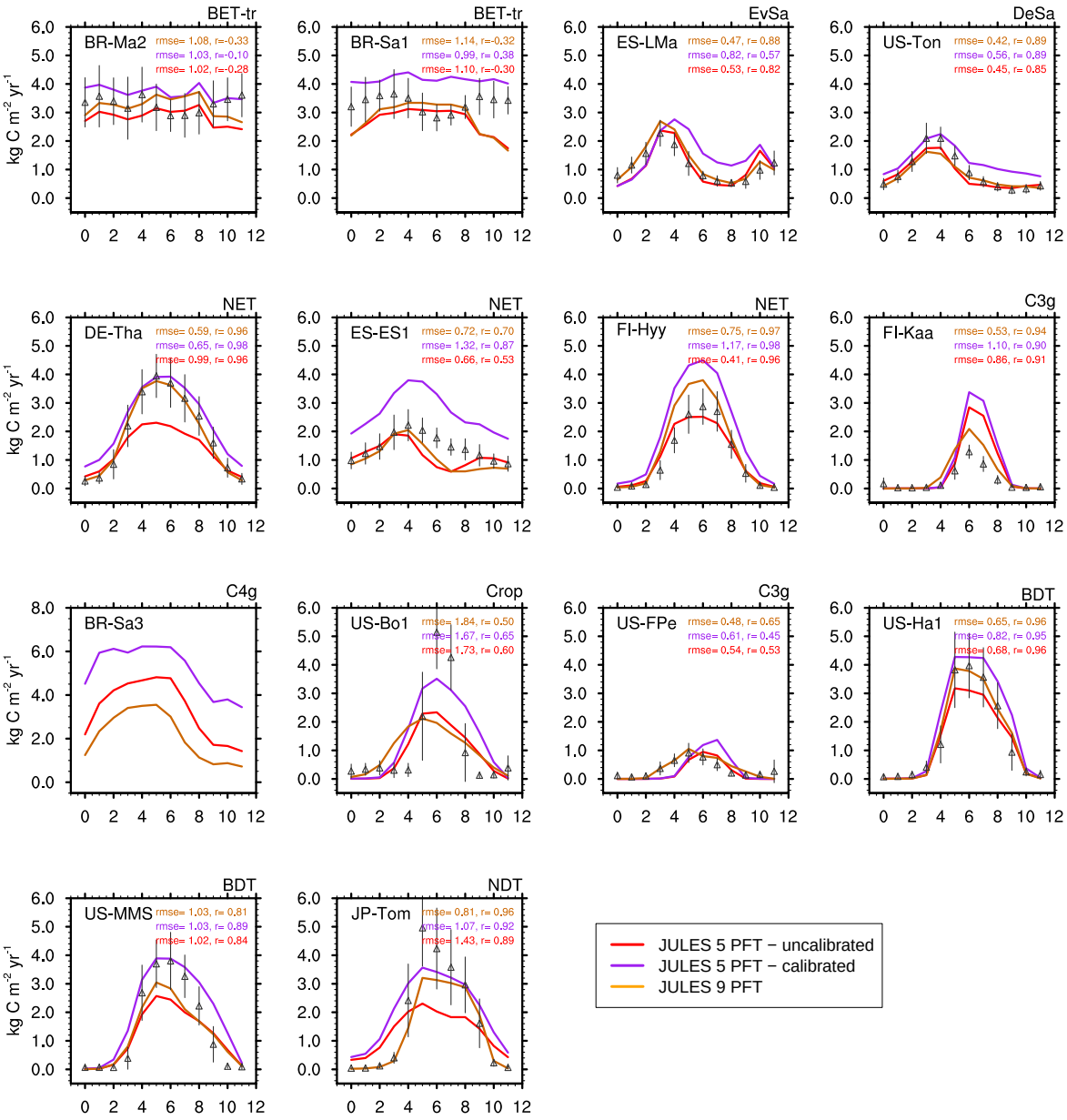



Figure 4.10.: As in Fig. 4.9 but for monthly GPP values.

additional sites chosen to represent more diversity in land cover types and climate [Harper et al., 2016]. The figures show the average seasonal cycle calculated at each site over a minimum of three continuous years (up to ten years when the data were available). The error bars represent the variation in these years.

First, consider the latent heat flux shown in Fig. 4.9. Using the RMSE metric, the parameters generated from adJULES can be seen to improve 9/14 of the sites compared to the default parameters. For eight of these sites, the adJULES changes give lower RMSE than the Harper changes. In contrast, the Harper runs have high correlations for the majority of the sites, the highest for 10/14 of the sites.

For the GPP flux (Fig. 4.10), the adJULES changes tend to give too great a seasonal cycle. The default runs in this configuration do not tend to underestimate the seasonal cycle to the same extent as the default runs used in the calibration. The correlation between the observations and the adJULES run however is high (over 85%) for 8/13 of the sites. This means that even though the cycle might have too high a magnitude, there can be confidence in the shape of the season cycle. The adJULES runs do well at optimising the deciduous broadleaf sites and the crop site, less well at the grass sites and evergreen broadleaf sites.

For the Harper runs, the GPP RMSE at half of the sites is reduced compared to the default runs. The correlation between the Harper runs and the observation is also over 85% for eight sites, however some of these sites differ from ones over which the adJULES runs fulfil the same criterion.

In this section, the 'big-leaf' canopy model has been utilised. However, Harper et al. [2016]'s changes pertain to a multi-layered canopy setup. The new adJULES parameter vectors and Harper et al. [2016]'s structural changes are responsible for different improvements in the model-data fit, highlighting the fact that both types of change are needed to advance model development. Due to TAF licensing issues, Harper et al. [2016] changes have not been integrated into the adJULES system. This is one of the main goals for future work, along with calibrating a multi-layered canopy version of JULES. Harper et al. [2016] also presents a new set of initial parameters for each of the PFTs. These values could be used to update $\mathbf{z}_0$ in the current adJULES system.

## 4.5. Redefining the PFTs through cluster analysis

In the experiments conducted so far, it has been assumed that all the sites belong to five predetermined PFT groups. This section challenges this assumption by attempting to reverse engineer the groupings using the single-site optimisation data. The optimised site-specific parameter vectors are used to group the sites through statistical clustering in a similar manner to Groenendijk et al. [2011].

In this section, two different experiments are considered. In the first instance, the parameter vectors themselves are clustered. This is done using two setups; first, the parameter

vectors found previously in the single-site optimisation are used, then new single-site optimisations are performed using the same starting vector $\mathbf{z}_0$ for all sites. For the second experiment, the clustering focuses on model-data fit improvement; sites are grouped if parameter vectors optimised at one site improve the fit at another.

### 4.5.1. Clustering techniques

Cluster analysis seeks to divide data into groups, known as clusters, in such a way that the data within each group are similar to one another (in some respect) and different from the data in the other groups. The greater the similarity (or homogeneity) within a group and the greater the difference between groups, the better the clustering. The clusters aim to capture the natural structure of the data and can be used to understand the relationship between the objects.

There are a number of different clustering algorithms, each differing depending on the type of clustering desired. For example, sets of clusters can be nested or distinct. This is the difference between *hierarchical* and *partitional* clustering [Tan et al., 2006]. In partitional clustering, the data are separated into distinct, non-overlapping groups. Each data point belongs to exactly one group. Hierarchical clustering, on the other hand, allows for subsets and groups data by creating a tree or dendrogram. The tree is a multilevel hierarchy, where clusters at one level are joined to clusters at the next level. Hierarchical clustering can be viewed as a sequence of partitional clustering. Similarly, a partitional clustering can be obtained by taking any member of that sequence i.e. by cutting the hierarchical tree at a particular level [Tan et al., 2006].

Cluster algorithms can also differ based on how strictly the group boundaries are defined. Exclusive clusters allow the data to belong to one cluster only, whereas overlapping clusters allow for multiple memberships. Fuzzy clustering lets every object belong to every cluster with a membership weight between 0 and 1. Some cluster algorithms also allow for partial clustering, i.e. can have outliers that do not belong to any group. Complete clustering algorithms on the other hand assign all the data to the groupings.

Cluster analysis is a wide and extensive field, with many different approaches and algorithms available. In this work, for simplicity, only complete exclusive partitional clustering is considered. However, there is a lot of potential for future work in this field.

The algorithm considered in this work is a version of $k$-means clustering (Hartigan and Wong [1979], `kmeans`: R Development Core Team [2015]) . This algorithm aims to partition the observations into $k$ clusters so that the within-cluster sum of squares (variance) is minimised. For a set of data $(\mathbf{x}_1, \ldots, \mathbf{x}_d)$ where each observation is a vector of dimension $n$, the algorithm aims to partition the $d$ observations into sets $\mathbf{S} = \{S_1, S_2, \ldots, S_k\}$ such that

$$\arg\min_{\mathbf{S}} \sum_{i=1}^{k} \sum_{\mathbf{x} \in S_i} ||\mathbf{x} - \boldsymbol{\mu}_i||^2 = \arg\min_{\mathbf{S}} \sum_{i=1}^{k} |S_i| \text{ Var } S_i \tag{4.3}$$

where $\boldsymbol{\mu}_i$ is the mean of points in $S_i$. This is equivalent to minimising the pairwise square deviations of the points in the same cluster [Kriegel et al., 2016].

The classic $k$-means algorithm is that of Lloyd [1982] (first proposed in 1957). Hartigan and Wong [1979] suggest an updated version of this algorithm, which is generally faster and has the ability to escape local optima by swapping points between clusters. The clusters found using $k$-means algorithms are as compact and well-separated as possible.

#### Choosing the number of clusters $k$

The number of clusters $k$ in the $k$-means algorithm is an input parameter. Determining its value can be a challenging problem, and the correct choice is often ambiguous. One way is to use external information or assumptions about the properties of the data set. For example, given that currently five PFTs exists in the JULES model, $k$ could be chosen to be five in clustering these data. Similarly, $k$ could be set to nine to mimic the nine PFT setup in Harper et al. [2016].

A more robust way to chose $k$ is to infer its value from the data. There are a number of methods to do this, however, if the data do not cluster in any obvious way, it is likely that the different methods will give different values of $k$. The Bayesian information criterion (BIC) is used in this thesis [Fraley and Raftery, 2002; Fraley et al., 2012].

### 4.5.2. Clustering the single-site parameter vectors

#### A look at the parameter values

In Sect. 4.1, single-site optimisations were performed at all sites and the improvement in model-data fit discussed. In this section, the parameter vectors themselves are considered. Figure 4.11 shows the pairwise relationship between the different optimised parameters. Since there are eight parameters in each vector, this is a representation of an eight dimensional space. This figure shows the results from two experiments, one where the background is included in the cost function, the other where the background term is omitted.

First, consider the parameter vectors found whilst including a background term (top part of Fig. 4.11). This means there is a belief placed in the prior values. For all sites, the optimised $\alpha$ values are found at the lower end of the prescribed range. The $\delta c/\delta L$ parameter on the other hand stays at the centre of the range, taking a value similar to its initial one.

From this first experiment, there are two parameters seen to separate the different PFTs; $\mathrm{T_{upp}}$ and $d_r$. $\mathrm{T_{upp}}$ is the parameter controlling the highest temperature at which photosynthesis can occur. This takes on low values for the needleleaf sites. These sites are found in the northern latitudes in colder climates. In contrast, the C4 grass sites opt

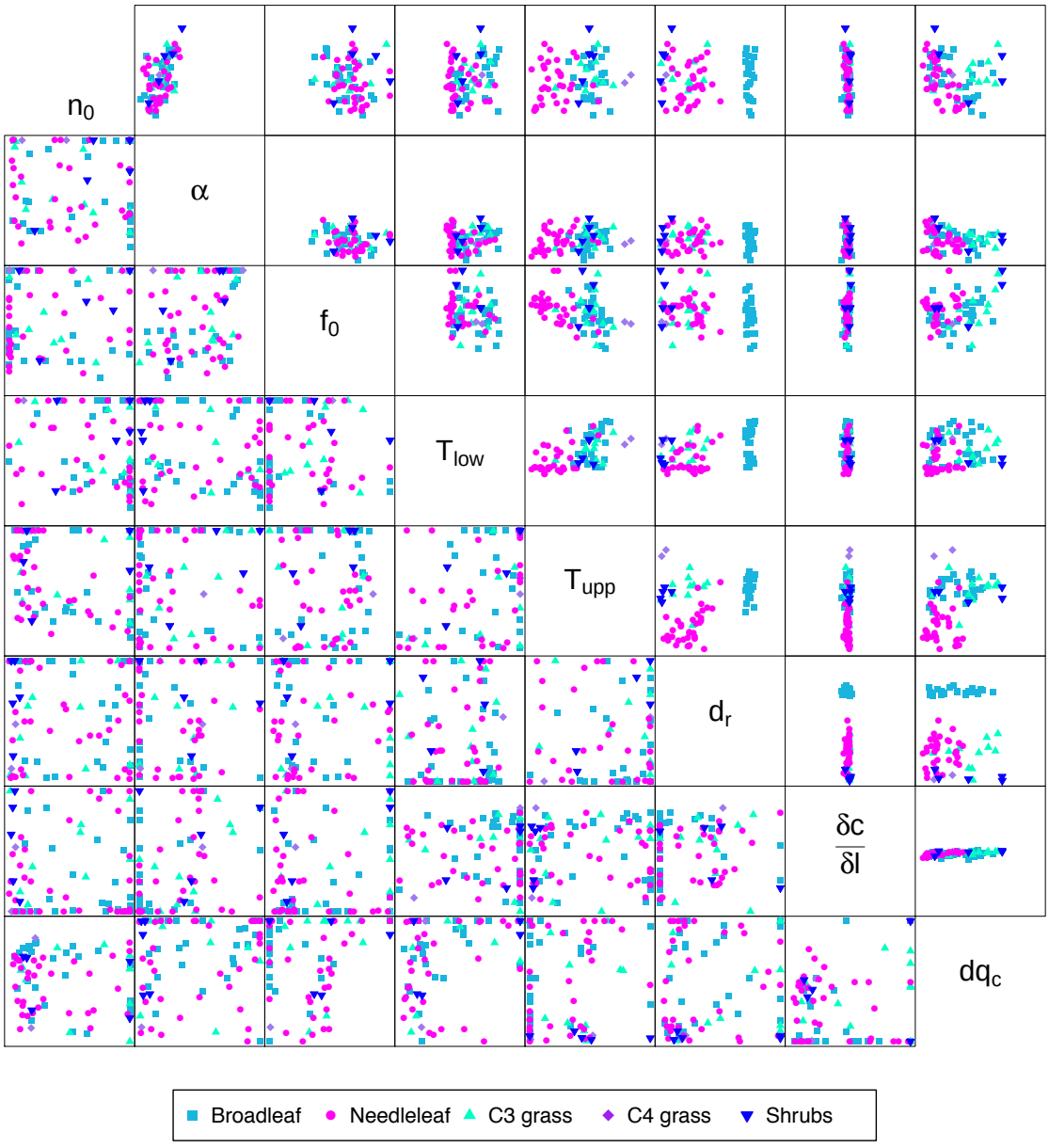
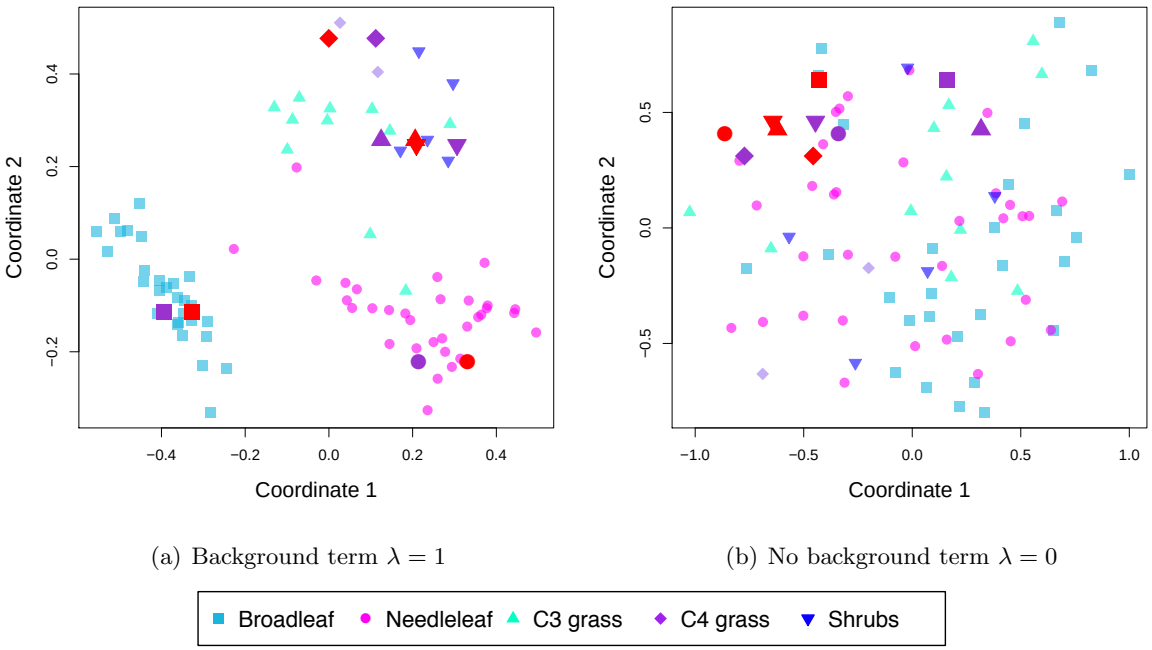



Figure 4.11.: Two-dimensional representation of the single-site parameter vectors. Optimised vectors found when $\lambda = 1$ are shown in the upper right-hand side triangle, and for $\lambda = 0$, the results are shown in lower left-hand side triangle. The dimensions of the boxes represent the prior range of each parameter.

for high values of $T_{upp}$. The two C4 grass sites are African sites where the weather is much warmer all year round. The $d_r$ parameter represents the root depth. Deep roots are needed for the tree sites, especially broadleaf trees, while shallow roots are needed for the grasses. Finally the $dq_c$ parameter takes on low values for needleleaf and C4 grass sites, whereas for C3 grass sites, the optimal value for this parameter tends to be higher. These optimised values do not differ too much from the initial PFT values given the parameters in Table 2.2, especially for these two defining parameters.

Now consider the lower half of Fig. 4.11. These are the results when the background term is omitted from the cost function. There are no obvious parameter clusters. The optimised values spread over most of the ranges with many parameters reaching the upper or lower limits of the boxes.

Figure 4.11 considers 2-D slices of parameter space. In order to visualise the eight dimensions more accurately, tools such as multidimensional scaling can be used. Classical multidimensional scaling, also known as principal coordinates analysis (PCoA; Gower [1966]), is used to generate the figures in Fig. 4.12. This algorithm maps a matrix of Euclidean distances between objects in a high-dimensional space to a lower-dimension coordinate matrix preserving the distances as well as possible.




(a) Background term $\lambda = 1$

(b) No background term $\lambda = 0$

Figure 4.12.: Classical multidimensional scaling used to scale the eight dimensions shown in Fig. 4.11 into two dimensional figures. The larger shapse represent the PFT-generic parameter vectors for each PFT: default shown in red and new multi-site values shown in purple.

In Fig. 4.12, the groupings identified in Fig. 4.11 for $\lambda = 1$ have been preserved. Similarly, when $\lambda = 0$, there are no visible groups. The initial parameter vector and the new PFT-generic parameter vectors have been added to these figures. The optimised parameters, both for the single-site and the PFT-generic cases, have not moved too far from the prior

in Fig. 4.12a. In contrast, the single-site optimised parameters in Fig. 4.12b spread far. Generally in both cases, the new PFT-generic vector remains closer to the prior than the majority of the single-site parameters.

So far the groupings, or lack of, have mainly been identified visually. There may still be some underlying structure that cannot be found with the human eye.

**Cluster analysis of the parameter vectors**

In this section, $k$-means clustering has been performed on the set of all normalised single-site parameter vectors. The clusters were then compared to other methods of partitioning the sites. First, with $k = 5$ to mimic the five JULES PFT groups, the results for experiments using the background term are shown in Fig. 4.13. The first row looks at clusters colour-coded by vegetation groups: PFT and IGBP classes. In Fig. 4.13a, it is clear that the $k$-mean clusters are partitioned in such a way that broadleaf sites find themselves separate to the other JULES PFTs. Clusters 3 and 5 mainly consist of the needleleaf sites, and cluster 4 contains a mix of all PFTs bar broadleaf sites. Figure 4.13b shows that even though the broadleaf sites have been put into two distinct clusters, these are not divided by deciduous and evergreen sites.

Figure 4.13c considers a different clustering to the rest of the plots in this figure. To run the $k$-means algorithm two input can be given, either the number of clusters $k$ or a set $k$ of initial (distinct) cluster centres. In the former case, these initial centres are then chosen at random from the data. In Fig. 4.13c, the five initial cluster centres have been set to the new PFT-generic values. Centering the clusters on these values as resulted in a clear broadleaf cluster and a clear needleleaf. This clustering is much closer to the PFT groupings defined in JULES.

Using the original clustering found by setting $k = 5$, the bottom row in Fig. 4.13 considers some of the physical features which can be used to distinguish between the different sites. Figure 4.13d shows the clustering coloured by the different climates and Fig. 4.13e shows the clustering coloured by the latitude of the site (rounded to the nearest ten degrees). No obvious pattern can be picked up in either plot. Each cluster covers a range of values in each case.

Different soil properties can be seen at each of the FluxNet sites. In running JULES, these have been read into the model from ancillary files to create soil parameter vectors of length nine. These in turn have been normalised and clustered with the $k$-means algorithm, BIC has been used to pick out the number of natural clusters. The eight soil clusters provide the colouring for Fig. 4.13f. Again, no trend can be picked out. The clustering cannot be explained by any of these three physical properties.

Using the BIC algorithm, the optimal value of $k$ for the vegetation parameter vectors was found to be five. Although not shown here, the optimised vegetation parameters in this experiment were also clustered with different $k$ values corresponding to the number of

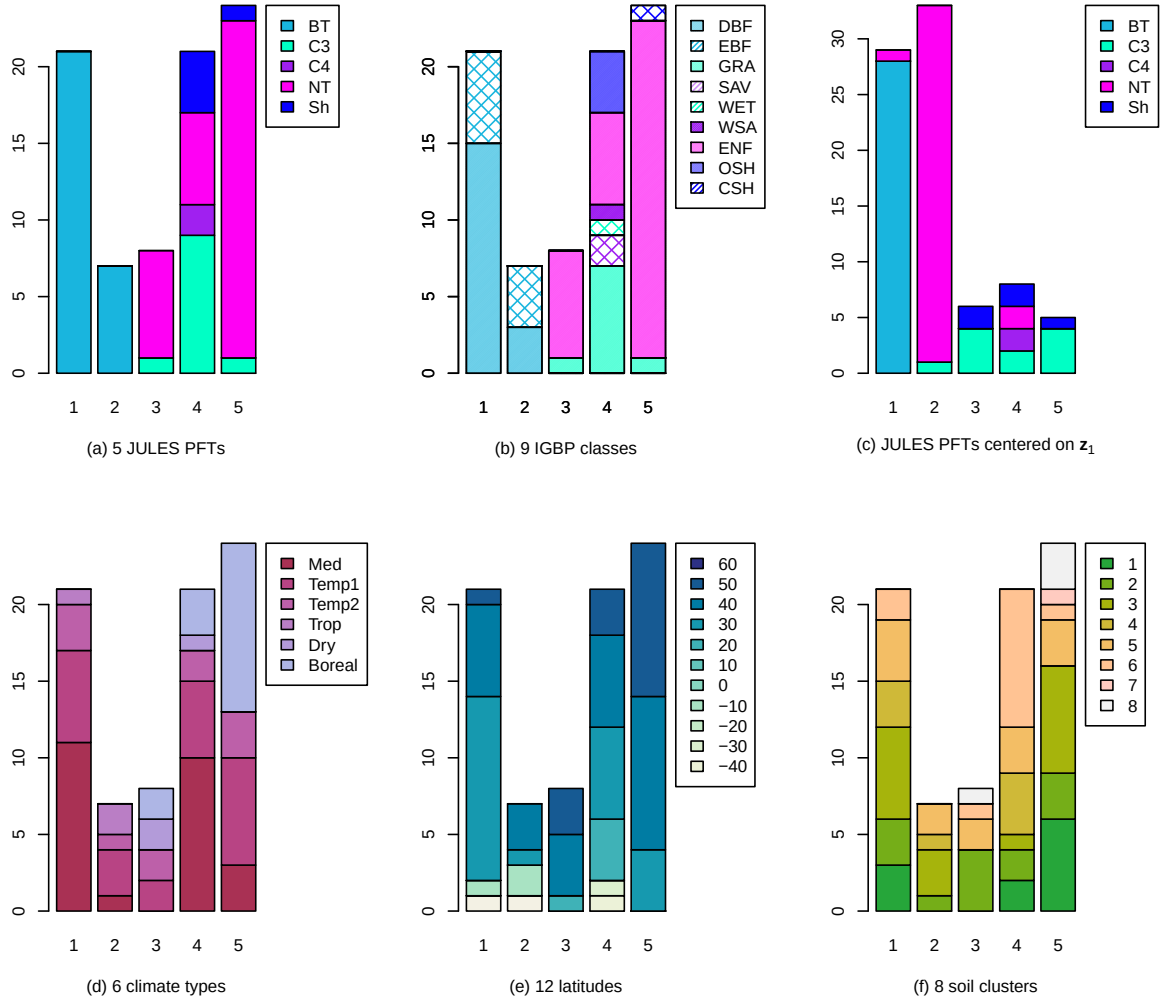



Figure 4.13.: FluxNet sites grouped into 5 clusters using $k$-means algorithm. The y-axis corresponds to the number of sites in each group. The first row shows the clusters further partitioned by vegetation groups: PFT and IGBP. The second row considers other physical features which could be used to distinguish between the different sites. For (d), the climate descriptions are shortened from subtropical-mediterranean (Med), temperate (Temp1), temperate-continental with hot/warm summers (Temp2), tropical (Trop).

different IGBP classes (9), climate types (6), latitudes (12), and soil clusters (8) covered by the FluxNet sites of this study. Similarly, with the exception of the broadleaf sites which separate into different clusters to the rest of the sites, no clear patterns could be picked out.

This experiment was run again for the parameter vectors found in the optimisations performed without the background term. When partitioned into five clusters, no patterns were identified in the clusters; vegetation or otherwise. Whereas previously the broadleaf sites were separate from the rest, the sites from each PFTs were now spread across the five clusters. The BIC algorithm suggests two clusters for these data, however, on closer inspection, this clustering consists of one main cluster and a second, much smaller cluster containing outliers.

### 4.5.3. Starting with the same initial conditions

When performing the single-site optimisations, each PFT has a slightly different starting $\mathbf{z}_0$ (see Table 2.2). To perform the experiments completely blind, i.e. with no knowledge about the PFT classification of the sites, the optimisations need to start from the same point in parameter space. This starting point is chosen to be the average of all the PFT $\mathbf{z}_0$ vectors.

The equations in JULES are programmed slightly differently for each PFT, however, this difference is minimal. The main difference is between C3 and C4 vegetation (trees and shrubs fall under the C3 photosynthetic equations), as such, the C4 grass sites were omitted from the analysis.

For both $\lambda = 0$ and $\lambda = 1$ experiments, no patterns in the clusterings could be found. The BIC algorithm suggests four and five clusters for each experiment respectively, but these clusters do not follow any of the potential classifications tested. There is too much variability within each PFT and too many similarities across all of the sites to be able to find any grouping resembling something meaningful.

It has been argued that roughness length is one of the most important parameters in distinguishing different PFTs. The `canht` parameter in adJULES can be used to measure this. However, adding this parameter to the optimisation did not help to partition the data in any meaningful way.

### 4.5.4. Grouping sites by parameter vector improvement

In one final clustering experiment, the parameter vectors found at each site in the original single-site experiments were transposed to the other sites. Let $s_i$ denote site $i$ and $\mathbf{z}_i$ the parameter vector locally optimised at that site. Recorded as a binary 'yes' or 'no', the experiment asks whether the model-data fit at site $i$ improves with the parameter vector found at site $j$, i.e. if $s_i$ improves with $\mathbf{z}_j$. This is displayed in Fig. 4.14 as a grid, with sites $i$ down the side and sites $j$ across the top.

No clear blocks can be identified in Fig. 4.14. If the optimised parameter vectors found at a given site improved only sites found in the same PFT, the coloured blocks alone would show. Instead, the majority of needleleaf sites can be seen to improve regardless what parameter vector is applied. The C3 grass sites show more sparsity, even when parameter vectors are transposed from the same PFT, these sites do not tend to improve.

These results are further summarised in Table 4.1. The fractional cover of each area shown in Fig. 4.14 is tabulated numerically as a conditional probability. The lowest probability shown is when the C3 grass sites use parameter vectors optimised over shrubs sites. The highest probability is when the C4 grass parameter vectors are transposed onto themselves. However this PFT only contains two sites so this probability is expected to be high (at least 0.5). The second highest probability is for the needleleaf sites using the parameter

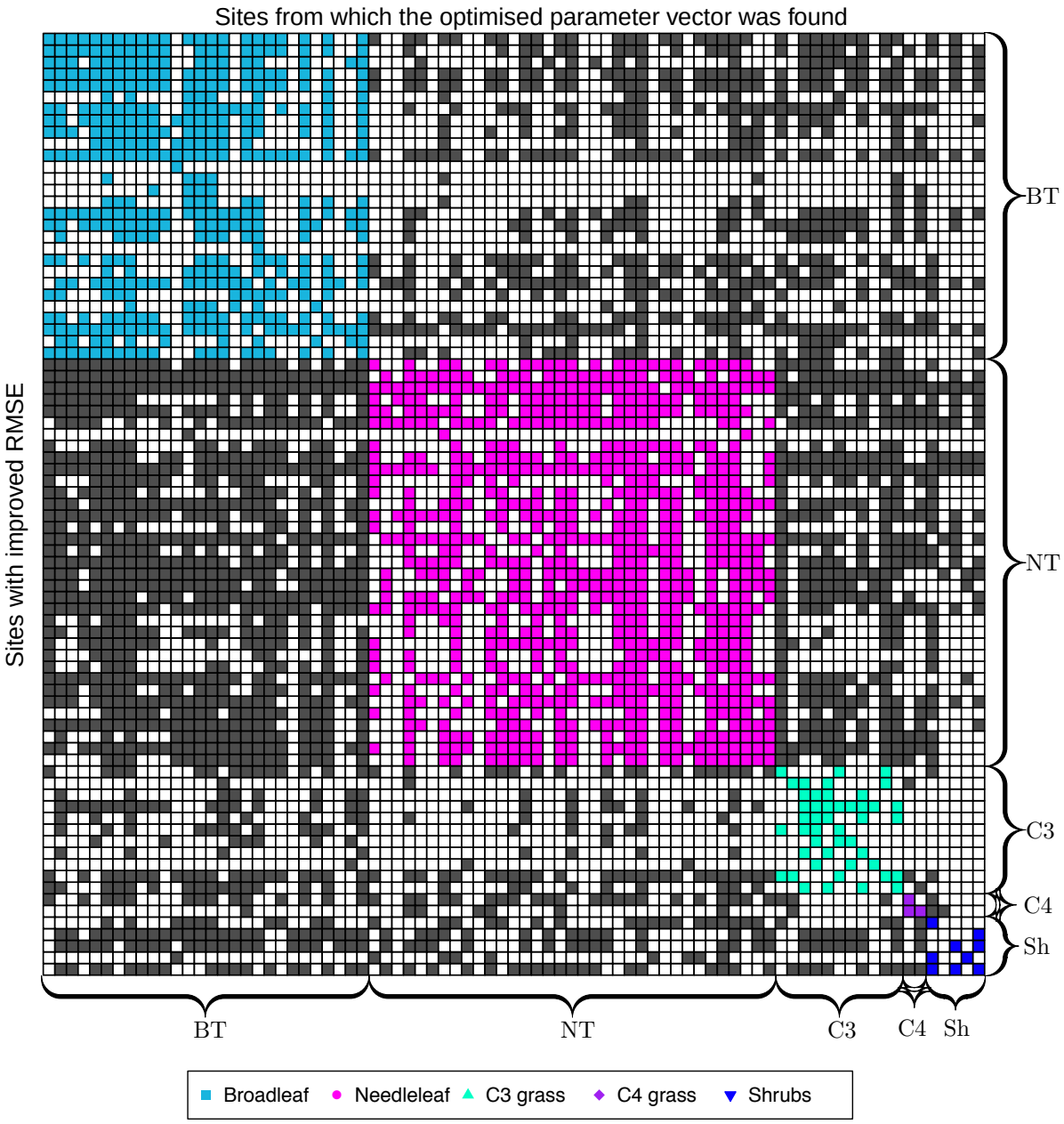



Figure 4.14.: Grid showing the successful transposing of parameter vectors. Parameter vectors found at the sites listed across the top are transposed to the sites listed downwards. The box is filled if the model-data fit is improved at the site compared to the default JULES parameters. Colours correspond to the PFT subsets.

vectors optimised at the broadleaf sites.

Surprisingly, the diagonal value for the needleleaf sites, i.e. the probability that a parameter vector optimised at a needleleaf site improves another needleleaf, is not the best - though it is still high. This shows that the default needleleaf parameters need to be reconsidered since the model-data fit at these sites is easily improved, especially when using parameter values optimised at the broadleaf sites.

Overall, half of the all sites improve regardless of where the transposed parameter vector was optimised.

| A \ B | BT | NT | C3 | C4 | Sh | All |
|---|---|---|---|---|---|---|
| BT | **0.542** | 0.340 | 0.458 | 0.357 | 0.229 | 0.419 |
| NT | 0.706 | **0.615** | 0.621 | 0.457 | 0.389 | 0.629 |
| C3 | 0.351 | 0.221 | **0.380** | 0.136 | 0.036 | 0.274 |
| C4 | 0.268 | 0.400 | 0.182 | **0.750** | 0.300 | 0.327 |
| Sh | 0.536 | 0.434 | 0.436 | 0.500 | **0.360** | 0.467 |
| All | 0.580 | 0.450 | 0.51 | 0.389 | 0.281 | 0.491 |

Table 4.1.: The probability that a site in subset $A$ improves given that the parameter vector used was optimised at a site in subset $B$, i.e. $\mathrm{P}(s_i \in A \text{ improves} \mid \mathbf{z}_j \text{ used where } s_j \in B)$.

The lack of symmetry in the grid means the relationship is not always bi-directional; $\mathbf{z}_j$ might improve the model-data fit at $s_i$, but optimal parameter vector $\mathbf{z}_i$ might not improve $s_j$. This is most apparent between the broadleaf and needleleaf subsets. The needleleaf sites are likely to improve when run with a parameter vector optimised at a broadleaf site, however, the broadleaf sites are much less likely to improve when using a locally optimised needleleaf parameter vector.

To filter these results down even further, sites were grouped only if they mutually improve each other. Each group $S$ is such that

$$s_i \in S \iff \forall s_j \in S,\ \mathbf{z}_j \text{ improves } s_i$$

When considering the whole 81 site set, 63 such $S$ groups were found with multiple overlaps. Similarly, 19 $S$ groups were found for the 28 broadleaf sites, again with multiple overlaps. Attempts to understand these grouping using graphs have not been informative. There is no apparent structure. There are many links within and between the different PFTs. Grouping sites in this manner has been unsuccessful. Instead sites, especially needleleaf ones, can be seen to improve for a number of parameter vectors, questioning the validity of the default parameter values in JULES.

In this section, the results from the $\lambda = 1$ experiments have been considered. The results are very similar for $\lambda = 0$.

## 4.6. Closing remarks

There is an apparent contradiction in this chapter. When PFTs are assumed, it is possible to find new generic parameter vectors which improve the majority of sites in each PFT. Indeed, for over 85% of the sites, PFT-specific optimal parameters perform better than default parameters when confronted with independent evaluation data. However, when sites are treated blindly, it is impossible to find such PFT groupings, or indeed any meaningful clusters. This lack of clustering was also found in Groenendijk et al. [2011].

As commented previously, there are many more clustering approaches that could be used. This chapter has only scratched the surface. For example, the optimised parameter vectors found at each site belong to multivariate probability distributions. One potential further avenue would be to cluster these distributions, perhaps by creating a similarity matrix between the distributions at sites $i$ and $j$.

The lack of informative clusters could be due to many other factors. For one, the choice of parameters used in the optimisation might not sufficiently represent the PFT. There may be other PFT specific parameters that need to be included. The fact that most of the needleleaf sites improve regardless of the parameter vector applied highlights the possibility that some of the parameters optimised in this study are not PFT specific at all.

The optimisation length of one year for each site might be too short. One year of observational data at each site was chosen for the optimisations. This was in order to include as many sites as possible in the analysis. However, an optimisation over multiple years might capture a general seasonal trend at each site, rather than just specifics of that year. These averaged seasonal cycles might be more similar over the PFTs and hence lead to clearer clustering.

Another reason for so much variability within each PFT might be down to the variability in the quality of the observation data. All sites were included in the analysis. However, some of the sites are better than others, for example having more complete time-series. The multi-site optimisation is able to improve the majority of sites in a set even if the set includes outliers. In contrast, the clustering algorithm used in this study will be more sensitive to these outliers.

The validity of each optimised parameter vector may also affect the clusterings. In single-site optimisations, it is possible that there are numerous local minima. These would result in a range of different optimal parameters within each PFT. The argument that multi-site optimisations are less susceptible to local minima might explain why a generic parameter vector can be found.

The single-site parameters may be too site-specific to be generalised or to be used in clustering experiments. The multi-site optimisations, on the other hand, offer different parameter vectors which can be used to better describe the PFTs. Most LSMs use PFT groups to describe vegetation and the number of PFTs is much less than the actually variation seen in nature. These PFTs are used to efficiently run the model globally or as part of climate simulations. Therefore, it is important to be able to improve these effectively. However, since the PFTs don't emerge naturally from the selection of observations and parameters used in these experiments, it is also possible the ecosystem modellers need to reconsider the use of PFTs and potentially look for alternative ways to represent vegetation in LSMs.

In the following chapter, land surface models (LSMs) are discussed as components of wider climate models. The responses of both the calibrated and uncalibrated JULES model to

atmospheric $CO_2$ and temperature changes are considered, in order to understand the effect of parameter changes on the sensitivities of the model.

# 5. Using adJULES to constrain future predictions

This chapter examines the effect that calibrating JULES has on the model's sensitivity to increases in atmospheric $CO_2$ and associated greenhouse warming. In Sect. 5.1, the responses of JULES to changing temperature and atmospheric $CO_2$ are considered; first, by focusing on photosynthesis in Sect. 5.1.1, and then by looking at Water Use Efficiency (WUE), which is the ratio of carbon gain from photosynthesis to water loss from transpiration, in Sect. 5.1.2.

## 5.1. The effects of calibration on $CO_2$-driven climate change

ESMs are computationally extremely expensive to run due to their highly complex nature. As a result, it is difficult to determine quickly the impacts of model changes on future predictions. Tools do exist to emulate the processes involved. For example, in Huntingford et al. [2010] a pattern-scaling approach to climate change is used to drive the JULES land-surface model. This computationally efficient model developed in Huntingford et al. [2010] (IMOGEN: Integrated Model Of Global Effects of climatic aNomalies) incorporates an analogue of the climatic response of the Hadley Centre GCM.

Here, in order to simulate $CO_2$ driven climate change, a simpler and more transparent approach is used to understand the effects of co-varying values of $CO_2$ and temperature. In very simple experiments, JULES runs were conducted at each of the FluxNet site with increasing atmospheric $CO_2$ concentrations ($c_a$) and temperatures ($T$) values. The same year at each site was repeated for ten years back-to-back in order to let the system spin up to the new atmospheric conditions. The different outputs of the JULES runs were averaged to find annual means for each quantity considered. This was done in order to create a grid in $c_a - T$ space from which contours could be interpolated (e.g. Fig. 5.1).

All of the sites start with the same $c_a$ value of 400ppm, fixed constant over the whole year. This value is treated like a parameter in the JULES code. Increasing values of $c_a$ were calculated by using different multiplication factors ranging from one (no change) to four. In contrast, the temperatures at each site are taken from the FluxNet database. These temperature data are driving data (see Sect. 2.1.3). The temperatures differ between sites in magnitude and season cycle. The increased temperature values were calculated by adding $\Delta T$ to the each data point in the driving data so that the annual temperature at

each site keeps its variability during the year run. As in previous experiments, the main focus is on the broadleaf sites.

### 5.1.1. Gross Primary Productivity

First consider the effects of changing atmospheric $CO_2$ concentration and annual temperature on the GPP flux. Since the GPP fluxes are calculated as part of the FluxNet data, the runs can be directly compared to the observations for present-day conditions: i.e. no $CO_2$ increase and no temperature increase. To measure sensitivity of the model to the atmospheric changes, $A$ and $B$ values are calculated where

$$A = \frac{\mathrm{GPP}(2 \times c_a) - \mathrm{GPP}(c_a)}{\mathrm{GPP}(c_a)}, \qquad B = \frac{\mathrm{GPP}(T+2) - \mathrm{GPP}(T)}{\mathrm{GPP}(T)}. \tag{5.1}$$

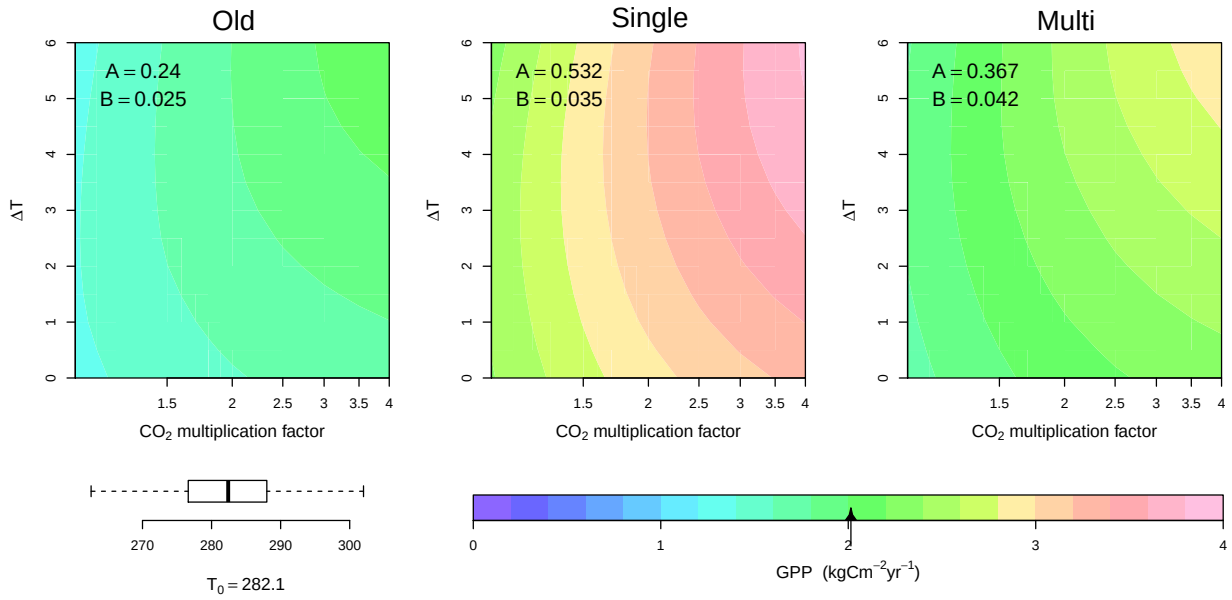
These measure the rate of the change of GPP in response to changing $c_a$ and $\Delta T$. The $A$ value measures the sensitivity to doubled $c_a$, and the $B$ value measures the sensitivity to 2 degrees of warming. These limits were chosen since they are commonly used in the literature.

Two broadleaf sites are shown in Fig. 5.1. In both cases, three contours plots are shown: one generated from the uncalibrated model (denoted 'old') and two generated from the calibrated models (denoted 'single' when locally-optimised at the single site and 'multi' when calibrated as part of the multisite experiments). Let $\mathrm{GPP}_0$ denote the initial value of GPP found at the current climate ($\Delta T = 0$, $CO_2$ multiplication factor = 1). For both sites, this value is closest to the observed value for the single case, followed by the multi case, and furthest away in the old case. This mirrors the results of the calibrations at these sites, where GPP is similarly underestimated for the uncalibrated runs.

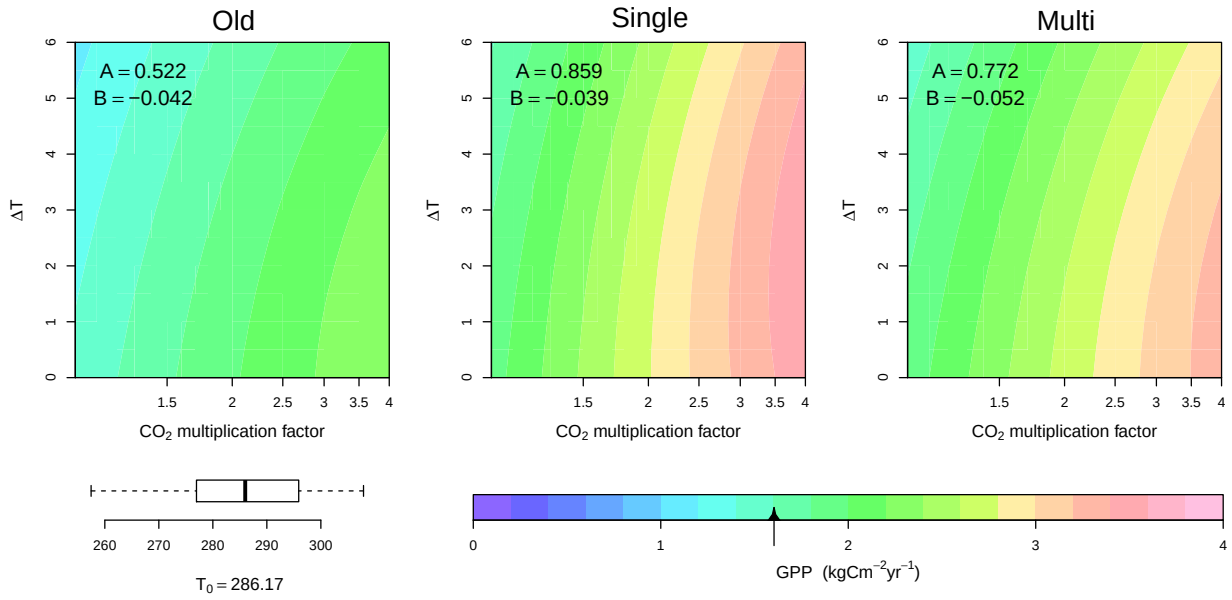
In Fig. 5.1(a) and (b), the rate at which GPP changes is greatly increased in the calibrated runs. For the runs using the locally optimised parameters, the increase in GPP is the most notable. This can be observed both visually, with the increase in the steepness of the colour gradient, and numerically by considering the $A$ and $B$ values.

For DK-Sor shown in Fig. 5.1(a), the $A$ value, which measures the sensitivity of GPP to doubled $c_a$, is over double for the runs with the locally optimised parameters than the unoptimised runs. The $B$ value, which measures the sensitivity of the GPP to 2 degrees of warming also increases for the calibrated runs. This value is highest for the runs in the multi case. Since $A > B$ in all cases, GPP is more sensitive to the doubling of atmospheric $CO_2$ concentrations than to a temperature increase of two degrees.

For US-MSS shown in Fig. 5.1(b), $A$ also increases for the calibrated runs, with the highest $A$ again for the contour plot generated from the locally-optimised parameter vector. Similarly, $|A| > |B|$. In contrast, this site has negative $B$ values for all parameter settings. This means that the rate of photosynthesis decreases with increasing temperature anomaly. These negative $B$ values are responsible for the different slopes of the contours seen at




(a) DK-Sor site




(b) US-MMS site

Figure 5.1.: The changes in GPP with increasing $CO_2$ and increasing temperature for different parameter settings at two different broadleaf sites: (a) DK-Sor and (b) US-MMS. The three panels in each case show the runs using different parameter settings: default JULES parameters (left), the parameters found optimising locally (middle) and the new PFT-generic parameters (right). The observed GPP value is indicated by the arrow on the colour scale. $A$ and $B$ represent the initial rate of change along the x- and y-axis respectively (taken over doubled $CO_2$ and a 2 degree increase in temperature). Below the panels is the colour scale for the contour plots and a box-plot representing the variation of the initial annual temperature cycle (K). $T_0$ refers to the mean annual value.

both sites.

The values of $A$ correspond to a doubling of $CO_2$ but with no temperature. In the FACE experiment results (Free-air $CO_2$ enrichment; Ainsworth and Long [2005]) the increase was found to be around 30%. The uncalibrated and multi-site values are similar to this value, whereas the single-site calibrated values suggest a higher increase. The negative sensitivity of the model to temperature, $B$, seen at sites that do not have a very high mean annual temperature is surprising. This may due in part the experiment design: there are other parts of the system responding to the atmospheric changes not considered here, for example the soil moisture at these sites.

**Optimal atmospheric temperature for photosynthesis**

The contours shown in Fig. 5.1 are shaped as different sections of right-facing semicircles. These contours can also be thought as the left-hand side of inlaid circles. For DK-Sor (Fig. 5.1(a)), the bottom left quadrant of the circles is shown and for US-MMS (Fig. 5.1(b)), the middle to top left-hand quadrant of the circles is shown.

Each arc has a peak which represents the optimal temperature at a given $CO_2$ concentration for the maximum rate of photosynthesis. The value of $\Delta T$ at the peak of a given $CO_2$ concentration, denoted $\Delta T^{\dagger}$, can be added to the annual mean temperature ($T_0$) to give a value for this optimal temperature ($T_{\text{opt}}$).

$$T_{\text{opt}} = T_0 + \Delta T^{\dagger} \tag{5.2}$$

For DK-Sor, the peak is positioned at relatively high temperature anomalies. For US-MMS, this peak is positioned for low (or even negative) values of $\Delta T$.

The different initial temperatures at each site shown in Fig. 5.1 could explain the difference in the contour maps and $T_{\text{opt}}$ trends. Unperturbed, DK-Sor has an annual mean temperature four degrees lower than US-MMS. When starting the contour plots for DK-Sor (Fig. 5.1(a)) at $\Delta T = 4$, a different section of the circles would be shown. In fact, the contours would have a similar shape to the contours shown in Fig. 5.1(b) for the US-MMS site. This is especially apparent when considering the contours generated using the locally optimised parameter values. Therefore, it is possible that the values of $T_{\text{opt}}$ will be comparable between sites.

A value of $T_{\text{opt}}$ can be calculated for all the sites. The position of each peak for a set of given $CO_2$ concentrations, calculated to the nearest half degree, is added to the respective mean temperature of each site. Due to time constraints, only values of $\Delta T \in \{-10, 10\}$ were tested. Figure 5.2 shows boxplots of the broadleaf $T_{\text{opt}}$ values.

The value of $T_{\text{opt}}$ can be seen to increase with increasing $c_a$. This means that with increasing $CO_2$ concentration, the optimal temperature for which photosynthesis can occur is pushed up. The median value (and the mean) can be seen to increase by at least two degrees for every doubling of atmospheric $CO_2$ concentration.

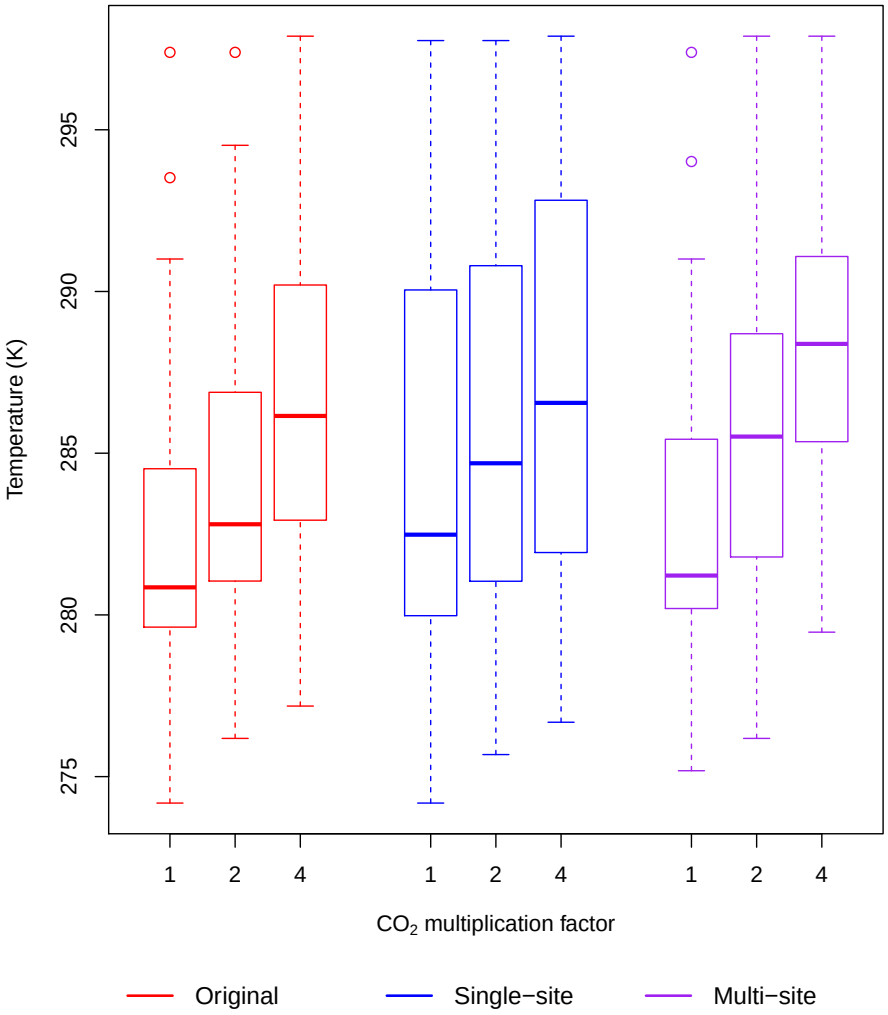



Figure 5.2.: Boxplots representing $T_{\mathrm{opt}}$ values for the broadleaf sites.

The spread of $T_{\mathrm{opt}}$ values found at the locally-optimised runs is larger than the range of $T_{\mathrm{opt}}$ values found at run conducted with the PFT-generic parameter vectors, both old and new. This variability between sites means that the value of $T_{\mathrm{opt}}$ is sensitive to one or more of the parameters used in the optimisation.

For the needleleaf sites (not shown), the values of $T_{\mathrm{opt}}$ also increase with increasing $c_a$. On the other hand, the spread of $T_{\mathrm{opt}}$ values remains more consistent across parameter settings. It is possible that the values of $T_{\mathrm{opt}}$ for needleleaf sites are not sensitive to the same parameters as in the broadleaf case. Alternatively, the parameters responsible for the range of $T_{\mathrm{opt}}$ values might not change to the same extent when calibrating over the needleleaf sites as when calibrating over the broadleaf sites. For C3 grass, the spread of $T_{\mathrm{opt}}$ values increases with increasing $c_a$. The C4 grass and shrubs sites were excluded from the analysis due to the small sample size.

There is also an optimal temperature for photosynthesis that can be derived from the underlying equations. This value corresponds to the maximum of the $V_{\mathrm{cmax}}$ curve (seen

in Fig. 4.4) and will be denoted $T_{\text{max}}$.

In order to find the value of this optimum, the equation for $V_{\text{cmax}}$ (Eq. 2.5) has to be differentiated, and with the use of the quadratic formula, the following expression for $T_{\text{max}}$ is found:

$$T_{\text{max}} = -\frac{10}{3}\ln\left(\frac{\ln 2\cdot(ab)^{-1}}{6+\ln 4}\left[\sqrt{\frac{36}{(\ln 2)^2}ab+(a-b)^2}-(a+b)\right]\right) \tag{5.3}$$

where $a = e^{\frac{3}{10}T_{\text{low}}}$ and $b = e^{\frac{3}{10}T_{\text{upp}}}$ are used to simplify the notation. Without loss of generality, let $T_{\text{upp}} > T_{\text{low}}$ so that

$$T_{\text{upp}} = T_{\text{low}} + \xi \text{ and } T_{\text{upp}} = T_{\text{max}} + \psi$$

for some positive $\psi, \xi \in \mathbb{R}$. After substituting these relationships into Eq. 5.3 and rearranging, the following expression between $\psi$ and $\xi$ is obtained:

$$\psi = \frac{10}{3}\ln\left(\frac{\ln 2}{6+\ln 4}\left[\sqrt{\left(\frac{6}{\ln 2}\right)^2 e^{\frac{3}{10}\xi}+(e^{\frac{3}{10}\xi}-1)^2}-(e^{\frac{3}{10}\xi}+1)\right]\right). \tag{5.4}$$

Taking the limit of this expression, the following value is found:

$$\lim_{\xi\to\infty}\psi(\xi) = \frac{10}{3}\ln\left(\frac{3}{\ln 2}-1\right) \tag{5.5}$$

$$\approx 4.00799. \tag{5.6}$$

This means that for a large enough difference between $T_{\text{upp}}$ and $T_{\text{low}}$, the maximum of the $V_{\text{cmax}}$ curve will be approximately four degrees lower than $T_{\text{upp}}$. A difference of 20°C is enough to ensure this relationship holds.

Using these formulae and the values of $T_{\text{upp}}$ and $T_{\text{low}}$ from the default and multisite parameter vectors, $T_{\text{max}}$ values can be calculated for each PFT. For the tree PFTs, the value of $T_{\text{max}}$ is increased by two degrees after the multisite calibration, whereas for the grass and shrub PFTs, the value of $T_{\text{max}}$ is decreased by one degree. The needleleaf PFT has the lowest value of $T_{\text{max}}$ which is consistent with these sites existing in colder parts of the planet. Similarly, the C4 grass PFT has the highest value of $T_{\text{max}}$ and these sites are found in the hotter parts of the planet.

In addition, using the results from the single-site optimisations, each individual site has a different value of $T_{\text{max}}$. These values of $T_{\text{max}}$ are compared to the values of $T_{\text{opt}}$ found from the contour plots.

The correlation between $T_{\text{max}}$ and $T_{\text{opt}}$ is found to be negligible, and the values of $T_{\text{max}}$ are found to be much higher than the values of $T_{\text{opt}}$ for any of the parameter settings. The magnitude of $T_{\text{max}}$ is much closer to the maximum temperature value of each site. $T_{\text{opt}}$ has been calculated by adding the position of the peak to mean annual temperature (Eq. 5.2). By changing the definition of $T_0$ to describe a different feature of the annual temperature

cycle at each site, such as its variability or magnitude, another value of $T_{\text{opt}}$ can be derived. Definitions of $T_0$ tested include: the maximum annual temperature, the average temperature when photosynthesis is occurring (values found between $T_{\text{upp}}$ and $T_{\text{low}}$), the average temperature ± standard derivation of the temperature, and the maximum annual temperature ± standard derivation of the temperature. Even with these alternative $T_{\text{opt}}$ values, no relationship between $T_{\text{opt}}$ and $T_{\text{max}}$ could be identified. The only link found was that both values increased by a couple of degrees for the tree sites after calibration.

In conclusion, the optimal temperature for photosynthesis derived from the equations is higher and seemingly unrelated to the optimal temperature for photosynthesis observed by changing the atmospheric temperature in the model. The value of $T_{\text{max}}$ is derived from the $V_{\text{cmax}}$ equation (Eq. 2.5). However, this equation is only one part of the limiting processes used to calculate photosynthesis. It is possible that the value of $T_{\text{opt}}$ is more sensitive to these other constraining factors, hence its lower value. The lack of relationship between the two values might also be due to the design of the experiment. For example a change in annual mean temperature would not happen in such a uniform manner.

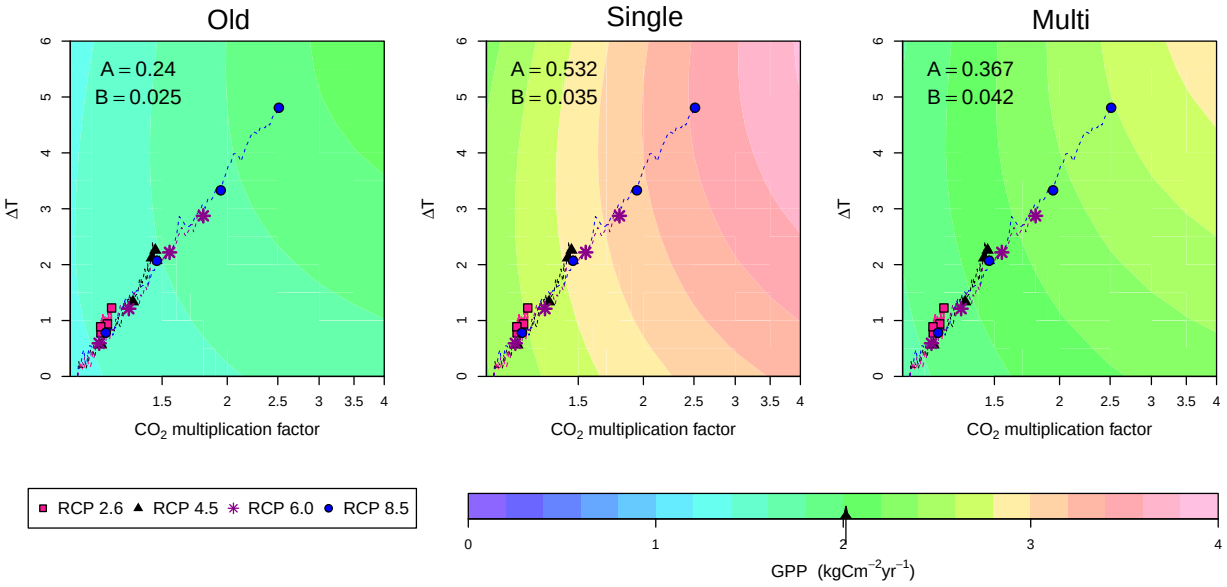
With increasing atmospheric $CO_2$ concentrations, $T_{\text{opt}}$ is found to increase, whereas $T_{\text{max}}$ is insensitive to $c_a$ in its derivation. Therefore $T_{\text{opt}}$ is probably a better indicator of the relationship between photosynthesis and temperature at high temperatures and different atmospheric $CO_2$ concentrations. Calibration of the model at tree sites suggests that $T_{\text{opt}}$ for different $c_a$ is one or two degrees higher than in the uncalibrated model. With increased values of $T_{\text{opt}}$, global simulation of the carbon cycle will suggest a stronger carbon sink by trees, especially the broadleaf ones. Booth et al. [2012] found that sensitivity of photosynthetic metabolism to temperature is one of the most important uncertainties in understanding the magnitude of future change. There remain open questions about the potential role of plant acclimation to increasing temperatures [Booth et al., 2012].

**Using RCPs for context**

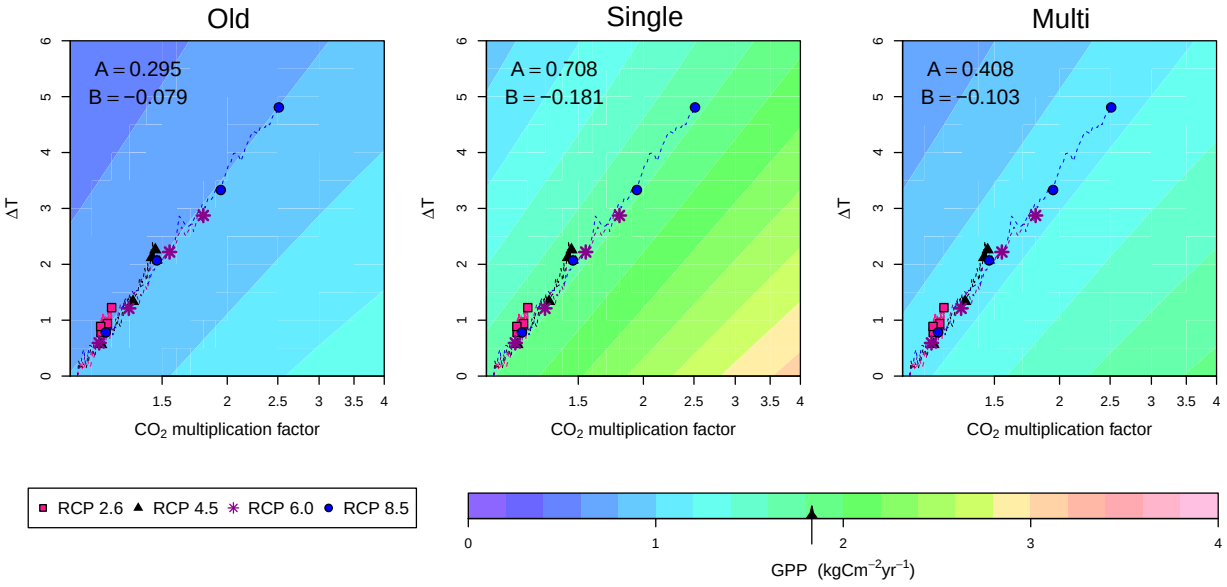
The RCPs described in Sect. 1.3.3 can be plotted over the contour maps to add context to the changes in the GPP (Fig. 5.3). It is important to remember that the pathways represent global averages, whereas the experiments here are all conducted at a site level. Nevertheless, they can still be informative.

The shape of the contours strongly influences how photosynthesis changes along the different trajectories. In Fig. 5.3(a), the pathways can be seen to cross multiple contours, whereas in Fig. 5.3(b), the trajectories are parallel to the contours. This means that the rate of change of GPP at the DK-Sor site is increased significantly as the years progress, whereas, at the FR-Fon site, the GPP value does not change much from its initial annual value. The temperature and $CO_2$ effect almost completely cancel each other out. This highlights the fact that at site level, there are very different responses possible to $CO_2$ induced climate change.

For DK-Sor in Fig. 5.3(a), the number of contours crossed can be seen to increase when the




(a) DK-Sor site




(b) FR-Fon site

Figure 5.3.: Four RCPs plotted on contour plots of GPP at (a) DK-Sor and (b) FR-Fon. Each pathway, taken from the HadGEM2ES runs shown in Fig. 1.1, is represented by a different coloured line. The shapes are placed every 25 years, starting from 2025 till 2100.

site is calibrated compared to the uncalibrated experiment. For the experiment run with the locally optimised parameters, the GPP values for the different trajectories at the year 2100 are approximately 1.5 times the corresponding values for the uncalibrated experiment. Note that not only is this factor similar between trajectories, but the underlying pattern of the contour plots remains consistent regardless of parameter setting. This is because whilst the calibration has greatly changed the magnitude of the GPP flux, the fractional response of GPP to changing atmospheric conditions, i.e. $\mathrm{GPP}/\mathrm{GPP}_0$, does not change as much between the different parameter settings

**The effect of individual parameters**

To better understand the role of the calibration on GPP predictions, this section considers the effect of each individual parameter on both the magnitude of the GPP, and fractional change in GPP when confronted with changing atmospheric temperatures and $CO_2$ concentrations.

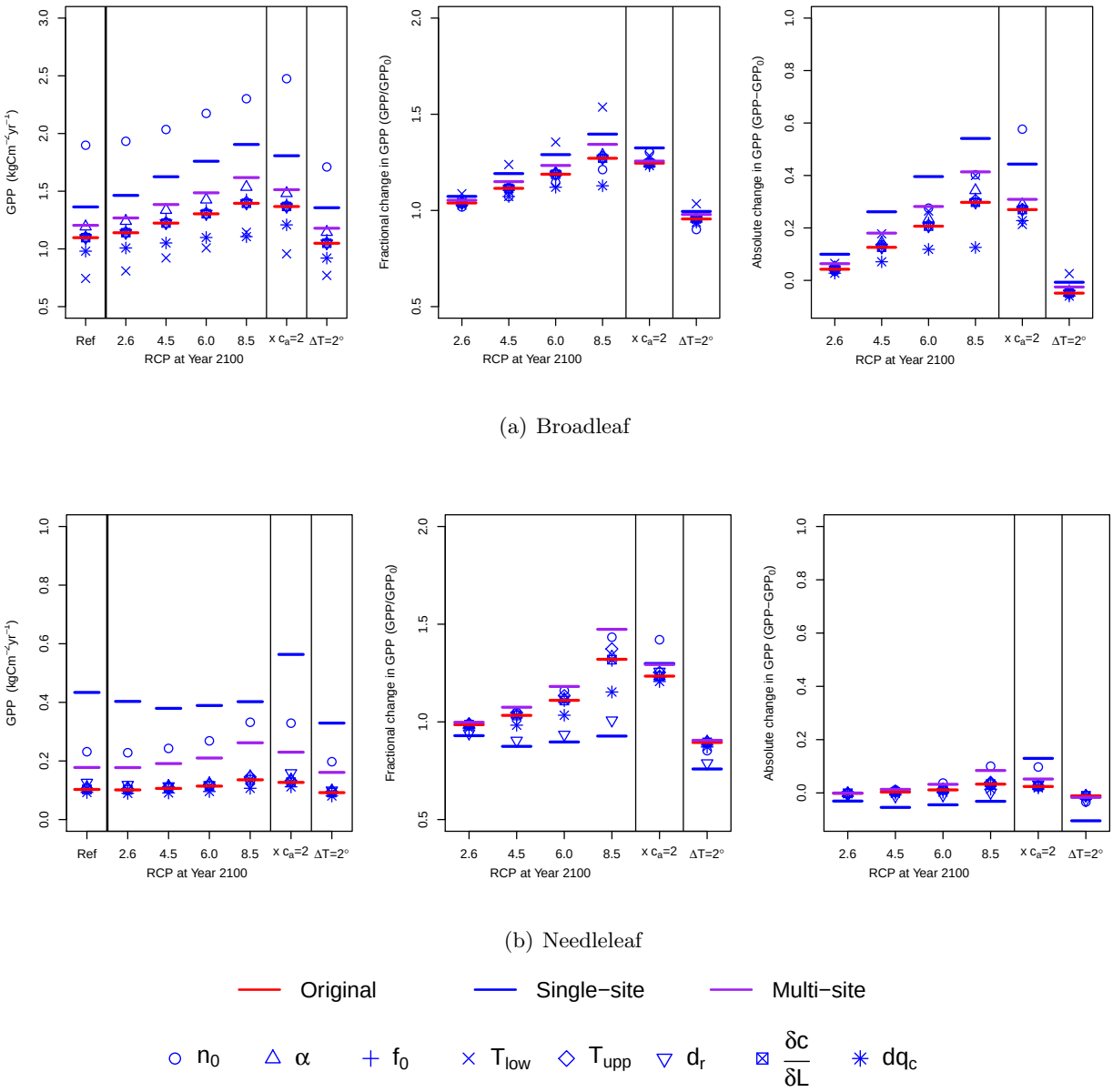
First consider the broadleaf site shown in Fig. 5.4(a). This site is representative of most of the broadleaf set. At each of the atmospheric perturbations, the spread of GPP values for the different parameter settings tested is much larger than the spread of fractional change. This means that the magnitude of the flux varies greatly for different parameter settings. The rate of change of GPP to atmospheric changes, however, is more stable between parameter settings; large changes in atmospheric temperature and $CO_2$ concentration are needed to see the effects of the difference in parameter vectors.

In this example, the magnitude of the GPP is mainly influenced by the $n_0$ and $T_{\mathrm{low}}$ parameters, and to a lesser extent, $\alpha$ and $dq_c$. The $n_0$ parameter inflates the GPP value by a factor of 1.5 when it is the only parameter changed in the parameter vector. This is to be expected as this parameter controls the height of the $V_{\mathrm{cmax}}$ curve (Eq. 2.5), which is use to calculate photosynthesis. When changed as part of the full set of parameters, its effect on the magnitude of the flux is lessened by $T_{\mathrm{low}}$ which acts in the opposite direction. Similarly, the effects of $\alpha$ and $dq_c$ cancel each other out. Across the different atmospheric changes, the effect of the alternative parameter vectors on the GPP value relative to its default position appears to be consistent.

If the actually values of each parameter are considered, the $\alpha$ parameter doubles its values in the optimisation, compared to $n_0$ which changes by a factor of 1.3. However, the $n_0$ parameter can be seen to change the GPP magnitude significantly more than $\alpha$.

The fractional change in GPP (right-hand side of Fig. 5.4) shows the sensitivity of the site to temperature and $c_a$ changes. $T_{\mathrm{low}}$ can be seen to be the most important parameter, followed by $dq_c$. Fractional changes are largest for the most pronounced atmospheric changes. Considering the right-hand columns of the plot were only $c_a$ or $T$ change, $T_{\mathrm{low}}$ is seen to change mainly as a response to increasing temperature.

For this site, $T_{\mathrm{low}}$ increases significantly compared to its uncalibrated value. As discussed




(a) Broadleaf

(b) Needleleaf

Figure 5.4.: Values of GPP (left), fractional change of GPP (middle), and absolute change of GPP (right) for different parameter settings and atmospheric perturbations. Runs using the default JULES parameter vector (horizontal red) are compared to runs using locally-optimised parameter vector (horizontal blue) and the multisite parameter vector (purple). JULES is run eight more times where each parameter is individually changed from the default JULES parameter vector to its locally optimised value (shapes). Six atmospheric changes are covered. In the first four cases, $c_a$ and $\Delta T$ are covaried to represent the RCP atmospheric state at year 2100. In the next case, $c_a$ is doubled with fixed temperature and finally $\Delta T = 2$ with $c_a$ fixed. Note the change of scales for the GPP flux (left) between (a) and (b).

previously, $T_{\text{low}}$ controls the steepness of the lower end of the $V_{\text{cmax}}$ curve (Fig. 4.4). By increasing $T_{\text{low}}$, the lower-end of the curve has become steeper. In order to fix the underestimation of the GPP flux, photosynthesis rates need to increase. By steepening the lower part of the curve, the model is more likely to be in the nitrogen-limited regime control by $V_{\text{cmax}}$, where $n_0$ can be used to increase the magnitude of the flux. The steeper the lower end of the $V_{\text{cmax}}$ curve, the more sensitive the model is to temperature changes. The big-leaf model used in these experiments tends to be light-saturated which is why the light-limited regime is less explored.

The final figure shown in Fig. 5.4(a) illustrates the absolute change in GPP. When only the $dq_c$ parameter is changed from the default vector, the additional GPP saturates with increasing atmospheric perturbations. The $n_0$ parameter in responsible for the increased GPP in response to the $c_a$ changes. Overall, the update of GPP increased with calibration suggests the site is storing more carbon, i.e. is becoming a stronger sink.

Figure 5.4(b) shows the same results but for a needleleaf site. The magnitude of the GPP does not change very much between parameter settings and atmospheric conditions. In a similar manner to the broadleaf site, $n_0$ can be seen to be responsible for the GPP increase when the locally-optimised parameters are used. In contrast, the fractional change in GPP is much more sensitive to the different parameters. The fractional change in GPP is especially sensitive to $n_0$ for carbon changes and $d_r$ for temperature changes. Combined, $d_r$ and $dq_c$ decrease the GPP response to these climate changes, whereas, to a lesser extent, $T_{\text{upp}}$ and $n_0$ increase the fractional change. At some of the other needleleaf sites, $T_{\text{upp}}$ is found also to influence the magnitude of the flux.

When looking at the fractional change of the parameter values between the default JULES vector and the locally-optimised set for this site, $n_0$ nearly triples, as does $\alpha$. The $d_r$ and $dq_c$ parameters change by a factor of 0.65 and 0.79 respectively. These do not translate into the sensitivity seen in the figure; parameters that change the most are not responsible for the most change.

The absolute change in GPP for the needleleaf site is minimal between atmospheric perturbations. The slight increase in the flux resulting from increased $c_a$ is cancelled by the decrease of the flux in response to increasing $\Delta T$.

Finally, note that the optimised parameter values used in this experiment were taken from multivariate distributions. Therefore, it would be possible to test a much larger range of parameter vectors generated from these distributions.

### 5.1.2. Water Use Efficiency

#### Definition

Plants assimilate atmospheric $CO_2$ through photosynthesis and, in doing so, lose water vapour through stomata, the small pores on leaf surfaces that regulate the diffusion of

these two gases between the leaf and the atmosphere. The rate of carbon uptake per unit of water lost is called the water-use efficiency (WUE) and is measured by the ratio

$$\mathrm{WUE} = \frac{\mathrm{GPP}}{E_T} \tag{5.7}$$

where $E_T$ is the transpiration flux, one of the fluxes that contributes to the latent heat flux. Since the calibrations performed in this study were against latent heat and GPP observations, WUE is of particular relevance to this thesis. Given ongoing global environmental issues, such as climatic change and ecosystem degradation, an improved understanding of WUE will help to model and predict the carbon and water cycles better, and to refine water management [Tang et al., 2014].

The question of how much water a plant uses relative to carbon gained is key in understanding the metabolism of terrestrial ecosystems. This closely relates to the interactions between the carbon and water cycles both at the leaf scale and watershed scale [Ito and Inatomi, 2012]. At leaf-level, WUE is controlled by the stomatal exchange of $CO_2$ and water vapour [Cowan, 1972]. At an ecosystem-level, WUE varies among PFTs and environmental conditions [Jarvis and McNaughton, 1986; Schulze et al., 1987]. This section focuses on the former.

Increasing carbon dioxide in the atmosphere tends to increase the rate of photosynthesis (GPP) in the absence of severe nutrient limitations. This process is known as the $CO_2$ fertilisation effect. Under elevated CO2, stomata also tend to close partially [Field et al., 1995], which can lead to reductions in transpiration (ET) and increases in runoff [Gedney et al., 2006].

However, other factors can influence these two rates. For example, $CO_2$ fertilisation of photosynthesis is often found to be limited by nutrient availability [Norby et al., 2010], and large-scale transpiration does not necessarily reduce with $CO_2$-induced stomatal closure if the plant leaf area index increases to counteract this effect [Piao et al., 2007].

Plant photosynthesis and transpiration are coupled through the behaviour of leaf stomatal pores [Dekker et al., 2016]. Both GPP and $E_T$ can be written as the product of a canopy conductance and a concentration gradient. For GPP, the concentration gradient is the difference between the atmospheric $CO_2$ concentration at the leaf surface ($c_a$) and the internal $CO_2$ concentration within plant leaves ($c_i$):

$$\mathrm{GPP} = g_c(c_a - c_i) \tag{5.8}$$

where $g_c$ is the canopy conductance for $CO_2$. This expression is similar to stomatal conductance for water vapour described in Eq. 2.10.

For $E_T$, the concentration gradient is the difference between the specific humidity of the atmosphere at the leaf surface ($q_a$) and the specific humidity inside the plant leaves, which is saturated at the leaf temperature ($q_{\mathrm{sat}}$). The canopy conductances for GPP and $E_T$ arise from diffusion through the leaf stomatal pores, and therefore, only differ by a constant

factor 1.6 [Dekker et al., 2016]. This is the square root of the ratio of the molecular masses of $CO_2$ and $H_2O$.

$$E_T = 1.6 g_c (q_{\text{sat}} - q_a) \tag{5.9}$$

By combining these two expressions, the following equation for WUE is derived:

$$\text{WUE} = \frac{(c_a - c_i)}{1.6(q_{\text{sat}} - q_a)} = \frac{(c_a - c_i)}{1.6\ dq} = \frac{c_a(1-f)}{1.6\ dq} \tag{5.10}$$

where $dq$ is the atmospheric humidity deficit $(q_{\text{sat}} - q_a)$ and $f$ is the ratio of the internal to the external $CO_2$ concentration $(c_i/c_a)$. Hence WUE is written in terms of atmospheric variables, $c_a$ and $dq$ (which itself depends on relative humidity and temperature), along with the factor $f$. From Eq. 2.4, $f$ is seen to take the following form.

$$f \approx f_0 \left(1 - \frac{dq}{dq_c}\right). \tag{5.11}$$

Changes in stomatal opening in response to changes in sunlight, atmospheric temperature and humidity, soil moisture, and $CO_2$ are complex and uncertain [Berry et al., 2010]. By deriving the equation for WUE in this manner, the canopy conductance term $g_s$ is eliminated. This means that WUE is insensitive to this term, and the uncertainties linked to stomatal opening changes are lessened.

**Fractional change**

The definition of WUE in Eq. 5.10 relies predominately on $c_a$ and $dq$. Following the example of Dekker et al. [2016], in this section the fractional changes in WUE in response to fractional changes in $c_a$ and $dq$ are considered. The following expression is used

$$\frac{\text{WUE}}{\text{WUE}(0)} = \left(\frac{c_a}{c_a(0)}\right)^a \left(\frac{dq}{dq(0)}\right)^b, \tag{5.12}$$

where (0) denotes the initial state of each variable, and $a$ and $b$ are dimensionless coefficients. This holds by assuming $f$ from Eq. 5.11 remains constant. By using the logarithmic form of this equation, the coefficients $a$ and $b$ can be calculated:

$$\ln\left(1 + \frac{\Delta\text{WUE}}{\text{WUE}(0)}\right) = a \ln\left(1 + \frac{\Delta c_a}{c_a(0)}\right) + b \ln\left(1 + \frac{\Delta dq}{dq(0)}\right). \tag{5.13}$$

Further to this, the fractional change in humidity deficit can be partitioned into a dependence on temperature change and relative humidity $(R_H)$ change. The equations for humidity deficit and change in humidity deficit are as follows:

$$dq = q_{\text{sat}}(1 - R_H) \tag{5.14}$$

$$\Delta dq = \frac{\partial q_{\text{sat}}}{\partial T}(1 - R_H)\Delta T - q_{\text{sat}} \Delta R_H \tag{5.15}$$

where $R_H$ is the relative humidity. These two expressions combine to give the following:

$$\frac{\Delta dq}{dq} = \frac{1}{q_{\text{sat}}}\frac{\partial q_{\text{sat}}}{\partial T}\Delta T - \frac{\Delta R_H}{(1-R_H)}. \tag{5.16}$$

Using the Clausius-Clapeyron Equation, which states that:

$$\frac{1}{q_{\text{sat}}}\frac{\partial q_{\text{sat}}}{\partial T} = \frac{Lm_w}{RT^2} \approx 0.07\text{K}^{-1} \tag{5.17}$$

for latent heat of vaporisation of water $L = 2.5 \times 10^6$ J kg$^{-1}$, molecular mass of water $m_w = 0.018$ kg mol$^{-1}$ and perfect gas constant $R = 8.31$ J K$^{-1}$ mol$^{-1}$, Eq. 5.16 can be written as

$$\frac{\Delta dq}{dq} \approx 0.07\Delta T - \frac{\Delta R_H}{(1-R_H)} \tag{5.18}$$

In the following experiments, $R_H$ is kept constant so that

$$\frac{\Delta dq}{dq} \approx 0.07\Delta T. \tag{5.19}$$

**Values of $a$ and $b$**

The dimensionless parameters $a$ and $b$ in Eq. 5.10 measure the sensitivity of WUE to $c_a$ and $dq$ respectively. Though working with different underlying assumptions, stomatal optimisation theories from Katul et al. [2010] and Medlyn et al. [2011] suggest that $a = 1$ and $b = -0.5$ (see Dekker et al. [2016] for derivation).

In contrast, experiments using observational data found that these values are higher than predicted (e.g. Keenan et al. [2013]; Dekker et al. [2016]). Dekker et al. [2016] found $a = 0.79 \pm 0.79$ when considering eddy covariance data, and $a = 1.61 \pm 0.54$ when using tree-ring records. The latter is more robust due to the longer length of tree-ring records. Overall, by combining these two experiments, Dekker et al. [2016] found $a = 1.51 \pm 0.57$ and $b = -0.72 \pm 0.16$. These values are about 50% larger than predicted by these optimisation theories [Dekker et al., 2016].

Using the GPP runs from Sect. 5.8 and the transpiration, which is calculated by the JULES model in parallel, values of WUE for changes in $c_a$ and $T$ are found (Fig. 5.5). Parameters $a$ and $b$ are only calculated over a doubling $c_a$ and a temperature anomaly of $\Delta T = 2$. In Fig. 5.5, the values of $a$ found in each of the configurations are slightly lower than expected (theory suggests $a = 1$). The values of $b$ are much more negative than expected, with an absolute value three times larger than the theory suggests. There is also a saturating effect at high temperature anomalies. This saturating effect can be understood by considering the definition of WUE more closely. Equations 5.10 and 5.11 combine to give

$$\text{WUE} = \frac{c_a}{1.6}\left[\frac{1-f_0}{dq} + \frac{f_0}{dq_c}\right]. \tag{5.20}$$

Under low $T$, and therefore low $dq$ conditions since $R_H$ is assumed to be fixed, WUE drops in a manner inversely proportional to $dq$. In contrast, under high T (high $dq$)

conditions, the first term becomes negligible, and so WUE saturates at a low value of $\mathrm{WUE}_{\mathrm{min}} \sim c_a f_0 / dq_c$.

When considering other sites, $a$ is found to be in a similar range of 0.8-0.9. For $b$, values range from $-0.4$ to $-1.7$. Half of the broadleaf sites give a value of $b$ around $-0.5$ (as the theory suggests). The other half have a much lower value of $b$, as illustrated in Fig. 5.5.

The $a$ values found at the different configurations shown in Fig. 5.5 are nearly identical, this is the case for all of the sites. The $b$ values do show more variation. The configuration using the single-site parameters tend to have $b$ values closest to $-0.5$.

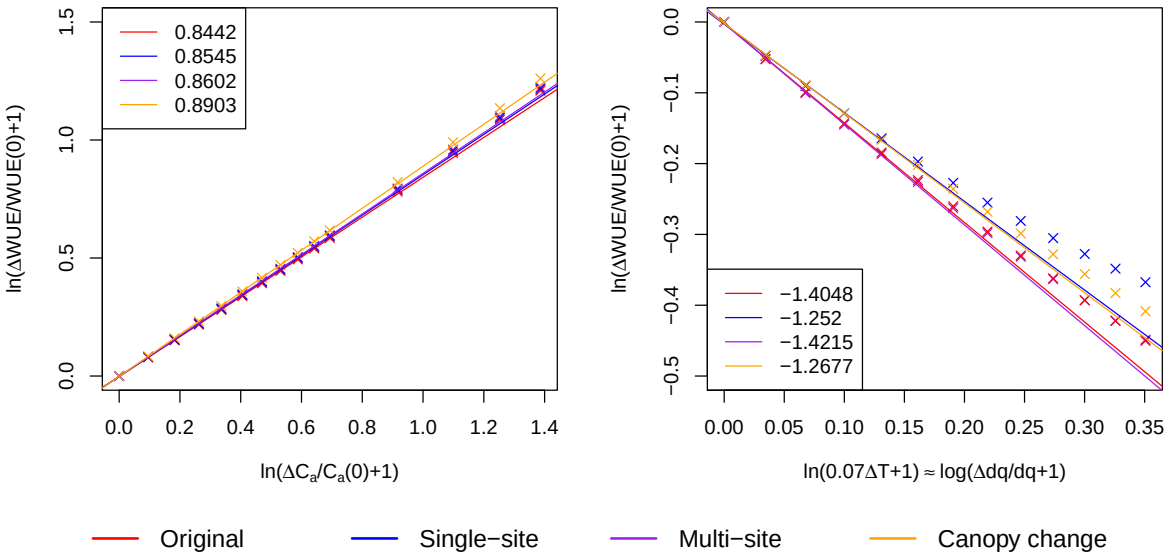



Figure 5.5.: The sensitivity of four different run configurations at broadleaf site US-UMB. The first three are runs with original parameters, locally-optimised parameters and generic-PFT optimised parameter (red, blue, purple). The fourth configuration is a change in canopy representation, moving from a big leaf model to a more complex light limited model (orange). For this configuration, the original parameters are used. Values shown in the left-hand plot are sensitivities to change in carbon $a$, and the values shown in the right-hand plot show the sensitivities to change in temperature $b$.

In trying to understand these trends, it was found that even when the atmospheric temperatures were fixed, increases in the atmospheric $CO_2$ concentration result in an increase of surface temperature ($T^*$). This happens in response to stomatal closure at high $c_a$, since less evaporation means less cooling. Since $q_{\mathrm{sat}}$ is temperature dependent, to take this increase into account, the equation for $E_T$ can be updated as follows:

$$E_T = g_s(\underbrace{q_{\mathrm{sat}}(T_a) - q_a}_{dq} + \frac{\partial q_{\mathrm{sat}}}{\partial T}(T^* - T_a)) \tag{5.21}$$

The definition of WUE described in Eq. 5.10 is the WUE at the atmospheric temperature $T_a$, $\mathrm{WUE}(T_a)$. To calculate WUE when the surface temperature has warmed due to increases in $c_a$, i.e. when $T^* > T_a$, GPP (Eq. 5.8) is divided by surface temperature

dependent value of $\mathrm{E}_T$ (Eq. 5.21) to obtain

$$\mathrm{WUE}(T^*) = \mathrm{WUE}(T_a)\left[\frac{dq}{dq + \frac{\partial q_{sat}}{\partial T}(T^* - T_a)}\right]. \tag{5.22}$$

Using the definition of $dq$ in Eq. 5.15 and the $q_{\mathrm{sat}}$ relationship in Eq. 5.17,

$$\mathrm{WUE}(T^*) = \mathrm{WUE}(T_a)\frac{q_{sat}(T_a)(1 - R_H)}{q_{sat}(T_a)(1 - R_H) + 0.07 q_{sat}(T_a)(T^* - T_a)} \tag{5.23}$$

$$= \mathrm{WUE}(T_a)\left[\frac{1}{1 + \frac{0.07(T^* - T_a)}{1 - R_H}}\right]. \tag{5.24}$$

Let $\Delta T^* = T^* - T_a$ and $\gamma = \frac{0.07}{1 - R_H}$,

$$\mathrm{WUE}(T^*) = \mathrm{WUE}(T_a)\left[\frac{1}{1 + \gamma \Delta T^*}\right]. \tag{5.25}$$

This additional factor on the right-side hand of the equation when $T^* > T_a$ may explain the underestimation in the value of $a$.

For a doubling of $c_a$, using this equation, the fractional change in WUE is expected to give

$$\frac{2}{1 + \gamma \Delta T^*} - 1 = \frac{1 - \gamma \Delta T^*}{1 + \gamma \Delta T^*} \tag{5.26}$$

Calculated over the broadleaf sites, this new theoretical value predicts value between 0.85 and 0.95, is approximately the same as the values calculated at the sites by manual increasing $c_a$. This calculation highlights the need to better understand the role of stomatal closure at high $CO_2$ concentrations.

Note that the calibration of the model barely changes $a$ compared to the default JULES model. Calibration of the model has not increased this sensitivity to that observed by Keenan et al. [2013] and Dekker et al. [2016]. This may be due to experimental design. Calibration of the model only used one year of data for each site compared to the minimum of six years per site used in Dekker et al. [2016]. Different parameters may be needed in the calibration, or the relevant parameters (i.e. $f_0$ and $dq_c$) might not be changing enough. However, the main reason the model gives a low value of $a$ is probably due to structural error.

In this example, calibration of the model has been insufficient in correcting a known fault of model sensitivity. Calibration can only do so much, the underlying model processes need to be changed in order to achieve a high value of $a$.

### Comparing structural and parameter changes

In a final experiment, structural changes are compared to parameter changes, in order to see which effect changes climate sensitivities the most.

To represent structural uncertainty, runs with the different parameter settings are also compared to a run with a different representation of the canopy (with default parameter settings). The two representations considered are the big leaf model and a light-limited canopy model [Mercado et al., 2009]. The default canopy setting is the big leaf model.

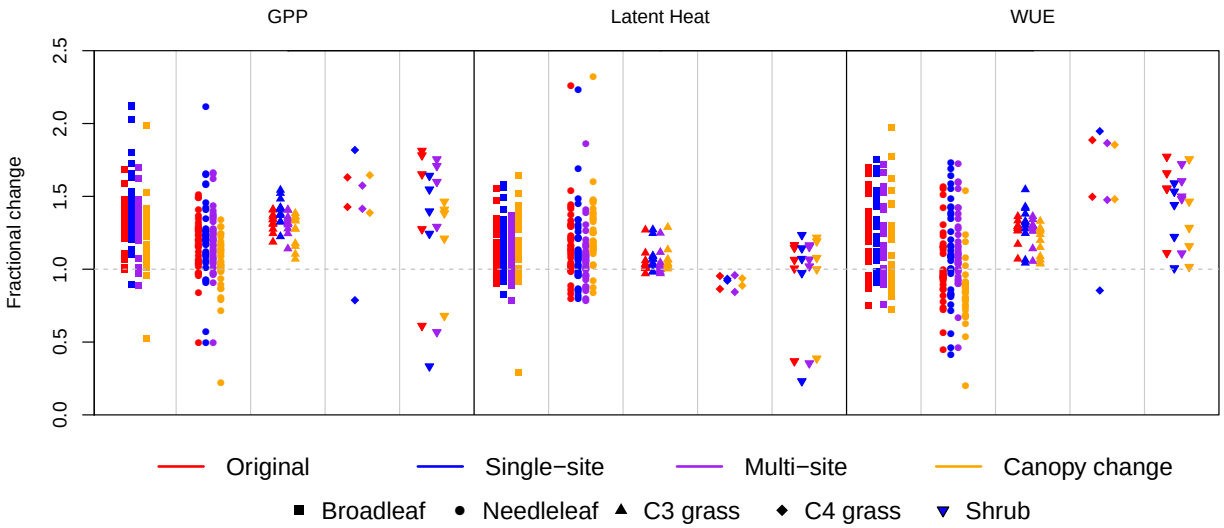



Figure 5.6.: Fractional change of sites at RCP8.5 year 2100. Each PFT is shown, and each site is run with four different JULES configurations: runs with original parameters, locally-optimised parameters and generic-PFT optimised parameters (red, blue, purple) and a fourth configuration where the canopy representation in the model has changes, moving from a big leaf model to a more complex light limited model (orange). For this 'canopy-change' configuration, the original parameters are used.

Figure 5.6 shows the fractional change of three different variables; GPP and latent heat which were used to calibrate the model, and WUE which is a ratio of the two fluxes, at the maximum atmospheric perturbations investigated in this study: RCP 8.5 at year 2100, i.e. $c_a = 2.5$ and $\Delta T = 4.8$. The spread of sites in each PFT is comparable between run configurations. This means that whilst calibration has improved the fit of the model to observations, it has not changed the sensitivity of these three fluxes to changes in the atmospheric temperature and $CO_2$ concentrations. Similarly, the change of canopy representation to a more sophisticated and complex one has not changed the sensitivity of the model.

It therefore seems that different structural changes are needed to change the fractional sensitivity of the model significantly. Since the canopy conductance term was eliminated in formulating WUE, it is possible that this variable is now relatively insensitive to a change in canopy representation.

## 5.2. Closing remarks

This chapter has presented simple sensitivity studies to assess how the calibration of JULES has affected the sensitivity of the model to $CO_2$-induced climate change.

In general, calibration was found to change the absolute size of the fluxes, but had a weaker impact on the fractional sensitivity of those fluxes to $CO_2$ and greenhouse warming. For some sites, GPP increases significantly upon calibration, which would imply larger feedbacks under $CO_2$-induced climate change (e.g. a larger carbon sink due to $CO_2$ fertilisation, a larger carbon source at very high warming levels).

The sensitivity of Water Use Efficiency (WUE) to $CO_2$-increase and warming was also assessed as a useful measure of the changing functioning of plants, which is less dependent on the uncertain behaviour of stomatal pores. Again, calibration was found to have little effect on the fractional sensitivity of WUE to $CO_2$ and warming, although for many sites the absolute value of WUE is changed significantly by the calibration. The calibrated and uncalibrated versions of JULES are less sensitive to $CO_2$ than has been suggested by recent observation-based estimates [Keenan et al., 2013; Dekker et al., 2016].

In the absence of surface temperature increases, JULES should give an increase in WUE that is proportional to $CO_2$, which is consistent with recent stomatal optimisation theories [Medlyn et al., 2011; Prentice et al., 2014]. However, a slightly weaker dependence is found due to increases in surface temperature (and therefore humidity deficit) associated with stomatal closure. A formula has been derived to estimate this slight suppression of WUE increase due to $CO_2$-induced stomatal closure. The humidity deficit dependence of WUE in JULES seems stronger than stomatal optimisation theories. The lower sensitivity of WUE to $CO_2$ than predicted means that in an Earth System simulation, the photosynthesis increase is slightly less than expected or the reductions in transpiration are slightly more.

Finally, the changes in JULES sensitivity due to calibration were compared to the changes in JULES sensitivity due to a major structural code change. The significantly more complex two-stream light-fleck canopy model [Mercado et al., 2009] was used for the latter structural change. In general, calibration was found to improve the performance of the model against observations much more effectively than the structural change, but has a weaker effect on the model's sensitivities. Therefore improvements in process representation are vital to produce more reliable projections, but better calibration is also required for credible models that reproduce contemporary observations reasonably well. Structural and parameter uncertainties therefore need to be reduced in parallel.

The following chapter concludes this thesis by providing discussion of the key findings and suggestions for future work.

# 6. Conclusions

This chapter describes how the analyses throughout this thesis have addressed the key research questions outlined in Chapter 1. Starting with Sect. 6.1, the key findings are summarised and given in response to the key questions. In Sect. 6.2, the results from each chapter are considered in greater depth, and in each case future work is suggested.

Section 6.3 concludes this thesis with a few closing remarks about the future of adJULES.

## 6.1. Key findings

In this section, a brief response is given to each of the key questions outlined in Chapter 1. These highlight the main achievements of this thesis.

**KQ1:** Can a (locally) optimal vector of generic parameters for each of the JULES PFT classes be found in a robust and repeatable manner?

Yes. The adJULES system now includes the option of calibrating over multiple sites simultaneously. This was shown to be robust over different timescales; parameters optimised over a small subset of sites could be generalised to a larger set and parameters found at different observation frequencies could be successfully transposed. In order to be robust and repeatable in the main experiments, optimisations included all available sites in a given set, instead of randomly choosing a subset or selecting the best ones. The multi-site extension was also shown to be necessary; the data from single-site optimisations did not improve as many sites as the data from multi-site optimisations.

**KQ2:** Are the PFT-definitions in JULES robust or do the observations suggest a different partitioning of the vegetation?

Yes and no. The multi-site optimisations over the different PFT groupings were extremely successful. For each PFT, it was possible to find an optimum vector that improved the model-data fit at the majority of the sites. 85% of all sites were improved by the new PFT-generic parameter vectors. Sites that did not improve were found to be outliers in the PFTs. However, using the single-site results in different clustering experiments, as it was not possible to find any sort of natural grouping. Similarly, given a single site parameter vector, it was not possible to determine to which PFT that site belonged.

**KQ3:** How do the parameter changes in JULES affect the model's response to $CO_2$ driven climate change?

Parameter changes affect the model's response under climate change mainly by changing the magnitude of each flux, but not the sensitivity. In this context, model sensitivity refers to the rate of change of each flux with respect to changes in atmospheric $CO_2$ concentrations and atmospheric temperature. The change in model sensitivity is negligible after calibration. Similarly, it is also negligible when changing the canopy representation in the model. This was done to illustrate a reduction in structural uncertainty. Changing this sensitivity requires a different structural change. A few individual parameters do affect the model's sensitivities, but their effects were dulled when part of a larger set. The absolute magnitude of the GPP flux at broadleaf sites is increased through calibration. This strengthens the carbon sinks found at these locations. For needleleaf sites, the absolute change in GPP was minimal.

Data assimilation is primarily used to improve models and so it is worth considering the main implication of the thesis on ecosystem modelling - especially related to JULES. The suggested changes to some of the parameters (primarily in the broadleaf case) highlighted the light-saturated nature of the big-leaf model. This is a known structural simplification - better more realistic results will be achieved using a multi-canopy model. The clustering experiments suggested that there is a need to possibly reexamine the PFT definitions in JULES, maybe consider alternative ways to group and model the different vegetation. Whilst the parameter did improve the fit, a few more experiments are probably needed before the optimised parameters become an integral part of JULES. A few parameters, e.g. $f_0$, were found to be too high in the default model - these could be tweaked to lower values initially. Working in parallel with model development, such as the 9 PFTs, is the most desirable. Also when working with model developer, a set of coding standards could be set up to facilitate the development of the adjoint.

## 6.2. Discussion and future work

In this section, each chapter is considered more throughly along with suggestions for future work.

### 6.2.1. Chapter 2

In Chapter 2, the different components of the adJULES system were introduced. Starting with the JULES model itself, one of the things highlighted was the fact that the version of JULES used is out-dated. Before it can become an integral part of the JULES distribution, the adJULES system needs to catch up to the newest version. In order to make sure the system then keeps up to date with the new releases of JULES, coding standards need to

be introduced so that the JULES model is built in a differentiable manner. Nevertheless, the framework created in this thesis remains relevant and informative.

The FluxNet eddy-covariance data used to constrain the experiments in this thesis were introduced in Sect. 2.2. There is a growing amount of different data available which could be used to constrain the carbon cycle, each spanning different temporal and spatial scales as illustrated in Fig. 6.1. This increased range of data offers many different avenues and possibilities. Each type of observation brings new information to calibrate against and opens up questions about how best to assimilate multiple data streams (see Peylin et al. [2016]).

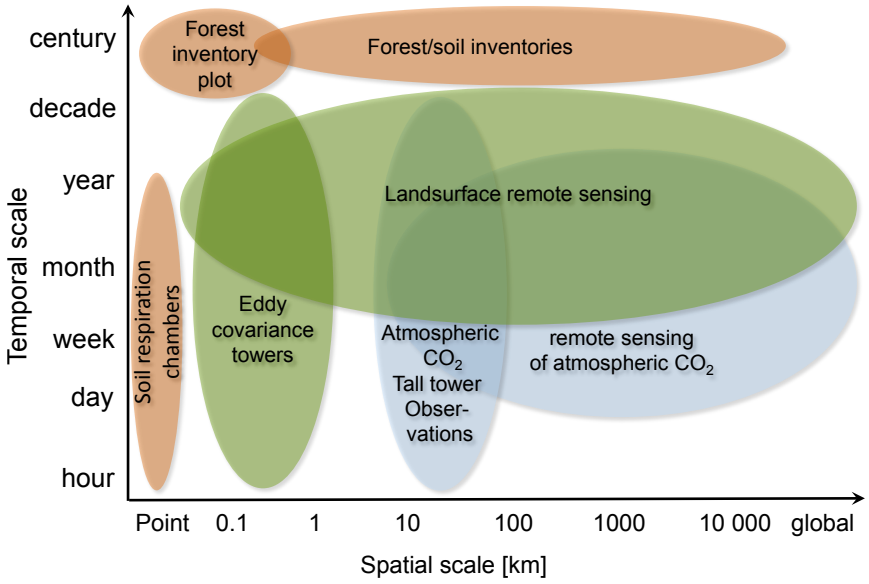



Figure 6.1.: Temporal and spatial scales spanned by available carbon cycle observations. Figure taken from Lasslop [2010].

In Sect. 2.3, data assimilation methods were discussed, focusing in particular on the adjoint method used in this study. There is always a risk of becoming stuck in local minima when optimising within a high-dimensional parameter space by gradient descent. When an optimisation finds a local minimum, the final optimised state depends on the initial conditions. Alternative methods, including ensemble methods, could avoid this issue, but are more computationally costly.

Chapter 2 concluded by discussing the inherited version of the adJULES system. The updated list is as follows, where changes are highlighted (∗). The adJULES system is set up to optimise

- 94 of the physical JULES parameters covering five PFTs and four soil layers
- ∗ simultaneously over a subset of 81 FluxNet sites

- simultaneously over a subset of the 6 different data streams (NEE, H, LE, T, GPP, Resp)

  * following the cost function:

$$J\left(\mathbf{z};\hat{\mathbf{z}},\mathbf{z}_0\right)=\frac{1}{2}\left[\sum_t\left(\mathbf{m}_t(\mathbf{z})-\mathbf{o}_t\right)^T\mathbf{R}\left(\hat{\mathbf{z}}\right)^{-1}\left(\mathbf{m}_t(\mathbf{z})-\mathbf{o}_t\right)+\lambda(\mathbf{z}-\mathbf{z}_0)^T\mathbf{B}^{-1}(\mathbf{z}-\mathbf{z}_0)\right] \tag{6.1}$$

  * by iteratively looping until $\hat{\mathbf{z}}$ converges

- using the adjoint generated from JULES version 2.2

- using the BFGS optimisation scheme

This chapter leads to the following suggestions for further work:

- Update the adJULES system to the newest version of JULES.

- Integrate new observations, such as satellite data, in the adJULES system.

### 6.2.2. Chapter 3

Chapter 3 discussed the cost function used in the adJULES system at great length. The mathematical theory used to define the **R** and **B** covariance matrices was covered, and alternative definitions of the cost function discussed.

In this chapter, the adJULES system is successfully extended to calibrate over multiple sites simultaneously. This multi-site optimisation is a relatively new feature in terrestrial data assimilation. The multi-site extension of adJULES was shown to be robust: a calibration performed over a subset of the sites can be generalised over a larger set of sites. Even with a clear outlier as part of the training set, the multi-site optimisation is able to find a best-fit set of parameters for the other sites.

When extending the cost function to optimise over multiple sites in Sect. 3.2, other implementations were also considered. One such feature allows for the calibration of a common set of parameters over multiple sites, while also allowing other parameters at the sites to improve individually.

The main challenge of this chapter was trying to understand the effect of a penalisation (background) term in the cost function on the calibration results, especially in a multi-site framework where the additional sites also add constraints to the cost function. The idea of conditioning was introduced and the addition of the background term was shown to reduce the condition number significantly from $\mathcal{O}(10^{13})$ to $\mathcal{O}(10^5)$. However, in most cases, the penalisation term was then found to dominate the cost function and the parameter values did not significantly change from their initial values, particularly when the model was calibrated with daily observations (i.e. with many data points). The question of how best to weight this background term remains open.

The two experiments in this chapter suggest a possible 'smoothing' of the cost function; an idea first proposed in Kuppel et al. [2014]. This is the idea that increasing the number of sites adds constraints on the parameters, thereby causing the cost function to become 'smoother' and thus the optimisation scheme may be less likely to become trapped in local minima. First, the parameter vectors found over multiple sites were sometimes found to outperform the locally optimised parameter vectors. In these cases, a different and better minimum had been discovered. Second, the addition of more sites to the optimisation meant that only a low-weighted background term was needed to help condition the problem. The additional data provided by the extra sites replaced the extra constraints given by the background term. However, a sensitivity experiment used to test this 'smoothing' hypothesis was inconclusive.

Finally, the adJULES system was found to be more successful in finding an optimal parameter vector than randomly sampling parameter space. However, the system was shown to be sensitive to the initial conditions.

This chapter leads to the following suggestions for further work:

- Explore different formulations of the cost function, potentially consider allowing additional parameters to vary locally in parallel to the main optimisation.

- Investigate the relative importance of the background term, number of sites, and observation frequency in conditioning the problem.

### 6.2.3. Chapter 4

In Chapter 4, the main results of the optimisations were covered. For individual FluxNet sites, adJULES was shown to have the ability to find local (site-specific) optimal parameter vectors that significantly improve the performance of the JULES model compared to runs generated using the default parameters. The data streams used in the calibration, LE and GPP, are both modelled more accurately with the optimal parameter vectors, with the GPP flux improving the most. The greater improvement in the GPP flux is largely due to the fact that the parameters considered in this study are mainly related to photosynthesis. For the LE flux to improve more significantly, more water and energy-related parameters would need to be considered in the optimisation.

When optimised locally to find site-specific parameters, all of the sites in this study were seen to improve the model-data fit for the calibration year. In addition, when confronted with independent data from an evaluation year, the locally optimised parameter vectors decreased the error in model-data fit for 84% of the sites. This evaluation of the site-specific parameter vectors is promising, and suggests that the adJULES system is robust. It also gives confidence that the parameter vectors found can be generalised over different locations.

Although the PFT-specific optimal parameters do not always fit the data as well as site-specific optimal parameters, they still offer significant improvements over the default

JULES parameters. For over 85% of the sites, PFT-specific optimal parameters perform better than default parameters when confronted with independent evaluation data. For some of the sites, the PFT-specific optimal parameters perform at least as well as site-specific optimal parameters. This implies that the multi-site methodology is less susceptible to over-tuning, both in terms of variability across sites (e.g. different overground biomass and tree ranges), and in terms of variability through time (e.g. unusually high rainfall in the calibration year).

The PFT-specific parameters found in this study represent a significant improvement on the default ones. That such parameters could be found implies robust parameterisations independent of geography, supporting the idea that it is possible to represent global vegetation with a relatively small number of PFTs.

A successful and robust multi-site optimisation assumes that sites can be grouped and parameter values can be applied to several sites at once. Whilst the PFT-specific parameters show great improvement, agreeing with the use of five PFTs in JULES, the latter half of the chapter challenged these groupings.

The alternative PFT definitions presented in Harper et al. [2016] were compared to the calibrated JULES model from this study. Both were shown to improve fluxes in different ways, highlighting the need for both calibration and improved process representation in the JULES model.

Finally, a cluster analysis in parameter space was performed in order to identify PFTs empirically. The data retrieved from the single-site optimisation did not suggest clustering of any kind.

This chapter leads to the following suggestions for further work:

- Consider a different parameter vector in the optimisation with the addition of LE-based parameters.
- Calibrate over the updated PFT groupings suggested by Harper et al. [2016].

### 6.2.4. Chapter 5

In Chapter 5, simple sensitivity studies were used to assess how the calibration of JULES affects the sensitivity of the model to $CO_2$-induced climate change. In general, calibration was found especially to change the absolute size of the quantities considered, but had a weaker impact on the fractional sensitivity of those quantities to $CO_2$ fertilisation and warming.

Two quantities were discussed in this chapter: GPP and WUE. For some sites, the GPP flux was found to increase significantly upon calibration, implying larger feedbacks under $CO_2$-induced climate change (e.g. a larger carbon sink due to $CO_2$ fertilisation, a larger carbon source at very high warming levels). An optimal temperature for photosynthesis

was derived at each site. This was different to the optimal temperature derived from the underlying equations and was found to increase with increasing $CO_2$.

WUE was considered since it provides a useful measure of the changing functioning of plants, and related to the two fluxes, GPP and LE, used in the calibration. Calibration was found little effect on the fractional sensitivity of WUE to $CO_2$ and warming, and the calibrated and uncalibrated versions of JULES are less sensitive to $CO_2$ than has been suggested by recent observation-based estimates [Keenan et al., 2013; Dekker et al., 2016].

JULES was found to give a slightly weaker dependence of WUE on $CO_2$ increase than expected; the increase in WUE is suggested to be proportional to $CO_2$ in recent stomatal optimisation theories [Medlyn et al., 2011]. This weaker dependence was found to be due to increases in surface temperature, and therefore humidity deficit, associated with stomatal closure.

JULES allows for different representation of the canopy, ranging from the simple big leaf model which is used throughout this thesis, to the the significantly more complex two-stream light-fleck canopy model [Mercado et al., 2009]. This difference in representation allowed for the changes in JULES sensitivity due to calibration to be compared to the changes in JULES sensitivity due to a major structural code change. Calibration was found to reduce errors in model-data more effectively than the structural change. However, calibration did not change the model's sensitivities. Improvements in process representation are needed to capture these sensitivities and therefore produce reliable climate projects. Calibration is still important to ensure credible models. This highlights the need for simultaneous reductions in both structural and parameter uncertainties.

Limitations of this study are linked to its simplicity. For one, the temperature anomaly was applied uniformly across the time-series at all sites. The global temperatures changes derived from the RCPs were assumed to mirror the local temperature changes. However, temperature response will differ spatially and temporally. Using IMOGEN patterns [Huntingford et al., 2010], localised temperature response can be derived and used to better understand $CO_2$ induced climate change at different locations.

The second simplification is due to optimal values themselves. These optimal values belong to the multivariate normal distribution and as such have uncertainties associated with them. In order for the experiments to be more informative, not only do the optimal values need to be considered but their associated uncertainties need to be also feed-through.

This chapter leads to the following suggestions for further work:

- Use IMOGEN patterns to find the localised temperature responses to $CO_2$ induced climate change.

- Run similar sensitivity studies including the uncertainties associated with each parameter.

## 6.3. The future of adJULES

The adJULES system is an extremely powerful tool with great potential. It provides a much needed framework to confront the JULES model with observational data. In addition to the future work discussed throughout this chapter, the adJULES system could also be used in identification of model structural errors and this presents opportunities to make real improvements to model parameterisation.

# Appendices

# A. Description of the FluxNet Data

Table A.1.: FluxNet sites used in this study, labelled by a country code (first two letters) and site name (last three letters). The period corresponds to the available years of data for each of the sites. See Groenendijk et al. [2011] for the site references.

| Site | Period | Calibration year | Evaluation year | Latitude | Longitude |
|---|---|---|---|---|---|
| Broadleaf sites (BT) | | | | | |
| DE-Hai | (2000, 2006) | 2005 | 2004 | 51.079 | 10.452 |
| DK-Sor | (1996, 2006) | 2006 | 2004 | 55.487 | 11.646 |
| FR-Fon | (2005, 2006) | 2006 | – | 48.476 | 2.780 |
| FR-Hes | (1997, 2006) | 2003 | 1998 | 48.674 | 7.065 |
| IT-Col | (1996, 2006) | 2005 | 2001 | 41.849 | 13.588 |
| IT-LMa | (2003, 2006) | 2006 | 2004 | 45.581 | 7.155 |
| IT-Non | (2001, 2006) | 2002 | 2003 | 44.690 | 11.089 |
| IT-PT1 | (2002, 2004) | 2003 | 2004 | 45.201 | 9.061 |
| IT-Ro1 | (2000, 2006) | 2006 | 2005 | 42.408 | 11.930 |
| IT-Ro2 | (2002, 2006) | 2004 | 2006 | 42.390 | 11.921 |
| UK-Ham | (2004, 2005) | 2005 | – | 51.121 | −0.861 |
| UK-PL3 | (2005, 2006) | 2006 | – | 51.450 | −1.267 |
| US-Bar | (2004, 2005) | 2005 | – | 44.065 | −71.288 |
| US-Ha1 | (1991, 2006) | 1996 | 1998 | 42.538 | −72.171 |
| US-MMS | (1999, 2005) | 2002 | 2003 | 39.323 | −86.413 |
| US-MOz | (2004, 2006) | 2006 | 2005 | 38.744 | −92.200 |
| US-UMB | (1999, 2003) | 2003 | 2002 | 45.560 | −84.714 |
| US-WCr | (1999, 2006) | 2005 | 2000 | 45.806 | −90.080 |
| AU-Tum | (2001, 2006) | 2003 | 2005 | −35.656 | 148.152 |
| AU-Wac | (2005, 2007) | 2006 | – | −37.429 | 145.187 |
| BR-Sa1 | (2002, 2004) | 2003 | 2004 | −2.857 | −54.959 |
| BR-Sa3 | (2000, 2003) | 2002 | 2003 | −3.018 | −54.971 |
| FR-Pue | (2000, 2006) | 2006 | 2005 | 43.741 | 3.596 |
| ID-Pag | (2002, 2003) | 2003 | – | 2.345 | 114.036 |
| IT-Cpz | (1997, 2006) | 2004 | 2006 | 41.705 | 12.376 |
| IT-Lec | (2005, 2006) | 2006 | – | 43.305 | 11.271 |
| PT-Esp | (2002, 2004) | 2004 | 2003 | 38.639 | −8.602 |
| PT-Mi1 | (2003, 2005) | 2005 | – | 38.541 | −8.000 |
| C3 grasses sites (C3G) | | | | | |
| DE-Gri | (2005, 2006) | 2006 | – | 50.950 | 13.512 |
| DK-Lva | (2005, 2006) | 2006 | – | 55.683 | 12.083 |
| ES-LMa | (2004, 2006) | 2006 | 2005 | 39.941 | −5.773 |
| HU-Bug | (2002, 2006) | 2006 | 2005 | 46.691 | 19.601 |
| HU-Mat | (2004, 2006) | 2006 | 2005 | 47.847 | 19.726 |
| IT-Amp | (2002, 2006) | 2006 | 2005 | 41.904 | 13.605 |
| PL-wet | (2004, 2005) | 2005 | – | 52.762 | 16.309 |
| PT-Mi2 | (2004, 2006) | 2006 | 2005 | 38.477 | −8.025 |
| US-Bkg | (2004, 2006) | 2006 | 2005 | 44.345 | −96.836 |
| US-FPe | (2000, 2006) | 2002 | 2004 | 48.308 | −105.101 |
| US-Goo | (2002, 2006) | 2006 | 2004 | 34.250 | −89.970 |

Table A.1.: Continued.

| Site | Period | Calibration year | Evaluation year | Latitude | Longitude |
|---|---|---|---|---|---|
| Needleleaf sites (NT) | | | | | |
| CA-Man | (1997, 2003) | 2001 | 2002 | 55.880 | −98.481 |
| CA-NS1 | (2002, 2005) | 2004 | 2003 | 55.879 | −98.484 |
| CA-NS2 | (2001, 2005) | 2002 | 2004 | 55.906 | −98.525 |
| CA-NS3 | (2001, 2005) | 2004 | 2002 | 55.912 | −98.382 |
| CA-NS4 | (2002, 2004) | 2004 | 2003 | 55.912 | −98.382 |
| CA-NS5 | (2001, 2005) | 2004 | 2002 | 55.863 | −98.485 |
| CA-Qcu | (2001, 2006) | 2005 | 2006 | 49.267 | −74.037 |
| CA-Qfo | (2003, 2006) | 2006 | 2005 | 49.693 | −74.342 |
| CA-SF1 | (2003, 2005) | 2004 | 2005 | 54.485 | −105.818 |
| CA-SF2 | (2003, 2005) | 2004 | 2005 | 54.254 | −105.878 |
| CA-SF3 | (2003, 2005) | 2005 | 2004 | 54.092 | −106.005 |
| DE-Bay | (1996,1999) | 1999 | 1998 | 50.142 | 11.867 |
| DE-Har | (2005, 2006) | 2006 | – | 47.934 | 7.601 |
| DE-Tha | (1996, 2006) | 2005 | 2004 | 50.964 | 13.567 |
| DE-Wet | (2002, 2006) | 2006 | 2004 | 50.453 | 11.457 |
| ES-ES1 | (1999, 2006) | 2005 | 2000 | 39.346 | −0.319 |
| FI-Hyy | (1996, 2006) | 2006 | 2004 | 61.847 | 24.295 |
| FR-LBr | (2003, 2006) | 2006 | 2005 | 44.717 | −0.769 |
| IL-Yat | (2001, 2006) | 2005 | 2006 | 31.345 | 35.051 |
| IT-Lav | (2000, 2002) | 2001 | 2002 | 45.955 | 11.281 |
| IT-Ren | (1999, 2006) | 2005 | 2006 | 46.588 | 11.435 |
| IT-SRo | (1999, 2006) | 2006 | 2005 | 43.728 | 10.284 |
| NL-Loo | (1996, 2006) | 2006 | 2003 | 52.168 | 5.744 |
| RU-Fyo | (1998, 2006) | 2005 | 2006 | 56.462 | 32.924 |
| RU-Zot | (2002, 2004) | 2003 | 2004 | 60.801 | 89.351 |
| SE-Fla | (1996,1998) | 1998 | 1997 | 64.113 | 19.457 |
| SE-Nor | (1996,1999) | 1997 | 1999 | 60.086 | 17.480 |
| SE-Sk2 | (2004, 2005) | 2005 | – | 60.130 | 17.840 |
| UK-Gri | (1997,1998) | 1998 | – | 56.607 | −3.798 |
| US-Blo | (1997, 2006) | 2006 | 2000 | 38.895 | −120.633 |
| US-Ho1 | (1996, 2004) | 2004 | 2003 | 45.204 | −68.740 |
| US-Me4 | (1996, 2000) | 2000 | – | 44.499 | −121.622 |
| US-SP1 | (2000, 2001) | 2001 | – | 29.738 | −82.219 |
| US-SP2 | (1998, 2004) | 2001 | 2004 | 29.765 | −82.245 |
| US-SP3 | (1999, 2004) | 2001 | 2002 | 29.755 | −82.163 |
| Shrubs sites (Sh) | | | | | |
| CA-Mer | (1998, 2005) | 2004 | 2005 | 45.409 | −75.519 |
| CA-NS6 | (2001, 2005) | 2003 | 2004 | 55.917 | −98.964 |
| CA-NS7 | (2002, 2005) | 2003 | 2004 | 56.636 | −99.948 |
| IT-Pia | (2002, 2005) | 2003 | 2004 | 42.584 | 10.078 |
| US-Los | (2001, 2005) | 2005 | 2003 | 46.083 | −89.979 |
| C4 grasses sites (C4G) | | | | | |
| BW-Ma1 | (1999, 2001) | 2000 | 2001 | −19.916 | 23.561 |
| ZA-Kru | (2001, 2003) | 2002 | 2003 | −25.020 | 31.497 |

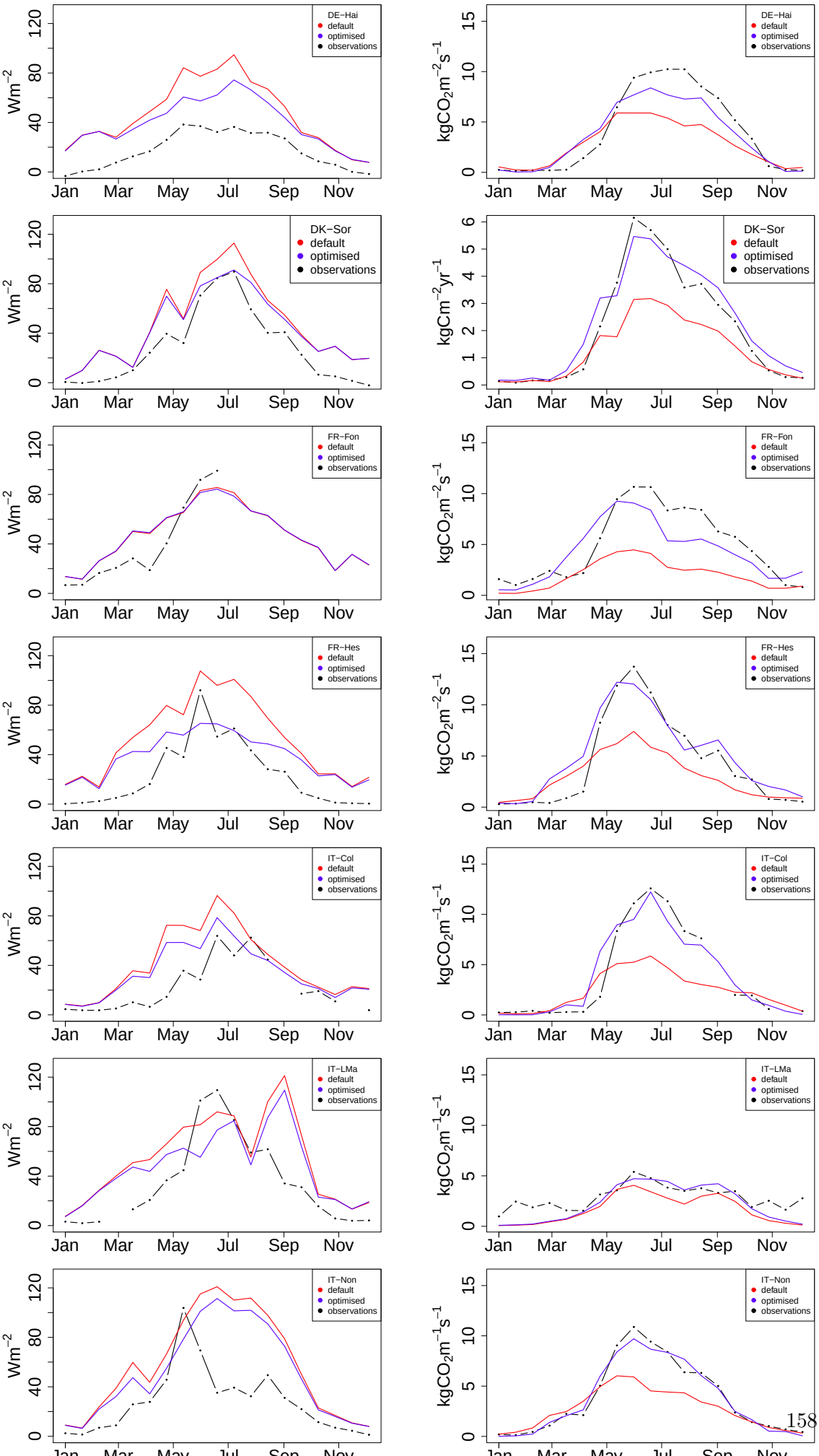
Table A.2.: IGBP land cover classification system. Table retrieved from www.eomf.ou.edu/static/IGBP.pdf

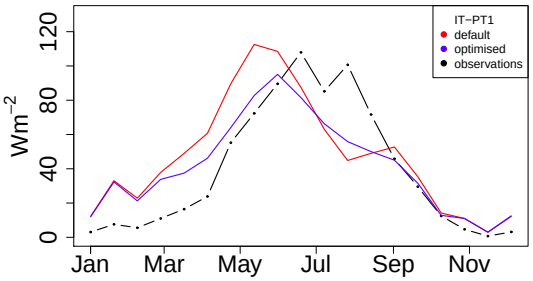
| | Class name | Description |
|---|---|---|
| 1 | Evergreen needleleaf forests | Lands dominated by needleleaf woody vegetation with a percent cover 60% and height exceeding 2m. Almost all trees remain green all year. Canopy is never without green foliage. |
| 2 | Evergreen broadleaf forests | Lands dominated by broadleaf woody vegetation with a percent cover 60% and height exceeding 2m. Almost all trees and shrubs remain green year round. Canopy is never without green foliage. |
| 3 | Deciduous needleleaf forests | Lands dominated by woody vegetation with a percent cover 60% and height exceeding 2m. Consists of seasonal needleleaf tree communities with an annual cycle of leaf-on and leaf-off periods. |
| 4 | Deciduous broadleaf forests | Lands dominated by woody vegetation with a percent cover 60% and height exceeding 2m. Consists of broadleaf tree communities with an annual cycle of leaf-on and leaf-off periods. |
| 5 | Mixed forests | Lands dominated by trees with a percent cover 60% and height exceeding 2m. Consists of tree communities with interspersed mixtures or mosaics of the other four forest types. None of the forest types exceeds 60% of landscape. |
| 6 | Closed shrublands | Lands with woody vegetation less than 2m tall and with shrub canopy cover 60%. The shrub foliage can be either evergreen or deciduous. |
| 7 | Open shrublands | Lands with woody vegetation less than 2m tall and with shrub canopy cover between 10% and 60%. The shrub foliage can be either evergreen or deciduous. |
| 8 | Woody savannas | Lands with herbaceous and other understory systems, and with forest canopy cover between 30% and 60%. The forest cover height exceeds 2m. |
| 9 | Savannas | Lands with herbaceous and other understory systems, and with forest canopy cover between 10% and 30%. The forest cover height exceeds 2m. |
| 10 | Grasslands | Lands with herbaceous types of cover. Tree and shrub cover is less than 10%. |
| 11 | Permanent wetlands | Lands with a permanent mixture of water and herbaceous or woody vegetation. The vegetation can be present either in salt, brackish, or fresh water. |
| 12 | Croplands | Lands covered with temporary crops followed by harvest and a bare soil period (e.g., single and multiple cropping systems). Note that perennial woody crops will be classified as the appropriate forest or shrub land cover type. |
| 13 | Urban and built-up lands | Land covered by buildings and other man-made structures. |
| 14 | Cropland/natural vegetation mosaics | Lands with a mosaic of croplands, forests, shrubland, and grasslands in which no one component comprises more than 60% of the landscape. |
| 15 | Snow and ice | Lands under snow/ice cover throughout the year. |
| 16 | Barren | Lands with exposed soil, sand, rocks, or snow and never have more than 10% vegetated cover during any time of the year. |
| 17 | Water bodies | Oceans, seas, lakes, reservoirs, and rivers. Can be either fresh or saltwater bodies. |

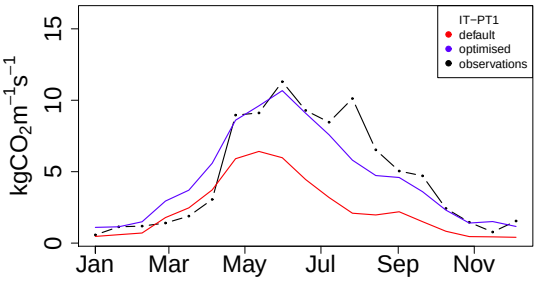
# B. Optimised time-series

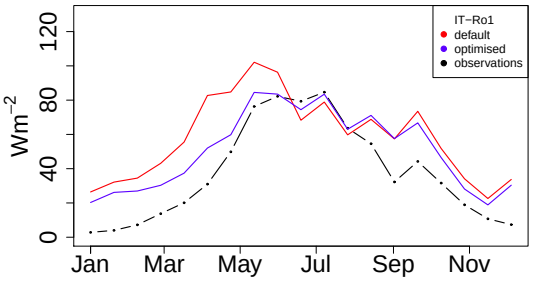
Time-series plots for the different site-specific evaluations showing LE (left) and GPP (right) for each of the different PFTs. Observations (black) are compared to JULES runs using default parameters (red) and site-specific optimal parameters (blue).

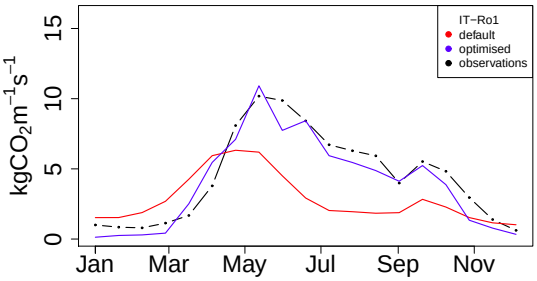
## B.1. Broadleaf sites

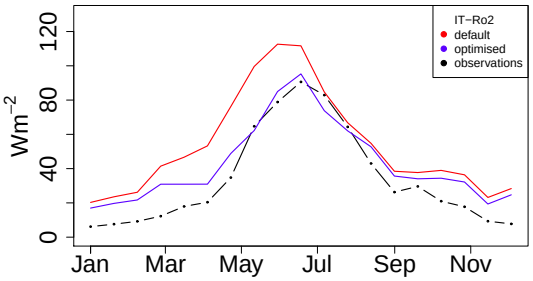


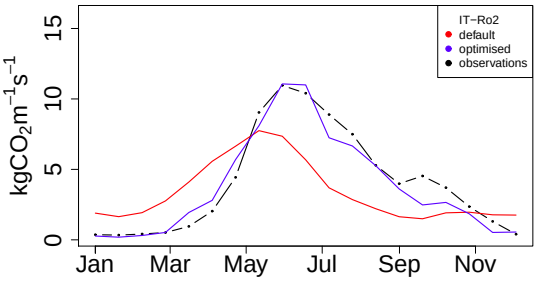

IT−PT1
default
optimised
observations
Wm−2
0
40
80
120
Jan
Mar
May
Jul
Sep
Nov

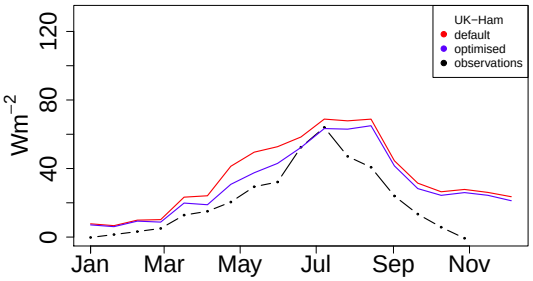

IT−PT1
default
optimised
observations
kgCO2m−1s−1
0
5
10
15
Jan
Mar
May
Jul
Sep
Nov

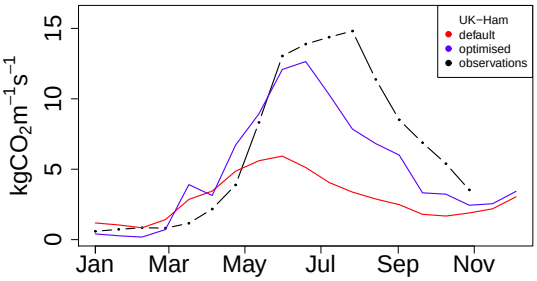

IT−Ro1
default
optimised
observations
Wm−2
0
40
80
120
Jan
Mar
May
Jul
Sep
Nov

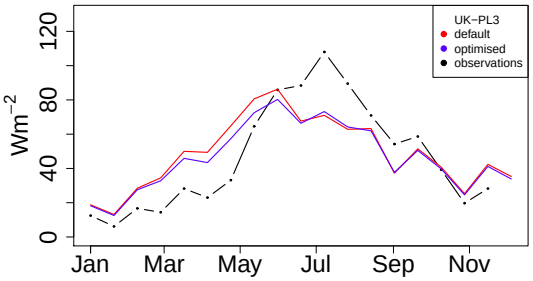

IT−Ro1
default
optimised
observations
kgCO2m−1s−1
0
5
10
15
Jan
Mar
May
Jul
Sep
Nov

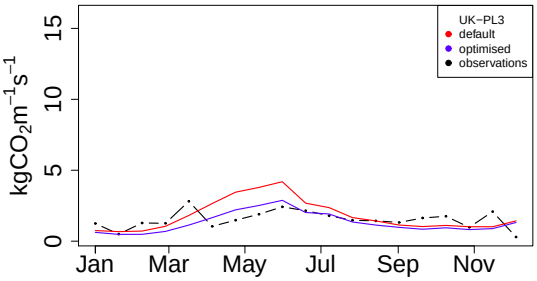

IT−Ro2
default
optimised
observations
Wm−2
0
40
80
120
Jan
Mar
May
Jul
Sep
Nov

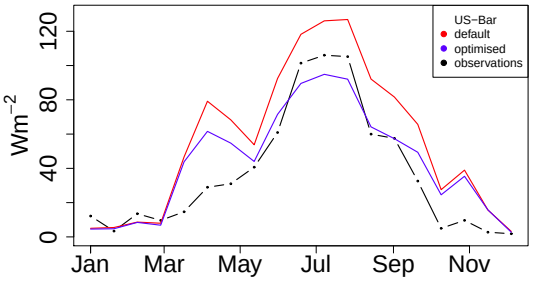

IT−Ro2
default
optimised
observations
kgCO2m−1s−1
0
5
10
15
Jan
Mar
May
Jul
Sep
Nov

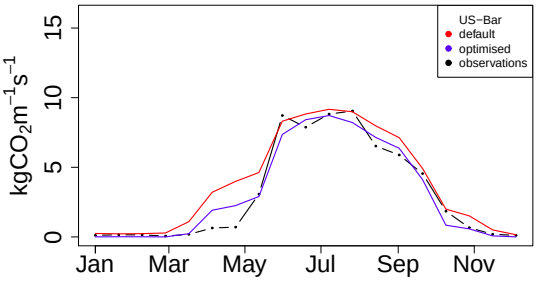

UK−Ham
default
optimised
observations
Wm−2
0
40
80
120
Jan
Mar
May
Jul
Sep
Nov

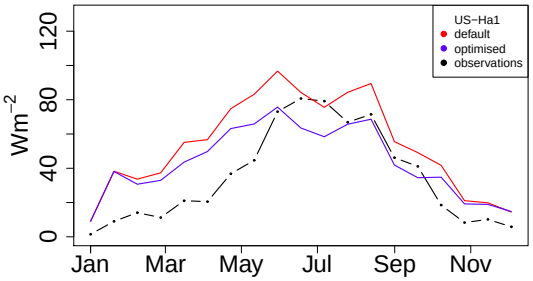

UK−Ham
default
optimised
observations
kgCO2m−1s−1
0
5
10
15
Jan
Mar
May
Jul
Sep
Nov

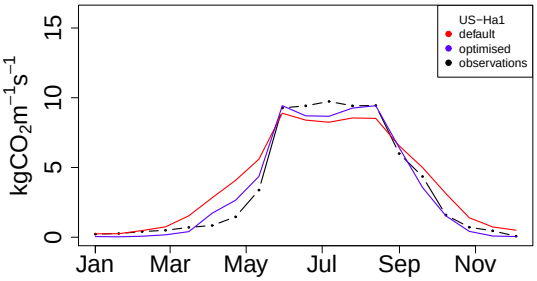

UK−PL3
default
optimised
observations
Wm−2
0
40
80
120
Jan
Mar
May
Jul
Sep
Nov

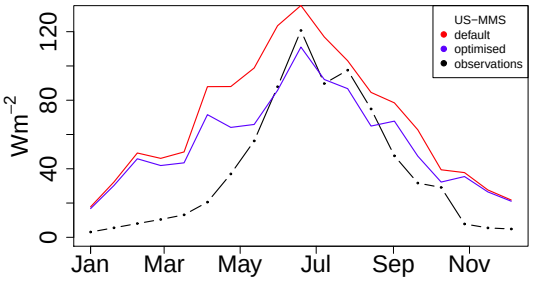

UK−PL3
default
optimised
observations
kgCO2m−1s−1
0
5
10
15
Jan
Mar
May
Jul
Sep
Nov

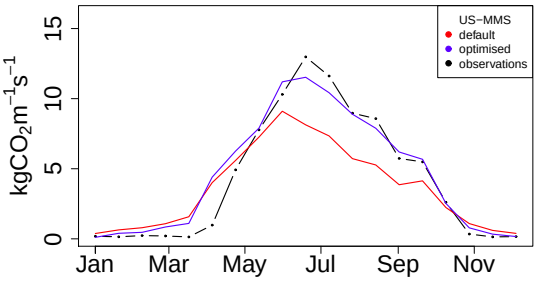

US−Bar
default
optimised
observations
Wm−2
0
40
80
120
Jan
Mar
May
Jul
Sep
Nov

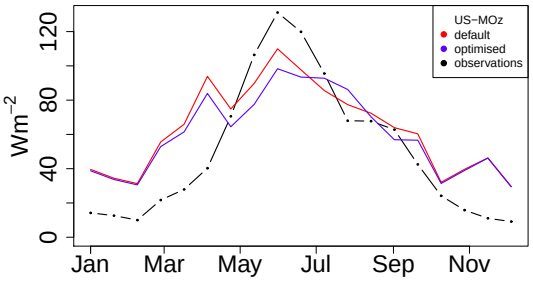

US−Bar
default
optimised
observations
kgCO2m−1s−1
0
5
10
15
Jan
Mar
May
Jul
Sep
Nov

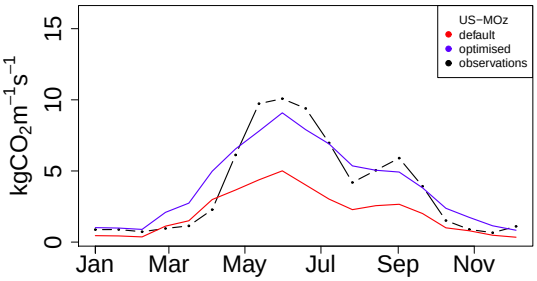

US−Ha1
default
optimised
observations
Wm−2
0
40
80
120
Jan
Mar
May
Jul
Sep
Nov

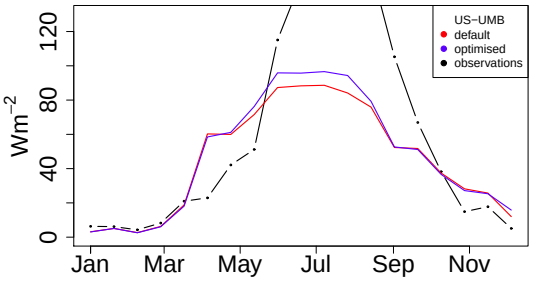

US−Ha1
default
optimised
observations
kgCO2m−1s−1
0
5
10
15
Jan
Mar
May
Jul
Sep
Nov

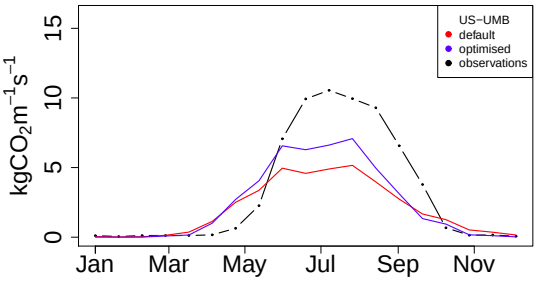

US−MMS
default
optimised
observations
Wm−2
0
40
80
120
Jan
Mar
May
Jul
Sep
Nov

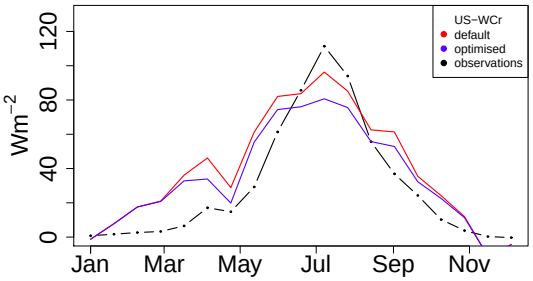

US−MMS
default
optimised
observations
kgCO2m−1s−1
0
5
10
15
Jan
Mar
May
Jul
Sep
Nov

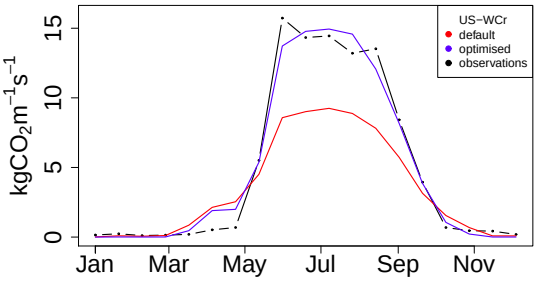

US−MOz
default
optimised
observations
Wm−2
0
40
80
120
Jan
Mar
May
Jul
Sep
Nov

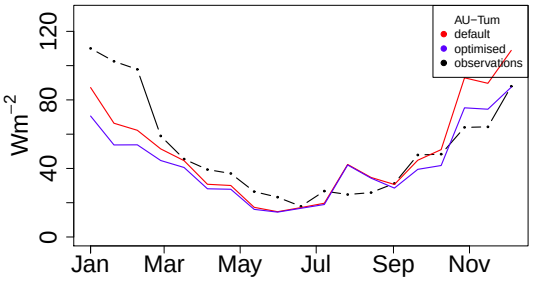

US−MOz
default
optimised
observations
kgCO2m−1s−1
0
5
10
15
Jan
Mar
May
Jul
Sep
Nov

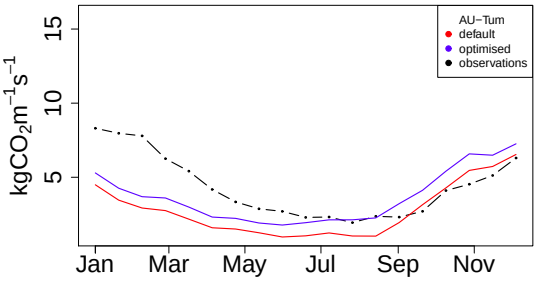

US−UMB
default
optimised
observations
Wm−2
0
40
80
120
Jan
Mar
May
Jul
Sep
Nov

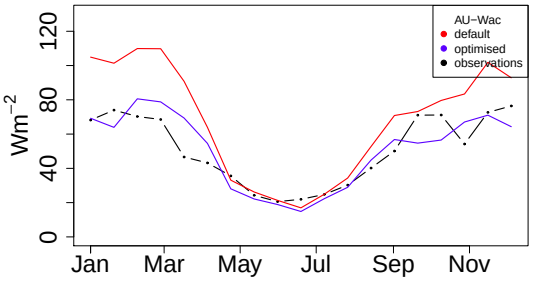

US−UMB
default
optimised
observations
kgCO2m−1s−1
0
5
10
15
Jan
Mar
May
Jul
Sep
Nov

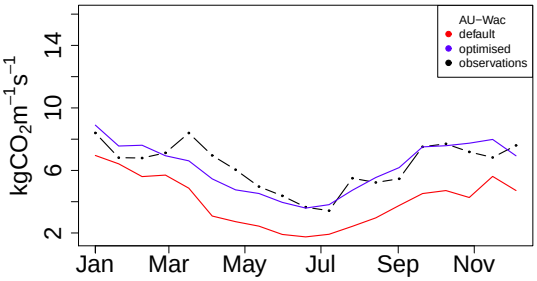

US−WCr
default
optimised
observations
Wm−2
0
40
80
120
Jan
Mar
May
Jul
Sep
Nov

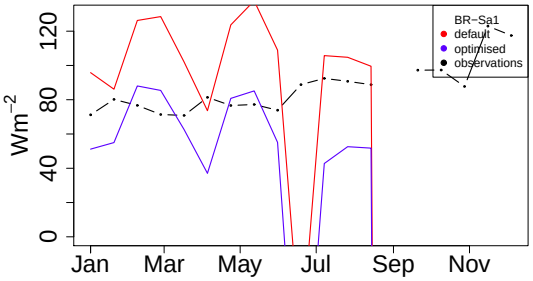

US−WCr
default
optimised
observations
kgCO2m−1s−1
0
5
10
15
Jan
Mar
May
Jul
Sep
Nov

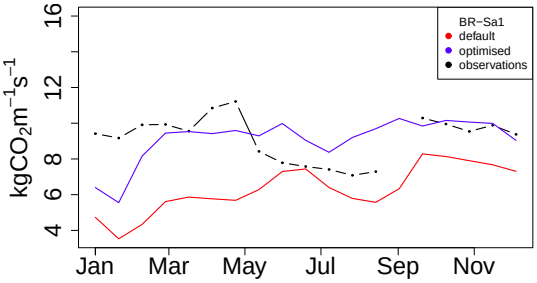

AU−Tum
default
optimised
observations
Wm−2
0
40
80
120
Jan
Mar
May
Jul
Sep
Nov

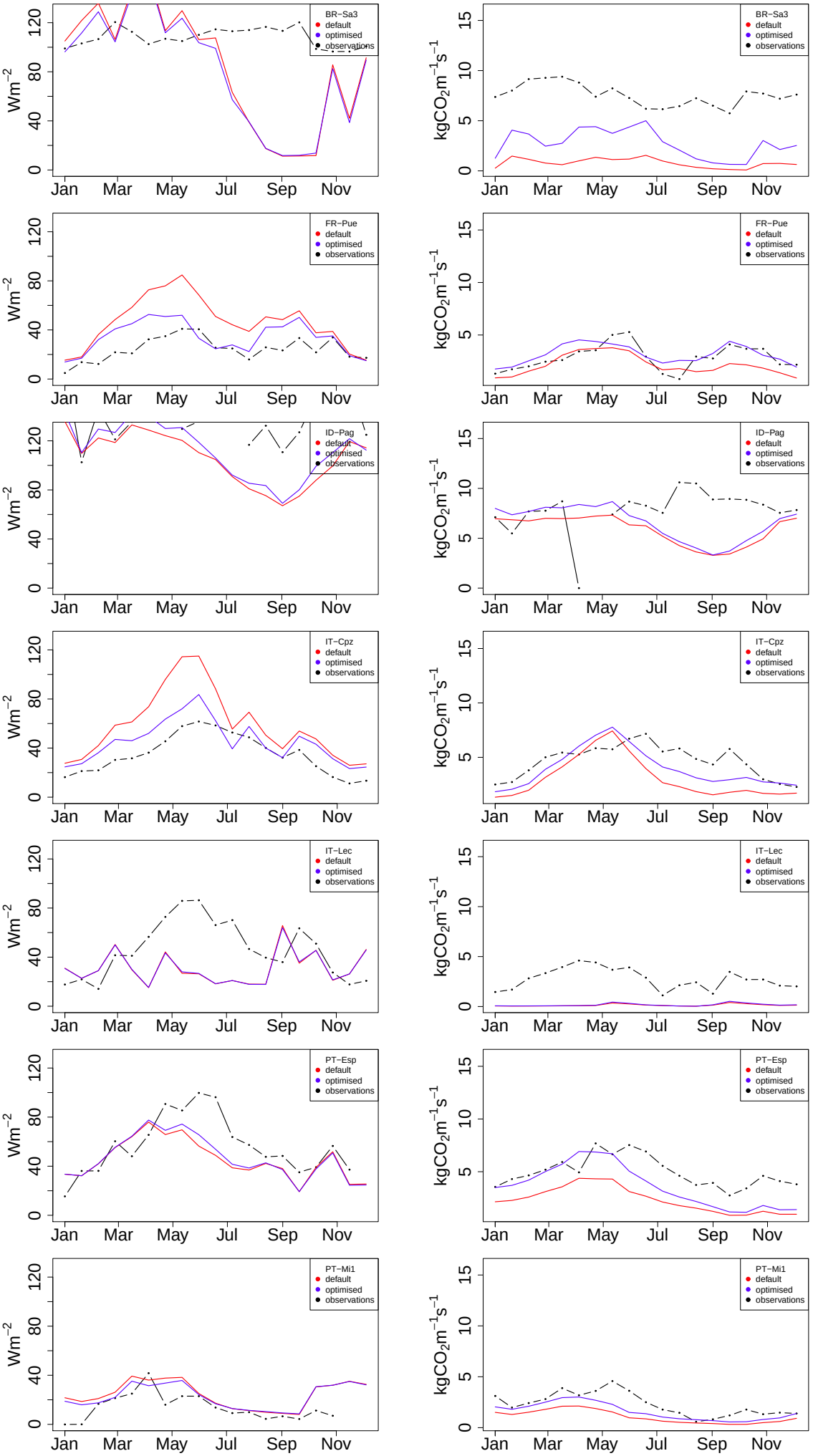

AU−Tum
default
optimised
observations
kgCO2m−1s−1
5
10
15
Jan
Mar
May
Jul
Sep
Nov


AU−Wac
default
optimised
observations
Wm−2
0
40
80
120
Jan
Mar
May
Jul
Sep
Nov

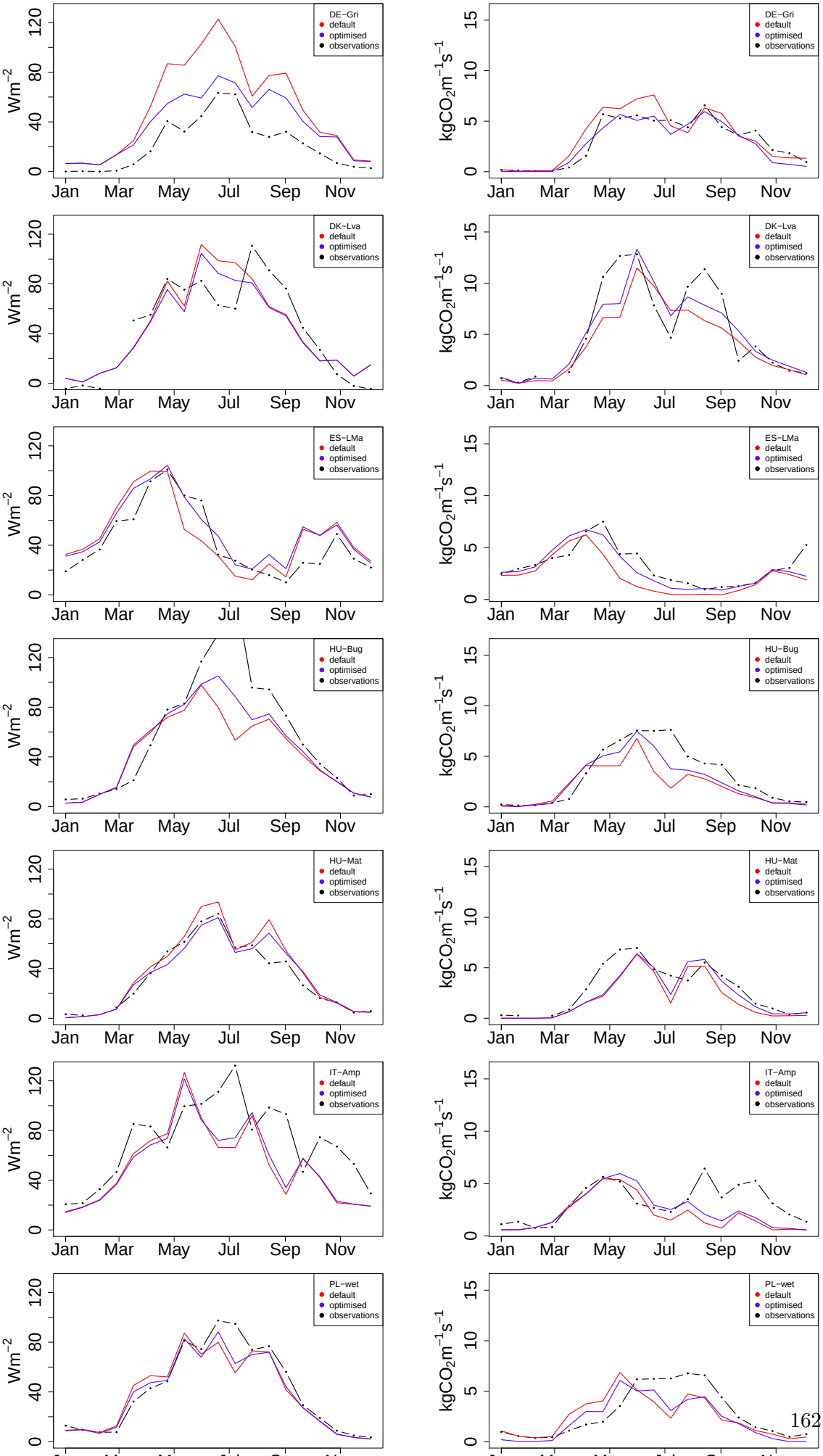

AU−Wac
default
optimised
observations
kgCO2m−1s−1
2
6
10
14
Jan
Mar
May
Jul
Sep
Nov

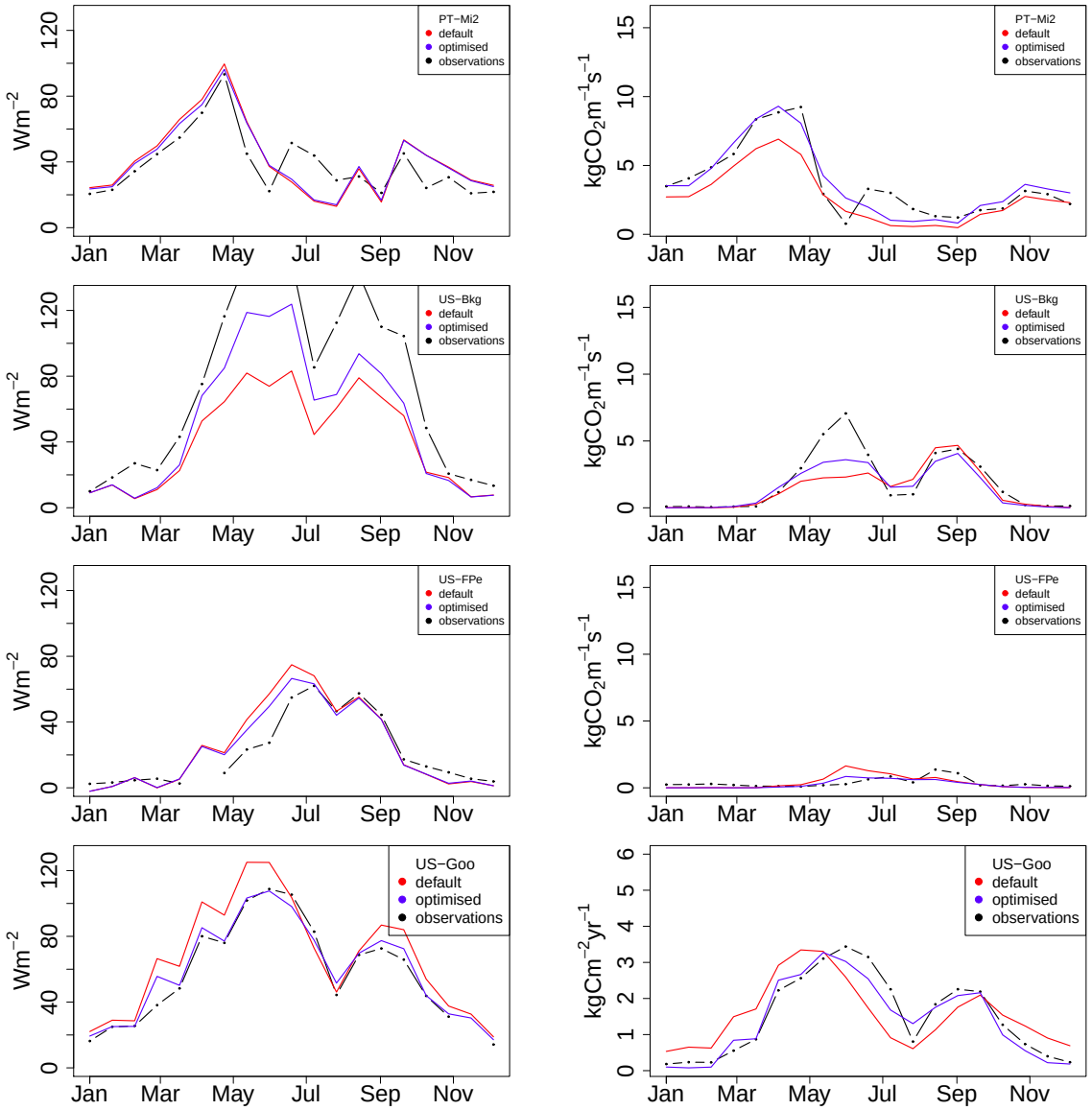

BR−Sa1
default
optimised
observations
Wm−2
0
40
80
120
Jan
Mar
May
Jul
Sep
Nov


BR−Sa1
default
optimised
observations
kgCO2m−1s−1
4
6
8
12
16
Jan
Mar
May
Jul
Sep
Nov

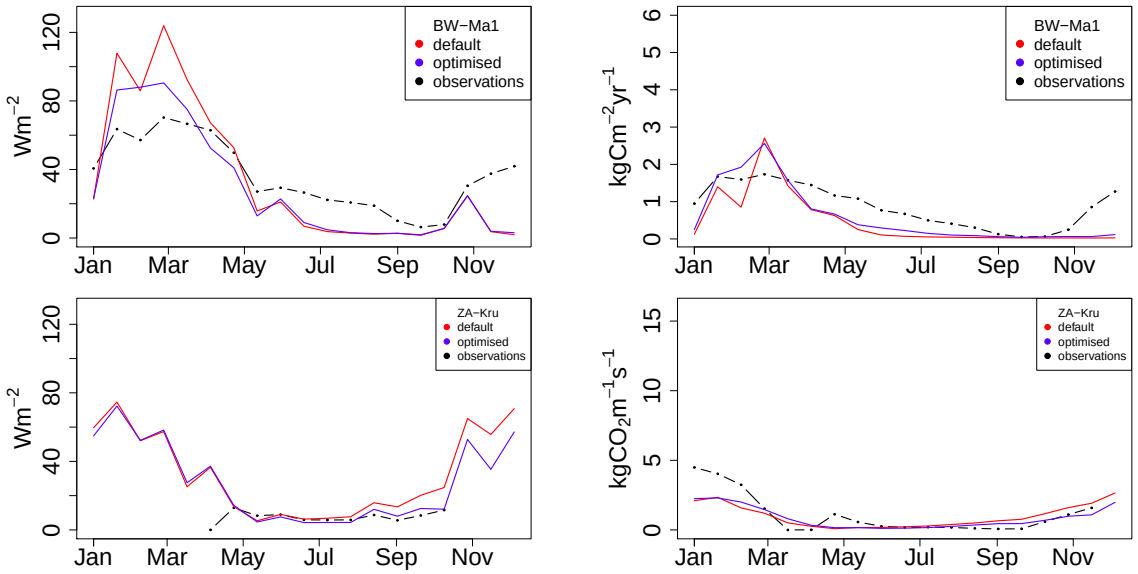

BR−Sa3
FR−Pue
ID−Pag
IT−Cpz
IT−Lec
PT−Esp
PT−Mi1
default
optimised
observations
Wm−2
kgCO2m−1s−1
0
40
80
120
5
10
15
Jan
Mar
May
Jul
Sep
Nov

## B.2. C3 grass sites

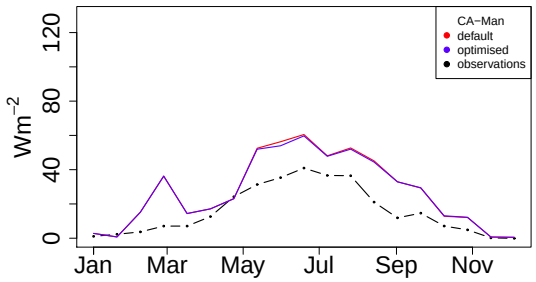


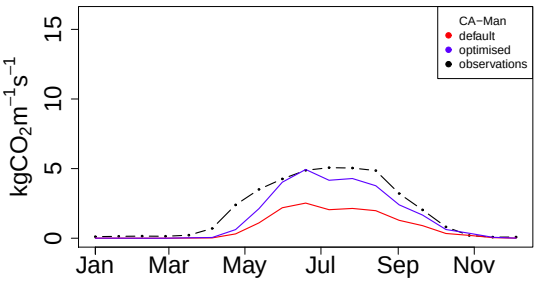


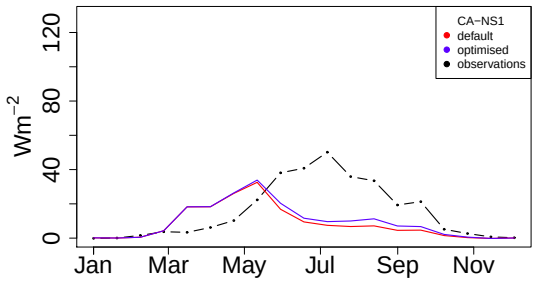

## B.3. C4 grass sites

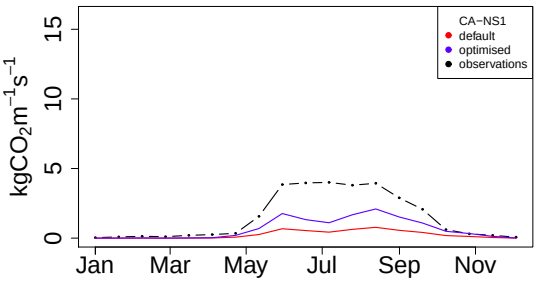


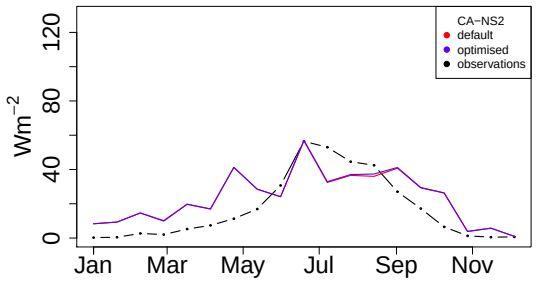
## B.4. Needleleaf sites

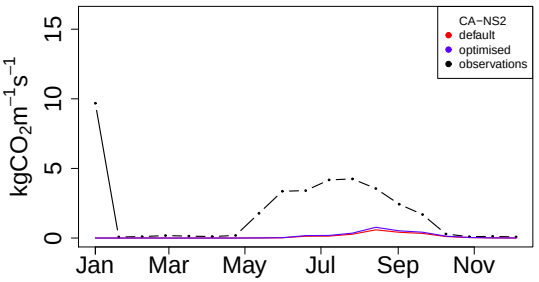


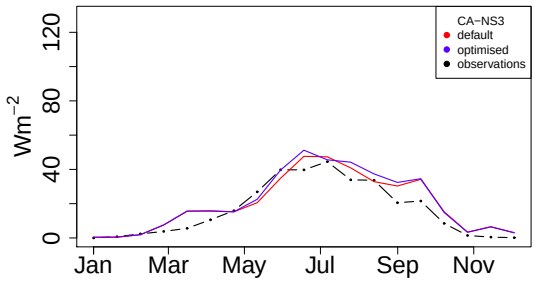



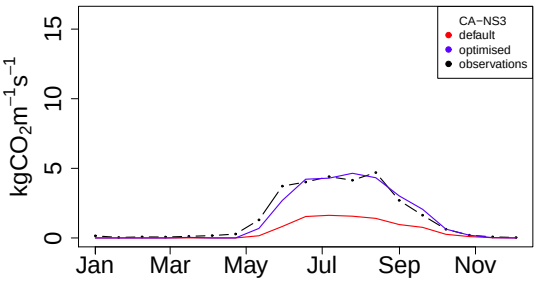



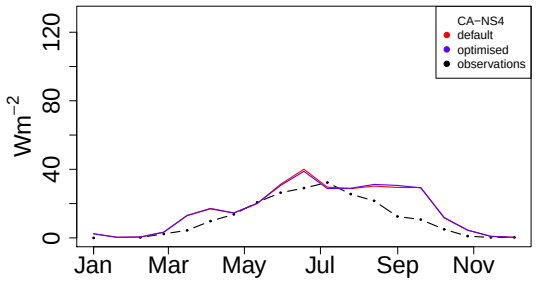



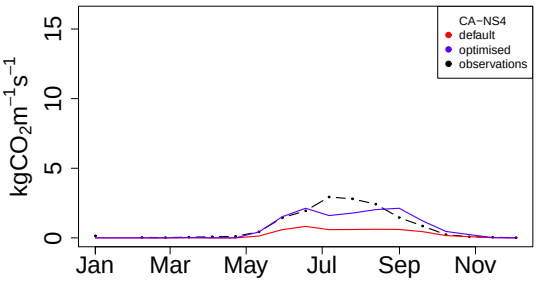



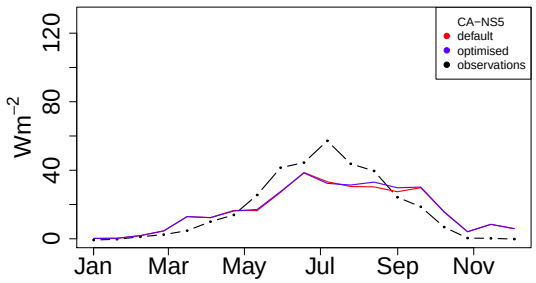



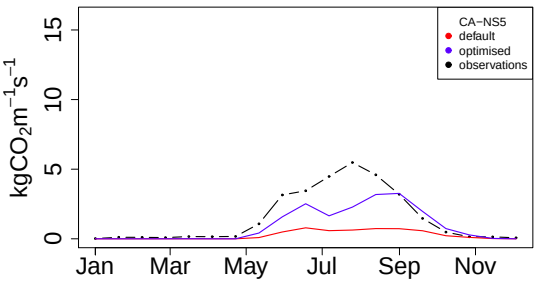



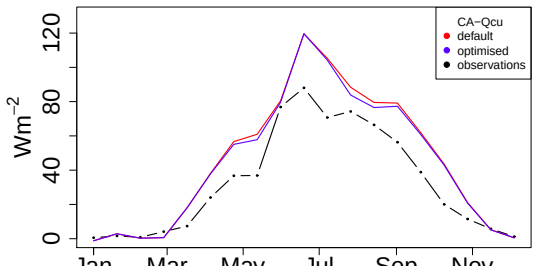



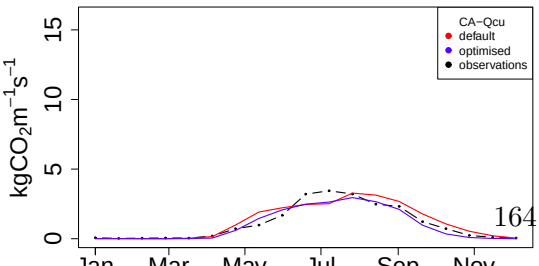



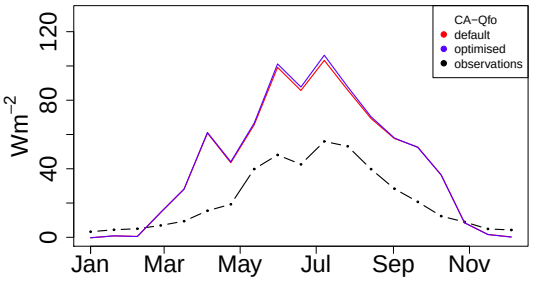



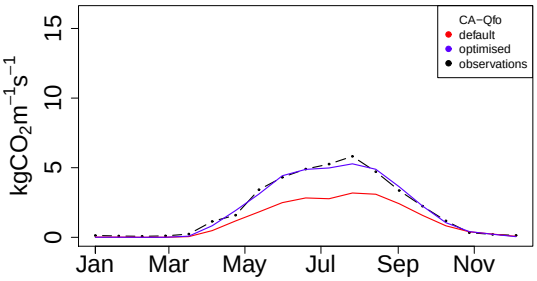



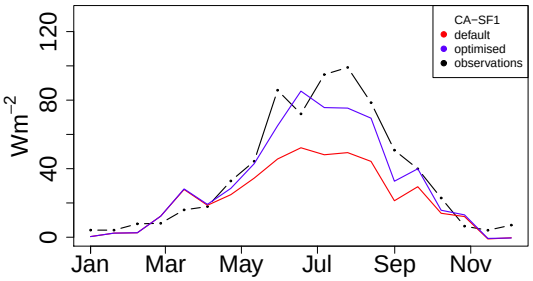



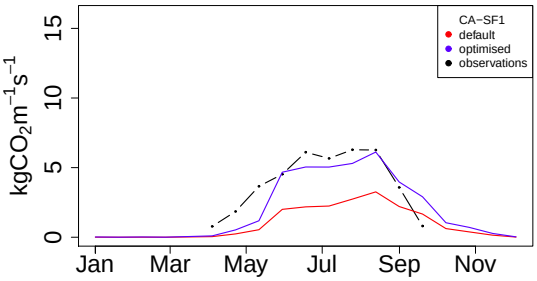



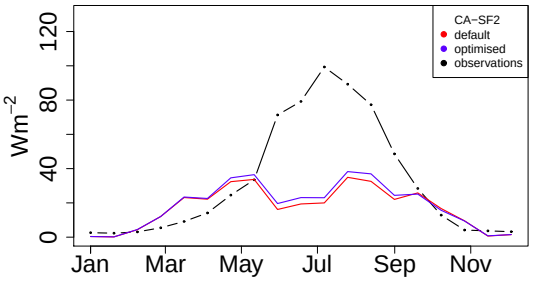



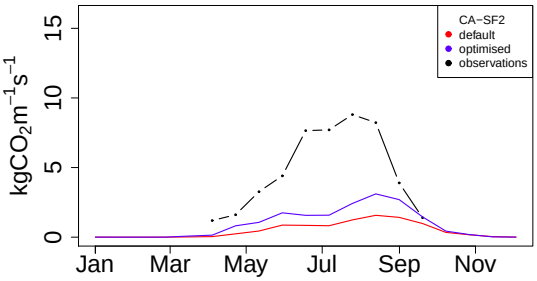

CA−Qfo
default
optimised
observations
Wm−2
0
40
80
120
Jan
Mar
May
Jul
Sep
Nov

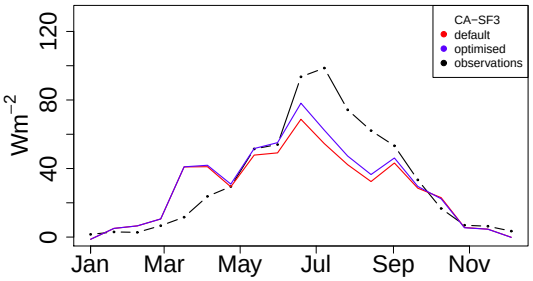

CA−Qfo
default
optimised
observations
kgCO2m−1s−1
0
5
10
15
Jan
Mar
May
Jul
Sep
Nov

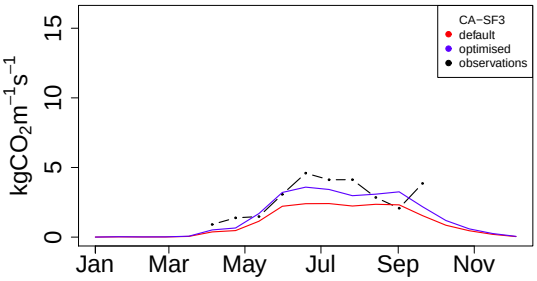

CA−SF1
default
optimised
observations
Wm−2
0
40
80
120
Jan
Mar
May
Jul
Sep
Nov

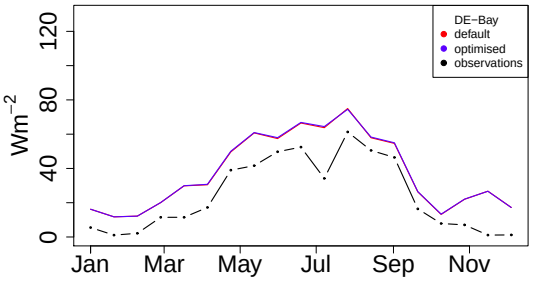

CA−SF1
default
optimised
observations
kgCO2m−1s−1
0
5
10
15
Jan
Mar
May
Jul
Sep
Nov

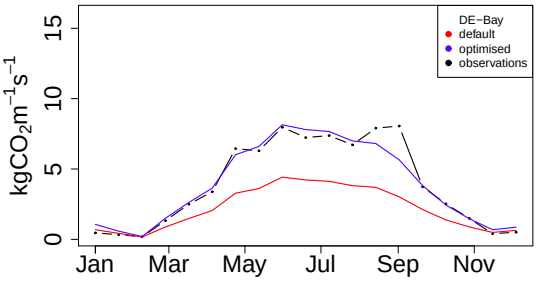

CA−SF2
default
optimised
observations
Wm−2
0
40
80
120
Jan
Mar
May
Jul
Sep
Nov

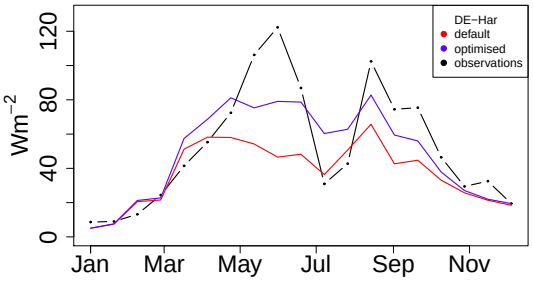

CA−SF2
default
optimised
observations
kgCO2m−1s−1
0
5
10
15
Jan
Mar
May
Jul
Sep
Nov

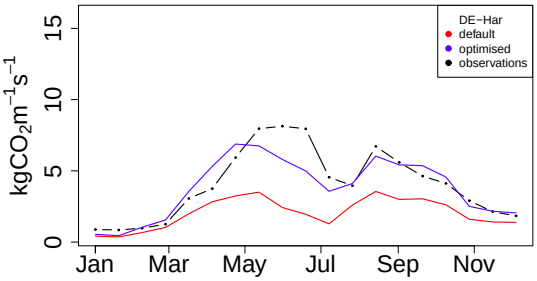

CA−SF3
default
optimised
observations
Wm−2
0
40
80
120
Jan
Mar
May
Jul
Sep
Nov

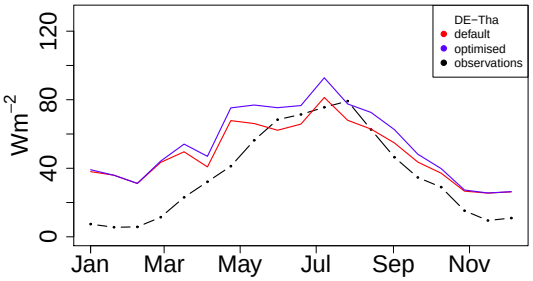

CA−SF3
default
optimised
observations
kgCO2m−1s−1
0
5
10
15
Jan
Mar
May
Jul
Sep
Nov

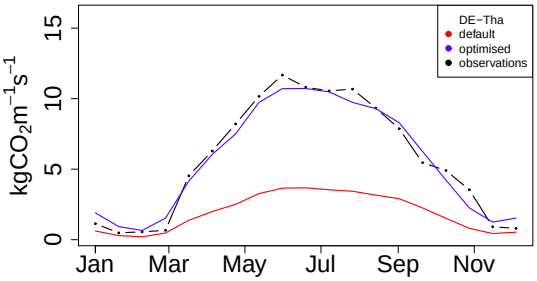

DE−Bay
default
optimised
observations
Wm−2
0
40
80
120
Jan
Mar
May
Jul
Sep
Nov

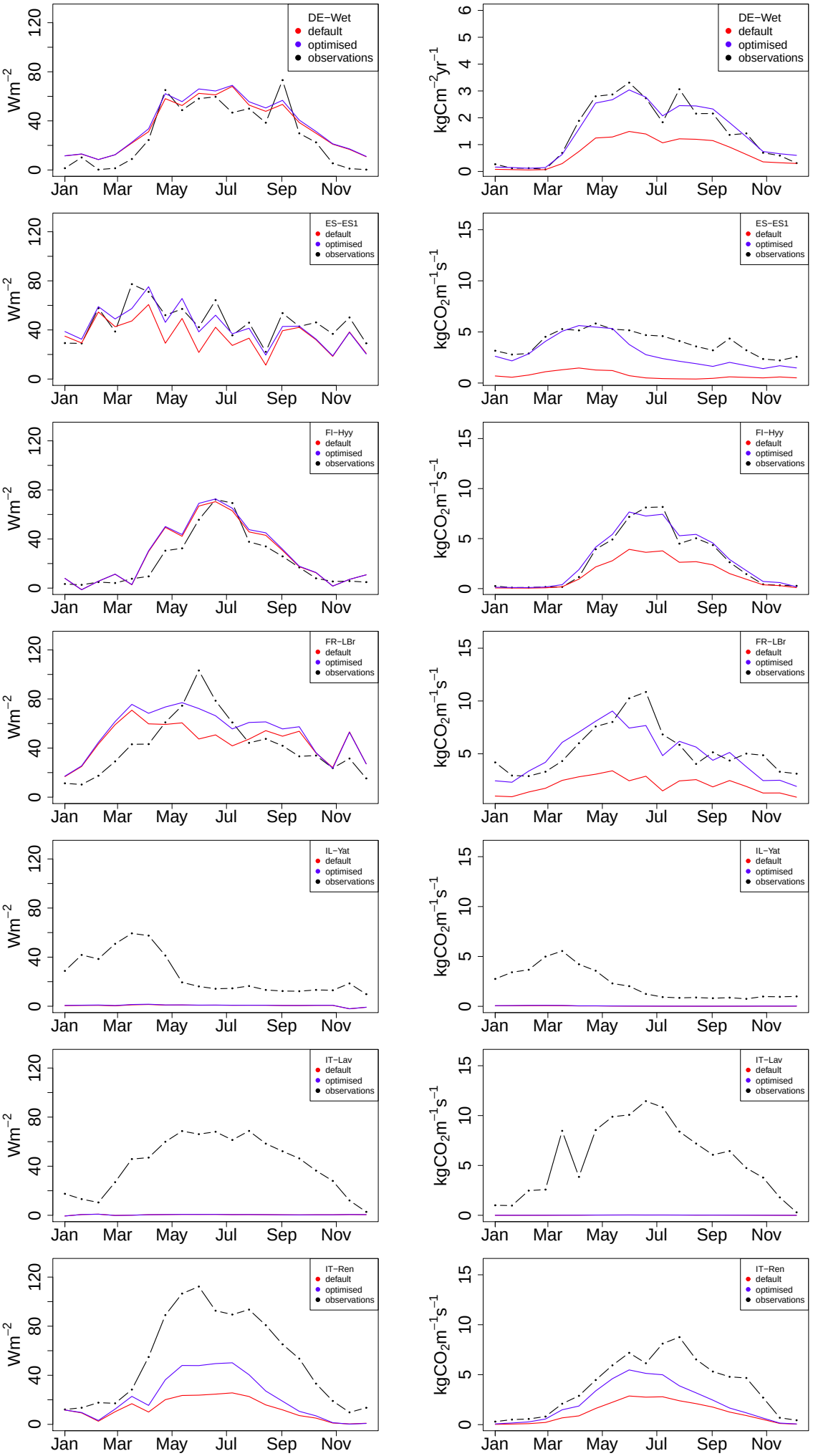

DE−Bay
default
optimised
observations
kgCO2m−1s−1
0
5
10
15
Jan
Mar
May
Jul
Sep
Nov

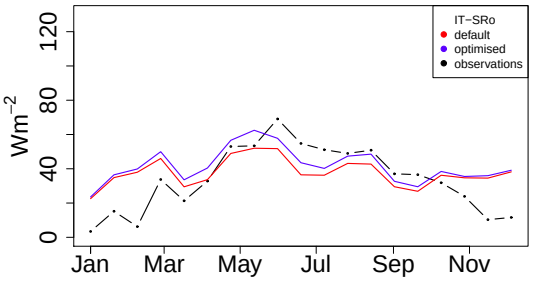

DE−Har
default
optimised
observations
Wm−2
0
40
80
120
Jan
Mar
May
Jul
Sep
Nov

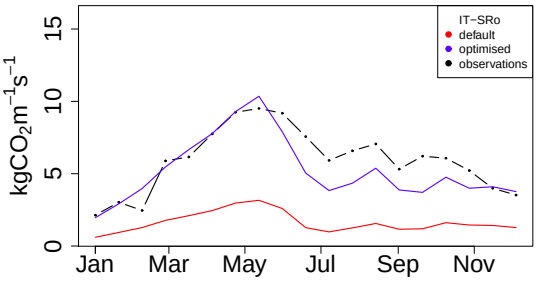

DE−Har
default
optimised
observations
kgCO2m−1s−1
0
5
10
15
Jan
Mar
May
Jul
Sep
Nov

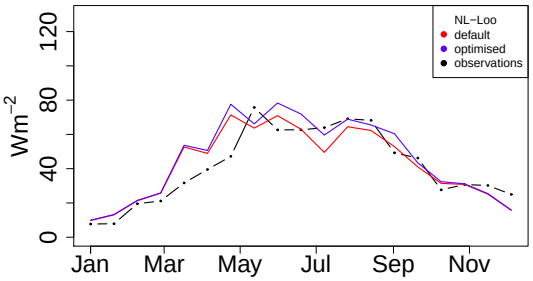

DE−Tha
default
optimised
observations
Wm−2
0
40
80
120
Jan
Mar
May
Jul
Sep
Nov

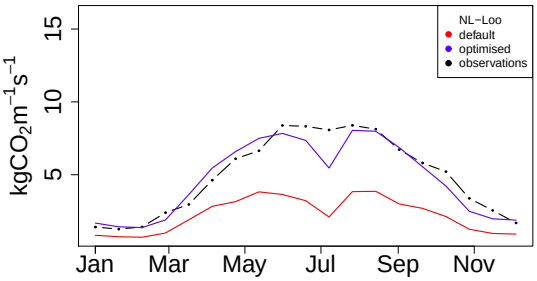

DE−Tha
default
optimised
observations
kgCO2m−1s−1
0
5
10
15
Jan
Mar
May
Jul
Sep
Nov

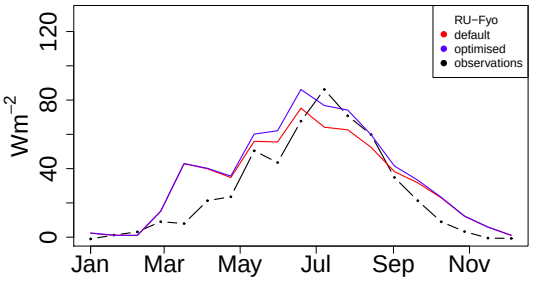

DE−Wet
default
optimised
observations
Wm$^{-2}$
kgCm$^{-2}$yr$^{-1}$
ES−ES1
kgCO$_2$m$^{-1}$s$^{-1}$
FI−Hyy
FR−LBr
IL−Yat
IT−Lav
IT−Ren
Jan
Mar
May
Jul
Sep
Nov

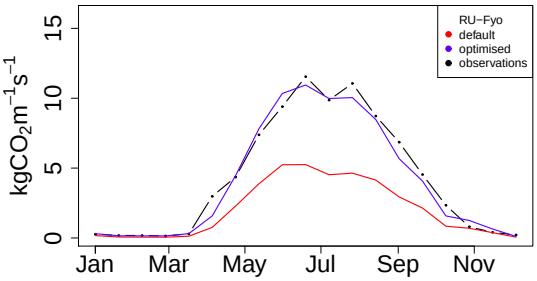

IT−SRo
default
optimised
observations
Wm−2
0
40
80
120
Jan
Mar
May
Jul
Sep
Nov

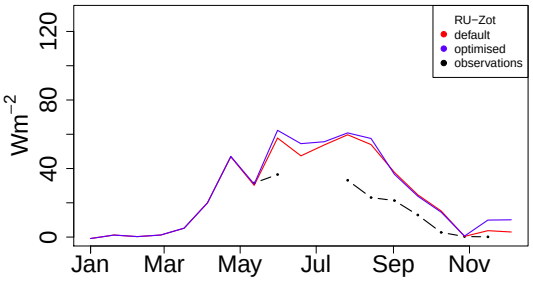

IT−SRo
default
optimised
observations
kgCO2m−1s−1
0
5
10
15
Jan
Mar
May
Jul
Sep
Nov

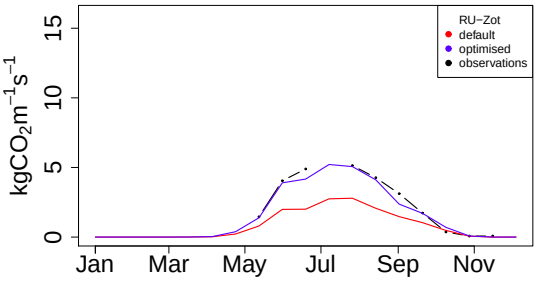

NL−Loo
default
optimised
observations
Wm−2
0
40
80
120
Jan
Mar
May
Jul
Sep
Nov

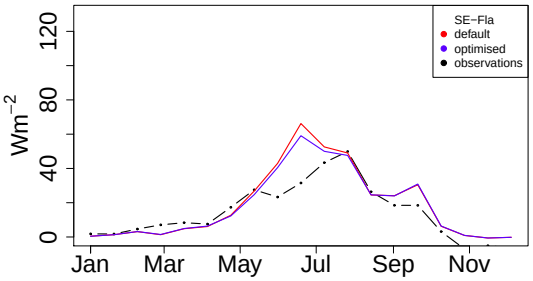

NL−Loo
default
optimised
observations
kgCO2m−1s−1
0
5
10
15
Jan
Mar
May
Jul
Sep
Nov

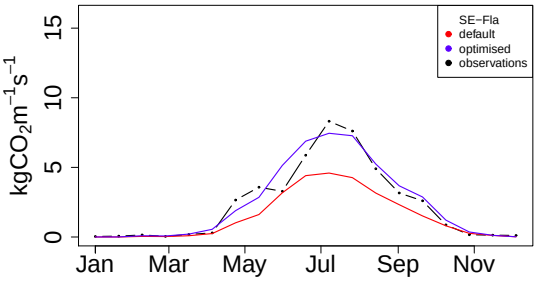

RU−Fyo
default
optimised
observations
Wm−2
0
40
80
120
Jan
Mar
May
Jul
Sep
Nov

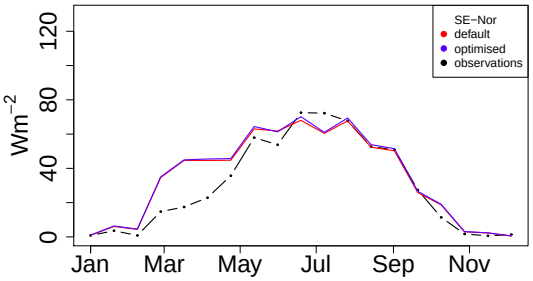

RU−Fyo
default
optimised
observations
kgCO2m−1s−1
0
5
10
15
Jan
Mar
May
Jul
Sep
Nov

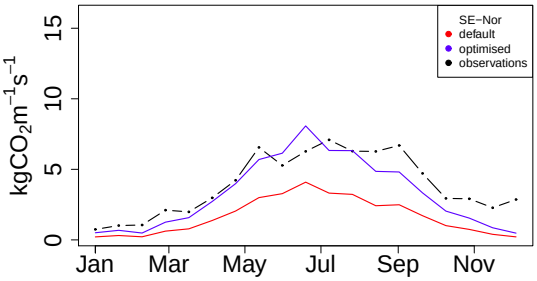

RU−Zot
default
optimised
observations
Wm−2
0
40
80
120
Jan
Mar
May
Jul
Sep
Nov

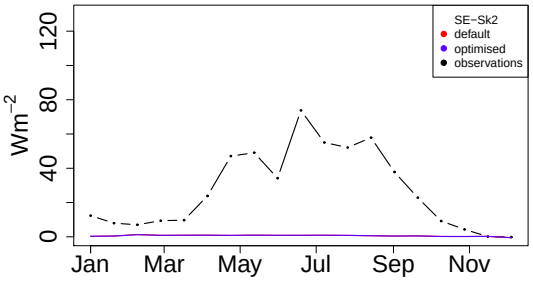

RU−Zot
default
optimised
observations
kgCO2m−1s−1
0
5
10
15
Jan
Mar
May
Jul
Sep
Nov

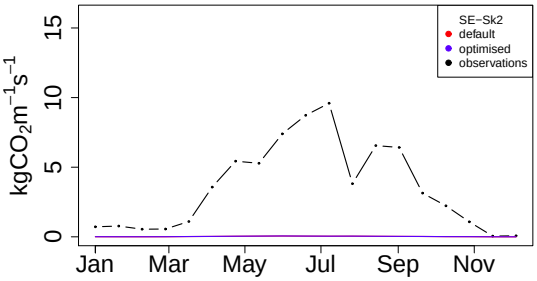

SE−Fla
default
optimised
observations
Wm−2
0
40
80
120
Jan
Mar
May
Jul
Sep
Nov

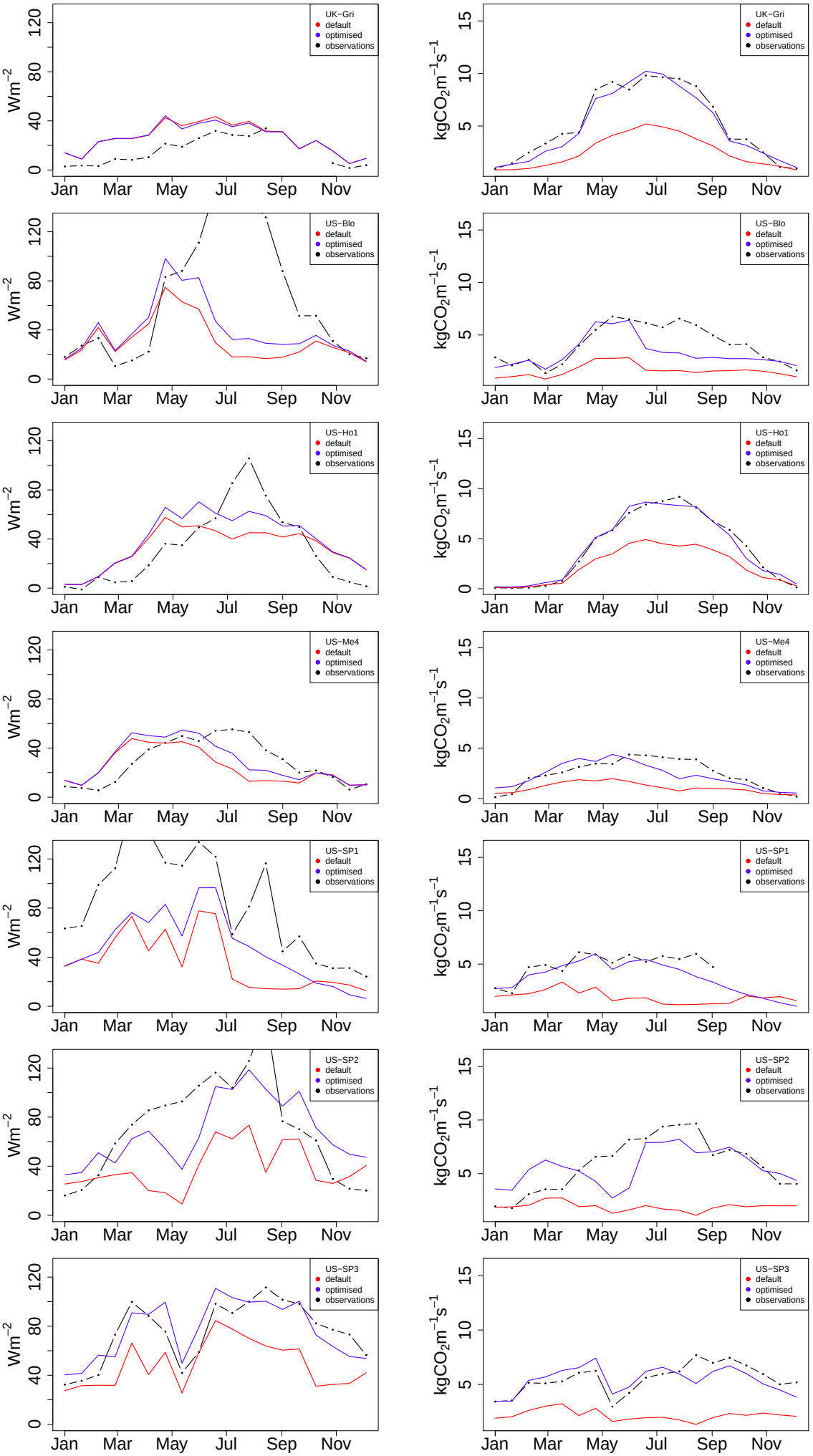

SE−Fla
default
optimised
observations
kgCO2m−1s−1
0
5
10
15
Jan
Mar
May
Jul
Sep
Nov


SE−Nor
default
optimised
observations
Wm−2
0
40
80
120
Jan
Mar
May
Jul
Sep
Nov

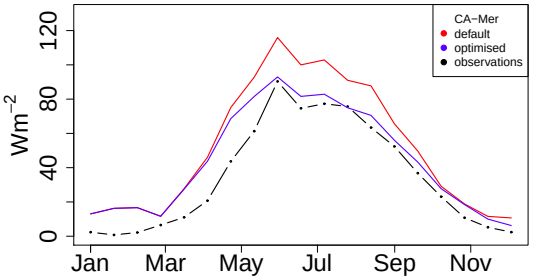

SE−Nor
default
optimised
observations
kgCO2m−1s−1
0
5
10
15
Jan
Mar
May
Jul
Sep
Nov

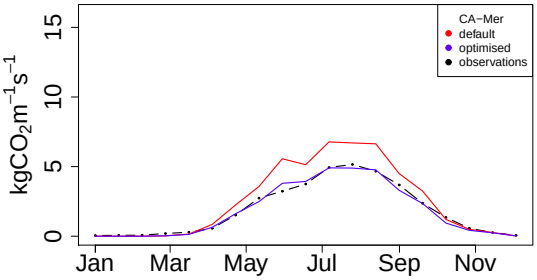

SE−Sk2
default
optimised
observations
Wm−2
0
40
80
120
Jan
Mar
May
Jul
Sep
Nov

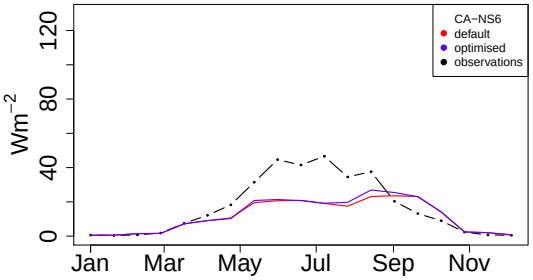

SE−Sk2
default
optimised
observations
kgCO2m−1s−1
0
5
10
15
Jan
Mar
May
Jul
Sep
Nov

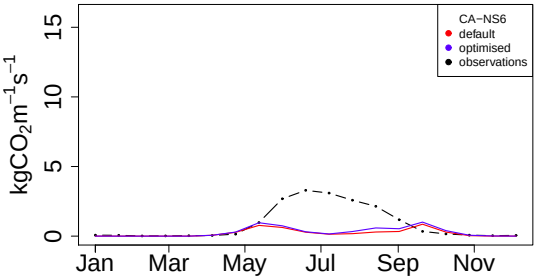

UK−Gri
US−Blo
US−Ho1
US−Me4
US−SP1
US−SP2
US−SP3
default
optimised
observations
Wm−2
kgCO2m−1s−1
0
40
80
120
5
10
15
Jan
Mar
May
Jul
Sep
Nov

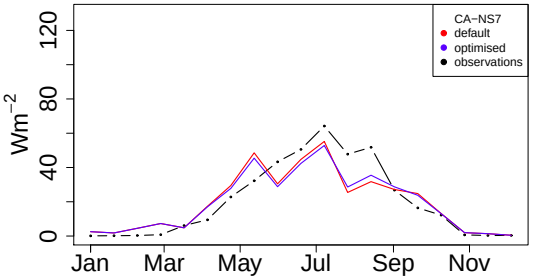
## B.5. Shrub sites

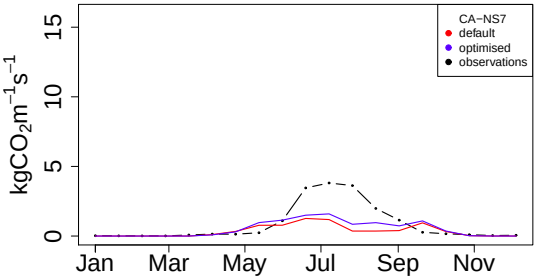


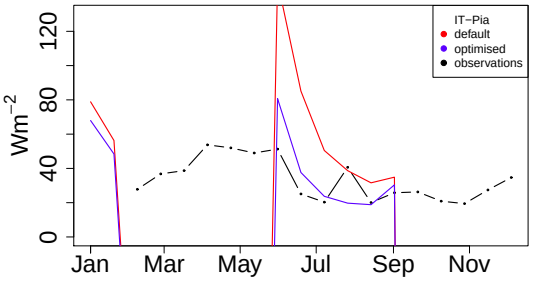



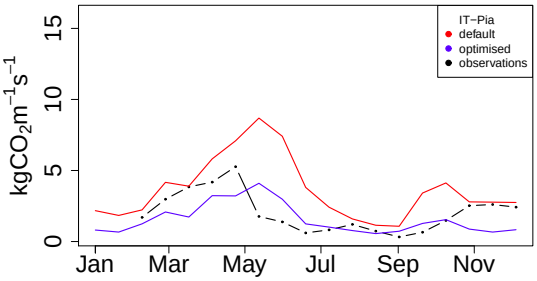



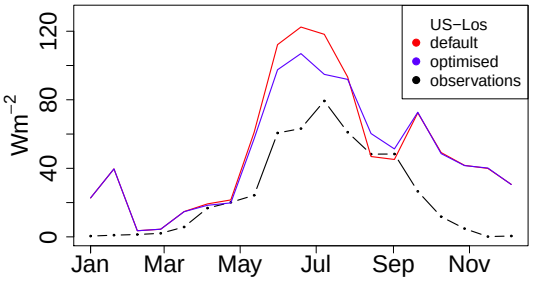



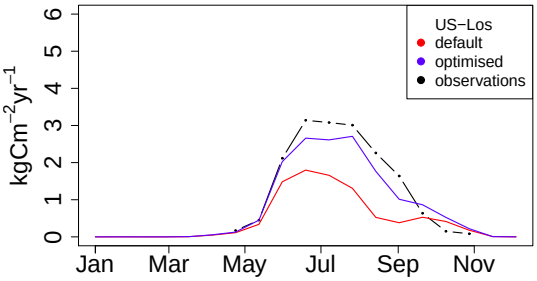



# Bibliography


Ainsworth, E. A. and Long, S. P. (2005). What have we learned from 15 years of free-air $CO_2$ enrichment (face)? a meta-analytic review of the responses of photosynthesis, canopy properties and plant production to rising co2. *New Phytologist*, 165(2):351–372.

Ajami, N. K., Duan, Q., and Sorooshian, S. (2007). An integrated hydrologic Bayesian multimodel combination framework: Confronting input, parameter, and model structural uncertainty in hydrologic prediction. *Water Resources Research*, 43(1).

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.

Bannister, R. (2004). Using Prior Knowledge in Data Assimilation. Oxford/RAL Spring School: http://www.met.reading.ac.uk/~ross/Documents/OxRAL04.html.

Bannister, R. N. (2008). A review of forecast error covariance statistics in atmospheric variational data assimilation. II: Modelling the forecast error covariance statistics. *Quarterly Journal of the Royal Meteorological Society*, 134(637):1971–1996.

Bartholomew-Biggs, M., Brown, S., Christianson, B., and Dixon, L. (2000). Automatic differentiation of algorithms. *Journal of Computational and Applied Mathematics*, 124(1):171–190.

Berger, A., Mesinger, F., and Sijacki, D. (2012). *Climate change: inferences from paleoclimate and regional aspects*. Springer Science & Business Media.

Berland, H. (2006). Automatic differentiation. Presentation, http://www.robots.ox.ac.uk/~tvg/publications/talks/autodiff.pdf.

Berry, J. A., Beerling, D. J., and Franks, P. J. (2010). Stomata: key players in the earth system, past and present. *Current Opinion in Plant Biology*, 13(3):232–239.

Best, M. (2005). JULES Technical Documentation. *Hadley Centre*.

Best, M. J., Pryor, M., Clark, D. B., Rooney, G. G., Essery, R. L. H., Ménard, C. B., Edwards, J. M., Hendry, M. A., Porson, A., Gedney, N., Mercado, L. M., Sitch, S., Blyth, E., Boucher, O., Cox, P. M., Grimmond, C. S. B., and Harding, R. J. (2011). The Joint

UK Land Environment Simulator (JULES), Model description - Part 1: Energy and water fluxes. *Geoscientific Model Development Discussions*, 4(1):595–640.

Blyth, E., Clark, D., Ellis, R., Huntingford, C., Los, S., Pryor, M., Best, M., and Sitch, S. (2011). A comprehensive set of benchmark tests for a land surface model of simultaneous fluxes of water and carbon at both the global and seasonal scale. *Geoscientific Model Development*, 4(2):255–269.

Blyth, E., Gash, J., Lloyd, A., Pryor, M., Weedon, G. P., and Shuttleworth, J. (2010). Evaluating the JULES land surface model energy fluxes using FLUXNET data. *Journal of Hydrometeorology*, 11(2):509–519.

Booth, B. B. B., Dunstone, N. J., Halloran, P. R., Andrews, T., and Bellouin, N. (2012). Aerosols implicated as a prime driver of twentieth-century North Atlantic climate variability. *Nature*, 484(7393):228–232.

Bouttier, F. and Courtier, P. (2002). Data assimilation concepts and methods March 1999. *Meteorological training course lecture series. ECMWF*.

Braswell, B. H., Sacks, W. J., Linder, E., and Schimel, D. S. (2005). Estimating diurnal to annual ecosystem parameters by synthesis of a carbon flux model with eddy covariance net ecosystem exchange observations. *Global Change Biology*, 11(2):335–355.

Brovkin, V., Boysen, L., Raddatz, T., Gayler, V., Loew, A., and Claussen, M. (2013). Evaluation of vegetation cover and land-surface albedo in MPI-ESM CMIP5 simulations. *Journal of Advances in Modeling Earth Systems*, 5(1):48–57.

Burba, G. and Anderson, D. (2010). *A Brief Practical Guide to Eddy Covariance Flux Measurements*. LICOR Biosciences, Lincoln, USA.

Byrd, R., Lu, P., Nocedal, J., and Zhu, C. (1995). A limited memory algorithm for bound constrained optimization. *SIAM Journal on Scientific Computing*, 16(5):1190–1208.

Cheney, E. and Kincaid, D. (2012). *Numerical mathematics and computing*. Nelson Education.

Chevallier, F. (2016). Personal communication.

Clark, D. B., Mercado, L. M., Sitch, S., Jones, C. D., Gedney, N., Best, M. J., Pryor, M., Rooney, G. G., Essery, R. L. H., Blyth, E., Boucher, O., Harding, R. J., and Cox, P. M. (2011). The Joint UK Land Environment Simulator (JULES), Model description - Part 2: Carbon fluxes and vegetation. *Geoscientific Model Development Discussions*, 4(1):641–688.

Clarke, L., Edmonds, J., Jacoby, H., Pitcher, H., Reilly, J., and Richels, R. (2007). Scenarios of greenhouse gas emissions and atmospheric concentrations. *US Department of Energy Publications*, page 6.

Collatz, G., Ball, J., Grivet, C., and Berry, J. (1991). Physiological and environmental

regulation of stomatal conductance, photosynthesis and transpiration: a model that includes a laminar boundary layer. *Agricultural and Forest Meteorology*, 54(2-4):107–136.

Collatz, G. J., Ribas-Carbo, M., and Berry, J. (1992). Coupled photosynthesis-stomatal conductance model for leaves of C4 plants. *Functional Plant Biology*, 19(5):519–538.

Cowan, I. (1972). Oscillations in stomatal conductance and plant functioning associated with stomatal conductance: observations and a model. *Planta*, 106(3):185–219.

Cox, P. (2001). Description of the TRIFFID dynamic global vegetation model. *Hadley Centre Technical Note*, 24.

Cox, P., Betts, R., Jones, C., Spall, S., and Totterdell, I. (2000). Acceleration of global warming due to carbon-cycle feedbacks in a coupled climate model. *Nature*, 408(6809):184–187.

Cox, P., Betts, R. A., Bunton, C. B., Essery, R. L. H., Rowntree, P. R., and Smith, J. (1999). The impact of new land surface physics on the GCM simulation of climate and climate sensitivity. *Climate Dynamics*, 15(3):183–203.

Cox, P., Harris, P., Huntingford, C., Betts, R., Collins, M., Jones, C., Jupp, T., Marengo, J., and Nobre, C. (2008). Increasing risk of Amazonian drought due to decreasing aerosol pollution. *Nature*, 453(7192):212–215.

Cox, P., Huntingford, C., and Harding, R. (1998). A canopy conductance and photosynthesis model for use in a GCM land surface scheme. *Journal of Hydrology*, 212:79–94.

Cox, P., Huntingford, C., and Jones, C. D. (2006). Conditions for sink-to-source transitions and runaway feedbacks from the land carbon cycle. Cambridge University Press.

de Laplace, P. S. (1820). Théorie analytique des probabilités - 3ème édition.

Dehuri, S., Jagadev, A. K., and Panda, M. (2015). *Multi-objective Swarm Intelligence: Theoretical Advances and Applications*, volume 592. Springer.

Dekker, S. C., Booth, B. B., and Cox, P. M. (2016). Spatial and temporal variations in plant water-use efficiency inferred from tree-ring, eddy covariance and atmospheric observations. *Earth System Dynamics*, 7(2):525.

Denman, K. L., Brasseur, G., Chidthaisong, A., Ciais, P., Cox, P. M., Dickinson, R. E., Hauglustaine, D., Heinze, C., Holland, E., Jacob, D., et al. (2007). *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. Couplings Between Changes in the Climate System and Biogeochemistry. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Deser, C., Phillips, A., Bourdette, V., and Teng, H. (2012). Uncertainty in climate change projections: the role of internal variability. *Climate Dynamics*, 38(3-4):527–546.

Errico, R. (1997). What is an adjoint model? *Bulletin of the American Meteorological Society*, 78(11):2577–2591.

Falge, E., Baldocchi, D., Olson, R., Anthoni, P., Aubinet, M., Bernhofer, C., Burba, G., Ceulemans, R., Clement, R., Dolman, H., et al. (2001). Gap filling strategies for defensible annual sums of net ecosystem exchange. *Agricultural and Forest Meteorology*, 107(1):43–69.

FastOpt (2010). *Transformation of Algorithms in Fortran.* Manual, http://www.FastOpt.com. Hamburg, Germany.

Field, C. B., Jackson, R. B., and Mooney, H. A. (1995). Stomatal responses to increased co2: implications from the plant to the global scale. *Plant, Cell & Environment*, 18(10):1214–1225.

Fischer, G. R., Costa, M. H., Murta, F. Z., Malhado, A. C., Aguiar, L. J., and Ladle, R. J. (2013). Multi-site land surface model optimization: An exploration of objective functions. *Agricultural and Forest Meteorology*, 182–183:168 – 176.

Fraley, C. and Raftery, A. E. (2002). Model-based clustering, discriminant analysis, and density estimation. *Journal of the American Statistical Association*, 97(458):611–631.

Fraley, C., Raftery, A. E., Murphy, T. B., and Scrucca, L. (2012). mclust Version 4 for R: Normal Mixture Modeling for Model-Based Clustering, Classification, and Density Estimation. Technical report, No. 597, Department of Statistics, University of Washington.

Friedlingstein, P., Bopp, L., Ciais, P., Dufresne, J.-L., Fairhead, L., LeTreut, H., Monfray, P., and Orr, J. (2001). Positive feedback between future climate change and the carbon cycle. *Geophysical Research Letters*, 28(8):1543–1546.

Friedlingstein, P., Bopp, L., Rayner, P., Cox, P. M., Betts, R., Jones, C., Von Bloh, W., Brovkin, V., Cadule, P., Doney, S., et al. (2006). Climate–carbon cycle feedback analysis: results from the C4MIP model intercomparison. *Journal of Climate*, 19:3337–3353.

Friedlingstein, P., Meinshausen, M., Arora, V. K., Jones, C. D., Anav, A., Liddicoat, S. K., and Knutti, R. (2013). Uncertainties in cmip5 climate projections due to carbon cycle feedbacks. *Journal of Climate*, 27(2):511–526.

Fujino, J., Nair, R., Kainuma, M., Masui, T., and Matsuoka, Y. (2006). Multi-gas mitigation analysis on stabilization scenarios using AIM global model. *The Energy Journal*, pages 343–353.

Gedney, N., Cox, P., Betts, R., Boucher, O., Huntingford, C., and Stott, P. (2006). Detection of a direct carbon dioxide effect in continental river runoff records. *Nature*, 439(7078):835–838.

Geman, S. and Geman, D. (1984). Pattern Analysis and Machine Intelligence, IEEE Trans-

actions on. *Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images*.

Genz, A., Bretz, F., Miwa, T., Mi, X., Leisch, F., Scheipl, F., and Hothorn, T. (2015). *mvtnorm: Multivariate Normal and t Distributions*. R package version 1.0-5. https://CRAN.R-project.org/package=mvtnorm.

Ghent, D., Kaduk, J., Remedios, J., Ardö, J., and H.Balzter (2010). Assimilation of land surface temperature into the land surface model JULES with an ensemble Kalman filter. *Journal of Geophysical Research*, 115.

Giering, R. and Kaminski, T. (2003). Applying TAF to generate efficient derivative code of Fortran 77-95 programs. *Proceedings in Applied Mathematics and Mechanics*, 2(1):54–57.

Giering, R., Kaminski, T., and Slawig, T. (2005). Generating efficient derivative code with TAF: Adjoint and tangent linear Euler flow around an airfoil. *Future Generation Computer Systems*, 21(8):1345 – 1355.

Göckede, M., Markkanen, T., Hasager, C. B., and Foken, T. (2006). Update of a footprint-based approach for the characterisation of complex measurement sites. *Boundary-Layer Meteorology*, 118(3):635–655.

Goulden, M. L., Munger, J. W., Fan, S.-M., Daube, B. C., and Wofsy, S. C. (1996). Measurements of carbon sequestration by long-term eddy covariance: methods and a critical evaluation of accuracy. *Global Change Biology*, 2(3):169–182.

Gower, J. C. (1966). Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika*, 53(3-4):325.

Groenendijk, M., Dolman, A., Van Der Molen, M., Leuning, R., Arneth, A., Delpierre, N., Gash, J., Lindroth, A., Richardson, A., Verbeeck, H., and Wohlfahrt, G. (2011). Assessing parameter variability in a photosynthesis model within and between plant functional types using global fluxnet eddy covariance data. *Agricultural and forest meteorology*, 151(1):22–38.

Gupta, H. V., Bastidas, L. A., Sorooshian, S., Shuttleworth, W. J., and Yang, Z. L. (1999). Parameter estimation of a land surface scheme using multicriteria methods. *Journal of Geophysical Research: Atmospheres*, 104(D16):19491–19503.

Gupta, H. V., Sorooshian, S., and Yapo, P. O. (1998). Toward improved calibration of hydrologic models: Multiple and noncommensurable measures of information. *Water Resources Research*, 34(4):751–763.

Haben, S. A. (2011). *Conditioning and Preconditioning of the Minimisation Problem in Variational Data Assimilation*. PhD thesis, University of Reading.

Hadamard, J. (1902). Sur les problèmes aux dérivés partielles et leur signification physique. *Princeton University Bulletin*, 13:49–52.

Hararuk, O., Xia, J., and Luo, Y. (2014). Evaluation and improvement of a global land model against soil carbon data using a Bayesian Markov chain Monte Carlo method. *Journal of Geophysical Research: Biogeosciences*, 119(3):403–417. 2013JG002535.

Harper, A. B., Cox, P. M., Friedlingstein, P., Wiltshire, A. J., Jones, C. D., Sitch, S., Mercado, L. M., Groenendijk, M., Robertson, E., Kattge, J., Bönisch, G., Atkin, O. K., Bahn, M., Cornelissen, J., Niinemets, Ü., Onipchenko, V., Peñuelas, J., Poorter, L., Reich, P. B., Soudzilovskaia, N. A., and Bodegom, P. V. (2016). Improved representation of plant functional types and physiology in the Joint UK Land Environment Simulator (JULES v4.2) using plant trait information. *Geoscientific Model Development*, 9:2415–2440.

Hartigan, J. and Wong, M. (1979). Algorithm AS 136: A K-means clustering algorithm. *Applied Statistics*, pages 100–108.

Hijioka, Y., Matsuoka, Y., Nishimoto, H., Masui, M., and Kainuma, M. (2008). Global GHG emissions scenarios under GHG concentration stabilization targets. *Journal of Global Environmental Engineering*, 13:97–108.

Hollinger, D. and Richardson, A. (2005). Uncertainty in eddy covariance measurements and its application to physiological models. *Tree Physiology*, 25(7):873–885.

Homescu, C. (2011). Adjoints and automatic (algorithmic) differentiation in computational finance.

Huntingford, C., Booth, B., Sitch, S., Gedney, N., Lowe, J., Liddicoat, S., Mercado, L., Best, M., Weedon, G., Fisher, R., et al. (2010). IMOGEN: an intermediate complexity model to evaluate terrestrial impacts of a changing climate. *Geoscientific Model Development*, 3(2):679.

Ito, A. and Inatomi, M. (2012). Water-use efficiency of the terrestrial biosphere: a model analysis focusing on interactions between the global carbon and water cycles. *Journal of Hydrometeorology*, 13(2):681–694.

Izui, K., Yamada, T., Nishiwaki, S., and Tanaka, K. (2015). Multiobjective optimization using an aggregative gradient-based method. *Structural and Multidisciplinary Optimization*, 51(1):173–182.

Jarvis, P. (1995). Scaling processes and problems. *Plant, Cell & Environment*, 18(10):1079–1089.

Jarvis, P. and McNaughton, K. (1986). Stomatal control of transpiration: scaling up from leaf to region. *Advances in Ecological Research*, 15:1–49.

Jogireddy, V., Cox, P., Huntingford, C., Harding, R., and Mercado, L. (2006). An improved description of canopy light interception for use in a GCM land-surface scheme: calibration and testing against carbon fluxes at a coniferous forest. Technical Report 63, Hadley Centre, Met Office, Exeter, UK.

Jones, G. S., Stott, P. A., and Christidis, N. (2013). Attribution of observed historical near-surface temperature variations to anthropogenic and natural causes using CMIP5 simulations. *Journal of Geophysical Research: Atmospheres*, 118(10):4001–4024.

Kaminski, T., Knorr, W., Rayner, P., and Heimann, M. (2002). Assimilating atmospheric data into a terrestrial biosphere model: A case study of the seasonal cycle. *Global Biogeochemical Cycles*, 16(4):1066.

Kaminski, T., Knorr, W., Schürmann, G., Scholze, M., Rayner, P. J., Zaehle, S., Blessing, S., Dorigo, W., Gayler, V., Giering, R., Gobron, N., Grant, J. P., Heimann, M., Hooker-Stroud, A., Houweling, S., Kato, T., Kattge, J., Kelley, D., Kemp, S., Koffi, E. N., Köstler, C., Mathieu, P.-P., Pinty, B., Reick, C. H., Rödenbeck, C., Schnur, R., Scipal, K., Sebald, C., Stacke, T., van Scheltinga, A. T., Vossbeck, M., Widmann, H., and Ziehn, T. (2013). The BETHY/JSBACH Carbon Cycle Data Assimilation System: experiences and challenges. *Journal of Geophysical Research: Biogeosciences*, 118(4):1414–1426.

Kattge, J., Diaz, S., Lavorel, S., Prentice, I., Leadley, P., Bönisch, G., Garnier, E., Westoby, M., Reich, P. B., Wright, I., et al. (2011). TRY–a global database of plant traits. *Global Change Biology*, 17(9):2905–2935.

Katul, G., Manzoni, S., Palmroth, S., and Oren, R. (2010). A stomatal optimization theory to describe the effects of atmospheric $CO_2$ on leaf photosynthesis and transpiration. *Annals of Botany*, 105:431–442.

Kavetski, D., Kuczera, G., and Franks, S. W. (2006a). Bayesian analysis of input uncertainty in hydrological modeling: 1. Theory. *Water Resources Research*, 42(3).

Kavetski, D., Kuczera, G., and Franks, S. W. (2006b). Bayesian analysis of input uncertainty in hydrological modeling: 2. Application. *Water Resources Research*, 42(3).

Keenan, T. F., Hollinger, D. Y., Bohrer, G., Dragoni, D., Munger, J. W., Schmid, H. P., and Richardson, A. D. (2013). Increase in forest water-use efficiency as atmospheric carbon dioxide concentrations rise. *Nature*, 499(7458):324–327.

Knorr, W. and Kattge, J. (2005). Inversion of terrestrial ecosystem model parameter values against eddy covariance measurements by Monte Carlo sampling. *Global Change Biology*, 11(8):1333–1351.

Kriegel, H.-P., Schubert, E., and Zimek, A. (2016). The (black) art of runtime evaluation: Are we comparing algorithms or implementations? *Knowledge and Information Systems*, pages 1–38.

Krinner, G., Viovy, N., de Noblet-Ducoudré, N., Ogée, J., Polcher, J., Friedlingstein, P., Ciais, P., Sitch, S., and Prentice, I. C. (2005). A dynamic global vegetation model for studies of the coupled atmosphere-biosphere system. *Global Biogeochemical Cycles*, 19(1).

Kuppel, S., Peylin, P., Chevallier, F., Bacour, C., Maignan, F., and Richardson, A. (2012). Constraining a global ecosystem model with multi-site eddy-covariance data. *Biogeosciences*, 9(10):3757–3776.

Kuppel, S., Peylin, P., Maignan, F., F.Chevallier, Kiely, G., Montagnani, L., and Cescatti, A. (2014). Mode-data fusion across ecosystems: from multi-site optimizations to global simulations. *Geoscientific Model Development Discussions*, 7:2961–3011.

Lasslop, G. (2010). *Model data fusion for terrestrial biosphere models with carbon and water cycle observations*. PhD thesis, Universität Hamburg.

Lasslop, G., Reichstein, M., Kattge, J., and Papale, D. (2008). Influences of observation errors in eddy flux data on inverse model parameter estimation. *Biogeosciences Discussions*, 5(1):751–785.

Leggett, J., Pepper, W. J., and Swart, R. J. (1992). Emissions scenarios for the IPCC: an update. *Climate Change*, pages 69–95.

Lewis, J. M., Lakshmivarahan, S., and Dhall, S. (2006). *Dynamic data assimilation: a least squares approach*, volume 13. Cambridge University Press.

Liu, Y. and Gupta, H. V. (2007). Uncertainty in hydrologic modeling: Toward an integrated data assimilation framework. *Water Resources Research*, 43(7).

Lloyd, S. (1982). Least squares quantization in PCM. *IEEE Transactions on Information Theory*, 28(2):129–137.

Luke, C. (2011). *Modelling aspects of land-atmosphere interaction: thermal instability in peatland soils and land parameter estimation through data assimilation*. PhD thesis, University of Exeter.

Luo, Y., Keenan, T. F., and Smith, M. (2015). Predictability of the terrestrial carbon cycle. *Global Change Biology*, 21(5):1737–1751.

MacBean, N., Peylin, P., Chevallier, F., Scholze, M., and Schürmann, G. (2016). Consistent assimilation of multiple data streams in a carbon cycle data assimilation system. *Geoscientific Model Development*, 9(10):3569.

Madsen, H. (2003). Parameter estimation in distributed hydrological catchment modelling using automatic calibration with multiple objectives. *Advances in Water Resources*, 26(2):205 – 216.

Manabe, S. (1969). Climate and the Ocean Circulation 1. *Monthly Weather Review*, 97(11):739–774.

McGuffie, K. and Henderson-Sellers, A. (2001). Forty years of numerical climate modelling. *International Journal of Climatology*, 21(9):1067–1109.

Medlyn, B. E., Duursma, R. A., Eamus, D., Ellsworth, D. S., Prentice, I. C., Barton, C. V.,

Crous, K. Y., De Angelis, P., Freeman, M., and Wingate, L. (2011). Reconciling the optimal and empirical approaches to modelling stomatal conductance. *Global Change Biology*, 17(6):2134–2144.

Medvigy, D. and Moorcroft, P. R. (2011). Predicting ecosystem dynamics at regional scales: an evaluation of a terrestrial biosphere model for the forests of northeastern North America. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 367(1586):222–235.

Medvigy, D., Wofsy, S. C., Munger, J. W., Hollinger, D. Y., and Moorcroft, P. R. (2009). Mechanistic scaling of ecosystem function and dynamics in space and time: Ecosystem Demography model version 2. *Journal of Geophysical Research: Biogeosciences*, 114(G1).

Mercado, L., Bellouin, N., Sitch, S., Boucher, O., Huntingford, C., Wild, M., and Cox, P. (2009). Impact of changes in diffuse radiation on the global land carbon sink. *Nature*, 458(7241):1014–1017.

Mercado, L., Huntingford, C., Gash, J. H., Cox, P. M., and Jogireddy, V. (2007). Improving the representation of radiation interception and photosynthesis for climate model applications. *Tellus B*, 59(3):553–565.

Moffat, A. M., Papale, D., Reichstein, M., Hollinger, D. Y., Richardson, A. D., Barr, A. G., Beckstein, C., Braswell, B. H., Churkina, G., Desai, A. R., Falge, E., Gove, J. H., Heimann, M., Hui, D., Jarvis, A. J., Kattge, J., Noormets, A., and Stauch, V. J. (2007). Comprehensive comparison of gap-filling techniques for eddy covariance net carbon fluxes. *Agricultural and Forest Meteorology*, 147(3–4):209 – 232.

Moncrieff, J., Malhi, Y., and Leuning, R. (1996). The propagation of errors in long-term measurements of land-atmosphere fluxes of carbon and water. *Global Change Biology*, 2(3):231–240.

Monsi, M. and Saeki, T. (1953). The light factor in plant communities and its significance for dry matter production. *Japanese Journal of Botany*, 14(1):22–52.

Moss, R. H., Edmonds, J. A., Hibbard, K. A., Manning, M. R., Rose, S. K., van Vuuren, D. P., Carter, T. R., Emori, S., Kainuma, M., Kram, T., Meehl, G. A., Mitchell, J. F. B., Nakicenovic, N., Riahi, K., Smith, S. J., Stouffer, R. J., Thomson, A. M., Weyant, J. P., and Wilbanks, T. J. (2010). The next generation of scenarios for climate change research and assessment. *Nature*, 463(7282):747–756.

Myneni, R., Hoffman, S., Knyazikhin, Y., Privette, J., Glassy, J., Tian, Y., Wang, Y., Song, X., Zhang, Y., Smith, G., et al. (2002). Global products of vegetation leaf area and fraction absorbed PAR from year one of MODIS data. *Remote Sensing of Environment*, 83(1):214–231.

Nakicenovic, N., Alcamo, J., Grubler, A., Riahi, K., Roehrl, R., Rogner, H.-H., and Victor, N. (2000). *Special Report on Emissions Scenarios (SRES), A Special Report of Working*

*Group III of the Intergovernmental Panel on Climate Change*. Cambridge University Press.

Nash, J. and Sutcliffe, J. (1970). River flow forecasting through conceptual models part I — A discussion of principles. *Journal of Hydrology*, 10(3):282 – 290.

Naumann, U. (2011). *The art of differentiating computer programs: an introduction to algorithmic differentiation*. SIAM.

Norby, R. J., Warren, J. M., Iversen, C. M., Medlyn, B. E., and McMurtrie, R. E. (2010). $CO_2$ enhancement of forest productivity constrained by limited nitrogen availability. *Proceedings of the National Academy of Sciences*, 107(45):19368–19373.

Papale, D. (2012). *Data Gap Filling*, pages 159–172. Springer Netherlands, Dordrecht.

Papale, D., Reichstein, M., Aubinet, M., Canfora, E., Bernhofer, C., Kutsch, W., Longdoz, B., Rambal, S., Valentini, R., Vesala, T., et al. (2006). Towards a standardized processing of Net Ecosystem Exchange measured with eddy covariance technique: algorithms and uncertainty estimation. *Biogeosciences*, 3(4):571–583.

Pearson, D., Jones, C. D., and Hughes, J. K. (2009). Estimation of carbon cycle parameters in jules. Presentation at JULES Science Meeting, https://jules.jchmr.org/community/meetings/january-2009/Pearson.pdf.

Peng, C., Guilot, J., Wu, H., Jiang, H., and Luo, Y. (2011). Integrating models with data in ecology and palaeoecology advances towards a model-data fusion approach. *Ecology Letters*, 14:522–536.

Peylin, P., Bacour, C., MacBean, N., Leonard, S., Rayner, P., Kuppel, S., Koffi, E., Kane, A., Maignan, F., Chevallier, F., et al. (2016). A new stepwise carbon cycle data assimilation system using multiple data streams to constrain the simulated land surface carbon cycle. *Geoscientific Model Development*, 9(9):3321.

Piao, S., Friedlingstein, P., Ciais, P., de Noblet-Ducoudré, N., Labat, D., and Zaehle, S. (2007). Changes in climate and land use have a larger direct impact than rising $CO_2$ on global river runoff trends. *Proceedings of the National Academy of Sciences*, 104(39):15242–15247.

Pinnington, E., Casella, E., Dance, S., Lawless, A., Morison, J., Nichols, N., Wilkinson, M., and Quaife, T. L. (2016). Investigating the role of prior and observation error correlations in improving a model forecast of forest carbon balance using Four Dimensional Variational data assimilation. *Agricultural and Forest Meteorology*, (228-229):299–314.

Pitman, A. (2003). The evolution of, and revolution in, land surface schemes designed for climate models. *International Journal of Climatology*, 23(5):479–510.

Poorter, H., Niinemets, Ü., Poorter, L., Wright, I. J., and Villar, R. (2009). Causes and consequences of variation in leaf mass per area (LMA): a meta-analysis. *New Phytologist*, 182(3):565–588.

Prentice, I. C., Dong, N., Gleason, S. M., Maire, V., and Wright, I. J. (2014). Balancing the costs of carbon gain and water transport: testing a new theoretical framework for plant functional ecology. *Ecology Letters*, 17(1):82–91.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. (2007). *Numerical Recipes 3rd Edition: The Art of Scientific Computing*. Cambridge University Press, New York, NY, USA, 3 edition.

Quaife, T. L., Brown, P., Black, E., and Lewis, J. (2014). Assimilating EO data into JULES. Presentation at JULES Science Meeting.

R Development Core Team (2015). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

Raoult, N. M., Jupp, T. E., Cox, P. M., and Luke, C. M. (2016). Land-surface parameter optimisation using data assimilation techniques: the adJULES system V1.0. *Geoscientific Model Development*, 9(8):2833–2852.

Raupach, M., Rayner, P., Barrett, D., Defries, R., Heimann, M., Ojima, D., Quegan, S., and Schmullius, C. (2005). Model-data synthesis in terrestrial carbon observation methods, data requirements and data uncertainty specifications. *Global Change Biology*, 11:378–297.

Rayner, P., Scholze, M., Knorr, W., Kaminski, T., Giering, R., and Widmann, H. (2005). Two decades of terrestrial carbon fluxes from a carbon cycle data assimilation system (CCDAS). *Global Biogeochemical Cycles*, 19.

Reichstein, M., Falge, E., Baldocchi, D., Papale, D., Aubinet, M., Berbigier, P., Bernhofer, C., Buchmann, N., Gilmanov, T., Granier, A., et al. (2005). On the separation of net ecosystem exchange into assimilation and ecosystem respiration: review and improved algorithm. *Global Change Biology*, 11(9):1424–1439.

Reichstein, M., Tenhunen, J., Roupsard, O., Ourcival, J.-M., Rambal, S., Miglietta, F., Peressotti, A., Pecchiari, M., Tirone, G., and Valentini, R. (2003). Inverse modeling of seasonal drought effects on canopy $CO_2/H_2O$ exchange in three Mediterranean ecosystems. *Journal of Geophysical Research: Atmospheres (1984–2012)*, 108(D23).

Reynolds, O. (1895). On the dynamical theory of incompressible viscous fluids and the determination of the criterion. *Philosophical Transactions of the Royal Society of London*, 186:123.

Riahi, K., Grübler, A., and Nakicenovic, N. (2007). Scenarios of long-term socio-economic and environmental development under climate stabilization. *Technological Forecasting and Social Change*, 74(7):887–935.

Riahi, K., Rao, S., Krey, V., Cho, C., Chirkov, V., Fischer, G., Kindermann, G., Nakicenovic, N., and Rafaj, P. (2011). RCP 8.5—A scenario of comparatively high greenhouse gas emissions. *Climatic Change*, 109(1):33.

Richardson, A. D., Hollinger, D. Y., Burba, G. G., Davis, K. J., Flanagan, L. B., Katul, G. G., Munger, J. W., Ricciuto, D. M., Stoy, P. C., Suyker, A. E., et al. (2006). A multi-site analysis of random error in tower-based measurements of carbon and energy fluxes. *Agricultural and Forest Meteorology*, 136(1):1–18.

Richardson, A. D., Mahecha, M. D., Falge, E., Kattge, J., Moffat, A. M., Papale, D., Reichstein, M., Stauch, V. J., Braswell, B. H., Churkina, G., Kruijt, B., and Hollinger, D. Y. (2008). Statistical properties of random $CO_2$ flux measurement uncertainty inferred from model residuals. *Agricultural and Forest Meteorology*, 148(1):38 – 50.

Rosolem, R., Gupta, H. V., Shuttleworth, W. J., de Gonçcalves, L. G. G., and Zeng, X. (2012). Towards a comprehensive approach to parameter estimation in land surface parameterization schemes. *Hydrological Processes*, 27:2075–2097.

Santaren, D., Peylin, P., Viovy, N., and Ciais, P. (2007). Optimizing a process-based ecosystem model with eddy-covariance flux measurements: A pine forest in southern France. *Global Biogeochemical Cycles*, 21.

Schulze, E.-D., Robichaux, R., Grace, J., Rundel, P., and Ehleringer, J. (1987). Plant Water Balance: In diverse habitats, where water often is scarce, plants display a variety of mechanisms for managing this essential resource. *BioScience*, 37(1):30–37.

Schwartz, P. (1996). *The art of the long view: planning for the future in an uncertain world*. New York: Doubleday.

Sellers, P. (1985). Canopy reflectance, photosynthesis and transpiration. *International Journal of Remote Sensing*, 6(8):1335–1372.

Sellers, P., Dickinson, R., Randall, D., Betts, A., Hall, F., Berry, J., Collatz, G., Denning, A., Mooney, H., Nobre, C., et al. (1997). Modeling the exchanges of energy, water, and carbon between continents and the atmosphere. *Science*, 275(5299):502.

Sitch, S., Cox, P., Collins, W., and Huntingford, C. (2007). Indirect radiative forcing of climate change through ozone effects on the land-carbon sink. *Nature*, 448(7155):791–794.

Smith, S. J. and Wigley, T. (2006). Multi-gas forcing stabilization with Minicam. *The Energy Journal*, pages 373–391.

Stauch, V. J., Jarvis, A. J., and Schulz, K. (2008). Estimation of net carbon exchange using eddy covariance $CO_2$ flux observations and a stochastic model. *Journal of Geophysical Research: Atmospheres*, 113(D3).

Stewart, L. M., Dance, S. L., and Nichols, N. K. (2013). Data assimilation with correlated observation errors: experiments with a 1-D shallow water model. *Tellus A: Dynamic Meteorology and Oceanography*, 65(1):19546.

Stocker, T., Qin, D., Plattner, G.-K., Tignor, M., Allen, S., Boschung, J., Nauels, A., Xia,

Y., Bex, V., and Midgley, P., editors (2013). *Summary for Policymakers*, pages 1–30. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Stöckli, R., Rutishauser, T., Dragoni, D., O'Keefe, J., Thornton, P. E., Jolly, M., Lu, L., and Denning, A. S. (2008). Remote sensing data assimilation for a prognostic phenology model. *Journal of Geophysical Research: Biogeosciences*, 113(G4).

Takashima, T., Hikosaka, K., and Hirose, T. (2004). Photosynthesis or persistence: nitrogen allocation in leaves of evergreen and deciduous Quercus species. *Plant, Cell & Environment*, 27(8):1047–1054.

Tan, P., Steinbach, M., and Kumar, V. (2006). *Introduction to Data Mining*. Always learning. Pearson Addison Wesley.

Tang, X., Li, H., Desai, A. R., Nagy, Z., Luo, J., Kolb, T. E., Olioso, A., Xu, X., Yao, L., Kutsch, W., et al. (2014). How is water-use efficiency of terrestrial ecosystems distributed and changing on Earth? *Scientific Reports*, 4:7483.

Tarantola, A. (1987). *Inverse Problem Theory: Methods for Data Fitting and Parameter Estimation*. Elsevier, New York.

Taylor, K. E. (2001). Summarizing multiple aspects of model performance in a single diagram. *Journal of Geophysical Research: Atmospheres*, 106(D7):7183–7192.

Tebaldi, C. and Knutti, R. (2007). The use of the multi-model ensemble in probabilistic climate projections. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 365(1857):2053–2075.

Thornton, P. E., Lamarque, J.-F., Rosenbloom, N. A., and Mahowald, N. M. (2007). Influence of carbon-nitrogen cycle coupling on land model response to $CO_2$ fertilization and climate variability. *Global Biogeochemical Cycles*, 21(4).

Thum, T., Aalto, T., Laurila, T., Aurela, M., Lindroth, A., and Vesala, T. (2008). Assessing seasonality of biochemical $CO_2$ exchange model parameters from micrometeorological flux observations at boreal coniferous forest. *Biogeosciences*, 5(6):1625–1639.

Trefethen, L. and Bau, D. (1997). *Numerical Linear Algebra*. Society for Industrial and Applied Mathematics.

Trudinger, C. M., Raupach, M. R., Rayner, P. J., Kattge, J., Liu, Q., Pak, B., Reichstein, M., Renzullo, L., Richardson, A. D., Roxburgh, S. H., Styles, J., Wang, Y. P., Briggs, P., Barrett, D., and Nikolova, S. (2007). OptIC project: An intercomparison of optimization techniques for parameter estimation in terrestrial biogeochemical models. *Journal of Geophysical Research*, 112.

Van Vuuren, D. P., Den Elzen, M. G., Lucas, P. L., Eickhout, B., Strengers, B. J., van Ruijven, B., Wonink, S., and van Houdt, R. (2007). Stabilizing greenhouse gas concentrations at low levels: an assessment of reduction strategies and costs. *Climatic Change*, 81(2):119–159.

Van Vuuren, D. P., Edmonds, J., Kainuma, M., Riahi, K., Thomson, A., Hibbard, K., Hurtt, G. C., Kram, T., Krey, V., Lamarque, J.-F., et al. (2011). The representative concentration pathways: an overview. *Climatic Change*, 109(1-2):5.

Verbeeck, H., Peylin, P., Bacour, C., Bonal, D., Steppe, K., and Ciais, P. (2011). Seasonal patterns of $CO_2$ fluxes in Amazon forests: Fusion of eddy covariance data and the ORCHIDEE model. *Journal of Geophysical Research*, 116(GO2018).

Vrugt, J. A., Diks, C. G. H., Gupta, H. V., Bouten, W., and Verstraten, J. M. (2005). Improved treatment of uncertainty in hydrologic modeling: Combining the strengths of global optimization and data assimilation. *Water Resources Research*, 41.

Wang, Y. P., Baldocchi, D., Leuning, R., Falge, E., and Vesala, T. (2007). Estimating parameters in a land-surface model by applying nonlinear inversion to eddy covariance flux measurements from eight Fluxnet sites. *Global Change Biology*, 13(3):652–670.

Wang, Y. P., Leuning, R., Cleugh, H. A., and Coppin, P. A. (2001). Parameter estimation in surface exchange models using nonlinear inversion: how many parameters can we estimate and which measurements are most useful? *Global Change Biology*, 7(5):495–510.

Wang, Y. P., Trudinger, C. M., and Enting, I. G. (2009). A review of applications of model–data fusion to studies of terrestrial carbon fluxes at different scales. *Agricultural and Forest Meteorology*, 149(11):1829–1842.

Weston, P., Bell, W., and Eyre, J. (2014). Accounting for correlated error in the assimilation of high-resolution sounder data. *Quarterly Journal of the Royal Meteorological Society*, 140(685):2420–2429.

Williams, M., Richardson, A., Reichstein, M., Stoy, P., Peylin, P., Verbeeck, H., Carvalhais, N., Jung, M., Hollinger, D., Kattge, J., et al. (2009). Improving land surface models with FLUXNET data. *Biogeosciences*, 6(7):1341–1359.

Williams, M., Schwarz, P. A., Law, B. E., Irvine, J., and Kurpius, M. R. (2005). An improved analysis of forest carbon dynamics using data assimilation. *Global Change Biology*, 11(1):89–105.

Wise, M., Calvin, K., Thomson, A., Clarke, L., Bond-Lamberty, B., Sands, R., Smith, S. J., Janetos, A., and Edmonds, J. (2009). Implications of limiting $CO_2$ concentrations for land use and energy. *Science*, 324(5931):1183–1186.

Xiao, J., Davis, K. J., Urban, N. M., Keller, K., and Saliendra, N. Z. (2011). Upscaling carbon fluxes from towers to the regional scale: Influence of parameter variability and land cover representation on regional flux estimates. *Journal of Geophysical Research: Biogeosciences*, 116(G3).

Yapo, P. O., Gupta, H. V., and Sorooshian, S. (1998). Multi-objective global optimization for hydrologic models. *Journal of Hydrology*, 204(1-4):83–97.

Zaehle, S., Friedlingstein, P., and Friend, A. D. (2010). Terrestrial nitrogen feedbacks may accelerate future climate change. *Geophysical Research Letters*, 37(1).